

# Phylogenetic Estimation of Timescales Using Ancient DNA: The Effects of Temporal Sampling Scheme and Uncertainty in Sample Ages

Martyna Molak,<sup>\*1</sup> Eline D. Lorenzen,<sup>2,3</sup> Beth Shapiro,<sup>4,5</sup> and Simon Y.W. Ho<sup>1</sup>

<sup>1</sup>School of Biological Sciences, University of Sydney, Sydney, Australia

<sup>2</sup>Department of Integrative Biology, University of California, Berkeley

<sup>3</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

<sup>4</sup>Department of Biology, Pennsylvania State University

<sup>5</sup>Department of Ecology and Evolutionary Biology, University of California, Santa Cruz

**\*Corresponding author:** E-mail: martyna.molak@sydney.edu.au.

**Associate editor:** Barbara Holland

## Abstract

In recent years, ancient DNA has increasingly been used for estimating molecular timescales, particularly in studies of substitution rates and demographic histories. Molecular clocks can be calibrated using temporal information from ancient DNA sequences. This information comes from the ages of the ancient samples, which can be estimated by radiocarbon dating the source material or by dating the layers in which the material was deposited. Both methods involve sources of uncertainty. The performance of Bayesian phylogenetic inference depends on the information content of the data set, which includes variation in the DNA sequences and the structure of the sample ages. Various sources of estimation error can reduce our ability to estimate rates and timescales accurately and precisely. We investigated the impact of sample-dating uncertainties on the estimation of evolutionary timescale parameters using the software BEAST. Our analyses involved 11 published data sets and focused on estimates of substitution rate and root age. We show that, provided that samples have been accurately dated and have a broad temporal span, it might be unnecessary to account for sample-dating uncertainty in Bayesian phylogenetic analyses of ancient DNA. We also investigated the sample size and temporal span of the ancient DNA sequences needed to estimate phylogenetic timescales reliably. Our results show that the range of sample ages plays a crucial role in determining the quality of the results but that accurate and precise phylogenetic estimates of timescales can be made even with only a few ancient sequences. These findings have important practical consequences for studies of molecular rates, timescales, and population dynamics.

**Key words:** sample-dating error, evolutionary rate, molecular clock, calibration.

## Introduction

The tempo and timescale of evolutionary and demographic processes are of considerable interest in biological research. These can be studied using molecular-clock models in phylogenetic analyses of DNA sequence data. Although some clock models assume a constant rate of evolution (strict clock), others allow the evolutionary rate to vary among lineages (relaxed clock). A common characteristic of all molecular-clock methods is that they require the use of age calibrations to convert units of genetic change into units of time. Calibrations are commonly based on paleontological or geological data, which can provide an estimate of the timing of divergence events (internal nodes) in the phylogenetic tree. However, because of difficulties in assigning fossils to branches of the evolutionary tree, the placement of calibrations is often highly uncertain (Lee et al. 2009). In addition, internal-node calibrations can carry a substantial amount of temporal uncertainty, and it is usually difficult to quantify this (Ho and Phillips 2009).

Identifying reliable calibrations for intraspecific analyses is particularly challenging. Estimating intraspecific timeframes is important when studying changes in population sizes and structure and associating these with abiotic and biotic factors such as climate change or human activity (Arbogast et al. 2002; Ramakrishnan and Hadly 2009; de Bruyn et al. 2011). The fossil record is usually uninformative with respect to the timing of intraspecific divergences (Ho, Lanfear, Bromham, et al. 2011), whereas calibrations based on geological events require a number of strong assumptions that might not be met by the data (e.g., Marko 2002). Furthermore, because of factors such as incomplete lineage sorting, the association between genetic divergence and population divergence is not always clear (Edwards and Beerli 2000). Although paleontological or geological calibrations can be used, there is evidence that these are inappropriate for intraspecific analyses because of the effects of saturation, purifying selection, and other factors (Ho and Larson 2006; Ho et al. 2008). Therefore, it is preferable to employ calibrations within the

intraspecific genealogy, using DNA sequences from dated material sampled at various points in time—time-stamped DNA sequences (Rambaut 2000; Drummond et al. 2004). Provided that these sequences are sufficiently variable, owing to either a high mutation rate or a broad temporal span, it is possible to estimate the rate of molecular evolution (Drummond et al. 2002, 2003).

Several phylogenetic methods can use temporal information from time-stamped DNA sequences to calibrate molecular clocks. These include dedicated maximum-likelihood and Bayesian methods, which have been implemented in a range of programs including BEAST (Drummond and Rambaut 2007), PAML (Yang 2007), r8s (Sanderson 2003), and Bayesian Serial SimCoal (Excoffier et al. 2000; Anderson et al. 2005). Some of these methods require a fixed tree, whereas others can coestimate substitution rates, node times, and phylogenetic relationships. Here, we focus on Bayesian phylogenetic methods (implemented in the software BEAST), which allow the uncertainty in sample ages to be included in the form of prior probability distributions.

The performance of Bayesian phylogenetic analysis of ancient DNA strongly depends on the information content of the data set (Ho, Lanfear, Phillips, et al. 2011). The number, genetic variation, and age range of the ancient sequences used for calibration of molecular clocks are expected to affect the ability to estimate rates and timescales accurately (Drummond et al. 2003). For this reason, the mitochondrial control region is a useful marker because its high mutation rate means that it can accumulate an appreciable amount of genetic change over a short period of time. Various sources of error, such as in determination of sample age, can also reduce the accuracy of phylogenetic estimates of timescales (e.g., Wertheim 2010).

When dealing with ancient DNA, sample ages are usually unknown and need to be estimated using direct or indirect dating methods. However, the common practice of treating the estimates of sample ages as point values for calibration is potentially problematic because it ignores the associated uncertainties (Ho and Phillips 2009; Shapiro et al. 2011).

Accelerator mass spectrometry radiocarbon dating is often used to directly date material <50,000 years old. This method measures the  $^{14}\text{C}$ -isotope content of a sample and assumes a constant rate of radioactive decay. There are several sources of error associated with radiocarbon dating, including estimating the number of stochastic  $^{14}\text{C}$  decay events within a finite time interval. Dating laboratories usually describe these errors using Gaussian (normal) distributions and estimate them along with the radiocarbon dates (Stuiver and Polach 1977; Bowman 1990).

Because of variation in the level of atmospheric  $^{14}\text{C}$  through time, the age estimates from radiocarbon dating, given in radiocarbon years, do not equal calendar years. It is sometimes useful to convert radiocarbon years into calendar years, for example, when the aim is to compare the estimated timescale of demographic events with records of climate change or other factors (Svensson et al. 2008). This conversion, which can be performed using calibration curves,

reshapes the distribution of estimated dating error and introduces additional sources of uncertainty (Bronk Ramsey 1995; Beavan-Athfield et al. 2001).

Other sources of dating error are less quantifiable. For example, samples can be contaminated with more recent sources of carbon, owing to either the dynamics of the depositional environment or during excavation and laboratory preparation (Mellars 2006). The resulting increase in  $^{14}\text{C}$  content can lead to underestimation of the true age of the sample. Such risks can be assessed and minimized by replicating the dating process using different sections of the sample and by checking concordance with archeological or geological context (Törnqvist et al. 1992).

As an alternative to radiocarbon dating, which is costly and involves destruction of the material being analyzed, samples can be dated indirectly. In indirect dating, the age of the sample's depositional layer is estimated by archeological or stratigraphic context or by directly dating organic remains within or at the boundary of the layer. Indirect dates are, however, associated with far greater uncertainties than direct dates. Apart from the errors associated with the dating of the layer boundaries, reburial or mixing of deposits can lead to substantial errors. Consequently, assigning a point value for the age of a layer-dated sample can be highly misleading. In studies of ancient DNA from environmental samples, there is an additional risk of DNA migrating between strata (Haile et al. 2007).

Sample-age uncertainties can be incorporated into Bayesian methods by specifying the prior age distribution of each ancient sample, rather than assigning a point value (Ho and Phillips 2009). For several reasons, however, they are usually ignored in phylogenetic analyses. Uncertain sample ages need to be estimated in the analysis, which increases the number of parameters, reduces overall estimation precision, and leads to the risk of overparameterization.

The possibility of modeling the age uncertainty in ancient DNA sequences is particularly useful for analyses of layer-dated samples, which can have very wide errors (Korsten et al. 2009). It also enables the inclusion of samples that are beyond the reach of radiocarbon dating and can only be given a minimum age constraint. The posterior age distributions can vary greatly from the prior distributions, likely reflecting the true age of the samples. This also presents a method for dating samples of unknown or highly uncertain ages (Shapiro et al. 2011).

Despite the growing use of ancient DNA in population genetics and phylogeographic research, it is not known whether uncertainty in the estimates of sample ages has an impact on molecular estimates of rates and timescales. In this study, we incorporate the estimated sample-age uncertainty into Bayesian phylogenetic analyses of 11 published ancient DNA data sets. We investigate the impact of this uncertainty on the precision and accuracy of estimates of substitution rates and divergence times. This allows us to determine whether estimates of evolutionary timescales can be improved by taking sample-age error into account. We also examine how the number and ages of the samples used, as

well as other properties of the data set, affect the power and reliability of timescale estimation.

## Materials and Methods

### Data Sets

Eleven ancient DNA data sets were analyzed in our study (table 1). These were chosen according to two criteria: 1) samples had been dated using radiocarbon and/or layer dates and 2) strength of the temporal signal in the data set has been confirmed using the date-randomization test described later. We chose to use uncalibrated radiocarbon dates rather than calendar dates, because the former are easier to model using simple parametric distributions. Undated samples, as well as those given only minimum radiocarbon ages (“infinite” dates), were excluded from all analyses.

In the case of  $^{14}\text{C}$ -dated samples, we define the error in sample-age estimation as the standard errors provided by the dating laboratory. In the case of layer-dated samples, the age bounds of the assigned layer are taken as the estimation error. These assume that the layers have been assigned correctly to the samples but that the exact position of the sample within its layer is unknown. This is often the case for published data (e.g., sample age assigned as “Late Pleistocene”). For the sake of simplicity, we assumed that all sample-dating errors are uncorrelated among ancient samples. This assumption would be violated if systematic biases are introduced by contamination or laboratory errors.

A variety of temporal spans are represented by the data sets, ranging from 5,400 years in boar to 120,000 years in brown bear. Most samples were directly dated using  $^{14}\text{C}$ . Only arctic fox (all except one sample), boar, and a single brown bear individual were layer dated. Estimation errors in the layer dates are generally higher (average 8% of the age of the sample) than those in the  $^{14}\text{C}$  dates (3%). The boar samples were dated according to cultural periods (Watanabe et al. 2001, 2004). The arctic fox samples were dated by stratigraphic horizons, the ages of which were estimated using  $^{14}\text{C}$  dating of associated organic material (Dalén et al. 2007). Sample sizes and alignment lengths vary among data sets, ranging from 23 to 182 cave lion and bison sequences, respectively, and 193 to 682 nucleotides in brown bear and musk ox, respectively (table 1).

### The Effect of Sample-Dating Errors on Bayesian Parameter Estimates

The best-fit model of nucleotide substitution was chosen for each data set using the Bayesian information criterion, calculated using ModelGenerator (Keane et al. 2006). Compared with other model-selection criteria, the Bayesian information criterion has been shown to perform well under a variety of scenarios (Luo et al. 2010). Because of the intraspecific character of the data, substitution models containing invariant sites were excluded. At the population level, the proportion of invariant sites is typically overestimated because these sites are difficult to distinguish from those that simply have not yet changed.

For each data set, we determined whether the temporal span of the sample dates and the DNA sequence information were sufficient for calibrating rate estimates. This was done using a date-randomization test, described in previous studies using time-stamped sequences (Ramsden et al. 2009; Firth et al. 2010; Ho, Lanfear, Phillips, et al. 2011). We performed the test using 10 replicates of each data set. Following Firth et al. (2010), the sampling times in a data set were considered to have sufficient temporal structure and spread when the mean estimate of the evolutionary rate was not included in any of the 95% highest posterior density (HPD) intervals of the rate estimates from the date-randomized replicates. Eleven data sets met this condition and were used for further analysis (results of the test for these data sets are given on supplementary fig. S1, Supplementary Material online). Data sets are listed in table 1.

To determine whether the incorporation of sample-age uncertainty reduces the precision of estimates of timescale parameters, Bayesian phylogenetic analyses were performed using BEAST v1.6.1 (Drummond and Rambaut 2007). Uninformative priors, given in the form of uniform distributions ranging from 0 to infinity, were assigned to the evolutionary rate and population size. For each data set, constant-size and Bayesian skyride models of population history were compared using Bayes factors (Suchard et al. 2001). Analyses were first performed with point values for the sample ages (“point calibrations”). For layer-dated samples, the midpoint of the source layer was used as a point estimate. A second analysis was performed with informative prior distributions (“non-point calibrations”) of the ages of ancient DNA sequences (Shapiro et al. 2011). For each radiocarbon-dated specimen, the age was assigned a normal prior with a standard deviation equal to the standard error of the  $^{14}\text{C}$  date. Uniform priors were specified for the ages of layer-dated samples, with minimum and maximum constraints chosen according to the time period spanned by the layer. For data sets containing only ancient samples, the age of the youngest sample was used as a point calibration. Median estimates and 95% HPD intervals of substitution rates and root ages were compared between the treatments involving point calibrations and nonpoint calibrations.

A strict-clock model was used in all analyses owing to the intraspecific level of study. Rate variability among intraspecific lineages is assumed to be stochastic rather than driven by evolutionary processes (Drummond et al. 2006; Ho 2009). For shallow phylogenies, relaxed-clock models generally do not outperform strict-clock models, whereas they reduce the precision of parameter estimates (Brown and Yang 2011).

Posterior distributions of parameters were estimated using Markov chain Monte Carlo (MCMC) sampling. Each analysis was run 10 times, with samples drawn every  $10^3$  steps over  $10^7$  steps. Results from the 10 replicates were combined using LogCombiner (Drummond and Rambaut 2007), with the first  $10^6$  steps of each run discarded as burn-in. Samples from the posterior were checked for acceptable effective sample sizes ( $>200$ ) and for convergence and mixing by visual inspection

**Table 1.** Details of the 11 Mitochondrial DNA Data Sets Used in This Study.

Species and Sequence Reference	Number of Samples (Ancient + Modern)	Sampling Time Span (Years)	Alignment Length (bp)	Dating Method	Average Dating Error (%)	Best-Fit Substitution Model <sup>a</sup>	Population Model <sup>b</sup>
Arctic fox ( <i>Alopex lagopus</i> ) <sup>c</sup> (Dalén et al. 2007)	8 + 41	0–16,000	291	Layer <sup>d</sup>	7.22	HKY + G	Constant
Bison ( <i>Bison priscus</i> ) <sup>e</sup> (Shapiro et al. 2004)	160 + 22	0–60,400	615	<sup>14</sup> C	3.42	TrN + G	Constant
Boar ( <i>Sus scrofa</i> ) <sup>f</sup> (Watanabe et al. 2001, 2004)	81 + 7	0–5,400	572	Layer	7.90	HKY + G	Skyride
Brown bear ( <i>Ursus arctos</i> ) <sup>g</sup> (Korsten et al. 2009; Lindqvist et al. 2010)	47 + 66	0–120,000	193	<sup>14</sup> C <sup>f</sup>	1.82	K80 + G	Constant
Cave lion ( <i>Panthera leo spelaea</i> ) <sup>h</sup> (Barnett et al. 2009)	23 + 0	11,925–58,200	213	<sup>14</sup> C	1.31	HKY	Constant
Horse ( <i>Equus ferus</i> ) <sup>g</sup> (Lorenzen et al. 2011)	128 + 0	2,220–43,900	349	<sup>14</sup> C	0.96	HKY + G	Constant
Muskox ( <i>Ovibos moschatus</i> ) <sup>g</sup> (Campos et al. 2010)	121 + 4	0–42,550	682	<sup>14</sup> C	2.47	HKY + G	Constant
Reindeer ( <i>Rangifer tarandus</i> ) <sup>g</sup> (Lorenzen et al. 2011)	137 + 25	0–37,500	435	<sup>14</sup> C	4.56	HKY + G	Skyride
Tuatara ( <i>Sphenodon punctatus</i> ) <sup>g</sup> (Hay et al. 2008)	33 + 41	0–8,478	470	<sup>14</sup> C	3.11	F81 + G	Constant
Tuco-tuco ( <i>Ctenomys sociabilis</i> ) <sup>g</sup> (Chan et al. 2006)	45 + 1	0–10,209	253	<sup>14</sup> C	6.22	HKY + G	Constant
Woolly rhinoceros ( <i>Coelodonta antiquitatis</i> ) <sup>g</sup> (Lorenzen et al. 2011)	55 + 0	12,460–43,850	546	<sup>14</sup> C	0.70	TrN + G	Constant

NOTE.—All DNA sequences were from the control region, except in tuco-tuco where data were from cytochrome *b*.

<sup>a</sup>Chosen according to Bayesian information criterion.

<sup>b</sup>Chosen using Bayes factors.

<sup>c</sup>Date-randomization tests of data sets were passed in Ho, Lanfear, Phillips, et al. (2011).

<sup>d</sup>One sample was <sup>14</sup>C dated.

<sup>e</sup>Date-randomization tests of data sets were passed in this study.

<sup>f</sup>One sample was layer dated.

of MCMC traces in Tracer v1.5 (Rambaut and Drummond 2007). When these were not satisfactory, the MCMC analysis was continued until adequate sampling had been achieved.

To determine whether the 95% HPD intervals of the estimated parameters differed significantly between point-calibrated analyses and those incorporating uncertainty in the sample ages, we performed Wilcoxon signed-rank nondirectional tests. Pairs of 95% HPD interval sizes for substitution rate and root age estimates for each data set were compared between the analyses using point and nonpoint calibrations.

To determine which features of the data sets had the greatest effect on parameter estimates, we conducted an analysis of variance for linear regression using the statistical software R (R Development Core Team 2012). Sizes of 95% HPD intervals for substitution rate and root age estimates were tested against the following features: range of sample ages, number of variable sites, alignment length, sequence variability (fraction of variable sites), number of sequences, fraction of ancient samples in the data set, mean age of all samples, and mean age of nonmodern samples (see [supplementary table S1, Supplementary Material](#) online, for all parameter values).

### The Effect of Different Prior Distributions for Sample Ages

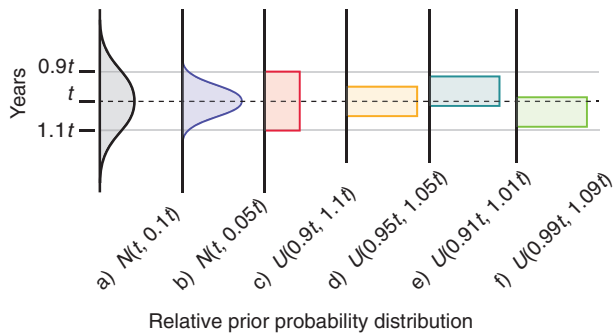
To provide further insight into the effect of including the uncertainty in sample ages on Bayesian estimates of evolutionary timescales, we performed additional analyses on two of the data sets. We focused on the largest available data set (bison) and the smallest data set that contains both ancient and modern samples (arctic fox). Thus, we covered the two extremes of the size range of data sets in this examination.

We assigned artificial dating errors to the ancient samples, including a range of normal and uniform prior distributions ([fig. 1](#)). Six analyses were performed for each data set. In each analysis, a single type of artificial error was applied to all the ancient samples. We used normal prior distributions to mimic radiocarbon-dating uncertainty, with standard deviations of 10% and 5% of the sample age ([fig. 1a](#) and [b](#)). Uniform distributions reflect layer dating where the width of the layer is either 10% or 5% of the sample age. The uniform distribution is centered on the true age of the sample ([fig. 1c](#) and [d](#)). The effect of different within-layer positions of the specimens was investigated by shifting the uniform prior, so that the true sample age was near either the minimum or the maximum bound ([fig. 1e](#) and [f](#)).

Bayesian phylogenetic analyses were performed using the various calibrations described earlier. The settings for the analyses were the same as those described in the previous section. We obtained posterior estimates of substitution rates, root ages, and sampling times.

### The Amount of Information Required to Calibrate the Molecular Clock

We investigated the relationship between the number and temporal spread of dated samples and the performance of Bayesian phylogenetic estimation of rates and timescales. Here, we focused on the brown bear data because they



**Fig. 1.** Arbitrary sample-age distributions used to examine the impact of sample-dating error on estimates of substitution rates and timescales. Distribution parameters are relative to each sample age ( $t$ ). Normal distributions =  $N(\text{mean}, \text{SD})$ ; uniform distributions =  $U(\text{minimum bound}, \text{maximum bound})$ . SD, standard deviation.

include a large number of both ancient and modern sequences (47 and 66, respectively). Apart from a single sample dated at 120,000 years, the ancient sequences are almost uniformly distributed over the past 50,000 years. These characteristics make the brown bear sequences a good data set for investigating whether and how the use of samples of varying ages for calibration influences parameter estimates.

Bayesian phylogenetic analyses were conducted using BEAST as described earlier, except that we excluded the majority of the ancient samples and used only one or three ancient sequences to calibrate the parameter estimates. The analysis with one ancient sequence included 1) the oldest sample (120,000 years), 2) the sample with age closest to 10% of the age of the oldest sample (11,940 years), or 3) the sample with age closest to 1% of the age of the oldest sample (1,550 years) (supplementary fig. S2, Supplementary Material online). The analysis with three ancient samples included 1) the three oldest sequences, 2) the three sequences of intermediate age, 3) the three youngest sequences, or 4) one sequence from each of the three age categories (supplementary fig. S2, Supplementary Material online). Samples from the posterior were drawn every 2,000 steps over  $2 \times 10^7$  steps. The first  $10^6$  steps of each run were treated as burn-in. Ten replicate analyses were performed, and the samples from the posterior were combined.

In practice, limits on the accuracy of radiocarbon dating mean that few data sets include DNA sequences that exceed 50,000 years. Accordingly, we repeated some of the analyses after excluding the 120,000-year sample. These additional analyses included those with only one or three ancient sequences.

In an additional round of analyses, we randomly removed an increasing number of ancient DNA sequences from the brown bear data set. This was done to examine the effect of using only a small number of sampling times for molecular clock calibration. We analyzed 17 data subsets of varying size, comprising the 66 modern samples and a decreasing number of ancient DNA sequences. To account for sampling effects, we performed three replicates of each analysis, with ancient sequences chosen randomly each time. Pruned data sets were analyzed using BEAST, with the settings described earlier.

To test whether the results from this analysis were applicable beyond the brown bear data, we performed a simulation study. Using BayeSSC (Excoffier et al. 2000; Anderson et al. 2005), we simulated sequence evolution to produce data sets comprising 50 modern and 50 ancient samples. Simulations were conducted using three different evolutionary rates ( $2 \times 10^{-7}$ ,  $5 \times 10^{-7}$ , and  $10^{-6}$  substitutions/site/year), constant population size, and the HKY substitution model with  $\kappa = 20$  (transition/transversion bias = 0.909). The 50 ancient samples were drawn at 1,000-year intervals from 1,000 to 50,000 years before present.

For each of the simulated data sets, we conducted phylogenetic analyses after randomly removing an increasing number of ancient DNA sequences. To account for sampling effects, we performed three replicates of each analysis, with ancient sequences chosen randomly each time. Data sets were analyzed using BEAST as described earlier.

## Results

### The Effect of Sample-Dating Errors on Bayesian Parameter Estimates

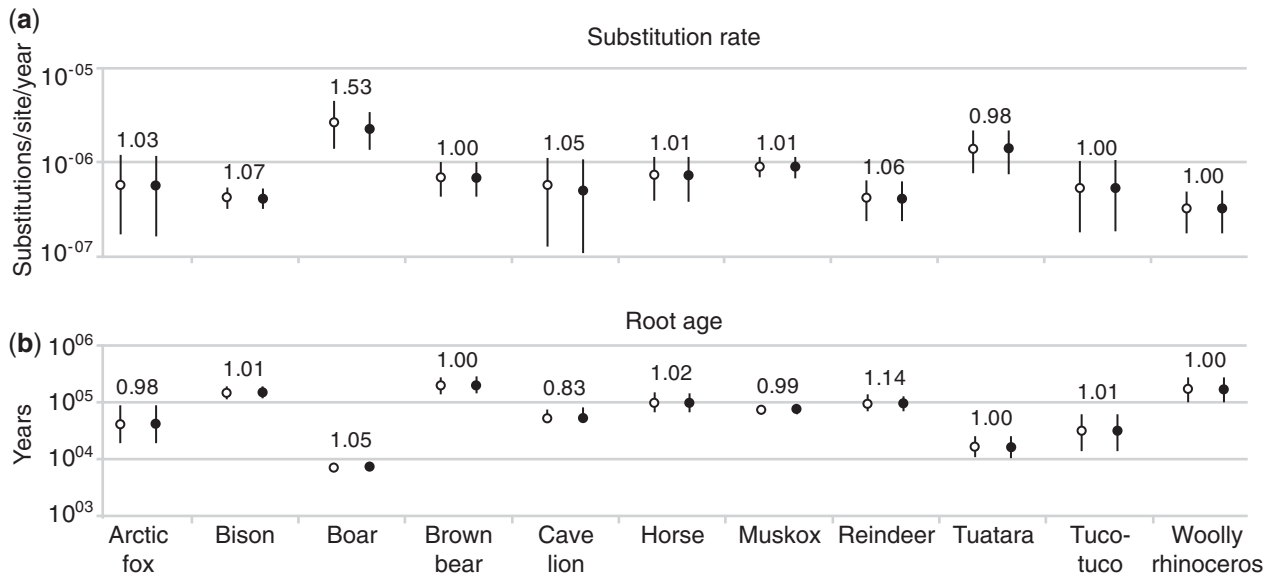
In most cases, incorporating the uncertainty in sample ages did not substantially affect estimates of either substitution rate (fig. 2a) or root age (fig. 2b). The 95% HPD intervals of the posterior rate estimates changed by more than 5% for only three of the 11 data sets (increase of 7% in bison, 53% in boar, and 6% in reindeer). The 95% HPD intervals for root age estimates changed noticeably in only two data sets (17% decrease in cave lion and 14% increase in reindeer). The Wilcoxon signed-rank test did not indicate any effect of incorporating sample age uncertainties on estimates of substitution rates ( $P = 0.054$ ) or root ages ( $P = 0.610$ ). A regression analysis revealed no significant relationships between the performance of Bayesian parameter estimation and the characteristics of the data sets (supplementary table S2, Supplementary Material online).

### The Effect of Different Prior Distributions for Sample Ages

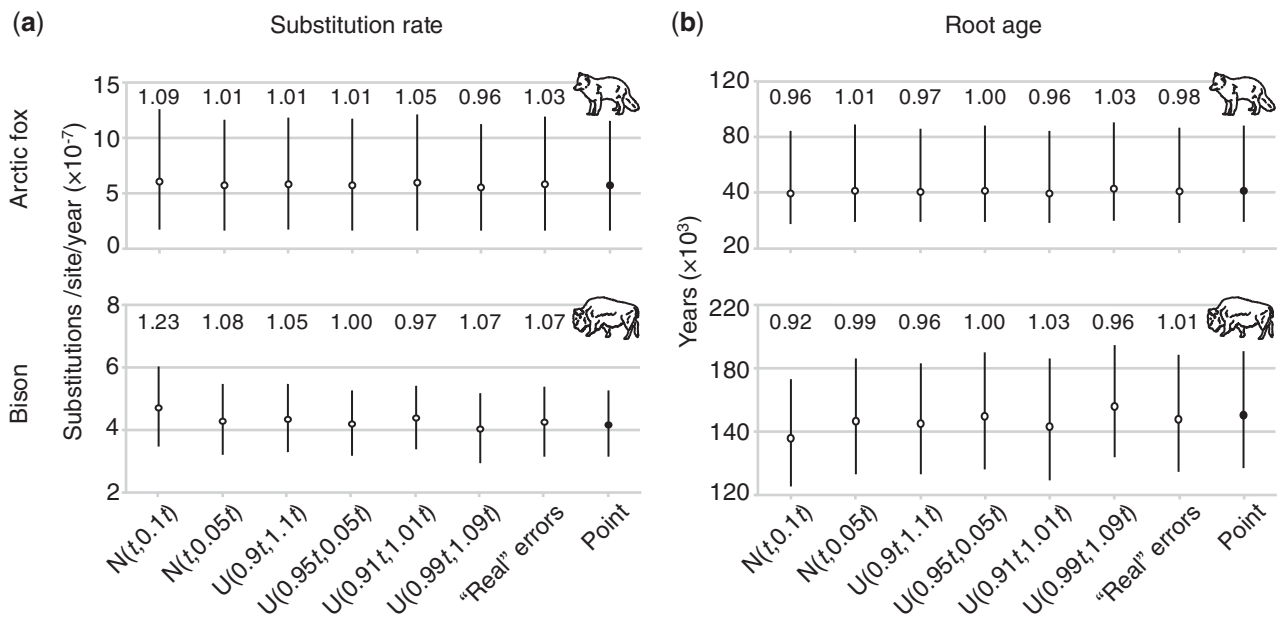
Assigning artificial errors to the ancient DNA sequence ages did not influence estimates of substitution rate and root age for the arctic fox data set (fig. 3a and b). The 95% HPD intervals of the rate estimate changed (relative to estimates made using point calibrations) by more than 5% only when a prior distribution of  $N(0.9t, 1.1t)$  was used for the sampling times. The 95% HPD intervals of root-age estimates did not change by more than 4% for any level of sample-dating error.

The bison data set was more affected by the introduction of uncertainties in the sampling times (fig. 3c and d). The 95% HPD interval of the estimate of the substitution rate increased for most error levels except for calibrations with prior distributions  $U(0.95t, 1.05t)$  and  $U(0.91t, 1.01t)$ . For the 95% HPD intervals of root-age estimates, the only significant change was a decrease when the sample-dating uncertainty was  $N(0.9t, 1.1t)$ .

We inspected the marginal posterior densities of sample-date estimates when arbitrary prior distributions



**Fig. 2.** Estimates of (a) substitution rate and (b) root age for the 11 data sets. For each species, age errors were either incorporated into the estimates (empty circles) or not (full circles). Circles denote medians, and vertical bars denote 95% HPD intervals. Values above bars indicate the relative size of the 95% HPD interval of the analysis including age errors compared with the analysis using point calibrations.



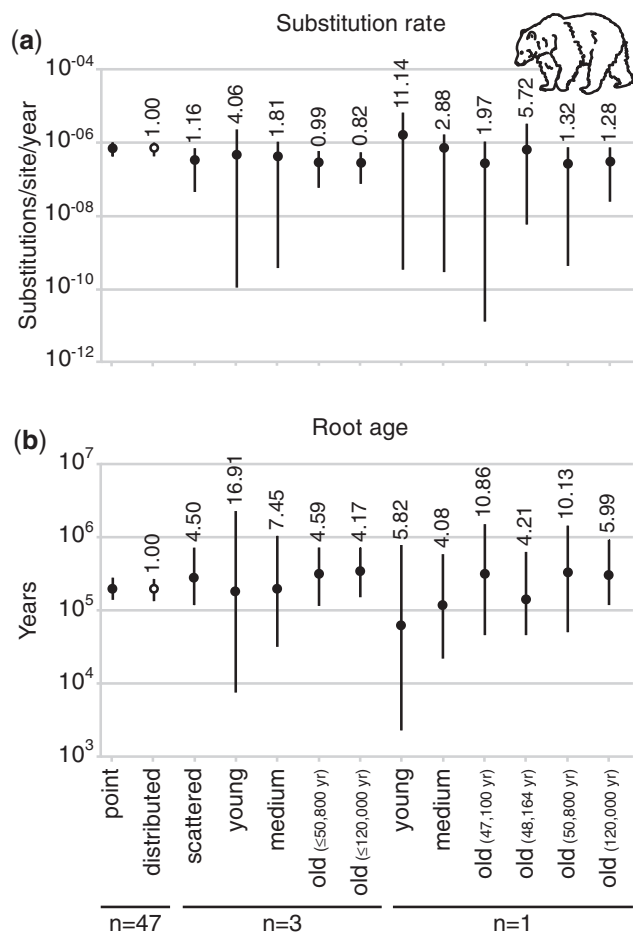
**Fig. 3.** Bayesian parameter estimates using arbitrary sample-age errors: (a) rate estimates and (b) root-age estimates for arctic fox and bison. Estimates for analyses using point calibrations (point) and published sample-age errors ("real" errors) are shown for comparison. The value above each bar indicates the relative size of the 95% HPD interval of this estimate compared with the analysis using point calibrations.

were used to model the uncertainty in sampling times. For some sequences, the posterior distributions of the sampling times differed noticeably from the specified prior distributions (supplementary fig. S3, Supplementary Material online).

### The Amount of Information Required to Calibrate the Molecular Clock

In Bayesian phylogenetic analyses of the brown bear data, changes in the number and age of ancient DNA sequences affected the performance of parameter estimation. Including

only one or three ancient sequences in the analysis led to a substantial reduction in the precision of most estimates of substitution rate (fig. 4a) and all root-age estimates (fig. 4b). However, the distribution of sequence dates had a crucial impact on the analysis. For some of the older samples (120,000 years and 50,800 years), the precision of substitution rate estimation with only one sequence was comparable to that obtained using the whole data set. Younger samples produced estimates with much wider 95% HPD intervals. Estimates using three old samples (regardless of whether



**FIG. 4.** Bayesian parameter estimates of sets of brown bear samples including a varying number of ancient samples. The graph presents the median estimates and 95% HPD intervals of (a) substitution rate and (b) root age. Analyses were calibrated using all 66 modern samples and  $n = 47$ ,  $n = 3$ , or  $n = 1$  ancient sequences. In schemes  $n = 3$  and  $n = 1$ , “old” denotes the three or one oldest sequence(s) (either including or excluding the 120,000-year sequence); “medium” denotes the three or one sequence(s) closest to 10% of the age of the oldest sample; “young” denotes the three or one sequence(s) closest to 1% of the age of the oldest sample; and “scattered” includes one sequence from each of the three age categories. Ages of sequences are given in [supplementary figure S2, Supplementary Material](#) online. For the comparison of estimation performance, estimates using all 47 ancient sequences, both with nonpoint (distributed) and point calibrations, are shown. Values above the bars indicate the relative size of the 95% HPD intervals when compared with that obtained using the complete data set (66 modern and 47 ancient) and using point calibrations for the ancient sequences.

the 120,000-year sequence was included or excluded) were of comparable precision to the one using samples from all three age categories. Root-age estimates were poor when there was a reduced number of ancient DNA sequences in the analysis.

When we sequentially pruned ancient sequences from the brown bear data set, the performance of rate estimation only started to decline when fewer than 16 of the initial 47 ancient samples remained in the analysis ([fig. 5a](#)). However, a substantial drop in estimation performance was found only when 4–5 ancient samples remained in the data set, with considerable variability among the three replicates. A similar pattern

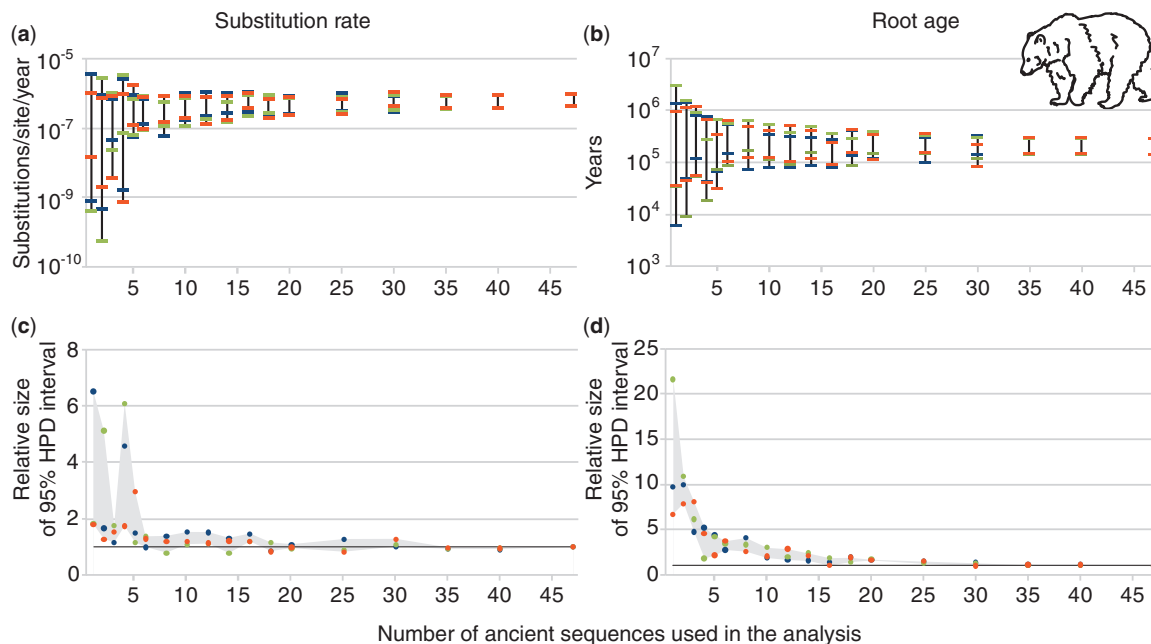
was observed for estimates of the root age, although the decline in performance was more gradual ([fig. 5b](#)).

Similar results were obtained from the simulation study ([supplementary figs. S4 and S5, Supplementary Material](#) online). The performance of the method in estimating rates and root ages did not substantially differ from the analysis of the full data set (50 ancient and 50 modern sequences) until fewer than six ancient sequences were included. This result was independent of the rate of evolution used for simulating sequence evolution. The temporal information in data sets containing fewer than five ancient samples, with a simulated rate of  $2 \times 10^{-7}$  substitutions per site per year, was in a few cases insufficient to allow parameter estimation. The simulated rate and root age fell within the 95% HPD intervals of the respective estimates in 99% and 93% cases.

## Discussion

Time-stamped DNA sequences offer a useful source of calibrating information for molecular clocks, especially for intra-specific analyses. However, age estimates of samples are not free from error. We hypothesized that this source of error might detrimentally affect the precision of timescale estimates. Our results show, however, that incorporating age uncertainty into these analyses has minimal impact on phylogenetic estimates of substitution rates and divergence times, at least for data sets comprising samples with a wide age range and small dating errors ([fig. 2](#)). The only data set to show a substantial reduction in estimation performance is the boar, which comprises young sequences (oldest sample dated at 5,400 years) with large uncertainties in the sample ages (8% on average and up to 20%). These uncertainties stem from the dating method employed; the boar samples were dated using cultural context with a broad age range for each horizon, spanning up to 1,500 years ([Watanobe et al. 2004](#)).

The minimal impact of incorporating sample-dating error was also evident when different arbitrary levels of dating error were introduced to the analysis ([fig. 3](#)). Although the artificial errors (up to 10% of the sample age) were generally higher than real sample-dating errors (on average 8% of the mid-point ages for layer dates and 3% for radiocarbon dates), they had only a negligible effect on the results. The precision of substitution rate estimates was generally reduced (wider 95% HPD intervals) when age uncertainties were incorporated; however, for the arctic fox data set, this decrease was smaller than the level of uncertainty in the prior distributions. The bison data set was more affected by introducing errors, probably owing to the much higher proportion of ancient sequences (88%) and larger temporal span (60,400 years) than in the arctic fox data set (16% and 16,000 years, respectively). The largest increase in the 95% HPD interval of the rate estimate was 23%, for a sample-dating error of  $N(0.9t, 1.1t)$ . In comparison, in our sample-pruning analysis of brown bear ([fig. 5](#)), choosing two random sets of 30 ancient sequences (from the 47 available; with 66 modern sequences) resulted in 95% HPD intervals differing by up to 27% in size. For data sets with 25 ancient sequences, there was an increase of up to 55%. These results show the effects of the choice (or, more typically, availability) of samples, which has a larger impact on



**Fig. 5.** Bayesian estimates of (a) substitution rate and (b) root age in brown bear using various numbers of ancient DNA sequences. All analyses included the 66 modern sequences. (a) and (b) show 95% HPD intervals, and (c) and (d) show the relative size of the 95% HPD interval compared with the analysis of the full data set (66 modern and 47 ancient sequences). Three replicates for each setting are shown in different shades. In (c) and (d), the shaded area indicates the difference between the highest and lowest values for each setting.

parameter estimates than incorporating sample-dating uncertainties.

When sample ages were estimated in the phylogenetic analysis, posterior mean values were not always consistent with the ages determined by  $^{14}\text{C}$  or layer dating. Some of the posterior distributions of the sample dates were skewed toward the boundaries of the prior distributions (supplementary fig. S3, Supplementary Material online). This could be caused by a number of factors, including erroneous sample dates, sequence contamination, damage-driven errors in the DNA, violation of the assumption of panmixia by the coalescent model, or inadequate modeling of among-lineage rate variation (Shapiro et al. 2011).

Converting radiocarbon years to calendar years is of interest to most research involving estimating phylogenetic timeframes. In view of the results of our analyses, the additional uncertainty introduced by converting dates might have little effect on substitution rate estimates. For example, in an analysis of 578 Late Pleistocene herbivore samples, the mean standard error for radiocarbon and calendar dates was 1.48% and 1.55%, respectively (Lorenzen et al. 2011). Minimum and maximum sample-dating errors in the data set were 0.25% and 10.75% for radiocarbon dating and 0.36% and 12.65% for the calendar ages, respectively. Therefore, we believe that our findings using radiocarbon dates can be extrapolated to analyses using calendar dates. Additional sources of uncertainty associated with dating samples, such as contamination or the choice between marine and terrestrial calibration curves, should still be taken into consideration.

A number of data set characteristics, such as the number of samples, length of alignment, and sampling time span, were

tested for relationships with the performance of phylogenetic timescale estimation. No significant relationships were found (supplementary table S2, Supplementary Material online). This might be due, in part, to the limited number of data sets included in our study but might also indicate that estimation performance is influenced by a combination of many factors (some of which might not have been taken into account in this analysis) rather than one single factor.

Our study shows that even a small number of ancient DNA sequences can produce estimates of substitution rates of comparable precision to those obtained from larger data sets, provided that the sequences are of sufficient age (figs. 4 and 5). Furthermore, provided that old sequences are included in the analysis, the age distribution of ancient samples used (i.e., whether they cluster around one time point or are of highly diverse ages) does not seem to influence the precision of substitution rate estimates; this is consistent with the findings of a previous simulation study (Ho et al. 2007). Compared with estimates of substitution rates, estimates of root age appear to be less robust to a reduction in calibration information and perform poorly with a restricted number of time-stamped sequences. Nevertheless, all our analyses (each comprising 50 or more modern samples) show that including six ancient DNA sequences is enough to calibrate the molecular clock (fig. 5 and supplementary figs. S4 and S5, Supplementary Material online). Increasing the number of sequences improves estimates of substitution rate and root age until a certain threshold, after which additional sequences do not lead to a noticeable improvement in estimates. A recent study by Dodge (2012) has also shown that a relatively small number of samples can be sufficient to estimate model parameters in a Bayesian phylogenetic framework.



However, the ages of the samples are important. In addition, the individual properties of each data set, such as the genetic variability of sequences, will affect the number of samples required for reliable estimates of substitution rates and timescales.

## Conclusion

Our study shows that incorporating sample-dating errors into phylogenetic estimates generally has a negligible impact on estimates of substitution rate and divergence times. However, high levels of sample-dating uncertainty, such as those arising from layer dating of relatively young samples, can severely decrease the performance of phylogenetic analysis. We have also shown that, to obtain accurate and precise estimates of molecular evolution rates and timescales, it is not strictly necessary to have a large data set. A modest number of samples with widely distributed and well-determined ages can be sufficient. Under these conditions, accounting for sample-dating errors might not be a critical step in the analysis.

## Supplementary Material

Supplementary tables S1 and S2 and figures S1–S5 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

The authors thank Barbara Holland and three anonymous reviewers for helpful comments and suggestions and Sebastián Duchêne for help with analyses using R. This work was supported by the University of Sydney International Scholarship to M.M., a travel grant from the School of Biological Sciences, University of Sydney, to M.M., a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme to E.D.L., the Packard Foundation to B.S., NSF ARC-0909456 to B.S., the Australian Research Council to S.Y.W.H., and start-up funds from the University of Sydney to S.Y.W.H.

## References

- Anderson CN, Ramakrishnan U, Chan YL, Hadly EA. 2005. Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics* 21:1733–1734.
- Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annu Rev Ecol Syst.* 33:707–740.
- Barnett R, Shapiro B, Barnes I, et al. (17 co-authors). 2009. Phylogeography of lions (*Panthera leo* ssp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Mol Ecol.* 18:1668–1677.
- Beavan-Athfield NR, McFadgen BG, Sparks RJ. 2001. Environmental influences on dietary carbon and C-14 ages in modern rats and other species. *Radiocarbon* 43:7–14.
- Bowman S. 1990. Radiocarbon dating. Berkeley: University of California Press. p. 38–39.
- Bronk Ramsey C. 1995. Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon* 37:425–430.
- Brown RP, Yang Z. 2011. Rate variation and estimation of divergence times using strict and relaxed clocks. *BMC Evol Biol.* 11:271.
- Campos PF, Willerslev E, Sher A, et al. (20 co-authors). 2010. Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proc Natl Acad Sci U S A.* 107:5675–5680.
- Chan YL, Anderson CN, Hadly EA. 2006. Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. *PLoS Genet.* 2:e59.
- Dalén L, Nyström V, Valdiosera C, Germonpré M, Sablin M, Turner E, Angerbjörn A, Arsuaga JL, Götherström A. 2007. Ancient DNA reveals lack of postglacial habitat tracking in the arctic fox. *Proc Natl Acad Sci U S A.* 104:6726–6729.
- de Bruyn M, Hoelzel AR, Carvalho GR, Hofreiter M. 2011. Faunal histories from Holocene ancient DNA. *Trends Ecol Evol.* 26:405–413.
- Dodge DR. 2012. A molecular approach to Neanderthal extinction. *Quatern Int.* 259:22–32.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.
- Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161:1307–1320.
- Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. 2004. Genealogies from time-stamped sequence data. In: Buck CE, Maillard AR, editors. Tools for constructing chronologies: crossing disciplinary boundaries. London: Springer-Verlag. p. 149–171.
- Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. 2003. Measurably evolving populations. *Trends Ecol Evol.* 18:481–488.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.
- Edwards SV, Beerli P. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54:1839–1854.
- Excoffier L, Novembre J, Schneider S. 2000. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *J Hered.* 91:506–509.
- Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. 2010. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol Biol Evol.* 27:2038–2051.
- Haile J, Holdaway R, Oliver K, Bunce M, Gilbert MTP, Nielsen R, Munch K, Ho SYW, Shapiro B, Willerslev E. 2007. Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor? *Mol Biol Evol.* 24:982–989.
- Hay JM, Subramanian S, Millar CD, Mohandesan E, Lambert DM. 2008. Rapid molecular evolution in a living fossil. *Trends Genet.* 24:106–109.
- Ho SYW. 2009. An examination of phylogenetic models of substitution rate variation among lineages. *Biol Lett.* 5:421–424.
- Ho SYW, Kolokotronis SO, Allaby RG. 2007. Elevated substitution rates estimated from ancient DNA sequences. *Biol Lett.* 3:702–705.
- Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Cooper A. 2011. Time-dependent rates of molecular evolution. *Mol Ecol.* 20:3087–3101.
- Ho SYW, Lanfear R, Phillips MJ, Barnes I, Thomas JA, Kolokotronis SO, Shapiro B. 2011. Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Syst Biol.* 60:366–375.
- Ho SYW, Larson G. 2006. Molecular clocks: when times are a-changin'. *Trends Genet.* 22:79–83.

- Ho SYW, Phillips MJ. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst Biol*. 58:367–380.
- Ho SYW, Saarma U, Barnett R, Haile J, Shapiro B. 2008. The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS One* 3:e1615.
- Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol*. 6:29.
- Korsten M, Ho SYW, Davison J, et al. (24 co-authors). 2009. Sudden expansion of a single brown bear maternal lineage across northern continental Eurasia after the last ice age: a general demographic model for mammals? *Mol Ecol*. 18:1963–1979.
- Lee MSY, Oliver PM, Hutchinson MN. 2009. Phylogenetic uncertainty and molecular clock calibrations: a case study of legless lizards (Pygopodidae, Gekkota). *Mol Phylogenet Evol*. 50:661–666.
- Lindqvist C, Schuster SC, Sun Y, et al. (14 co-authors). 2010. Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proc Natl Acad Sci U S A*. 107:5053–5057.
- Lorenzen ED, Nogues-Bravo D, Orlando L, et al. (55 co-authors). 2011. Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* 479:359–364.
- Luo A, Qiao H, Zhang Y, Shi W, Ho SYW, Xu W, Zhang A, Zhu C. 2010. Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. *BMC Evol Biol*. 10:242.
- Marko PB. 2002. Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. *Mol Biol Evol*. 19:2005–2021.
- Mellars P. 2006. A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature* 439:931–935.
- R Development Core Team. 2012. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: <http://www.R-project.org>, last accessed May 2, 2012.
- Ramakrishnan U, Hadly EA. 2009. Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Mol Ecol*. 18:1310–1330.
- Rambaut A. 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* 16:395–399.
- Rambaut A, Drummond AJ. 2007. Tracer v1.5. Available from: <http://beast.bio.ed.ac.uk/Tracer>, last accessed May 2, 2011.
- Ramsden C, Holmes EC, Charleston MA. 2009. Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. *Mol Biol Evol*. 26:143–153.
- Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19:301–302.
- Shapiro B, Drummond AJ, Rambaut A, et al. (27 co-authors). 2004. Rise and fall of the Beringian steppe bison. *Science* 306:1561–1565.
- Shapiro B, Ho SYW, Drummond AJ, Suchard MA, Pybus OG, Rambaut A. 2011. A Bayesian phylogenetic method to estimate unknown sequence ages. *Mol Biol Evol*. 28:879–887.
- Stuiver M, Polach HA. 1977. Reporting of C-14 data. *Radiocarbon* 19:355–363.
- Suchard MA, Weiss RE, Sinsheimer JS. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol*. 18:1001–1013.
- Svensson A, Andersen KK, Bigler M, et al. (14 co-authors). 2008. A 60,000-year Greenland stratigraphic ice core chronology. *Clim Past*. 4:47–57.
- Törnqvist TE, De Jong AFM, Oosterbaan WA, Van der Borg K. 1992. Accurate dating of organic deposits by AMS C-14 measurement of macrofossils. *Radiocarbon* 34:566–577.
- Watanobe T, Ishiguro N, Nakano M, Matsui A, Hongo H, Yamazaki K, Takahashi O. 2004. Prehistoric Sado Island populations of *Sus. scrofa* distinguished from contemporary Japanese wild boar by ancient mitochondrial DNA. *Zool Sci*. 21:219–228.
- Watanobe T, Ishiguro N, Okumura N, Nakano M, Matsui A, Hongo H, Ushiro H. 2001. Ancient mitochondrial DNA reveals the origin of *Sus. scrofa* from Rebun Island, Japan. *J Mol Evol*. 52:281–289.
- Wertheim JO. 2010. The re-emergence of H1N1 influenza virus in 1977: a cautionary tale for estimating divergence times using biologically unrealistic sampling dates. *PLoS One* 5:e11184.
- Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.