informa
healthcare

Improved characterisation of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks

SIMON Y. W. HO[1,2] & ROBERT LANFEAR[1]

[1]*Research School of Biology, Centre for Macroevolution and Macroecology, Australian National University, Canberra, ACT, 0200, Australia, and* [2]*School of Biological Sciences, University of Sydney, Sydney, NSW 2006, Australia*

Abstract

Background and aims. There is extensive variation in substitution rates among metazoan mitochondrial genomes, including differences among lineages and among sites. Differences in substitution rates among lineages are routinely taken into account in molecular phylogenetic and dating analyses. However, this is rarely the case for differences in the lineage-specific patterns of rate variation among sites.

Materials and methods. We analysed an alignment of 32 cetacean mitogenomes using a partitioned relaxed-clock approach, in which a separate relaxed-clock model was applied to each of the three codon positions of the protein-coding genes.

Results. By comparing Bayes factors, we found overwhelming support for this model compared with a strict-clock model and less-complex models comprising fewer relaxed clocks. Our analyses reveal the presence of substantial rate heterogeneity among cetacean lineages, and that these patterns of variation differ between codon positions. However, we find no evidence of mitochondrial rate autocorrelation throughout the cetacean phylogeny.

Conclusions. Our study confirms the feasibility of applying a partitioned relaxed-clock model to mitogenomic data, and suggests that molecular phylogenetic and dating analyses may be improved by the application of these models.

Keywords: *Mitochondrial DNA, whales, molecular clock, rate heterogeneity*

Introduction

Extensive studies of substitution rates in mitochondrial DNA (mtDNA) have produced convincing evidence of rate variation among lineages and across time-scales (García-Moreno 2004; Ho et al. 2005a; Brown et al. 2008; Nabholz et al. 2009; Papadopoulou et al. 2010). This rate heterogeneity has important consequences because of the widespread use of mtDNA in studies of animal evolution and biogeography (Zink and Barrowclough 2008; Galtier et al. 2009). In particular, accommodation needs to be made for rate variation when estimating evolutionary time-scales using mtDNA data.

Molecular date estimates are typically made using phylogenetic methods that are based on the premise of a molecular clock. The molecular clock can be 'strict', whereby rates of evolution are assumed to be constant among all lineages, or 'relaxed', whereby rates are allowed to vary among lineages in a controlled manner. Several relaxed-clock models have been developed in the past decade, some of which have been implemented in Bayesian settings (Thorne et al. 1998; Aris-Brosou and Yang 2002; Drummond et al. 2006; Lepage et al. 2007; Rannala and Yang 2007). These models vary in their assumptions concerning the nature of among-lineage rate variation; for example, some assume that rates change instantaneously at internal nodes in the phylogeny (e.g. Drummond et al. 2006), while others allow rates to 'evolve' continuously over time (e.g. Thorne et al. 1998; Kishino et al. 2001).

Correspondence: S. Y. W. Ho, School of Biological Sciences, University of Sydney, Macleay Building A12, Sydney NSW 2006, Australia. Tel: 61 2 93518681. Fax: 61 2 91140979. E-mail: simon.ho@sydney.edu.au

RIGHTS LINK

Several aspects of mtDNA rate variation can be examined using relaxed-clock methods. First, the degree of departure from a strict molecular clock can be assessed using measures of among-lineage rate heterogeneity, such as the coefficient of rate variation or the index of rate dispersion (Ohta and Kimura 1971; Drummond et al. 2006). Second, hypotheses about the evolutionary processes driving variation in rates can be evaluated by measuring rate correlation between adjacent branches. For instance, molecular rates should show significant autocorrelation if they are closely tied to heritable life-history characteristics such as generation time, body size, metabolic rate, longevity and population size (Sanderson 1997; Lanfear et al. 2007; Smith and Donoghue 2008). The presence of rate autocorrelation can be investigated by testing correlations between rates in neighbouring branches of the tree or through statistical comparisons between models of autocorrelated and uncorrelated rate change (Lepage et al. 2007; Ho 2009).

A third aspect of rate variation has received less attention: the variation in patterns of among-lineage rate heterogeneity across sites of a sequence alignment. For instance, a common assumption in relaxed-clock analyses is that a single model of branch-specific rates applies across all partitions of the alignment. In other words, the rates of different partitions are usually assumed to co-vary perfectly throughout the phylogeny, such that they remain at a fixed ratio across all branches of the tree. However, separate partitions of an alignment are usually defined to reflect the existence of different substitution processes (reviewed by Simon et al. 2006), and it is known that the nature and strength of selection will vary among lineages, leading to patterns such as heterotachy (Lopez et al. 2002). If patterns of heterotachy differ among partitions of a single alignment, then it might be beneficial to relax the assumption that a single model of branch-specific rates applies across all partitions.

The coding sequences of mitogenomes represent a good example of sites that are likely to have evolved under different substitution processes, and so might have also experienced different patterns of heterotachy. For example, rates at the third codon positions might be largely influenced by the rate of mutation, whereas rates at the first and second codon positions are typically subject to higher levels of selection. This argues for a decoupling of not only models of nucleotide substitution (Shapiro et al. 2006), but also models of among-lineage rate heterogeneity, among different codon positions (Seo et al. 2004; Lemey et al. 2007; Ho 2009). Codon-substitution models have been proposed (Goldman and Yang 1994; Muse and Gaut 1994), but these have typically been intractable for large datasets (but see Suchard and Rambaut 2009).

In the present study, we investigate some of these questions using a cetacean mitogenome dataset. We assess the feasibility of employing a partitioned relaxed-clock approach to analyse patterns of rate variation and autocorrelation at the three codon positions of mitochondrial protein-coding genes. This approach takes advantage of the fact that all of the sites in a mitochondrial alignment should share the same underlying phylogeny, but permits disparate patterns of rate heterogeneity among codon positions.

Materials and methods

Complete mitogenome sequences were downloaded from GenBank for 32 species of Cetacea, including 14 baleen whales (mysticetes) and 18 toothed whales (odontocetes). Protein-coding genes were extracted from these mitogenomes and manually aligned at the amino-acid level. The *nd6* gene, which is situated on the light strand opposite the other 12 protein-coding genes, was discarded because of its distinctive base composition and patterns of substitution. The final alignment of 10,845 nucleotide sites is available as Supplementary material.

The dataset was analysed in a Bayesian relaxed-clock phylogenetic framework, as implemented in the software BEAST 1.5.3 (Drummond and Rambaut 2007). The sequence alignment was partitioned by codon position, with 3615 nucleotide sites in each of the three data categories. Each partition was assigned a separate general time-reversible (GTR + I + Γ) model of nucleotide substitution (Lanave et al. 1984; Yang 1994), allowing for different rate matrices, levels of among-site rate heterogeneity, proportions of invariant sites, and equilibrium base frequencies among codon positions. Comparison of values of the Bayesian information criterion confirmed that this model was always the best or second-best substitution model for each data partition.

Four analyses were conducted on the sequence alignment. First, the alignment was analysed using a strict-clock model across all codon positions. Subsequently, three relaxed-clock models were tested: an unpartitioned model with one relaxed clock for all three codon positions, with a simple estimated ratio of rates among the three codon positions; a codon-partitioned model with one relaxed clock for the first and second codon positions, with a simple estimated ratio of rates between the two codon positions, and a separate relaxed clock for the third codon positions; and a codon-partitioned model with a separate relaxed clock for each of the three codon positions. The four models were compared using Bayes factors, computed using the harmonic-mean estimator implemented in Tracer 1.5 (Suchard et al. 2001; Rambaut and Drummond 2007).

The phylogeny, divergence times and other model parameters were co-estimated from the sequence data.

The position of the root was fixed by enforcing mutual monophyly on Mysticeti and Odontoceti. In all analyses, the different data partitions shared the same underlying tree topology and branch lengths (measured in units of time), with the prior tree distribution generated using a Yule process. Posterior distributions were estimated by Markov chain Monte Carlo sampling, with samples drawn every 3×10^3 steps over a total of 3×10^7 steps. Posterior samples were inspected to determine the number of burn-in steps to be discarded, while convergence and acceptable mixing were checked using Tracer. For each of the four models, two independent runs were performed and the samples combined. Effective sample sizes for estimated parameters were almost all >500, with most being >1000; values >200 are regarded as being indicative of sufficient sampling from the posterior.

To allow the estimation of absolute substitution rates, independent calibrating information is required. We specified a uniform prior of $34-36$ million years (Myr) for the age of the root node. This was a key calibration used in previous studies of cetacean divergence times (Jackson et al. 2009; Steeman et al. 2009). We also specified five minimum-age constraints: crown Mysticeti (28 Myr), crown Balaenopteridae (7.3 Myr), Inioidea–Delphinoidea split (23.5 Myr), crown Inioidea (12 Myr) and crown Delphinoidea (10 Myr). For details and justifications of these age constraints, see Steeman et al. (2009).

To test whether improvements in likelihood gained under the three-partition relaxed clock were simply the result of increasing the number of model parameters, we performed 10 replicate analyses in which the sites of the alignment were randomly assigned to the three relaxed-clock models. This was done by shuffling the sites in the sequence alignment (equivalent to sampling sites randomly without replacement). The 10 shuffled datasets were analysed using BEAST, with the same settings as for the analysis of the three-partition relaxed clock above. Shuffling was performed using novel software, SiteSampler 1.0 (written in Java by S.Y.W.H.; available upon request as a Java application). SiteSampler is able to generate pseudo-replicate datasets by sampling with replacement (bootstrapping) or without replacement (shuffling or permutation) from an existing sequence alignment and is able to perform this site-sampling within predefined data partitions.

To estimate non-synonymous (dN) and synonymous (dS) substitution rates for each branch of the tree, we used HyPhy 2.0 (Kosakovsky Pond et al. 2005) with the MG94 model of codon evolution (Muse and Gaut 1994), codon frequencies estimated from the data in a 3×4 matrix and the tree fixed to the maximum-clade-credibility tree from the tripartitioned relaxed-clock Bayesian analysis. The MG94 model can be paired with any model of sequence

evolution from the six-parameter GTR family. There are 203 such models, and we chose the most appropriate model by fitting all possible models to the data and comparing them using the Akaike information criterion (AIC) (Akaike 1974). This was done using a custom script in the HyPhy batch language (codon_model_compare_local.bf, written by R.L.; available upon request), which calculates the likelihood of each model when dN/dS is free to vary among all branches of the tree. The best model had four parameters such that all substitution rates were distinct, except for two pairs of rates that were constrained to be equal ($\theta_{AC} = \theta_{GT}$, $\theta_{AT} = \theta_{CG}$). In HyPhy 2.0 notation, this model can be described as MG94_3×4_012230. We used this approach to calculate the likelihood of two models of molecular evolution: a 'free-ratio' model in which a separate dN/dS was estimated for each branch of the tree, and a nested 'single-ratio' model in which all branches share a single dN/dS.

Results

Calculation of Bayes factors indicated decisive support for the employment of relaxed clocks compared with a strict-clock model (Table I). Furthermore, among the three different partitioned relaxed-clock models, there was overwhelming support for the model in which each of the three codon positions was assigned a separate relaxed clock (Table I). Assigning sites to clock partitions on the basis of codon position resulted in far greater gains in likelihood (e.g. a gain of 1097 log-likelihood units for the three-partition relaxed clock over the one-partition relaxed clock) than randomly assigning sites to partitions (an average gain of 15.06 log-likelihood units for the three-partition relaxed clock over the one-partition relaxed clock).

There were some differences among estimates of topology made under the different clock models. The tree estimated using a three-partition relaxed clock is shown in Figure 1. In the strict-clock analysis, the pygmy right whale (*Caperea marginata*) was found to be the sister taxon to all other mysticetes (posterior probability = 1.00), but exchanged this position with Balaenidae (comprising *Balaena mysticetus*, *Eubalaena australis* and *Eubalaena japonica*) in all of the relaxed-clock analyses (posterior probability = 1.00). The latter is consistent with recent phylogenetic analyses of Cetacea that used combined nuclear and mitochondrial supermatrices (McGowen et al. 2009; Steeman et al. 2009). The placements of *Platanista minor* and *Lipotes vexillifer* showed some variation among the different clock models, but posterior probabilities were relatively low.

Estimates of divergence times varied slightly among clock models, possibly tempered by the employment of multiple informative calibrations (Table II).

RIGHTSLINK

Table I. Bayes-factor comparison of molecular-clock schemes in a phylogenetic analysis of cetacean mitogenomes.

Molecular clock	Codon partitioning for clocks	Harmonic mean of marginal log-likelihood	$\log_{10}$ Bayes factor (vs. three-partition model)
Strict clock	[1st 2nd 3rd]	−91,536	−678.4
One relaxed clock	[1st 2nd 3rd]	−91,071	−476.4
Two relaxed clocks	[1st 2nd][3rd]	−90,142	−73.0
Three relaxed clocks	[1st][2nd][3rd]	−89,974	−

The date estimates made using two-partition and three-partition relaxed clocks were very similar, but both of these differed from the estimates made using a one-partition relaxed clock. A clear trend can be seen in the sizes of the 95% highest posterior density (HPD) intervals on the divergence-time estimates, with the strict-clock model yielding estimates with the smallest amount of uncertainty. This is consistent with the findings of previous studies comparing strict-clock and relaxed-clock models (Ho et al. 2005b). Among the three relaxed-clock models, the sizes of the 95% HPDs decreased as the number of partitions increased, such that the three-partition relaxed-clock model gave the smallest 95% HPDs.
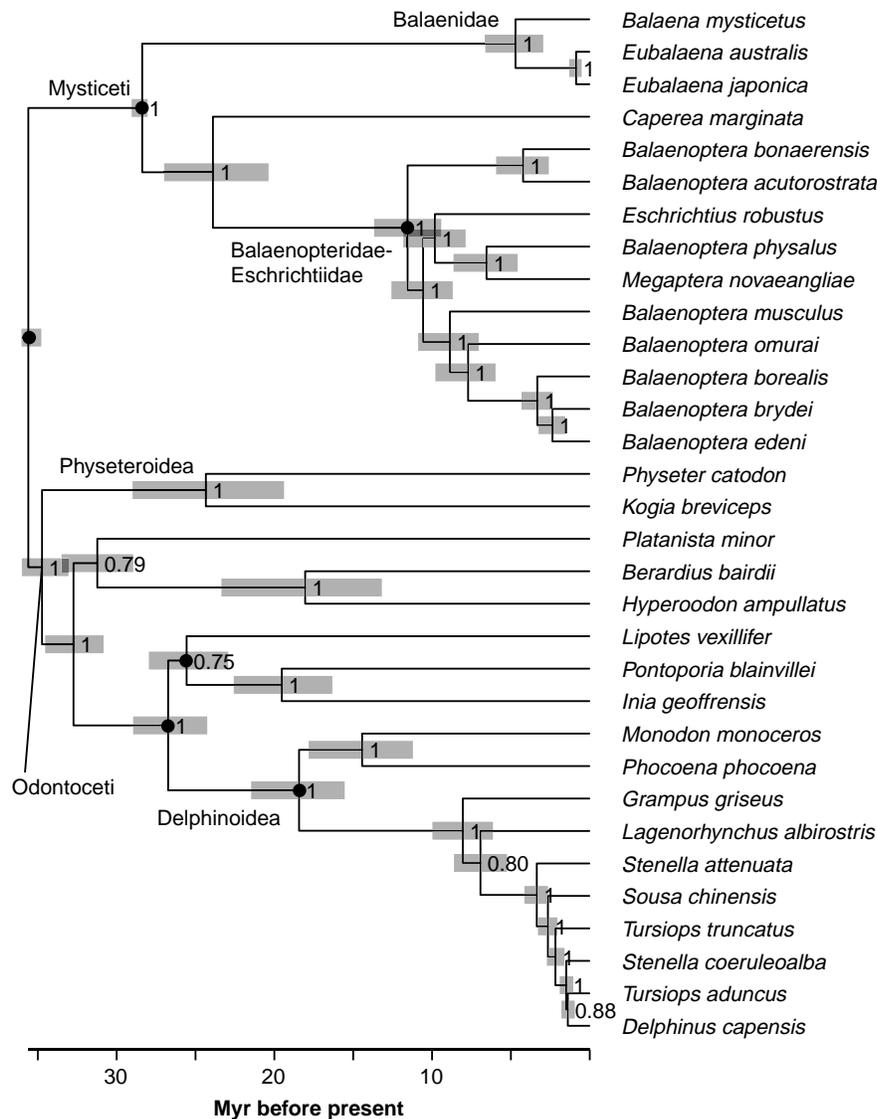
Figure 1. Maximum-clade-credibility chronogram estimated from an alignment of 12 mitochondrial protein-coding genes from 32 cetacean species. Nodal ages reflect mean posterior estimates. Nodes are labelled with posterior probabilities, while grey bars represent 95% HPD intervals for date estimates. Black circles denote nodes used for calibration. Estimates were made using a Bayesian approach in which a separate relaxed clock was assigned to each of the three codon positions.

Table II. Age estimates for key nodes in the cetacean tree, obtained under different molecular-clock treatments.

| | Age estimate (Myr) | | | | | | | |
| | Strict clock | | One relaxed clock | | Two relaxed clocks | | Three relaxed clocks | |
Node	Mean	95% HPD	Mean	95% HPD	Mean	95% HPD	Mean	95% HPD
Cetacea	35.95	(35.84–36.00)	35.26	(34.25–36.00)	35.48	(34.56–36.00)	35.57	(34.74–36.00)
Mysticeti	28.04	(28.00–28.13)	29.01	(28.00–30.89)	28.44	(28.00–29.29)	28.35	(28.00–29.01)
Balaenidae	4.35	(3.93–4.79)	5.49	(2.50–9.00)	4.97	(2.88–7.39)	4.68	(2.92–6.59)
Balaenopteridae–Eschrichtiidae	10.47	(9.86–11.12)	12.56	(8.49–16.84)	11.62	(9.45–14.07)	11.52	(9.38–13.64)
Odontoceti	35.87	(35.69–36.00)	33.65	(31.06–35.82)	34.40	(32.41–35.95)	34.70	(33.02–35.97)
Physeteroidea	25.22	(23.75–26.78)	21.62	(14.05–29.16)	25.27	(17.39–28.51)	24.31	(19.35–28.96)
Delphinoidea	20.01	(18.95–21.07)	17.69	(13.10–21.99)	18.11	(14.89–21.45)	18.40	(15.51–21.44)

The listed clades are shown in Figure 1.

Rate estimates were obtained for different partitions of the mitogenome alignment (Table III). In the unpartitioned relaxed-clock model, the mean substitution rate across all three codon positions was estimated at 2.00×10^{-2} subs/site/Myr, with a 95% HPD interval of 1.80×10^{-2}–2.20×10^{-2} subs/site/Myr. This was slightly higher than the mean substitution rate estimated under the assumption of a strict clock, which was 1.73×10^{-2} subs/site/Myr (95% HPD interval: 1.65×10^{-2}–1.81×10^{-2} subs/site/Myr). In the tripartitioned relaxed-clock model, the mean rate differed at each codon position—with the third codon position evolving the most quickly, followed by the first and second codon positions, respectively (Table III).

For all three codon positions, the 95% HPD interval of the coefficient of variation of branch-specific rates excluded zero, indicating substantial departure from the assumption of a strict clock (Table III and Figure 2a). This supports the results from the comparison of Bayes factors (Table I). Substitution rates at the second codon positions were the least clock-like, with a coefficient of variation of 0.545 (95% HPD interval: 0.433–0.672). The mean covariance of rates was positive for all partitions, but associated 95% HPD intervals included zero (Table III and Figure 2b).

Branch-specific comparisons of rates at the three codon positions revealed only weak correlations between rates at different sites (Figure 3). Examination of the variance in the ratios of rates between codon positions, shown in Figure 4, confirms that the variance in rate ratios is substantially higher than zero. These results are in contrast with the assumption commonly made in molecular-clock analyses that branch-specific rates are completely correlated between data partitions (and thus have zero variance in a given tree).

The mean dS (0.179) was much larger than the mean dN (0.009), as expected when non-synonymous substitutions experience higher constraint than synonymous substitutions. In view of the substantial differences in mean dS and dN, we used the coefficient of variation (the SD divided by the mean) to compare the variability of these quantities. The coefficient of variation of dS (1.052) is smaller than that of dN (1.209), in agreement with the results of the three-partition relaxed clock, in which the variability of rates in the third codon positions (for which most substitutions are synonymous) was lower than that of the second codon positions (for which most substitutions are non-synonymous).

In order to determine whether dN and dS tend to vary independently of one another, we calculated dN/dS for each branch of the tree. dN/dS showed a high degree of heterogeneity among lineages (mean = 0.144, coefficient of variation = 5.344), suggesting that dN and dS vary independently.

Table III. Posterior estimates of substitution rates, coefficients of rate variation and rate covariances in the protein-coding genes of cetacean mitogenomes.

Data partition	Mean substitution rate (subs/site/Myr)		Coefficient of variation		Covariance	
	Mean	95% HPD	Mean	95% HPD	Mean	95% HPD
Three-partition relaxed clock						
First, codon positions	3.52×10^{-3}	$(3.20 \times 10^{-3}$–$3.85 \times 10^{-3})$	0.382	(0.314–0.456)	0.041	$(-0.208$–$0.292)$
Second, codon positions	1.10×10^{-3}	$(9.65 \times 10^{-4}$–$1.23 \times 10^{-3})$	0.545	(0.433–0.672)	0.192	$(-0.082$–$0.466)$
Third codon positions	2.40×10^{-2}	$(2.20 \times 10^{-2}$–$2.61 \times 10^{-2})$	0.415	(0.342–0.489)	0.141	$(-0.116$–$0.397)$
One-partition relaxed clock						
All codon positions	2.00×10^{-2}	$(1.80 \times 10^{-2}$–$2.20 \times 10^{-2})$	0.502	(0.391–0.626)	0.061	$(-0.221$–$0.342)$

All estimates were obtained using the uncorrelated lognormal relaxed-clock model. Plots of the posterior distributions of the coefficients of variation and covariance are given in Figure 2.

Comparison of a free-ratio model (in which *dN/dS* is free to vary among branches) with a single-ratio model (in which all branches share a single *dN/dS* value) indicates that the former is significantly better (likelihood-ratio test, $P = 4.8 \times 10^{-26}$, ΔAIC = 136), confirming that *dN/dS* varies significantly among different branches of the tree.

Figure 2. Marginal posterior densities of (a) coefficient of variation of rates and (b) covariance of rates, estimated from cetacean mitogenomes using a three-partition relaxed clock in a Bayesian phylogenetic framework. Means and 95% HPD intervals are provided in Table III.

Discussion

In the present study, we have examined the application of partitioned relaxed-clock models to a single alignment of cetacean mtDNA sequences. This approach allows the user to account for situations in which the ratio of rates between data partitions, such as different codon positions, is not constant across all lineages. Most molecular-clock analyses assume that rates are perfectly correlated between data partitions, such that the partitions can evolve at different rates but at a fixed ratio across all branches of the tree. This is equivalent to assuming that the variance of rate ratios between partitions is equal to zero for any given tree. However, when this assumption is relaxed (i.e. the rate ratios between partitions are allowed to vary among branches), it can be seen that the variance of rate ratios between all partitions is substantially higher than zero for most trees (Figure 4).

Some of this variance might reflect the stochastic nature of the substitution process and may stem from the difficulty in measuring rates accurately in different partitions. However, there are two lines of evidence that suggest a substantial proportion of the variance in rate ratios has a biological underpinning. First, the variance in rate ratios is highest between data partitions that are known to differ most strongly in selective constraint (i.e. second vs. third codon positions; Figures 3c and 4), and lowest for those partitions that are known to be most similar (i.e. second vs. first codon positions; Figures 3a and 4). This interpretation is strongly supported by analyses of synonymous (*dS*) and non-synonymous (*dN*) substitution rates, which show that these quantities vary independently across the phylogeny. Second, the use of partitioned relaxed clocks leads to a dramatic improvement in the fit of the model to the data (Table I), and the results of the randomisation analyses confirm that the improvement in likelihood is not merely the result of increasing the number of parameters in the model. These results suggest that molecular dating studies may be improved by the application of such models, as evidenced by the
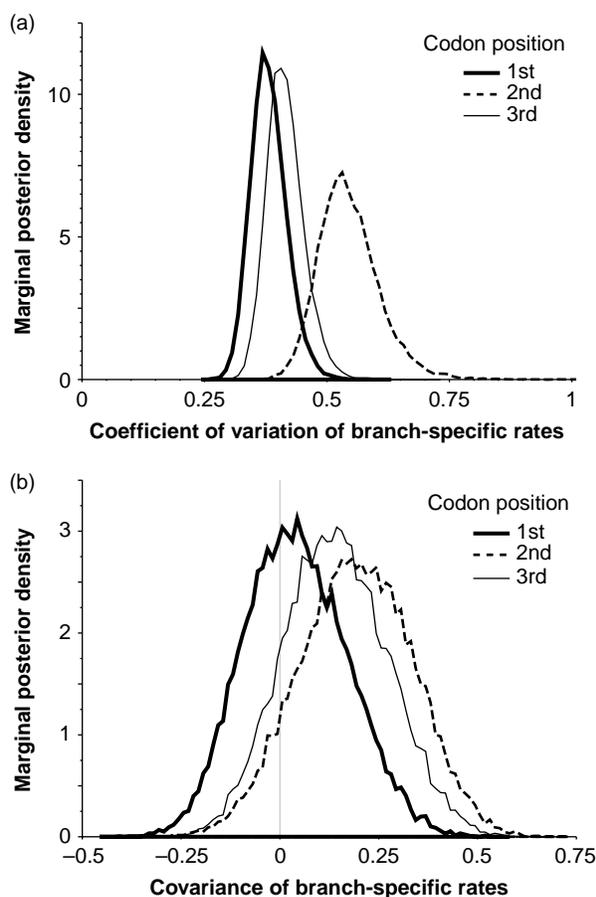
(a)

(b)

(c)
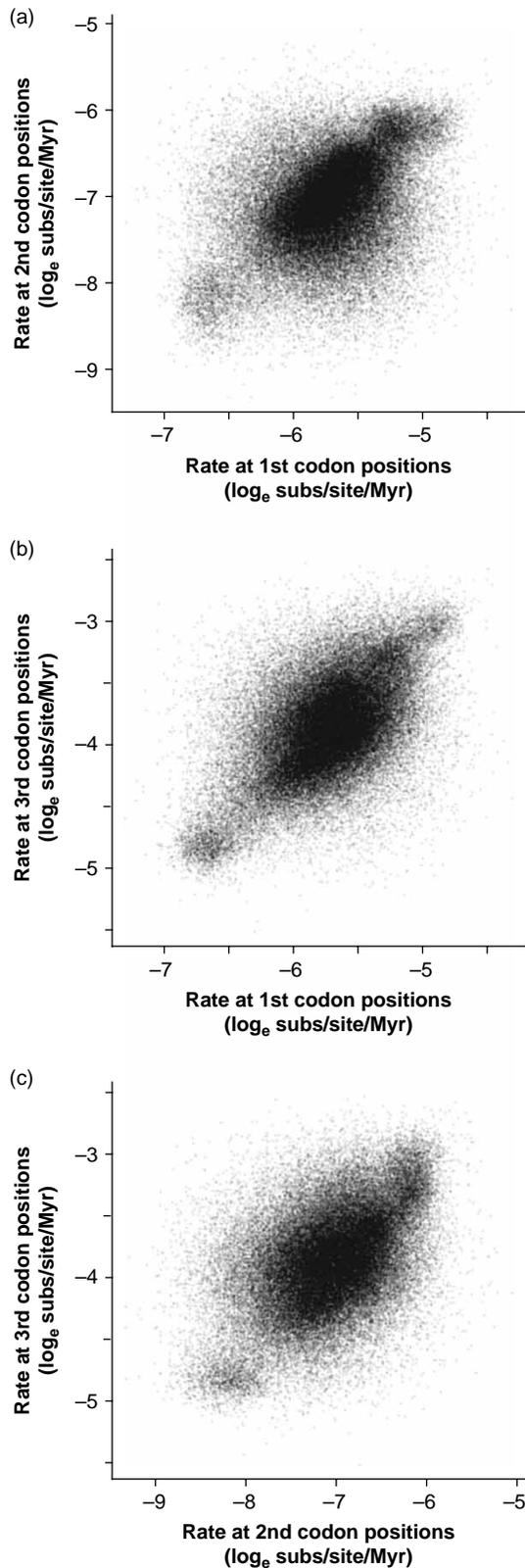
Figure 3. Plots of branch-specific rates for the three codon positions in a complete alignment of cetacean mitochondrial protein-coding genes. Data are shown for 1000 trees sampled from the posterior. Each data point represents a comparison of two rates along a single branch. Rates are plotted for (a) second against the first codon position, (b) third against the first codon position and (c) third against the second codon position.
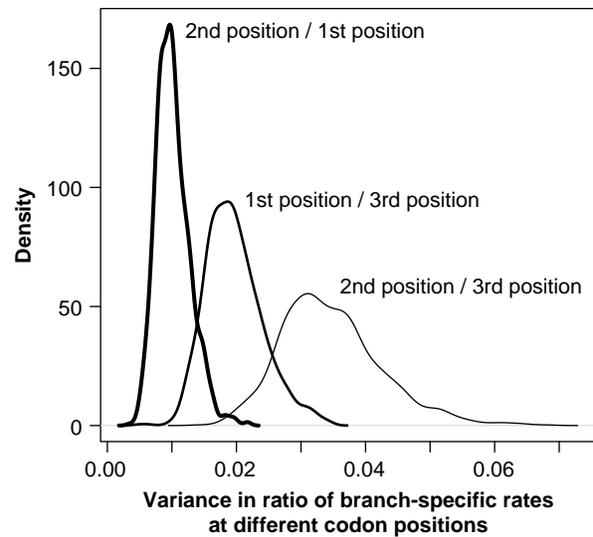
Figure 4. Plot showing the posterior distributions of variance in the ratios of rates between codon positions. Ratios of log-transformed rates were calculated for each pair of codon positions, for each branch in a tree. The variance in rate ratios across all 62 branches of each tree was calculated for 1000 trees sampled from the posterior. In all three pair-wise comparisons, the rate with the higher variance was chosen as the numerator.

reduction in the uncertainty of date estimates when partitioned relaxed-clock models are applied (Table II).

The implementation of multiple relaxed-clock models potentially represents a more biologically realistic approach to modelling rate variation among lineages. By allowing different patterns of among-lineage rate heterogeneity across codon positions, some accommodation can be made for heterotachy and variation in the ratio of non-synonymous to synonymous substitutions (Seo et al. 2004; Lemey et al. 2007). However, the employment of such models would need to be balanced against the cost of overparameterisation, which might present a legitimate problem in analyses of uninformative datasets. For example, it is not clear that such a model could be applied to intraspecific sequence data. Studies based on simulated data will be able to provide further insights into the performance of a partitioned relaxed-clock model in these settings. If feasible, such an approach could address the growing concerns about the impact of purifying selection on estimates made at the intraspecific level (e.g. Endicott and Ho 2008; Endicott et al. 2009; Subramanian et al. 2009).

Apart from the issue of variation in rate ratios among partitions, our estimates of rate covariance indicate no evidence for rate correlation between adjacent branches of the tree (Table2 III and Figure 2b). That is, the rate in one branch is no more similar to the rate in its ancestral or descendent branches than to rates in other branches of the tree. Similar conclusions were reached in previous Bayesian

analyses of viral and marsupial datasets (Drummond et al. 2006), whereas analyses of three protein-coding genes in avian mtDNA yielded weak evidence of a low level of autocorrelation (Brown et al. 2008). A study of substitution rates in insects found evidence of rate autocorrelation in a single nuclear gene, *Mp20*, but not in the other nuclear or mitochondrial genes included in the analysis (Papadopoulou et al. 2010).

The general lack of evidence for rate autocorrelation in cetacean mitogenomes is surprising, given the increasing evidence that rates of evolution are strongly linked to life-history traits across the tree of life (e.g. Smith and Donoghue 2008). However, a recent study revealed strong support for autocorrelated models of rate change and suggested that dense taxonomic sampling might be necessary to discriminate between autocorrelated and uncorrelated models of rate heterogeneity (Lepage et al. 2007). Moreover, it appears that the covariance statistic in the Bayesian uncorrelated lognormal relaxed-clock model might be a weak indicator of rate autocorrelation (Ho 2009).

Our estimate of the mean cetacean mtDNA rate is higher than the rate of 1×10^{-2} subs/site/Myr that is often assumed in mitogenomic studies of mammals and birds. This is somewhat surprising because a recent study of mysticete mtDNA yielded comparatively low estimates of substitution rates (Jackson et al. 2009), although this discrepancy might be due to our inclusion of the faster-evolving odontocetes. However, such a focus on the mean rate masks the true extent of rate variation among lineages. Our results show strong evidence for non-clock-like evolution in cetacean mtDNA, particularly at the second codon positions. This has implications for mitochondrial studies of cetacean phylogeny (McGowen et al. 2009; Steeman et al. 2009; Yang 2009). Complete sampling of mitogenomes from all ∼89 extant cetacean species will offer further insights into the patterns of mtDNA rate variation in this mammalian order.

It should be noted that our findings are dependent on the choice of age calibrations. Increasing the amount of calibrating information, either by adding new calibrations or by reducing the uncertainty on the calibrations already included, is likely to affect parameter estimates. For instance, the variance in rates among branches will be estimated more precisely if the number of calibrations is increased (Ho and Phillips 2009). It might be productive to investigate these properties in future studies, perhaps involving simulated datasets.

Substitution rates in metazoan mtDNA have been widely studied, but our understanding of rate variation stands in need of further improvement. By applying partitioned relaxed clocks to a wide range of datasets, it will be possible to characterise and quantify the extent of rate heterogeneity among lineages and among codon positions.

References

Akaike H. 1974. A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723.

Aris-Brosou S, Yang Z. 2002. Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. Syst Biol 51:703–714.

Brown JW, Rest JS, García-Moreno J, Sorenson MD, Mindell DP. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biol 6:6.

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

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol 4:e88.

Endicott P, Ho SYW. 2008. A Bayesian evaluation of human mitochondrial substitution rates. Am J Hum Genet 82:895–902.

Endicott P, Ho SYW, Metspalu M, Stringer C. 2009. Evaluating the mitochondrial timescale of human evolution. Trends Ecol Evol 24:515–521.

Galtier N, Nabholz B, Glemin S, Hurst GD. 2009. Mitochondrial DNA as a marker of molecular diversity: A reappraisal. Mol Ecol 18:4541–4550.

García-Moreno J. 2004. Is there a universal mtDNA clock for birds? J Avian Biol 35:465–468.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736.

Ho SYW. 2009. An examination of phylogenetic models of substitution rate variation among lineages. Biol Lett 5:421–424.

Ho SYW, Phillips MJ. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst Biol 58:367–380.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005a. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol 22:1561–1568.

Ho SYW, Phillips MJ, Drummond AJ, Cooper A. 2005b. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. Mol Biol Evol 22:1355–1363.

Jackson JA, Baker CS, Vant M, Steel DJ, Medrano-Gonzalez L, Palumbi SR. 2009. Big and slow: Phylogenetic estimates of molecular evolution in baleen whales (Suborder Mysticeti). Mol Biol Evol 26:2427–2440.

Kishino H, Thorne JL, Bruno WJ. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol Biol Evol 18:352–361.

Kosakovsky Pond SL, Frost SD, Muse SV. 2005. HyPhy: Hypothesis testing using phylogenies. Bioinformatics 21:676–679.

Lanave C, Preparata G, Saccone C, Serio G. 1984. A new method for calculating evolutionary substitution rates. J Mol Evol 20:86–93.

Lanfear R, Thomas JA, Welch JJ, Brey T, Bromham L. 2007. Metabolic rate does not calibrate the molecular clock. Proc Natl Acad Sci USA 104:15388–15393.

Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, Taveira N, Rambaut A, Rodrigo A. 2007. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. PLoS Comput Biol 3:e29.

Lepage T, Bryant D, Philippe H, Lartillot N. 2007. A general comparison of relaxed molecular clock models. Mol Biol Evol 24:2699–2680.

Lopez P, Casane D, Philippe H. 2002. Heterotachy, an important process of protein evolution. Mol Biol Evol 19:1–7.

McGowen MR, Spaulding M, Gatesy J. 2009. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. Mol Phylogenet Evol 53:891–906.

Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol Biol Evol 11: 715–724.

Nabholz B, Glemin S, Galtier N. 2009. The erratic mitochondrial clock: Variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. BMC Evol Biol 9:54.

Ohta T, Kimura M. 1971. On the constancy of the evolutionary rate of cistrons. J Mol Evol 1:18–25.

Papadopoulou A, Anastasiou I, Vogler AP. 2010. Revisiting the insect mitochondrial molecular clock: The mid-Aegean trench calibration. Mol Biol Evol 27:1659–1672.

Rambaut A, Drummond AJ. 2007. Tracer. Oxford: University of Oxford.

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

Sanderson MJ. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol Biol Evol 14:1218–1231.

Seo TK, Kishino H, Thorne JL. 2004. Estimating absolute rates of synonymous and nonsynonymous nucleotide substitution in order to characterize natural selection and date species divergences. Mol Biol Evol 21:1201–1213.

Shapiro B, Rambaut A, Drummond AJ. 2006. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Mol Biol Evol 23:7–9.

Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT. 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Ann Rev Ecol Syst 37: 545–579.

Smith SA, Donoghue MJ. 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322:86–89.

Steeman ME, Hebsgaard MB, Fordyce RE, Ho SYW, Rabosky DL, Nielsen R, Rahbek C, Glenner H, Sørensen MV, Willerslev E. 2009. Radiation of extant cetaceans driven by restructuring of the oceans. Syst Biol 58:573–585.

Subramanian S, Denver DR, Millar CD, Heupink T, Aschrafi A, Emslie SD, Baroni C, Lambert DM. 2009. High mitogenomic evolutionary rates and time dependency. Trends Genet 25: 482–486.

Suchard MA, Rambaut A. 2009. Many-core algorithms for statistical phylogenetics. Bioinformatics 25:1370–1376.

Suchard MA, Weiss RE, Sinsheimer JS. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol 18:1001–1013.

Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol 15: 1647–1657.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J Mol Evol 39:306–314.

Yang X-G. 2009. Bayesian inference of cetacean phylogeny based on mitochondrial genomes. Biologia 64:811–818.

Zink RM, Barrowclough GF. 2008. Mitochondrial DNA under siege in avian phylogeography. Mol Ecol 17:2107–2121.