

Bayesian Estimation of Substitution Rates from Ancient DNA Sequences with Low Information Content

SIMON Y. W. HO^{1,2,*}, ROBERT LANFEAR¹, MATTHEW J. PHILLIPS¹, IAN BARNES³,
JESSICA A. THOMAS^{1,3}, SERGIOS-ORESTIS KOLOKOTRONIS⁴, AND BETH SHAPIRO⁵

¹Centre for Macroevolution and Macroecology, Research School of Biology, Australian National University, Canberra, ACT 0200, Australia;

²School of Biological Sciences, University of Sydney, Sydney, NSW 2006, Australia;

³School of Biological Sciences, Royal Holloway University of London, Egham, Surrey TW20 0EX, UK;

⁴Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY 10024, USA; and

⁵Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA;

*Correspondence to be sent to: School of Biological Sciences, Macleay Building A12, University of Sydney, Sydney, NSW 2006, Australia; E-mail: simon.ho@sydney.edu.au.

Received 17 September 2009; reviews returned 28 February 2010; accepted 3 November 2010

Associate Editor: Kelly Zamudio

Rates of molecular evolution have been shown to vary significantly among nucleotide sites, loci, and taxa. In addition to these forms of rate heterogeneity, there is evidence that molecular rates vary with the timescale over which they are estimated. One of the most striking observations has been that of elevated mutation rates over very short timescales, such as those presented in studies of pedigrees (e.g., [Howell et al. 2003](#); [Millar et al. 2008](#)) and mutation accumulation lines (e.g., [Denver et al. 2000](#); [Haag-Liautard et al. 2008](#)). In contrast, much lower rates are observed over evolutionary timescales, as estimated in phylogenetic analyses calibrated with reference to paleontological or geological data.

The disparity between rates of spontaneous mutation and evolutionary substitution can exceed an order of magnitude. Intermediate rates are expected between these two ends of the spectrum, but there has been disagreement over the exact form of the decline from the mutation rate to the substitution rate. Some authors have suggested that elevated mutation rates are very short-lived, perhaps persisting for only a small number of generations ([Macaulay et al. 1997](#); [Gibbons 1998](#)). More recently, it was proposed that the estimated rate decays exponentially over tens to hundreds of thousands of years, producing a “time dependence” of rates, whereby the magnitude of the inferred rate depends on the age of the calibration used in the analysis ([Ho et al. 2005, 2007c](#); [Penny 2005](#); [Ho and Larson 2006](#)). Although some of the original evidence for this hypothesis has been challenged ([Emerson 2007](#); [Bandelt 2008](#)), there has been a steady accumulation of empirical and theoretical support for a prolonged elevation of short-term rates (e.g., [Genner et al. 2007](#); [BurrIDGE et al. 2008](#); [Henn et al. 2009](#); [Peterson and Masel 2009](#); [Soares et al. 2009](#)). This has included compelling evidence from analyses of ancient DNA (aDNA) in which the sampling times of the heterochronous sequences are able to provide cali-

brating information for estimating rates (e.g., [Lambert et al. 2002](#); [Barnes et al. 2007](#); [Ho et al. 2007b](#); [Hay et al. 2008](#); [Subramanian et al. 2009a](#)).

In a recent critique, [Debruyne and Poinar \(2009\)](#) have claimed that the high rate estimates obtained in Bayesian analyses of aDNA data are an unintended consequence of analyzing short sequences. According to their “signal-dependent artifact” hypothesis, aDNA-based rate estimates depend almost entirely on the information content in the sequence alignment. The essence of the criticism is that the posterior distribution of the rate becomes so wide that the posterior mean becomes an upwardly biased estimator. This behavior has been noted in previous studies of heterochronous data with low information content (e.g., [Ho et al. 2007c](#); [Firth et al. 2010](#)). However, [Debruyne and Poinar](#) go on to state that “the [rate] acceleration phenomenon is certainly of much lower magnitude than has been previously reported by [Ho et al. \(2005\)](#)” (p. 358). This is a misleading comparison because our study was based almost exclusively on analyses of modern DNA (isochronous sequences) using internal-node calibrations which, as argued by [Debruyne and Poinar](#), are able to overcome the signal-dependent artifact. In fact, much of the evidence for time-dependent rates has come from analyses of isochronous data (e.g., [Genner et al. 2007](#); [BurrIDGE et al. 2008](#); [Henn et al. 2009](#); [Soares et al. 2009](#); [Papadopoulou et al. 2010](#)).

Resolving the concerns over Bayesian rate estimates from aDNA is important for several reasons. First, aDNA sequences typically range in age from 10² to 10⁵ years, thereby filling a crucial calibration gap between the time periods covered by pedigrees (usually < 10² years) and fossil-calibrated species phylogenies (usually > 10⁶ years). Second, because the ages of aDNA sequences can provide sufficient calibrating information

for estimating rates (Drummond et al. 2003), these data make it possible to circumvent problems associated with choosing and implementing calibrations at internal nodes (Emerson 2007; Ho and Phillips 2009; Firth et al. 2010). In particular, terminal-node calibrations remove the need for assumptions about genetic divergence being correlated with the divergence of species or populations, which can be dubious because of the uncertainty posed by ancestral polymorphism (Charlesworth et al. 2005; Peterson and Masel 2009). Consequently, if rates can be accurately estimated from aDNA data, some insight can be gained into the underlying causes of time-dependent rates (Ho et al. 2007c).

There is still uncertainty regarding the factors driving the time dependence of rates. Previous studies have considered the possibility of contributions from incomplete purifying selection, calibration error, sequencing error, aDNA damage, ancestral polymorphism, saturation, and model misspecification, among others (Ho et al. 2005, 2007c; Woodhams 2006; Henn et al. 2009; Loogväli et al. 2009; Peterson and Masel 2009; Soares et al. 2009; Subramanian et al. 2009a). Debruyne and Poinar have added to this list with their suggestion that mean posterior rate estimates are upwardly biased for data sets with low information content. Distinguishing among these various factors is crucial to future studies of recent divergence times, evolutionary rates, and the molecular evolutionary process in general.

Below, we investigate the two major aspects of the critique by Debruyne and Poinar. The first of these is that the posterior mean provides a biased measure of the rate in Bayesian analyses of data sets with low information content. To examine this issue, we perform new analyses of sequence data simulated using known evolutionary parameters. We assess the relationship between sequence variation and estimated rate under a range of simulation conditions, including various rates and sequence lengths.

The second major aspect of the critique by Debruyne and Poinar is that the artifactual rate estimates from aDNA data are governed by the information content of the alignments, as measured by the number of variable sites. Indeed, Debruyne and Poinar base their entire signal-dependence model on analyses of alignments of varying length, which they regard as a suitable proxy for information content. Although this might be appropriate for isochronous data, we argue that it does not provide the full picture for heterochronous data because the ages of the tips represent a crucial part of the phylogenetic and temporal signal. We propose that the amount of information contained in these ages depends on their structure and spread, including the length of the sampling interval in relation to the period spanned by the genealogy of the sequences (Drummond et al. 2003; Firth et al. 2010). To investigate this, we perform new analyses of 18 published aDNA alignments to assess whether the ages of the sequences in these data sets provide sufficient calibrating information for estimating rates. The results of these analyses show that most real

data sets appear to have satisfactory temporal structure and signal.

The results of our new analyses indicate that the “signal dependence” hypothesis has limited relevance to the majority of real aDNA data sets. Our results also suggest that the signal dependence cannot be regarded as an analogue to time dependence, unless one is willing to accept the validity of equating alignment length with temporal depth in aDNA data. Moreover, our results highlight the importance of other factors, including the distribution of sampling times, choice of population size prior, and the use of appropriate summary statistics in analyses of heterochronous sequence data that exhibit low variation.

NEW ANALYSES IN RESPONSE TO DEBRUYNE AND POINAR

Here, we build upon a simulation study that was presented in one of our previous evaluations of Bayesian rate estimation using aDNA data (Ho et al. 2007b). Debruyne and Poinar have challenged the results of this study, criticising two aspects of our analyses. First, they argue that the rates estimated from the simulated data are more precise than those obtained from real aDNA data. Although this observation is correct, these results are an expected consequence of simulation-based analysis: the evolutionary models for nucleotide substitution and demographic history used in the analysis of the simulated data are chosen to match the conditions under which the data were generated. This is adopted as standard practice to make it easier to isolate the effects of the factor(s) of interest.

The second criticism of the simulation study of Ho et al. (2007b) is that the substitution rate used in the simulations is too high, with Debruyne and Poinar stating that the rate is “25-fold the estimate of the substitution rate for the mt genome of vertebrates” (p. 350). However, this simulation rate was inspired by published estimates from the mitochondrial D-loop (Lambert et al. 2002; Shapiro et al. 2004), whereas Debruyne and Poinar compare this rate to that estimated from their elephantid data, which is based on whole mitochondrial genomes analyzed over a phylogenetic timeframe. Indeed, the vast majority of published aDNA data sets comprise sequences from the D-loop, which exhibits much higher mutation and substitution rates than does the rest of the mitochondrial genome in vertebrates. This also calls into question the design of the main analysis presented in their critique, in which subsamples from the complete mitochondrial genomes of woolly mammoths were taken to be representative of real aDNA data sets.

Nevertheless, the high rate used in our simulation could be viewed as a legitimate problem if short-term rates were not actually elevated. This led Debruyne and Poinar to pose the question: “what would the accuracy and precision of the posterior rate of change be if a slower rate of substitution, in the range of the interspecific mitochondrial substitution rates (between 1

and 2×10^{-8} substitutions/site/year) were applied to simulate the same sequence data?" (p. 350). In response to this question, and to address some of their other concerns, we present the results of a detailed simulation study below.

Simulation Study

We conducted analyses of simulated aDNA data to investigate the performance of Bayesian rate estimation. The amount of rate estimation bias is quantified under various combinations of simulation rate and sequence length, including conditions that might match those commonly encountered in real aDNA research. We investigate the impact of varying the population-size prior, and we compare the performance of different posterior measures of the rate.

Materials and methods.—Sequence evolution was simulated using Seq-Gen (Rambaut and Grassly 1997) on random trees generated according to a coalescent model with a constant population size of 10^5 . Each simulated data set comprised 31 time-stamped, nonrecombining sequences, with ages of 0, 1000, 2000, ..., 30,000 years. All sequences were generated according to the Jukes–Cantor model of nucleotide substitution (Jukes and Cantor 1969), with rate homogeneity among sites and among branches. Simulations were performed with 3 different substitution rates (1×10^{-8} substitutions/site/year, 5×10^{-8} substitutions/site/year, and 1×10^{-7} substitutions/site/year) and 2 sequence lengths (100 and 1000 bp), representing the range of characteristics of typical aDNA data sets and encompassing conditions expected to generate sequence alignments with low information content. One thousand replicate data sets were generated for each combination of sequence length and rate. Apart from the substitution rate and sequence length, the simulations are identical to those described in the “uniform sampling regime” in our previous study (Ho et al. 2007b).

Substitution rates were estimated from the simulated data sets using the Bayesian phylogenetic software BEAST 1.4.8 (Drummond and Rambaut 2007). To match the simulation conditions, the Jukes–Cantor substitution model was assumed and a constant-size coalescent prior was chosen for the tree. A uniform prior of $[0, \infty)$ was chosen for the substitution rate. Posterior distributions of parameters were obtained by Markov chain Monte Carlo (MCMC) sampling, with samples drawn every 500 steps over a total of 2×10^7 steps, with the first 10% of samples discarded as burn-in. To compare different posterior measures of the substitution rate, the mean, median, and mode of the posterior rate distribution were calculated for each analysis. Effective sample sizes of parameters were examined to check for acceptable MCMC mixing and sufficient sampling from the posterior.

For any given data set, the estimates of rate and population size are closely tied. The population size prior can be influential in the estimation of rates, particularly

when the data set is relatively uninformative. We investigated this issue by performing three sets of analyses, differing only in the population size prior: 1) population size fixed to its true (simulation) value of 10^5 ; 2) population size given a uniform prior of $[0, \infty)$; and 3) population size given a uniform prior of $[10^0, 10^9]$, representing a range of values that could be considered biologically plausible for vertebrates. Note that in all these analyses, “population size” is actually given as $N_e\tau$, the product of the effective population size (N_e) and generation time in years (τ).

Results.—The performance of rate estimation varied considerably among the three sets of simulations, providing a strong indication of the influence of the population size prior (Table 1). When the population size is fixed to its true (simulation) value of 10^5 , estimates of rates are accurate and precise. The 95% highest posterior density (HPD) interval of the substitution rate included the simulation value at least 95% of the time. As noted by Debruyne and Poinar, the mean posterior rate estimates reveal that there is considerable overestimation of the rate when there is low information content or little sequence variability in the data set (low substitution rate and/or short sequence length). However, this bias disappears in the more informative data sets. As the posterior rate distributions are leptokurtic, the medians are less biased than the means. The posterior mode, which represents the maximum *a posteriori* estimate of the rate, appears to provide an unbiased measure across all combinations of substitution rate and sequence length.

A different pattern emerges when the population size is given an unbounded uniform prior distribution (Table 1). Many of the MCMC analyses failed to converge, yielding posterior samples with effective sample sizes not exceeding 100 and with the population size tending toward infinity and the rate tending toward zero. The percentage of analyses that failed to converge ranged from 10.2% to 99.2% across the 6 simulation settings (Fig. 1). If these problematic replicates are removed, the remaining replicates appear to yield reasonable estimates of the substitution rate (Table 1). The 95% HPD interval of the rate, although considerably wider than when the population size was fixed to its correct value, included the simulation value at least 96% of the time. Plausible, unimodal estimates of the population size were obtained in the MCMC analyses that showed signs of convergence. However, in almost all the simulation settings, the rate was overestimated by the mean, median, and mode. This could be a direct consequence of removing the replicates that produced unconverged MCMC analyses because those would have been the data sets with stochastically lower information content (i.e., driven by a smaller number of substitutions and thus producing lower rate estimates). Taking this into consideration, it is difficult to establish whether the estimation bias is genuine or whether it results from taking a biased sample of the simulation replicates.

TABLE 1. Summary of results from the simulation study, averaged across 1000 replicates. For the simulations with a population size prior of Uniform[0,∞), results were only summarized from the replicates that exhibited acceptable MCMC convergence. Further details are given in the text

Prior on population size	True rate (substitutions/site/year)	Length (bp)	Posterior rate estimate (substitutions/site/year)			Mean size of 95% HPD interval (substitutions/site/year)	95% HPD coverage ^a
			Mean	Median	Mode		
Fixed to 10 ⁵	1.00 × 10 ⁻⁸	100	2.32 × 10 ⁻⁸	1.87 × 10 ⁻⁸	1.05 × 10 ⁻⁸	5.63 × 10 ⁻⁸	0.98
Fixed to 10 ⁵	1.00 × 10 ⁻⁸	1000	1.20 × 10 ⁻⁸	1.14 × 10 ⁻⁸	1.01 × 10 ⁻⁸	1.67 × 10 ⁻⁸	0.96
Fixed to 10 ⁵	5.00 × 10 ⁻⁸	100	6.74 × 10 ⁻⁸	6.17 × 10 ⁻⁸	5.10 × 10 ⁻⁸	1.15 × 10 ⁻⁷	0.96
Fixed to 10 ⁵	5.00 × 10 ⁻⁸	1000	5.31 × 10 ⁻⁸	5.20 × 10 ⁻⁸	4.97 × 10 ⁻⁸	4.51 × 10 ⁻⁸	0.96
Fixed to 10 ⁵	1.00 × 10 ⁻⁷	100	1.20 × 10 ⁻⁷	1.13 × 10 ⁻⁷	1.00 × 10 ⁻⁷	1.67 × 10 ⁻⁷	0.97
Fixed to 10 ⁵	1.00 × 10 ⁻⁷	1000	1.04 × 10 ⁻⁷	1.03 × 10 ⁻⁷	9.99 × 10 ⁻⁸	7.18 × 10 ⁻⁸	0.95
Uniform[0,∞)	1.00 × 10 ⁻⁸	100	1.68 × 10 ⁻⁷	1.21 × 10 ⁻⁷	3.57 × 10 ⁻⁸	4.80 × 10 ⁻⁷	1.00
Uniform[0,∞)	1.00 × 10 ⁻⁸	1000	2.82 × 10 ⁻⁸	2.56 × 10 ⁻⁸	2.08 × 10 ⁻⁸	5.49 × 10 ⁻⁸	0.97
Uniform[0,∞)	5.00 × 10 ⁻⁸	100	1.92 × 10 ⁻⁷	1.65 × 10 ⁻⁷	1.07 × 10 ⁻⁷	4.32 × 10 ⁻⁷	0.96
Uniform[0,∞)	5.00 × 10 ⁻⁸	1000	5.89 × 10 ⁻⁸	5.71 × 10 ⁻⁸	5.35 × 10 ⁻⁸	8.10 × 10 ⁻⁸	0.98
Uniform[0,∞)	1.00 × 10 ⁻⁷	100	2.66 × 10 ⁻⁷	2.40 × 10 ⁻⁷	1.86 × 10 ⁻⁷	5.28 × 10 ⁻⁷	0.97
Uniform[0,∞)	1.00 × 10 ⁻⁷	1000	1.03 × 10 ⁻⁷	1.01 × 10 ⁻⁷	9.78 × 10 ⁻⁸	1.10 × 10 ⁻⁷	0.97
Uniform[10 ⁰ ,10 ⁹]	1.00 × 10 ⁻⁸	100	3.53 × 10 ⁻⁸	7.34 × 10 ⁻⁹	8.31 × 10 ⁻⁹	1.68 × 10 ⁻⁷	1.00
Uniform[10 ⁰ ,10 ⁹]	1.00 × 10 ⁻⁸	1000	9.07 × 10 ⁻⁹	6.71 × 10 ⁻⁹	2.48 × 10 ⁻⁹	2.52 × 10 ⁻⁸	0.81
Uniform[10 ⁰ ,10 ⁹]	5.00 × 10 ⁻⁸	100	5.20 × 10 ⁻⁸	2.75 × 10 ⁻⁸	8.48 × 10 ⁻⁹	1.84 × 10 ⁻⁷	0.83
Uniform[10 ⁰ ,10 ⁹]	5.00 × 10 ⁻⁸	1000	4.66 × 10 ⁻⁸	4.48 × 10 ⁻⁸	3.53 × 10 ⁻⁸	7.26 × 10 ⁻⁸	0.87
Uniform[10 ⁰ ,10 ⁹]	1.00 × 10 ⁻⁷	100	8.44 × 10 ⁻⁸	5.80 × 10 ⁻⁸	2.18 × 10 ⁻⁸	2.49 × 10 ⁻⁷	0.78
Uniform[10 ⁰ ,10 ⁹]	1.00 × 10 ⁻⁷	1000	9.79 × 10 ⁻⁸	9.63 × 10 ⁻⁸	9.11 × 10 ⁻⁸	1.09 × 10 ⁻⁷	0.91

^aProportion of simulations in which the 95% HPD interval of the rate contained the true (simulation) value.

When the population size is constrained to a range of biologically plausible values (10⁰–10⁹), yet another picture materialises. Coverage by the 95% HPD intervals was poorer, with the simulation value being excluded from the 95% HPD interval up to 22% of the time (Table 1). The mean size of the 95% HPD interval is smaller than in the analyses without any restrictions on the population size, although the disparity disappears as the number of variable sites in the alignment increases. The posterior mode is no longer the best summary of the rate, probably because the constraints on population size also impose restrictions on the values that can be taken by the substitution rate. In some cases, the posterior distribution of the rate is implicitly constrained, leading to a distorted mode. On the other hand, the posterior mean appears to provide a reasonably accurate estimate of the true substitution rate (Table 1), although it is possible that this is partly an unintended consequence of the population size constraints. That is, the mean posterior rate might only be accurate as a result of the population size priors constraining the substitution rate to reasonable values, even in the absence of real information on rates in the data. This effect could potentially explain some of the published rate estimates from uninformative aDNA sequence alignments, which have taken seemingly plausible values in spite of the low information content of the data.

aDNA Data Sets

Published aDNA data sets vary considerably in terms of their sequence lengths and underlying substitution rates as well as the temporal structure and spread of the

samples. It would be useful to evaluate the information content in these data sets to determine whether they can produce reliable estimates of substitution rates and divergence times. One significant facet of heterochronous data that is overlooked by the use of diversity statistics (Depaulis et al. 2009), and in the analyses of information content performed by Debruyne and Poinar, is that the ages of the sequences form an important component of the information content (e.g., Firth et al. 2010). This stems from the fact that the sequence ages are used for calibrating estimates of substitution rates. A potential problem in analyses of heterochronous data is that rate estimates could be an artifact of the sampling ages.

Here, we use a date randomization test to investigate temporal structure in 18 published aDNA data sets. This test involves reanalyzing each data set after randomly shuffling the ages of the sequences and follows several previous studies of heterochronous data (de Bruyn et al. 2009; Miller et al. 2009; Subramanian et al. 2009b; Firth et al. 2010). The date randomization analysis is able to provide some insight into whether the structure and spread of the sequence ages are sufficient to provide reliable information on the rate underlying the evolution of the data set. If the original rate estimate is recovered in the date-randomized data sets, then there is insufficient temporal structure in the original data set and the rate estimate cannot be supported (Firth et al. 2010).

Materials and methods.—Using the Bayesian phylogenetic method implemented in BEAST v1.5.4 (Drummond and Rambaut 2007), we analyzed 18 published aDNA alignments: 16 of the 19 aDNA data sets analyzed by Ho et al. (2007b), the 11 mitogenome alignment of woolly mammoths examined by Debruyne and Poinar, and

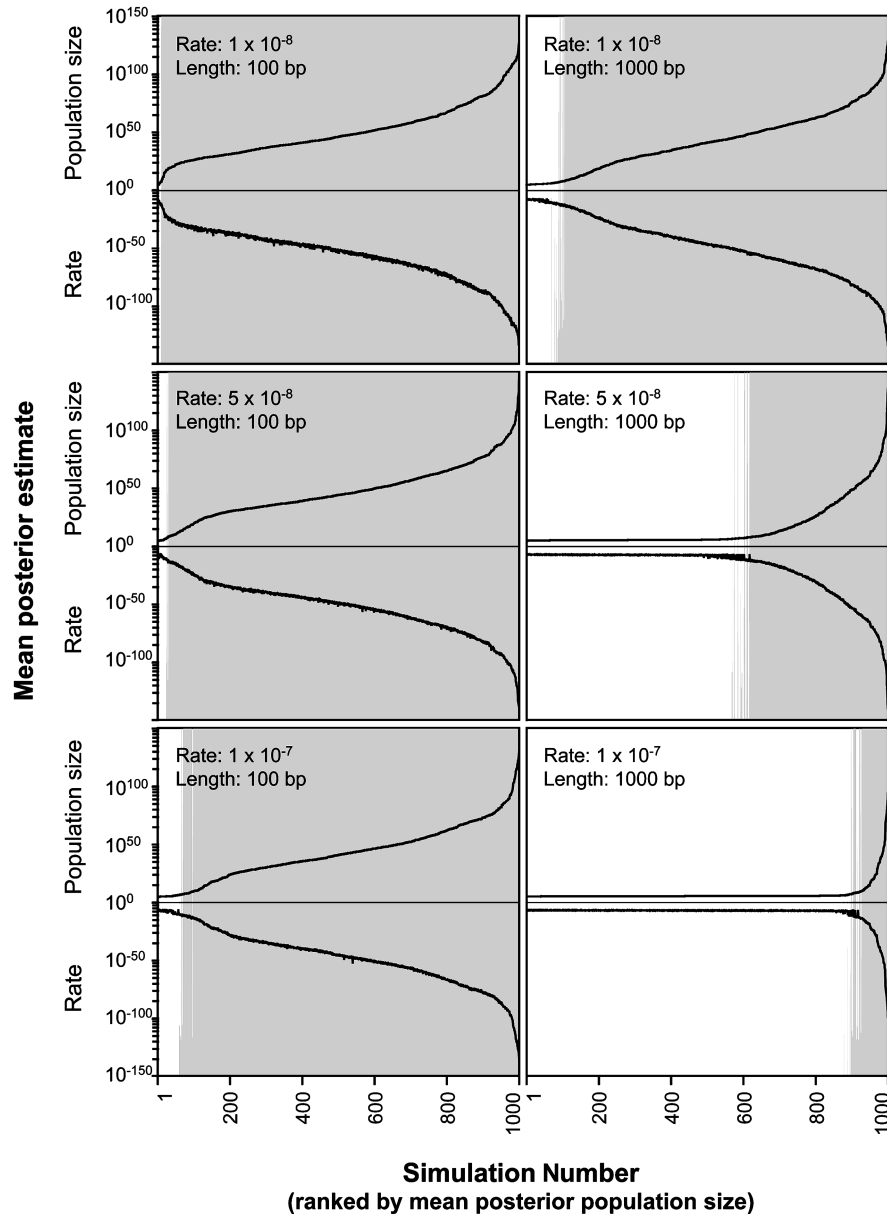


FIGURE 1. Graphs showing the correspondences between mean posterior population size, mean posterior rate, and MCMC convergence for Bayesian analyses of data generated under 6 different simulation conditions (3 different rates and 2 different sequence lengths). The results were obtained using an uninformative population size prior (uniform from 0 to ∞). Each panel shows the results from analyzing 1000 replicates, ranked from left to right by ascending mean posterior population size (top curve). The mean posterior rate estimate for the corresponding data set is also displayed on the same scale (lower curve), showing a close relationship with the estimated population size. Each simulation is given a gray vertical line in the background if the effective sample size for the posterior likelihood is below 100, which suggests a lack of convergence to the stationary distribution. For each MCMC analysis, samples were drawn from the posterior every 500 steps over a total of 2×10^7 steps, with the first 10% of samples discarded as burn-in.

a muskox D-loop alignment (Campos et al. 2010). We excluded three data sets from the study by Ho et al. (2007b): the *Chlorobium* and nene alignments contained too few ancient sequences for the randomization test, whereas the muskox alignment is superseded by the larger data set published by Campos et al. (2010). The basic characteristics of the 18 data sets are outlined in Table 2, with further details available in the original publications.

Substitution models were selected by comparison of Bayesian information criterion scores, with the number of aligned sites taken as the sample size for the penalty term. Owing to the intraspecific nature of the data sets, models that allowed a proportion of invariable sites were excluded. All data sets were treated as unpartitioned, and a constant-size coalescent prior was specified for the topology and divergence times. All analyses were repeated using a Bayesian skyride

TABLE 2. Details of aDNA alignments analyzed using the date randomization test described in the text

Species		Region	Sequences (ancient + modern)	Age range ^a (years)	Length (bp)	Variable sites	Result of date randomization test
Adélie penguin	<i>Pygoscelis adeliae</i>	D-loop	96 + 380	6424	347	159	✓
Arctic fox	<i>Alopex lagopus</i>	D-loop	8 + 41	16,000	291	23	✓
Aurochs	<i>Bos primigenius</i>	D-loop	41 + 0	10,300	360	34	✓
Bison	<i>Bison priscus</i>	D-loop	150 + 32	60,400	615	170	✓
Boar	<i>Sus scrofa</i>	D-loop	81 + 7	5400	572	47	✓
Bowhead whale	<i>Balaena mysticetus</i>	D-loop	99 + 68	51,000	453	72	Fail
Brown bear	<i>Ursus arctos</i>	D-loop	36 + 57	59,000	193	69	✓
Cave bear	<i>Ursus spelaeus</i>	D-loop	26 + 0	53,470	288	31	Fail
Cave hyaena	<i>Crocota crocuta spelaea</i>	D-loop	10 + 0	13,140	366	27	Fail
Cave lion	<i>Panthera leo spelaea</i>	D-loop	23 + 0	46,275	213	12	✓
Cow	<i>Bos taurus</i>	D-loop	36 + 91	8065	410	65	✓
Horse	<i>Equus caballus</i>	D-loop	12 + 33	28,340	348	70	✓
Maize	<i>Zea mays</i>	<i>adh2</i>	9 + 11	4500	190	26	Fail
Moa	<i>Pachyornis mappini</i>	D-loop	14 + 0	4912	241	20	Fail
Muskox	<i>Ovibos moschatus</i>	D-loop	114 + 16	45,740	682	203	✓
Tuco-tuco	<i>Ctenomys sociabilis</i>	<i>cytb</i>	45 + 1	10,208	253	13	✓
Woolly mammoth	<i>Mammuthus primigenius</i>	D-loop	32 + 0	35,970	741	42	Fail
Woolly mammoth	<i>Mammuthus primigenius</i>	Mitogenome	11 + 0	38,030	16,484	112	Fail

^aAge of oldest sequence minus age of youngest sequence.

demographic model (Minin et al. 2008). The better demographic model (constant size or Bayesian skyride) was chosen on the basis of visual inspection of the results. In each analysis, samples from the posterior were drawn every 5×10^3 steps from a total of 5×10^7 steps, with the first 10% being discarded as burn-in. Where necessary, the number of MCMC steps was doubled or tripled in order to achieve an effective sample size >100 for the rate estimate.

The sequence ages in each of the 18 aDNA data sets were then randomly reassigned. This randomization was performed 20 times for each data set using the Java application SiteSampler v1.1 (Ho and Lanfear 2010). Bayesian phylogenetic analyses were performed using the same settings as described above for the original data. For each date-randomized data set, the demographic model was chosen to match that selected for the original data.

Results.—The posterior rate estimates from the 18 data sets are shown in Figure 2. It is interesting to note that among the 7 data sets that failed the date randomization test not all produced rate estimates with wide 95% HPD intervals. In these cases, the modal posterior rate was similar to the mean posterior rate (results not shown).

To investigate the potential presence of signal-dependent biases in these estimates, we considered the mean posterior rates in relation to the characteristics of the data sets from which they were estimated. Debruyne and Poinar hypothesize that the mean posterior rate estimate should be exponentially related to the amount of information in the data set, as reflected by the alignment length. We examined 4 measures of information content: the number of aligned sites, the number of variable sites, the number of sequences, and the product of the number of sites and sequences in the alignment. Excluding the mitogenome alignment of woolly mammoths, which represents an outlier and is

nonindependent of the D-loop alignment from the same species, we find no evidence that any of these measures are related to the mean posterior rate estimate in the remaining 17 aDNA data sets ($r^2 < 0.1$ and $P > 0.2$ in all cases). However, more than 40% of the variation in rate estimates could be explained by an exponential relationship with the age range of the sequences in each data set ($r^2 = 0.431$ and $P = 0.004$).

Further insight into the temporal structure within the data sets was gained through the date randomization analyses. Eleven alignments passed the randomization test and seven failed (Fig. 2; Table 2). In addition to the results presented in this study, previous date randomization analyses of aDNA from tuatara (Subramanian et al. 2009b) and elephant seals (de Bruyn et al. 2009) have indicated that these two data sets contain sufficient temporal information to produce meaningful estimates of substitution rates. Among the data sets that failed the date randomization test, the bowhead whale alignment is noted for its low sequence diversity, with the observed variation dominated by singleton mutations (Borge et al. 2007). The maize alignment is a small data set comprising sequences sampled over a short time frame (Freitas et al. 2003). Notably, both of the mammoth alignments (D-loop and complete mitochondrial genome) failed the date randomization test.

DISCUSSION

Our analyses of simulated and real data show that the signal-dependent artifact highlighted by Debruyne and Poinar is unlikely to have contributed substantially to the published rate estimates from aDNA data sets. Our simulated data sets cover a range of sequence lengths and substitution rates, including those seen in real aDNA alignments. Regardless of the prior on population size, the posterior mean provides an unbiased estimate of the rate for the 1000 bp data sets simulated

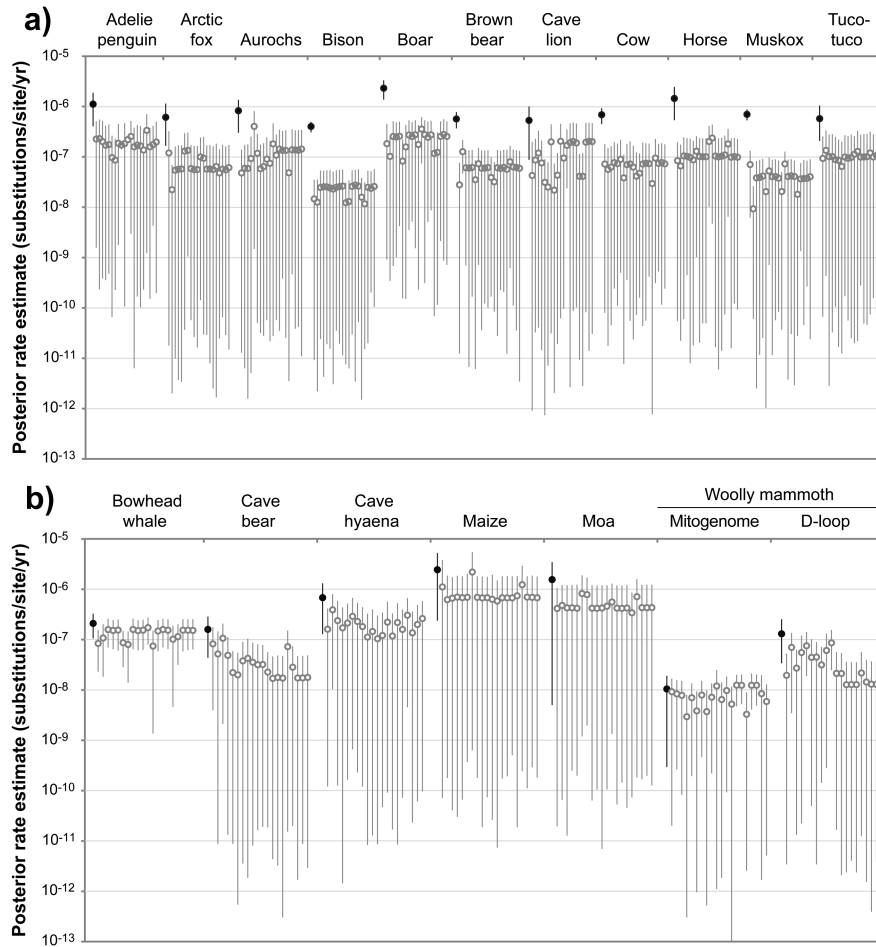


FIGURE 2. Estimates of substitution rates from a variety of aDNA alignments. For each data set, the first data point represents the rate estimated from the original data set (filled circles), whereas the remaining 20 data points (unfilled circles) represent the rates estimated from replicates in which the ages of the tips were randomly shuffled. Alignments were deemed to “pass” the date-randomization test if the mean posterior rate estimate from the original data set is not included in any of the 95% HPD intervals from the date-randomized replicates. a) Rate estimates from alignments that passed the date randomization test. b) Rate estimates from alignments that failed the date randomization test.

using a rate of 1×10^{-7} substitutions/site/year. These parameters are broadly similar to those of the mitochondrial D-loop in vertebrates. For the less informative alignments investigated here, including those simulated using lower rates, there is some degree of estimation bias unless the population size is fixed to its simulation value. However, the rates used for the simulations in this study are conservatively low because they are based on phylogenetic estimates. If short-term rates are actually elevated, as posited by the hypothesis of time-dependent rates, then the particular estimation biases observed in this study might be irrelevant to the majority of real aDNA alignments.

The results of our simulation analyses confirm that the posterior mean can be a biased measure of the substitution rate, as indicated by Debruyne and Poinar. However, the posterior estimates provide acceptable coverage because the 95% HPD intervals on the rates usually included the simulation value. The posterior mode, which is equivalent to the maximum *a posteriori* estimate of the rate, appears to be the best mea-

sure when the data set has low information content. Nevertheless, it can become distorted when a strongly bounded, informative prior is specified for the population size (or, presumably, for the substitution rate or age of the root). In view of these results, it might be most appropriate to report various summaries of the posterior distribution of rates and other parameters of interest. The three measures examined here, the mean, median, and mode, converge to the same value for the most informative data sets.

Contrary to the claim by Debruyne and Poinar, our analyses suggest that sequence variability is not the sole factor determining the performance of rate estimation. Other factors, such as the ages of the sequences, and probably the structure of the underlying genealogy, are also very important features of any aDNA data set. Evidently, the population size prior is highly influential in some of the analyses performed here. From a practical viewpoint, it is usually more feasible to use an informative prior for the age of the root rather than the population size. This is because effective population

size and generation time are often difficult to estimate reliably, whereas the age of the root can sometimes be inferred from independent palaeontological or biogeographic sources.

Taking into account the results of all the date randomization analyses, the sampling ages of seven real aDNA data sets (bowhead whale, cave bear, cave hyaena, maize, moa, woolly mammoth D-loop, and woolly mammoth mitogenomes) were found to produce artifactual rate estimates using the date randomization test. This result suggests that some of the published aDNA alignments do not contain sufficient temporal information to support reliable estimation of rates and timescales. Randomization of sequence ages represents a potentially useful technique for investigating the validity of rate estimates from heterochronous data, including those from aDNA and serially sampled viruses (de Bruyn et al. 2009; Miller et al. 2009; Subramanian et al. 2009b; Firth et al. 2010).

It is interesting to note that the alignment of complete mitochondrial genomes from woolly mammoths failed the date randomization test. This suggests that the analyses performed by Debruyne and Poinar might be misleading, being based on a data set that is unable to yield plausible posterior estimates without strong prior information on the population size or root age. There are probably several reasons for the poor performance of the mammoth mitogenomic data set. First, the alignment comprises only a small number of sequences. Second, the mitochondrial tree of mammoths has a highly unusual structure, with a very deep split separating the two major clades (Gilbert et al. 2008). This is reflected in the imprecision of the date estimates that are obtained when only the ages of the tips are used for calibration (Debruyne et al. 2008; Gilbert et al. 2008; Debruyne and Poinar 2009). Third, mammoth mitochondrial DNA has evolved at an exceptionally low rate, a phenomenon that is mirrored in the mammoth nuclear genome (Hofreiter 2008; Miller et al. 2008). For example, the substitution rate in elephantids has been much lower than that in hominoid primates, which in turn have been evolving more slowly than other primates (Steiper et al. 2004). This also calls into question the analyses performed by Debruyne and Poinar in which subsamples of the mammoth mitogenomes were assumed to be representative of typical aDNA alignments. In practice, short aDNA alignments have almost exclusively come from the D-loop, which is the most variable portion of the vertebrate mitochondrial genome. The subsampling procedure used by Debruyne and Poinar will tend to include portions of the mitogenome that are evolving much more slowly, leading to exceptionally uninformative data sets.

Debruyne and Poinar recommend that the bias due to signal dependence can be overcome through the employment of "deep" calibrations, for example, at the root of the tree. Often, this is neither possible nor appropriate in analyses of heterochronous data. If rates were truly time-dependent, then such analyses would need to be performed in a relaxed-clock framework to allow

the rate to vary between younger and older branches (Korsten et al. 2009). If a strict molecular clock is assumed, as in the analyses done by Debruyne and Poinar, then rate homogeneity across different timescales is invoked as an a priori assumption. Therefore, although this would seemingly address the problem posed by time-dependent rates, it only does so by assuming that the problem does not exist (Ho et al. 2007c). A suggested solution to this problem is to limit the analysis to third codon sites or synonymous sites, which are putatively subject to a much lesser degree of selective constraint (Briggs et al. 2009; Subramanian et al. 2009a; Endicott et al. 2010). In any case, reliable internal-node calibrations are rarely available in population-level analyses (Ho and Phillips 2009).

Although we have demonstrated that time-dependent rates are unlikely to be driven by a signal-dependent artifact, the findings obtained in the present study do not necessarily validate published estimates of rates from aDNA data. Such estimates can be detrimentally affected by a variety of other confounding factors, including misspecification of the demographic model (Emerson 2007; Ho et al. 2007c; Miller et al. 2009; Navascués and Emerson 2009; Subramanian et al. 2009b). Furthermore, postmortem damage can produce spurious polymorphisms in aDNA sequences, which can lead to biased estimates of rates (Ho et al. 2005, 2007a). Samples in several of the data sets have not been directly radiocarbon dated, but their ages have been inferred by stratigraphic correlation (layer dating). Rate estimates from these data sets, including the arctic fox, Adelie penguin, aurochs, boar, maize, and tuco-tuco, will be somewhat less reliable than those from data sets with directly dated samples. However, higher rate estimates have been obtained from a wide range of aDNA data sets, sourced from a variety of taxa with different demographic histories and biological characteristics, indicating that they should not be dismissed lightly. Combined with the exceptionally high rates estimated in studies of pedigrees and mutation accumulation lines, these results suggest that further empirical and theoretical investigations into the nature of time-dependent rates could be productive.

By their very nature, most aDNA data sets have low information content. Although the situation is changing, as high-throughput sequencing techniques allow complete mitochondrial genomes to be sequenced from conspecific individuals (Gilbert et al. 2008; Briggs et al. 2009; Stiller et al. 2009; Ho and Gilbert 2010), short alignments are likely to remain a common feature of aDNA studies in the near future. In these studies, the important question is not whether the information content is low, but whether it is sufficient for performing the analyses of interest.

FUNDING

This work was supported by the Australian Research Council and the Defense Advanced Research Projects Agency.

ACKNOWLEDGMENTS

The authors thank K.Z., C.B., and R.D. for helpful comments and suggestions that considerably improved the paper.

REFERENCES

- Bandelt H.J. 2008. 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.
- Borge T., Bachmann L., Björnstad G., Wiig Ø. 2007. Genetic variation in Holocene bowhead whales from Svalbard. *Mol. Ecol.* 16:2223–2235.
- Briggs A.W., Good J.M., Green R.E., Krause J., Maricic T., Stenzel U., Lalueza-Fox C., Rudan P., Brajkovic D., Kucan Z., Gusic I., Schmitz R., Doronichev V.B., Golovanova L.V., de la Rasilla M., Fortea J., Rosas A., Pääbo S. 2009. Targeted retrieval and analysis of five Neanderthal mtDNA genomes. *Science*. 325:318–321.
- Burridge C.P., Crow 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.
- Campos P.F., Willerslev E., Sher A., Orlando L., Axelsson E., Tikhonov A., Aaris-Sorensen K., Greenwood A.D., Kahlke R.D., Kosintsev P., Krakhmalnaya T., Kuznetsova T., Lemey P., MacPhee R., Norris C.A., Shepherd K., Suchard M.A., Zazula G.D., Shapiro B., Gilbert M.T. 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.
- Charlesworth B., Bartolomé C., Noël V. 2005. The detection of shared and ancestral polymorphisms. *Genet. Res.* 86:149–157.
- de Bruyn M., Hall B.L., Chauke L.F., Baroni C., Koch P.L., Hoelzel A.R. 2009. Rapid response of a marine mammal species to holocene climate and habitat change. *PLoS Genet.* 5:e1000554.
- 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.
- Debruyne R., Poinar H.N. 2009. Time dependency of molecular rates in ancient DNA data sets, a sampling artifact? *Syst. Biol.* 58:348–360.
- Denver D.R., Morris K., Lynch M., Vassilieva L.L., Thomas W.K. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science*. 289:2342–2344.
- Depaulis F., Orlando L., Hänni C. 2009. Using classical population genetics tools with heterochronous data: time matters! *PLoS ONE*. 4:e5541.
- Drummond A.J., Pybus O.G., Rambaut A., Forsberg R., Rodrigo A.G. 2003. Measurably evolving populations. *Trends Ecol. Evol.* 18:481–488.
- 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.W., Stringer C. Forthcoming 2010. Using genetic evidence to evaluate four palaeoanthropological hypotheses for the timing of Neanderthal and modern human origins. *J. Hum. Evol.* 59:87–95.
- Firth C., Kitchen A., Shapiro B., Suchard M.A., Holmes E.C., Rambaut A. 2010. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol. Biol. Evol.* 27:2038–2051.
- Freitas F.O., Bendel G., Allaby R.G., Brown T.A. 2003. DNA from primitive maize landraces and archaeological remains: implications for the domestication of maize and its expansion into South America. *J. Archaeol. Sci.* 30:901–908.
- Genner M.J., Seehausen O., Lunt D.H., Joyce D.A., Shaw P.W., Carvalho G.R., Turner G.F. 2007. Age of cichlids: new dates for ancient lake fish radiations. *Mol. Biol. Evol.* 24:1269–1282.
- Gibbons A. 1998. Calibrating the mitochondrial clock. *Science*. 279:28–29.
- Gilbert M.T.P., Drautz D.I., Lesk A.M., Ho S.Y.W., Qi J., Ratan A., Hsu C.-H., Sher A., Dalén L., Götherström A., Tomsho L.P., Rendulic S., Packard M., Campos P.F., Kuznetsova T., Shidlovskiy F., Tikhonov A., Willerslev E., Iacumin P., Buigues B., Ericson P.G., Germonpré 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. U.S.A.* 105:8327–8332.
- Haag-Liautard C., Coffey N., Houle D., Lynch M., Charlesworth B., Keightley P.D. 2008. Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biol.* 6:e204.
- Hay J.M., Subramanian S., Millar C.D., Mohandesan E., Lambert D.M. 2008. Rapid molecular evolution in a living fossil. *Trends Genet.* 24:106–109.
- Henn B.M., Gignoux C.R., Feldman M.W., Mountain J.L. 2009. Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Mol. Biol. Evol.* 26:217–230.
- Ho S.Y.W., Gilbert M.T.P. 2010. Ancient mitogenomics. *Mitochondrion*. 10:1–11.
- Ho S.Y.W., Heupink T.H., Rambaut A., Shapiro B. 2007a. Bayesian estimation of sequence damage in ancient DNA. *Mol. Biol. Evol.* 24:1416–1422.
- Ho S.Y.W., Kolokotronis S.-O., Allaby R.G. 2007b. Elevated substitution rates estimated from ancient DNA. *Biol. Lett.* 3:702–705.
- Ho S.Y.W., Lanfear R. 2010. Improved characterisation of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks. *Mitochondrial DNA*. 21:138–146.
- Ho S.Y.W., Larson G. 2006. Molecular clocks: when times are a-changin'. *Trends Genet.* 22:79–83.
- Ho S.Y.W., Phillips M.J. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* 58:367–380.
- Ho S.Y.W., 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.W., Shapiro B., Phillips M., Cooper A., Drummond A.J. 2007c. Evidence for time dependency of molecular rate estimates. *Syst. Biol.* 56:515–522.
- Hofreiter M. 2008. DNA sequencing: Mammoth genomics. *Nature*. 456:330–331.
- 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.
- Jukes T.H., Cantor C.R. 1969. Evolution of protein molecules. In: Munro H.N., editors. *Mammalian protein metabolism*. New York: Academic Press. p. 21–132.
- Korsten M., Ho S.Y.W., Davison J., Pahn B., Vulla E., Roht M., Tumanov I.L., Kojola I., Andersone-Lilley Z., Ozolins J., Pilot M., Mertzanis Y., Giannakopoulos A., Vorobiev A.A., Markov N.I., Saveljev A.P., Lyapunova E.A., Abramov A.V., Mannil P., Valdmann H., Pazetnov S.V., Pazetnov V.S., Rokov A.M., Saarma U. 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.
- Lambert D.M., Ritchie P.A., Millar C.D., Holland B., Drummond A.J., Baroni C. 2002. Rates of evolution in ancient DNA from Adélie penguins. *Science*. 295:2270–2273.
- Loogväli E.-L., Kivisild T., Margus T., Villems R. 2009. Explaining the imperfection of the molecular clock of hominid mitochondria. *PLoS ONE*. 4:e8260.
- Macaulay V.A., Richards M.B., Forster P., Bendall K.E., Watson E., Sykes B., Bandelt H.-J. 1997. mtDNA mutation rates—no need to panic. *Am. J. Hum. Genet.* 61:983–985.
- Millar C.D., Dodd A., Anderson J., Gibb G.C., Ritchie P.A., Baroni C., Woodhams M.D., Henty M.D., Lambert D.M. 2008. Mutation and evolutionary rates in adélie penguins from the Antarctic. *PLoS Genet.* 4:e1000209.

- Miller H.C., Moore J.A., Allendorf F.W., Daugherty C.H. 2009. The evolutionary rate of tuatara revisited. *Trends Genet.* 25:13–15.
- Miller W., Drautz D.I., Ratan A., Pusey B., Qi J., Lesk A.M., Tomsho L.P., Packard M.D., Zhao F., Sher A., Tikhonov A., Raney B., Patterson N., Lindblad-Toh K., Lander E.S., Knight J.R., Irzyk G.P., Fredrikson K.M., Harkins T.T., Sheridan S., Pringle T., Schuster S.C. 2008. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature.* 456:387–390.
- Minin V.N., Bloomquist E.W., Suchard M.A. 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* 25:1459–1471.
- Navascués M., Emerson B.C. 2009. Elevated substitution rate estimates from ancient DNA: model violation and bias of Bayesian methods. *Mol. Ecol.* 18:4390–4397.
- Papadopoulou A., Anastasiou I., Vogler A.P. 2010. Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Mol. Biol. Evol.* 27:1659–1672.
- Penny D. 2005. Relativity for molecular clocks. *Nature.* 426:183–184.
- Peterson G.I., Masel J. 2009. Quantitative prediction of molecular clock and K_a/K_s at short timescales. *Mol. Biol. Evol.* 26:2595–2603.
- Rambaut A., Grassly N.C. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* 13:235–238.
- 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.
- Soares P., Ermini L., Thomson N., Mormina M., Rito T., Röhl A., Salas A., Oppenheimer S., Macaulay V., Richards M.B. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* 84:1–20.
- Steiper M.E., Young N.M., Sukarna T.Y. 2004. Genomic data support the hominoid slowdown and an Early Oligocene estimate for the hominoid-cercopithecoid divergence. *Proc. Natl. Acad. Sci. U.S.A.* 101:17021–17026.
- Stiller M., Knapp M., Stenzel U., Hofreiter M., Meyer M. 2009. Direct multiplex sequencing (DMPS)—a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. *Genome Res.* 19:1843–1848.
- Subramanian S., Denver D.R., Millar C.D., Heupink T., Aschrafi A., Emslie S.D., Baroni C., Lambert D.M. 2009a. High mitogenomic evolutionary rates and time dependency. *Trends Genet.* 25:482–486.
- Subramanian S., Hay J.M., Mohandesan E., Millar C.D., Lambert D.M. 2009b. Molecular and morphological evolution in tuatara are decoupled. *Trends Genet.* 25:16–18.
- Woodhams M. 2006. Can deleterious mutations explain the time dependency of molecular rate estimates? *Mol. Biol. Evol.* 23:2271–2273.