

Big and Slow: Phylogenetic Estimates of Molecular Evolution in Baleen Whales (Suborder Mysticeti)

J. A. Jackson,* C. S. Baker,*† M. Vant,† D. J. Steel,* L. Medrano-González,‡ and S. R. Palumbi§

*Marine Mammal Institute, Hatfield Marine Science Center, Oregon State University; †School of Biological Sciences, Auckland University, Auckland, New Zealand; ‡Facultad de Ciencias, Universidad Nacional Autónoma de México, Circuito Exterior, Ciudad Universitaria, México, Mexico; and §Department of Biological Sciences, Hopkins Marine Station, Stanford University, Pacific Grove, CA

Baleen whales are the largest animals that have ever lived. To develop an improved estimation of substitution rate for nuclear and mitochondrial DNA for this taxon, we implemented a relaxed-clock phylogenetic approach using three fossil calibration dates: the divergence between odontocetes and mysticetes ~34 million years ago (Ma), between the balaenopterids and balaenopterids ~28 Ma, and the time to most recent common ancestor within the Balaenopteridae ~12 Ma. We examined seven mitochondrial genomes, a large number of mitochondrial control region sequences (219 haplotypes for 465 bp) and nine nuclear introns representing five species of whales, within which multiple species-specific alleles were sequenced to account for within-species diversity (1–15 for each locus). The total data set represents >1.65 Mbp of mitogenome and nuclear genomic sequence. The estimated substitution rate for the humpback whale control region (3.9%/million years, My) was higher than previous estimates for baleen whales but slow relative to other mammal species with similar generation times (e.g., human–chimp mean rate > 20%/My). The mitogenomic third codon position rate was also slow relative to other mammals (mean estimate 1%/My compared with a mammalian average of 9.8%/My for the cytochrome *b* gene). The mean nuclear genomic substitution rate (0.05%/My) was substantially slower than average synonymous estimates for other mammals (0.21–0.37%/My across a range of studies). The nuclear and mitogenome rate estimates for baleen whales were thus roughly consistent with an 8- to 10-fold slowing due to a combination of large body size and long generation times. Surprisingly, despite the large data set of nuclear intron sequences, there was only weak and conflicting support for alternate hypotheses about the phylogeny of balaenopterid whales, suggesting that interspecies introgressions or a rapid radiation has obscured species relationships in the nuclear genome.

Introduction

Estimated neutral rates of DNA substitution in mysticete whales are notably slow relative to other mammals (Martin and Palumbi 1993; Nabholz et al. 2008). In a survey of mitochondrial cytochrome *b* (third codon position) substitution rates, Nabholz et al. (2008) found that baleen whales (Suborder Mysticeti) have a substitution rate of 0.7–0.8% $\text{bp}^{-1} \text{My}^{-1}$ (million years), compared with a mammalian average of 9.8% $\text{bp}^{-1} \text{My}^{-1}$. This is slower than 97.5% of the species surveyed. Nuclear genomic estimates also suggest much slower rates among baleen whales (0.048% $\text{bp}^{-1} \text{My}^{-1}$, Schlötterer et al. 1991; Alter et al. 2007) compared with the average for mammals (0.21–0.37% $\text{bp}^{-1} \text{My}^{-1}$, Bulmer et al. 1991; Li 1997; Makalowski and Boguski 1998; Kumar and Subramanian 2002; Hardison et al. 2003). Reported substitution rates for both nuclear and mitochondrial markers (table 1) are therefore approximately one order of magnitude slower than the average estimate within mammals. These rates have been attributed to the low weight-specific metabolic rates (correlated by large body sizes) and relatively long generation times of baleen whales (Schlötterer et al. 1991; Martin and Palumbi 1993; Nabholz et al. 2008), as well as protection from cosmic radiation in the marine environment (Schlötterer et al. 1991). A negative relationship between neutral substitution rates and both generation time and body size has also been reported for other taxa (e.g., Li et al. 1987; Bromham et al. 1996; Gillooly et al. 2005).

Key words: mutation rate, genomic, mitochondrial, whale, cetacean.

E-mail: jacksonjennifera@gmail.com.

Mol. Biol. Evol. 26(11):2427–2440. 2009
doi:10.1093/molbev/msp169
Advance Access publication July 31, 2009

© The Author 2009. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org

Because the Tertiary fossil record for extant mysticete taxa is sparse, the majority of molecular rate studies have used calibration points dating to the origin of the cetaceans. To estimate more recent divergence dates within the order, Sasaki et al. (2005) and Árnason et al. (2000) used a calibration point at 55 Ma (million years ago; the date of divergence between hippopotami and cetaceans). In a few estimates, an early Oligocene calibration point dating to the origins of Suborder Odontoceti (beaked whales) has also been considered (Sasaki et al. 2005; Nikaïdo et al. 2006). However, relative rate tests suggest that there may be differences in rates between odontocetes and mysticetes (Kimura and Ozawa 2002). In view of this, the use of fossil calibrations specific to baleen whales (e.g., Rooney et al. 2001; Alter et al. 2007), or models allowing for a relaxation of rates across lineages (e.g., Sasaki et al. 2005; Nabholz et al. 2008), are likely to improve estimates of the molecular timescale of baleen whale evolution.

Estimation of substitution rates would be further aided by an accurate phylogeny for baleen whales. However, there is still much uncertainty regarding the species-level phylogeny of the family Balaenopteridae (rorqual whales). Complete mitogenome sequences (Sasaki et al. 2005), short interspersed nuclear element (SINE) characterizations (Nikaïdo et al. 2006), and nuclear loci (Rychel et al. 2004) have not fully resolved the early branching order within balaenopterids, perhaps due to a rapid species radiation occurring at the origin of the balaenopterids. If so, genetic histories could differ across loci with respect to the branching order within the balaenopterids. A final consideration is biases introduced by taxon and intraspecific sampling. To date, substitution rate estimates in baleen whales have been based on a variety of different fossils and taxon sampling. Therefore, estimates are vulnerable to uncertainty associated with both the dates and the topology underlying a given rate.

Table 1
Published Estimates of Evolutionary Substitution Rates for the Baleen Whales

Locus	Average Rate (%) bp ⁻¹ My ⁻¹	Full Range Reported (%)	Fossil Calibration	Dates (Ma)	Reference
Mitochondrial					
Restriction fragment length polymorphism mtDNA	0.40–1.00	N/A	Balaenopterid radiation	6–15	Martin and Palumbi (1993)
Control region	0.75	0.50–1.00	TMRCAs Mysticetes/ Odontocetes	30–40	Hoelzel et al. (1991)
			Delphinid radiation	5–10	
			TMRCAs <i>Orcinus</i> / delphinids	10	
Control region	0.70–1.00	0.40–1.80	TMRCAs Balaenids/ Balaenopterids	20–25	Baker et al. (1993)
			TMRCAs Eschrichtiids/ Balaenopterids	10–15	
			Balaenopterid radiation	6–15	
Control region	1.80–2.20	1.20–3.70	TMRCAs Eubalaena/ Balaena	3.4	Rooney et al. (2001)
Cytochrome <i>b</i>	0.40	0.37–0.43	TMRCAs Hippopotamids/ Cetacea	52–58	Alter et al. (2007)
Cytochrome <i>b</i> third codon positions	0.70–0.80	0.30–2.00	TMRCAs Hippopotamids/ Cetacea	52–58	Nabholz et al. (2008)
Genomic					
Intronic loci	0.048	0.015–0.100	TMRCAs Hippopotamids/ Cetacea	52–58	Alter et al. (2007)

Here, we compare substitution rate estimates over a variety of genomic and mitogenomic loci constrained by the same fossil calibration dates for a set of five species considered sufficient to represent a skeletal phylogeny of the mysticetes. Southern right whales (*Eubalaena australis*) were used to root the phylogeny, whereas blue (*Balaenoptera musculus*), fin (*Balaenoptera physalus*), Antarctic minke (*Balaenoptera bonaerensis*), and humpback (*Megaptera novaeangliae*) whales were chosen to represent the balaenopterid ingroup species. Because of their uncertain taxonomic affinity (Rychel et al. 2004; Sasaki et al. 2005; Nikaido et al. 2006), we did not include gray whales (*Eschrichtius robustus*) in our full data set, although some nuclear intronic loci were available to test the impact of their inclusion on a subset of loci (Alter et al. 2007). Uniquely among phylogenetic studies of cetaceans, we also accounted for intraspecific variation by using multiple phased nuclear alleles obtained by sequencing multiple individuals representing each species. The intention of this approach is to account for potential bias due to heterogeneity in branch lengths and increase precision of the rate estimate by sampling across the available range of alleles for each species. Mitochondrial control region substitution rates were also estimated using a large intraspecies data set for humpback whales (2,979 sequences), rooted with fin whale and blue whale outgroups. As the most widely sequenced and most rapidly evolving segment of the mitochondrial genome, the control region has been able to illuminate the most recent evolutionary history of many animals (e.g., Vigilant et al. 1991; Baker et al. 1993). Finally, we implement a Bayesian approach toward estimating phylogenetic rates to account for uncertainty in fossil dates, branch lengths, and rate variation between lineages (Thorne et al. 1998; Drummond et al. 2006). Together, the nuclear and mitochondrial data sets presented in this study

utilize the largest number of independent loci to date for the baleen whales.

Materials and Methods

Sample Collection and DNA Extraction

For most species of whales, small samples of skin tissue were obtained at sea using a stainless-steel biopsy dart (Lambertsen 1987) or from stranded whales during necropsy. For Antarctic minke whales, tissue was purchased from whale meat markets, as described by Baker et al. (2000). Tissue samples were stored in 70% ethanol while in the field and subsequently kept in long-term storage at -80°C . Genomic DNA was extracted from tissue samples using standard organic extraction procedures (Sambrook et al. 1989) as modified for small samples by Baker et al. (1994) or from whale meat products by Whole-Genome Amplification (Lasken and Egholm 2003).

Mitogenomes and Control Region Sequences—from Genbank and Direct Sequencing

Mitogenome sequences (13 protein-coding genes) for the five mysticete species were obtained from published sources (Árnason et al. 1991, 1993; Sasaki et al. 2005). To account for intraspecific diversity, we included a further two humpback mitogenomes representing three of the major haplogroups of this species (*AE*, *CD*, and *IJ*, Baker et al. 1993). Amplification and sequencing of multiple humpback whale mitogenomes were performed using multiple primer pairs after an initial long polymerase chain reaction amplification procedure. Five primer pairs were used to amplify the whole mitochondrial genome (Carragher 2004). Internal primers were designed to sequence in both directions across each protein-coding gene. All sequences were edited and

aligned in Sequencher (version 4.1.2). Sequence quality was evaluated using “Phred” or ABI quality control scores (Ewing and Green 1998; Ewing et al. 1998). Electropherograms were manually checked for sequencing errors and variable positions were confirmed by reference to the corresponding reverse sequence. The concatenated mitogenome data set was aligned to humpback mitogenome AP006467 (Sasaki et al. 2005) and edited to remove all stop codons. Regions of protein-coding gene overlap (between ATP8/ATP6, ND4L/ND4, and ND5/ND6) were represented only once, to a combined alignment length of 11,271 bp.

Mitochondrial control region sequences were amplified from individual humpbacks sampled across all oceans (see supplementary table S1, Supplementary Material online). Mitochondrial DNA (MtDNA) control region primers were designed to include the majority of variable nucleotide positions in the whale control region (Baker and Medrano-González 2002). Sequence alignment and quality control (Phred) scoring were carried out as described above. These sequences were then aligned with other published sequences (Baker et al. 1998; Baker and Medrano-González 2002; Rosenbaum et al. 2002; Olavarria et al. 2007; Engel et al. 2008; Rosenbaum et al. in press), representing all oceanic and most regional populations of humpbacks worldwide over an aligned length of 465 bp (219 haplotypes; supplementary table S1, Supplementary Material online). The assembled data set was aligned to blue and fin whale sequences as outgroups (*B. physalus*, AY582748 and NC_001321; *B. musculus*, NC_001601 and AY582748).

Nuclear Introns

Ten nuclear introns were amplified using standard protocols (Saiki et al. 1988; Palumbi 1995) and published primers (Lyons et al. 1997): actin (*ACT*), catalase (*CAT*), cholinergic receptor nicotinic alpha 1 (*CHRNA*), esterase D-formylglutathione hydrolase (*ESD*), fibrinogen gamma (*FGG*), glucosidase beta acid (*GBA*), lactose (*LAC*), parathyroid hormone (*PTH*), rhodopsin (*RHO*), and glucose-6-diphosphate (*G6PD*). A survey of *ACT* alleles was also conducted by single-strand conformation polymorphism following the methods of Friesen et al. (1997). Primer sequences and optimal annealing temperatures used for each locus are shown in supplementary table S2, Supplementary Material online. Sequence alignment and quality control (Phred) scoring was carried out as described in the section above.

We did not attempt a complete taxonomic survey of baleen whales, choosing instead to account for intraspecific diversity in five taxa sufficient to represent the primary structure of the mysticete phylogeny. Introns were amplified from 70 to 80 humpback whales, 50 to 60 Antarctic minke whales, 1 to 3 blue whales, 1 to 3 fin whales, and 20 to 25 Southern right whales. Four introns (*ACT*, *CAT*, *CHRNA*, and *GBA*) were also amplified for a small number of odontocete species (see supplementary table S3, Supplementary Material online), whereas a fifth (*LAC*) was aligned to odontocete species available on GenBank. Outgroup sequences for a subset of loci were obtained from Palumbi and Baker (1994); Milinkovitch et al. (1998); Caballero et al. (2008)

Unique sequences (alleles) of the 10 intronic loci were aligned for four representatives of the Balaenopteridae

(humpback, fin, blue, and minke whales) and one balaenid (right whale). These will henceforth be referred to as the “mysticete” (MYST) data sets. Five of these loci were also aligned with additional odontocete taxa (see supplementary table S3, Supplementary Material online) and will be referred to as the “cetacean” (CET) data sets. Three of the “MYST” data sets (*ACT*, *ESD*, and *LAC*) were extended to include gray whale introns (described in Alter et al. 2007) for sensitivity analysis. Intronic sequences possessed variable (heterozygote) sites for which gametic (haplotype) phase was unknown. Phasing of loci (estimating the gametic haplotypes) was performed with the program “FastPhase” (Scheet and Stephens 2006) and checked manually (see supplementary table S4, Supplementary Material online). Variable sites were identified in MacClade 4.0 (Maddison DR and Maddison WP 2005). A small number of sequences of questionable phase were excluded from subsequent phylogenetic analysis.

Phylogenetic Framework and Fossil Constraints

We chose to constrain some nodes within each phylogeny to obtain a rate consistent with the known relationships within the mysticetes. As there is no consensus agreement regarding interspecies relationships within the balaenopterids, the taxonomic constraints we imposed on the mysticete phylogeny were limited to the most strongly supported taxonomic groupings; species monophyly and balaenopterid monophyly (Deméré et al. 2005; Sasaki et al. 2005). We also explored 1) the effect of constraining the phylogeny to a sister-group relationship between humpbacks and fin whales as this has substantial support from many molecular data sets (Sasaki et al. 2005; Nikaido et al. 2006, 2007; Deméré et al. 2008) 2) the effect of including gray whale taxa on estimates of rates.

We considered three fossil nodes to be most suitable for the rate estimation.

“Node A”: The earliest fossil currently described within the Balaenopteridae is *Balaenoptera ryani* (Deméré et al. 2005). This fossil is assigned to the Tortonian (early late Miocene), an epoch currently dated at 7.25–11.61 Ma (Gradstein and Ogg 2004). An absolute minimum origin date for the balaenopterids is therefore 7.25 Ma, or more likely around 8 Ma, assuming morphological acquisition of taxonomically distinguishing characters will lag behind molecular divergence (see Jablonski and Bottjer 1990; Cooper and Fortey 1998). Consequently, we used a normally distributed constraint on the time to most recent common ancestor (TMRCA) within this family, with mean 12 Ma and 95% posterior probabilities (PPs) from 8 to 16 Ma \pm 2SD (standard deviation).

“Node B”: The earliest stem-group balaenid has been found in New Zealand (23–28.4 Ma; lower Oligocene) and is described by Fordyce (2002). This time period is consistent with the balaenid–balaenopterid divergence time of 27.3 \pm 1.9 Ma estimated by Sasaki et al. (2005) for mitochondrial sequences using the hippopotamid–cetacean and odontocete–mysticete divergence dates as calibrants. Consequently, we set the date of divergence between balaenids (right whales) and balaenopterids (node B) as a normal distribution of 25–31 Ma, mean estimate 28 Ma \pm 1.5SD (fig. 1).

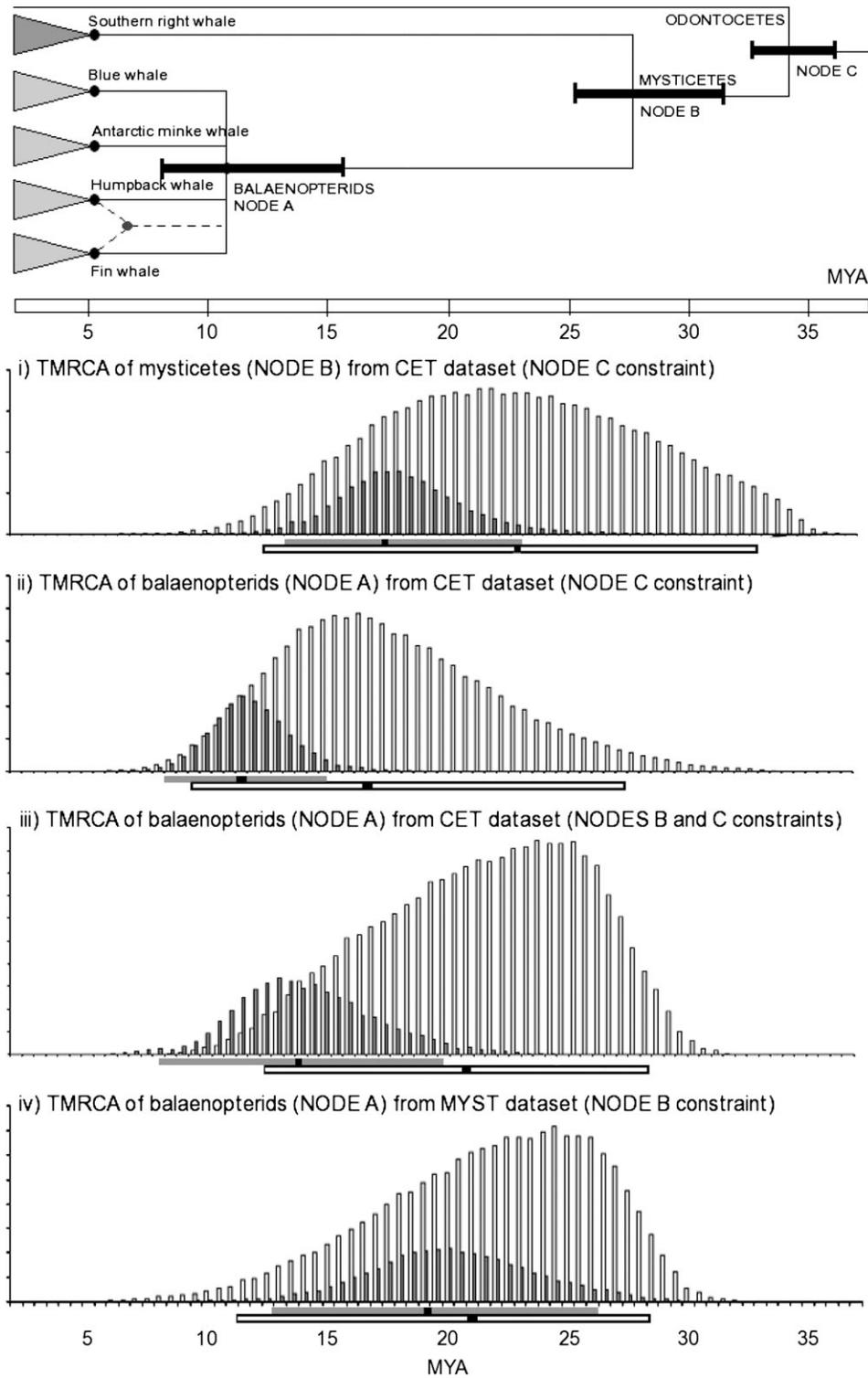


FIG. 1.—Phylogenetic framework for estimating divergence times (top) and distribution plots of substitution rates among the baleen whales (below). Filled black circles at nodes represent phylogenetic constraints, whereas the dashed line and filled circle uniting fin and humpback whales represents an additional constraint explored in a subset of analyses. Prior constraints on divergence times at nodes A, B, and C (95% range) are shown as black bars along branches. Posterior distributions are rooted either (i–iii) with an odontocete outgroup (CET data sets) or (iv) a southern right whale outgroup (MYST data sets). Beneath each distribution plot, the 95% probability interval for mitogenome third position sites (above, filled bar) and nuclear loci (below, white bar) are given, with mean values represented by a black square.

“Node C”: The earliest fossil mysticete is thought to be *Llanocetus denticrenatus* (Fordyce 1989; Mitchell 1989), found in the lower Eocene. Archaeocete sister taxa

to odontocetes and mysticete are only slightly older than this (e.g., Uhen 1999). Given that mysticetes are absent from the dense Eocene cetacean fossil record (Uhen

1998), it has been argued that a mysticete–odontocete divergence time close to the appearance of *Llanocetus* is likely (Nikaido et al. 2001). In view of this, we set the date of divergence between mysticetes and odontocetes (node C) as a normal distribution with a mean estimate of 34.5 Ma \pm 0.75SD.

Phylogenetic Reconstruction by Locus

The consensus phylogeny of the mitogenomic locus was estimated in a Bayesian framework using the program BEAST (v.1.4.7, Drummond and Rambaut 2007) and run for 10,000,000 generations (initial burn-in 1,000,000 generations) under a codon-partitioned model (general time reversible model; GTR+ gamma [G] + I parameters and branch lengths estimated separately for each codon position). Mitogenomic third codon sites were also analyzed as a separate data set as these sites are under less evolutionary constraint and are therefore expected to reflect a neutral rate. A control region phylogeny was reconstructed in the same Bayesian framework under an evolutionary model incorporating differential rates for transitions and transversions (Hasegawa–Kishino–Yano model, Hasegawa et al. (1985)) and a gamma model of rate variation. Posterior summary distributions were inspected for convergence and mixing in Tracer v1.4 (Rambaut and Drummond 2007).

Both mitogenomic and nuclear data sets were evaluated in a parsimony framework using MacClade 4.0 (Maddison DR and Maddison WP 2005), with gaps weighted equally to variable sites (each gap event was treated as one character irrespective of length). Maximum parsimony (MP) trees were obtained by heuristic searching with the Tree-bisection and reconnection branch swapping algorithm (100,000 replicates initiated with randomly constructed initial trees). Retention, homoplasy and consistency indices (CIs, Farris 1989) were estimated in PAUP 4.0b10 (Swofford 2002) and inspection of parsimony informative (PI) sites was performed in MacClade 4.0. Tree lengths were calculated for each hypothesis of interspecies relationships by summing tree lengths across all individual loci.

Rate Estimation

As inappropriate DNA substitution models can have a substantial effect on estimates of substitution rate in BEAST (Emerson 2007), we selected a model of DNA substitution for each genetic locus using Modeltest v3.7 (Posada and Crandall 1998). A second-order Akaike Information Criterion (AIC_c) score was calculated for each locus with branch lengths included as additional parameters in likelihood ratio testing and a correction for small sample sizes employed (Hurvich and Tsai 1989). The best fitting models were then used to choose the model of sequence evolution employed in each Bayesian analysis. A data set containing a concatenation of all nuclear loci was also constructed. For each locus, the allelic consensus from each species was constructed (i.e., all polymorphisms were collapsed) in MacClade 4.0, and loci were concatenated to a combined length of 4,517 bp. Divergence time analysis

was performed in a Bayesian framework using the program BEAST v1.4.7 (Drummond and Rambaut 2007). Monte Carlo Markov chains (MCMC) were used to coestimate phylogeny and substitution rates. After the chain has undergone an initial “burn-in” period (10% of the chain length in this analysis), posterior estimates of substitution rates and evolutionary parameters are averaged across tree space, with each tree weighted proportionally to its PP. A “Yule” process of speciation, which assumes a constant speciation rate per lineage, adapted from Yang and Rannala (1997), was chosen to describe branch splitting for the nuclear and mitogenomic phylogenies. Because the phylogenies include both intraspecies and interspecies divergences, substitution rates for two loci (*ACT* and *FGG*) were also estimated with alternate tree priors: the birth–death process and constant coalescent models. A constant coalescent prior was used to describe branch splitting in the control region data set. In order to test the sensitivity of this assumption, analyses were also repeated with a Skyline coalescent prior that integrates the rates analysis over a variety of demographic histories, sampled according to their likelihood (Drummond et al. 2005).

Taxonomic and divergence time constraints were imposed as described in the “Phylogenetic framework and fossil constraints” section. Data were analyzed with a relaxed uncorrelated lognormal clock. This model is able to accommodate data that are close to “clock-like” with a relatively high accuracy (Drummond et al. 2006). Sensitivity of estimates to relaxed-clock and constraint assumptions was evaluated by analyzing all introns under a strict clock and with relaxed taxonomic constraints. In order to evaluate the influence of within-species polymorphisms (multiple alleles for each species) on rate estimates, we constructed 10 one-allele-per-species MYST data sets for the *ACT* and *FGG* loci, randomly choosing each allele from among the alleles available for each species. Data sets were analyzed in BEAST under a strict clock scenario, with the Yule speciation prior and node B divergence time constraint imposed, under the same models of sequence evolution as were chosen for the multiple-allele data sets.

Two-cluster tests of rate variation were performed on the five CET data sets (LINTREE, Takezaki et al. 1995) with artiodactyl outgroups included (see supplementary table S3, Supplementary Material online) in order to identify possible rate variation between odontocetes and mysticetes. Trees tested for rate variation were first constructed in a maximum likelihood framework, with taxonomic constraints as described in the “Phylogenetic framework and fossil constraints” section. Two-cluster tests compared the differences in average substitutions per site between odontocetes and mysticetes relative to an artiodactyl outgroup, performing a two-tailed normal deviation test for every branch. For the control region data set, only the most recent divergence time constraint (node A, below) was explored. All Bayesian analyses were run for 5–10 million generations and were inspected for stationarity (effective mixing and convergence to the posterior distribution) using the program “Tracer” (Rambaut and Drummond 2007), with the initial 10% of these generations (parameters estimated prior to effective mixing and convergence of the MCMC chain) discarded as burn-in.

Table 2
Summary Sequence Characteristics by Locus

	Aligned Length (bp)	Variable Sites	PI Sites	Indels	MP Tree Length	CI	RI	HI	No. of Gene Copies ^a					Total bp Sequenced
									Humpback	Right	Minke	Blue	Fin	
Mitogenome	11,307	2296	951	12	2933	0.82	0.62	0.182	4 (3)	1 (1)	1 (1)	1 (1)	1 (1)	90,456
<i>ACT</i>	848	35	23	3	44	0.91	0.95	0.09	636 (8)	40 (3)	104 (7)	2 (2)	2 (2)	664,832
<i>CHRNA</i>	331	30	17	3	41	0.83	0.90	0.17	160 (5)	42 (5)	122 (15)	4 (4)	4 (1)	109,892
<i>GBA</i>	287	11	4	0	12	0.92	0.92	0.08	160 (3)	46 (1)	118 (5)	4 (2)	4 (1)	95,284
<i>PTH</i>	267	12	8	0	15	0.87	0.88	0.13	156 (5)	46 (4)	108 (4)	N/A	N/A	82,770
<i>CAT</i>	520	39	29	6	49	0.88	0.95	0.83	142 (4)	46 (5)	100 (10)	4 (4)	4 (2)	153,920
<i>LAC</i>	596	30	19	1	36	0.89	0.92	0.11	148 (4)	44 (2)	116 (10)	4 (2)	6 (2)	189,528
<i>RHO</i>	190	15	10	2	18	0.74	0.85	0.63	158 (12)	40 (4)	112 (4)	2 (1)	4 (2)	60,040
<i>ESD</i>	782	37	27	6	49	0.90	0.96	0.10	142 (8)	46 (5)	102 (5)	2 (1)	4 (2)	231,472
<i>FGG</i>	1,008	36	21	3	43	0.91	0.96	0.09	144 (7)	44 (4)	100 (13)	4 (2)	2 (2)	296,352
<i>G6PD</i>	257	10	7	1	11	1.00	1.00	1.00	50 (3)	19 (2)	46 (3)	1 (1)	2 (1)	30,326
Total nuclear	5,086	255	165	25										1,914,416
Concatenated nuclear ^b	4,517	100	25	25	112	0.88	0.44	0.39	1	1	1	1	1	

NOTE.—RI, retention index; HI, homoplasy index for MP tree.

^a The number of gene copies is shown for each species, with minimum numbers of alleles (obtained using PHASE) given in parentheses.

^b *G6PD* not included in the concatenated data set.

Results

Mitogenomic and Control Region Substitution Rates

There was no evidence for significant differences in base composition of the control region, mitogenome, or third codon position data sets (χ^2 test for homogeneity, mitogenome $\chi^2 = 6.0$, control region $\chi^2 = 15$), although base composition of the third codon position data set was skewed (A = 42%, C = 34%, G = 4%, T = 19%, $\chi^2 = 19.8$). High rate heterogeneity was estimated for the control region data set ($\alpha = 0.14$), whereas heterogeneity was predictably lower for third codon positions ($\alpha = 4.43$). Model-test estimates of the best fitting sequence models for control region and third codon mitogenome sequences are shown in supplementary table S5, Supplementary Material online.

Two alternative divergence time constraints were explored for the mitochondrial control region at the origin of the balaenopterids (node A). The node A constraint described in the previous section (12 My \pm 2SD) gave the highest mean rate estimate (5.85% $\text{bp}^{-1} \text{My}^{-1}$). The Skyline prior (Drummond et al. 2005) provided similar rate estimates under both divergence time constraints (supplementary table S6, Supplementary Material online). The balaenopterid divergence time obtained by the mitogenomic third codon sites under the balaenid (node B) constraint (19.8 My \pm 4SD) was also employed as a divergence constraint, and this gave a 1.5 times slower rate of 3.94% $\text{bp}^{-1} \text{My}^{-1}$ (table 3). Among codon positions in the mitogenomic data set, the highest substitution rates were estimated for third codon positions (1–1.5% $\text{bp}^{-1} \text{My}^{-1}$; table 3). The two-cluster test of rate variation (Takezaki et al. 1995) within the third codon positions found significantly slower rates among mysticetes compared with odontocete outgroups ($P < 0.05$; supplementary table S7, Supplementary Material online). Branch-length tests also determined significant departures from the average rate across the phylogeny ($\chi^2 = 62$; $P < 0.01$), supporting previous reports for the *cytb* gene (Kimura and Ozawa 2002).

Intronic Data Sets and Substitution Rates

Intronic sequences varied in length between 190 and 1,008 bp (table 2). Numbers of variable sites ranged between 10 and 35 per locus, of which 4–23 per locus were PI. For each of the 10 genomic loci, between 1 and 15 alleles were identified for each species (table 2, GenBank Accession numbers GQ407272–GQ408882). Homologous intronic artiodactyl sequences were obtained by Blast searching on GenBank (www.ncbi.nlm.nih.gov) and are listed in supplementary table S3, Supplementary Material online. Optimal alignments resulted in 24 insertions and deletions (indels) within the MYST data set, of which 14 were deletions (average length 3.1 bp) and 10 were insertions (average length 1.1 bp) relative to balaenid–odontocete outgroups. The locus *G6PD* is assumed to be X-linked as it is in other mammals, for example, Lyons et al. (1997). This was the most conserved among the chosen loci, with one allele shared among humpback, minke, blue, and fin whales. Because X-linked loci are assumed to be subject to a different evolutionary tempo from other genomic regions (e.g., McVean and Hurst 1997), this locus was excluded from subsequent evolutionary rate analyses.

The combined individual autosomal substitution rate estimates obtained using the three divergence time constraints were all of very similar magnitude (table 4). The fastest mean rates were obtained for the CET data sets (five loci) when all divergence time constraints were imposed (0.057% $\text{bp}^{-1} \text{My}^{-1}$), whereas the slowest rates were obtained by the MYST data sets when node B was constrained alone (0.045% $\text{bp}^{-1} \text{My}^{-1}$) and by the CET data sets when nodes B and C were constrained (0.049% $\text{bp}^{-1} \text{My}^{-1}$). The concatenated autosomal substitution rate estimate (node B, table 4) was slower (0.032% $\text{bp}^{-1} \text{My}^{-1}$) than that obtained from the combined individual sets, although confidence intervals overlapped. This suggests that the removal of within-species polymorphisms substantially reduces estimated genetic distance across the phylogeny. This difference could also be a function

Table 3
Mitochondrial Substitution Rate Estimates for Baleen Whales

Locus	Length (bp)	Species ^a	Node Date ^b	Substitution Rate ^c % bp ⁻¹ My ⁻¹			
				Lower 2.5% PP	Mean	Median	Upper 2.5% PP
Control region							
Constant ^d	465	1,2,3	A	2.969	5.853	5.558	9.321
		1,2,3	A'	1.666	3.943	3.502	7.087
Mitogenome	11,307	1,2,3,4,5	A/B	0.340	0.395	0.393	0.455
		1,2,3,4,5	B	0.263	0.312	0.310	0.367
		1,2,3,4,5,6	B/C	0.516	0.578	0.576	0.643
		1,2,3,4,5,6	C	0.548	0.635	0.629	0.726
		1,2,3,4,5,6	A/B/C	0.538	0.603	0.599	0.681
		COV ^e					
Third codon ^f	3,769	0.36	A/B	1.012	1.222	1.210	1.445
		0.27	B	0.802	0.996	0.987	1.208
		0.53	B/C	1.673	2.033	2.018	2.417
		0.22	C	1.269	1.459	1.445	1.671
		0.49	A/B/C	1.740	2.086	2.070	2.457

^a Species key 1) humpbacks, 2) fin whale, 3) blue whale, 4) minke whale, 5) right whale (outgroup for MYST data sets), 6) odontocetes (outgroup for CET data sets).

^b Divergence date constraints as indicated in the text. Estimates labeled A' are constrained to the estimate of balaenopterid divergence reported under the node B scenario for mitogenome third codon positions.

^c Substitution rates are expressed as percentage change (%) per base pair (bp) per million years (My).

^d Coalescent prior describing the distribution of branch bifurcations in the control region data set. A constant prior assumes a fixed population size over time.

^e COV (covariance) estimates describe deviation of each data set from a strict molecular clock.

^f Species used for the third codon site analyses are the same as the mitogenome analyses for the corresponding node dates.

of poor evolutionary model specification for the concatenated data set, each component of which is subject to slightly different evolutionary model fits under the AIC_c criterion (supplementary table S5, Supplementary Material online).

No significant rate variation between odontocetes and mysticetes was detected by cluster or branch-length comparisons for autosomal loci (supplementary table S7, Supplementary Material online). Summary and locus-specific estimates of rate variation ("COV"; table 4) suggested that when odontocetes were included in the data set, phylogenies were more clock-like. Within the MYST data set, the largest coefficient of variation was estimated for *CAT* (1.71) under the node B constraint, rejecting a strict molecular clock (table 4).

Sensitivity of estimates to alternate constraint and rates scenarios was explored for the node B-constrained divergence estimate. Genomic substitution rates were identical with gray whales included in the alignment (supplementary table S8, Supplementary Material online). Under a strict clock (node B) scenario (supplementary table S9, Supplementary Material online), genomic substitution rates were slightly higher in the taxonomically constrained framework (mean 0.044%/My) than when unconstrained (0.041% bp⁻¹ My⁻¹). Relaxing the assumption of fin whale and humpback whale monophyly had no effect. Substitution rates estimated under a coalescent tree prior were slightly faster than those estimated with speciation (Yule or birth-death) priors for both *ACT* and *FGG* loci under a strict clock (supplementary fig. 1, Supplementary Material online). Because the data include a mixture of coalescent and species-level divergences, neither of these priors was considered optimal and the true rate probably lies between the two estimates presented. When intraspecific polymorphisms were excluded from these data sets, estimates of substitution rate across all single-allele data sets were very similar in magnitude but less precise (supplementary fig. 1, Supplementary Material online). We interpreted this as reflecting greater

uncertainty in the probabilistic estimates of sequence model parameters across these single-allele phylogenies.

Dates of Divergence

Balaenopterid Radiation (Node A)

Posterior distributions provided an insight into the agreement of molecular rates and fossil dates (fig. 1). For the combined autosomal data set, mean time to the most recent common ancestor of balaenopterids (node A) was 21.5 Ma (95% PP 11.6–28.6 Ma) when only node B was constrained. A similar genomic estimate for this radiation was also obtained when nodes B and C were constrained (21.4 Ma, 95% PP 12.5–28.7 Ma). More recent estimates were obtained by the CET genomic data set when constrained at node C only (mean 17.0 Ma, 95% PP 9.7–26.7 Ma).

Mitogenomic third codon site estimates of divergence ranged more widely, with the node B constraint on the MYST data set giving a mean estimate of 19.8 Ma (95% PP 13.0–26.8 Ma). For the CET data set, constraining node C alone produced an estimate of 11.6 Ma (95% PP 8.5–15.6 Ma), whereas nodes B and C produced estimates of 14.0 Ma (95% PP 8.8–20.4 Ma).

Balaenid Divergence (Node B)

When constrained only by node C, estimated balaenid divergence dates from the CET data sets varied between autosomal loci (mean 22.3 Ma, 95% PP 12.7–32.9 Ma) and mitogenomic third codon sites (17.9 Ma, 95% PP 13.1–23.8 Ma). Both mean estimates gave divergence times more recent than indicated by the fossil record (28 Ma, Fordyce 2002).

Prior Constraints on Divergence Times

When both nodes A and B were constrained, the posterior distribution of divergence dates for the origin of the

Table 4
Genomic Substitution Rate Estimates by Locus for Baleen Whales

Locus	COV ^a	Species ^b	Node Date ^c	Substitution Rate ^d % bp ⁻¹ My ⁻¹				
				Lower 2.5% PP	Mean	Median	Upper 2.5% PP	
Nuclear								
<i>ACT</i>	0.38	1,2,3,4,5	A/B	0.039	0.063	0.061	0.089	
	0.32	1,2,3,4,5	B	0.030	0.054	0.054	0.080	
	0.26	1,2,3,4,5,6	B/C	0.049	0.069	0.068	0.090	
	0.27	1,2,3,4,5,6	C	0.048	0.071	0.070	0.094	
<i>CHRNA</i>	0.21	1,2,3,4,5,6	A/B/C	0.054	0.094	0.072	0.073	
	0.39	1,2,3,4,5	A/B	0.037	0.064	0.062	0.094	
	0.32	1,2,3,4,5	B	0.026	0.048	0.047	0.072	
	0.27	1,2,3,4,5,6	B/C	0.033	0.050	0.049	0.069	
<i>CAT</i>	0.26	1,2,3,4,5,6	C	0.035	0.054	0.052	0.075	
	0.26	1,2,3,4,5,6	A/B/C	0.040	0.059	0.058	0.081	
	1.28	1,2,3,4,5	A/B	0.039	0.063	0.061	0.089	
	1.71	1,2,3,4,5	B	0.045	0.078	0.076	0.113	
<i>LAC</i>	0.63	1,2,3,4,5,6	B/C	0.046	0.067	0.066	0.089	
	0.46	1,2,3,4,5,6	C	0.051	0.076	0.076	0.106	
	0.66	1,2,3,4,5,6	A/B/C	0.006	0.079	0.078	0.104	
	0.69	1,2,3,4,5	A/B	0.028	0.045	0.044	0.067	
<i>GBA</i>	0.45	1,2,3,4,5	B	0.020	0.035	0.034	0.052	
	1.23	1,2,3,4,5,6	B/C	0.015	0.026	0.025	0.038	
	0.38	1,2,3,4,5,6	C	0.048	0.066	0.065	0.087	
	1.55	1,2,3,4,5,6	A/B/C	0.021	0.035	0.034	0.050	
<i>RHO</i>	0.73	1,2,3,4,5	A/B	0.017	0.039	0.037	0.062	
	0.80	1,2,3,4,5	B	0.012	0.031	0.029	0.053	
	0.58	1,2,3,4,5,6	B/C	0.019	0.034	0.033	0.053	
	0.46	1,2,3,4,5,6	C	0.019	0.038	0.036	0.060	
<i>PTH</i>	0.56	1,2,3,4,5,6	A/B/C	0.022	0.039	0.039	0.058	
	0.41	1,2,3,4,5	A/B	0.035	0.065	0.066	0.102	
	0.43	1,2,3,4,5	B	0.027	0.055	0.053	0.086	
	0.56	1,4,5	A/B	0.022	0.053	0.050	0.088	
<i>ESD</i>	0.61	1,4,5	B	0.018	0.049	0.046	0.087	
	0.91	1,2,3,4,5	A/B	0.025	0.043	0.042	0.061	
	0.62	1,2,3,4,5	B	0.019	0.034	0.033	0.049	
	0.33	1,2,3,4,5	A/B	0.016	0.026	0.025	0.037	
<i>FGG</i>	0.44	1,2,3,4,5	B	0.011	0.020	0.019	0.030	
	Combined							
	0.65	1,2,3,4,5	A/B	0.021	0.051	0.049	0.095	
	0.80	1,2,3,4,5	B	0.015	0.045	0.041	0.077	
Concatenated	0.59	1,2,3,4,5,6	B/C	0.019	0.049	0.049	0.086	
	0.42	1,2,3,4,5,6	C	0.023	0.055	0.054	0.097	
	0.65	1,2,3,4,5,6	A/B/C	0.025	0.057	0.057	0.096	
	0.56	1,2,3,4,5	B	0.022	0.032	0.032	0.044	

^a COV (covariance) estimates describe deviation of each data set from a strict molecular clock.

^b Species key 1) humpbacks, 2) fin whale, 3) blue whale, 4) minke whale, 5) right whale (outgroup for MYST data sets), 6) odontocetes (outgroup for CET data sets).

^c Divergence date constraints as indicated in the text.

^d Substitution rates are expressed as percentage change (%) per bp per million years (MY).

balaenopterids (node A) was skewed older than the constraint prior (mean divergence time 13.7 Ma; 95% PP 10.1–17.3 Ma), whereas the posterior distribution of divergence of balaenids from balaenopterids (node B) was skewed toward the present (27.3 Ma; 95% PP 24.3–30.3 Ma, supplementary fig. 2, Supplementary Material online). Nearly identical estimates were obtained from the mitogenome third codon position analysis: node A 13.8 Ma (95% PP 10.2–17.3 Ma) and node B (27.3 Ma, 95% PP 24.3–30.2 Ma).

Phylogenetic Concordance

The mitogenomic data set, including multiple genomes for the humpback whale, provided 100% Bayesian posterior support for a basal position for minke whales and a sister-group relationship between humpbacks and fin

whales, with blue whales as a sister-group proximal to these two (fig. 2A). The same taxonomic relationship was reconstructed under MP, with 100% bootstrap support (1,000 replicates) for all relationships.

This high consistency was not reflected in the nuclear introns. Within the balaenopterids, combined autosomal tree lengths (320 steps) strongly supported a sister-group relationship for fin and blue whales (100% Bayesian PP), with only weak support (<50% PP) for the humpback as basal (fig. 2B). However, three loci supported a sister-group relationship among humpbacks and fin whales in a parsimony framework (fig. 3A and B), a result that is also supported by mitogenomes (fig. 2A).

CIs among genomic loci were higher (0.829–1.000) than for the mitogenomes (0.818). However, plots of the PI characters supporting various hypotheses of

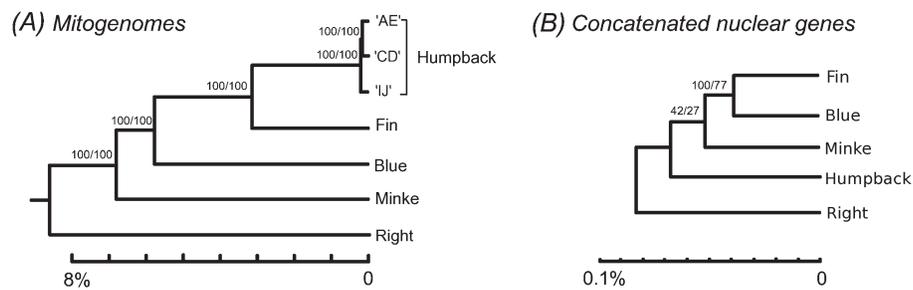


FIG. 2.—Phylograms depicting the relationships among baleen whales based on (A) mitogenome (protein-coding sequences) and (B) concatenated nuclear intronic sequences. Percentage support from Bayesian PPs and parsimony bootstraps, respectively, are shown above branches.

relationships within the Balaenopteridae (fig. 3) revealed that despite the small number of variable sites across all loci (table 2), both *LAC* and *CAT* showed conflicting character support. Two *LAC* characters supported a monophyletic clade uniting humpback, blue, and minke whales (excluding fin whales), whereas one supported the monophyly of fin and blue whales and another supported the monophyly of minke and fin whales. Two *CAT* characters supported a monophyletic clade containing fin, blue, and minke whales, whereas one insertion supported the monophyly of humpback and blue whales. Synapomorphic characters in *FGG*, *GBA*, *RHO*, and *ACT* were concordant with the mitogenome phylogeny (fig. 2A), whereas *ESD* and *CAT* supported an alternative where humpbacks fall basal within balaenopterids (fig. 3D).

All nuclear genomic loci possessed at least two characters uniting balaenopterids relative to the right whale (although only one right whale allele was available for *GBA*). Blue and fin whale alleles (1–2 individuals each) were identified by diagnostic sites from four loci (shown in fig. 3). Diagnostic sites unique to minke whales (>50 individual samples) were provided by four loci, whereas only two loci provided diagnostic sites for humpbacks (>70 individual samples).

Discussion

Slow Substitution Rates

Our analysis provides the most comprehensive and systematic evidence to date for a slow rate of molecular evolution in baleen whales. Rate estimates from third codon positions of the baleen whale mitogenome ($1\% \text{ bp}^{-1} \text{ My}^{-1}$ across 13 genes) were 10 times slower relative to other mammals (mean mammalian *cytb* third codon position rate is $9.8\% \text{ bp}^{-1} \text{ My}^{-1}$, Nabholz et al. 2008) but consistent with previously published estimates for the *cytb* gene in mysticetes (Kimura and Ozawa 2002; Nabholz et al. 2008, fig. 4). Substitution rate estimates from whole mitogenomes of baleen whales were significantly slower ($0.3\% \text{ bp}^{-1} \text{ My}^{-1}$, table 3) and similar to those estimated for elephants ($0.4\% \text{ bp}^{-1} \text{ My}^{-1}$, Rohland et al. 2007). Evidence from genomic studies suggests that organismal longevity operates as a decelerating factor on both mitochondrial and nuclear genomes (Nikolaev et al. 2007; Nabholz et al. 2008) but acts more strongly on mitochondria than on nuclear DNA. Body mass (as an inverse correlate for

per-gram metabolic rates) also exerted a weaker effect on both genomes (Nabholz et al. 2008) but again may have a greater influence in the mitochondria, where the oxidative metabolic processes take place.

The estimated mitochondrial control region substitution rate ($3.94\% \text{ bp}^{-1} \text{ My}^{-1}$) is two to three times faster than previously reported for baleen whales (Baker et al. 1993; Rooney et al. 2001). This could be due to the additional numbers of segregating sites sampled from the extremely large data set used in our analysis. Under comparable divergence time constraints, the control region rate is significantly faster than the third codon position rate, a result consistent with other studies and observations of mutational “hotspots” in the control region (Endicott and Ho 2008). Comparison of rate estimates from different priors on population size suggests that this is unlikely to have a large effect relative to the interspecies distance among humpbacks and the balaenopterid outgroups.

Our estimates of substitution rates for the nuclear genome in mysticetes ($0.045\% \text{ bp}^{-1} \text{ My}^{-1}$) are nearly one order of magnitude slower than those estimated for mammals across multiple neutral loci ($0.21\text{--}0.37\% \text{ bp}^{-1} \text{ My}^{-1}$ Bulmer et al. 1991; Li 1997; Makalowski and Boguski 1998; Kumar and Subramanian 2002; Hardison et al. 2003) and are consistent with those previously estimated for the gray whale ($0.048\% \text{ bp}^{-1} \text{ My}^{-1}$, Alter et al. 2007, fig. 4).

The consistent methodology and parallel analysis of mitochondrial and nuclear genomes provide an accurate comparison of the relative rate of evolution in the two genomes. The roughly 20-fold difference in these rates is one of the smallest ratios among mammals (which can vary up to 100-fold) and is similar to that seen in other long-lived animals, including primates and humans (Nabholz et al. 2008). This observation is consistent with the mitochondrial theory of aging, which predicts strong negative selection on mtDNA in organisms with long generation times (expressed as the average age of sexually mature females), driven by increasing free radical exposure over time for gametes (Harman 1957; Barja and Herrero 2000). Differences in rates between nuclear and mitochondrial regions could also be driven by differences in DNA repair machinery (e.g., Croteau and Bohr 1997; Croteau et al. 1999; Liu et al. 2008), which may have evolved differently between taxonomic groups, and differences in numbers of gamete cell replications between males and females (Li et al. 2002). Whether long generation times act as a predictor

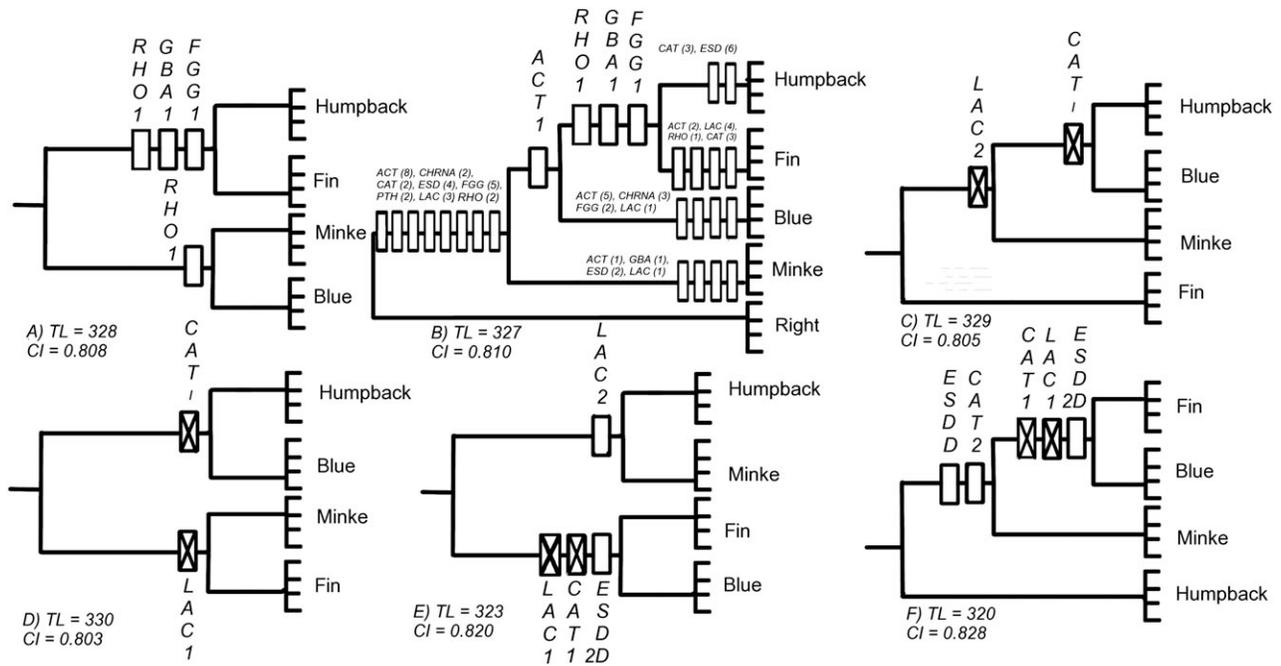


FIG. 3.—Phylogenetic hypotheses of relationships among the baleen whales, showing the numbers of supporting nuclear characters by locus on each branch. *D* indicates a deletion event, *I* an insertion event (e.g., *ESD 2D* describes two supporting variable sites plus one deletion event). *TL* indicates the combined genomic tree length supporting each phylogenetic hypothesis. Loci supporting each hypothesis are denoted by a clear rectangular box; if there is support for a conflicting hypothesis within that same locus, the box is crossed. Loci possessing variable sites supporting species-level and balaenopterid monophyly are shown in figure 3b (slender rectangles) and described above branches.

for a low ratio of mitochondrial to nuclear rates in other mammals has yet to be investigated and represents an interesting avenue for further research.

It is notable that genomic human nuclear mutation rates are roughly double that of cetaceans ($\sim 2.5 \times 10^{-8} \text{ gen}^{-1}$, Nachman and Crowell 2000) despite similar average generation times (25–28 years in humans; Fenner 2005, and 18–52 years in mysticete whales, respectively; Taylor et al. 2007). Generation times for chimpanzees (commonly used as outgroups for human rate estimates) are roughly estimated at 19–24 years (Matsumura and Forster 2008), suggesting that generation time has not changed substantially in human prehistory. This rate difference could therefore reflect an additional influence of scaled differences in per-gram metabolic rates between humans and cetaceans, because metabolic rates scale directly with body mass (West et al. 1997). However, such an inference is difficult to test directly, as there are no estimates of basal metabolic rates for baleen whales under field conditions.

Divergence Times within the Mysticetes

A lack of morphological characters and the scarcity of key specimens have made fossil assignment within mysticetes particularly challenging (Fordyce and Barnes 1994; Milinkovitch 1995). To identify sources of conflict and the extent of rate sensitivity to potential assignment error, we compared substitution rate estimates across multiple calibration points. We note several points of agreement and disagreement. Our estimates of divergence times obtained from the *CET* data sets (odontocete taxa included) are con-

sistent with the first appearance of balaenopterids in the fossil record (Deméré et al. 2005) but inconsistent with the first appearance of balaenids (Fordyce 2002), which predates the estimate of divergence. Both mitochondrial and nuclear data sets support this conclusion, but mitochondrial divergence estimates can be biased toward the present by rate variation between odontocete and mysticete lineages, as

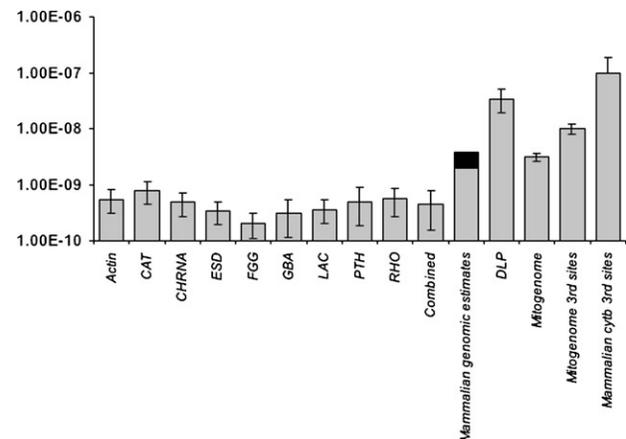


FIG. 4.—Mean substitution rates ($\% \text{ bp}^{-1} \text{ My}^{-1}$) of nuclear and mitochondrial genomes in baleen whales, shown on a log scale. Rates estimated using the balaenid–balaenopterid divergence time constraint (node B) and mysticete data sets (MYST). Ninety-five percent probability intervals are shown in black. Average *cytb* gene third codon site rates are as reported in Nabholz et al. (2008). Average mammalian genomic estimates are reported from a range of sources (Bulmer et al. 1991; Li 1997; Makalowski and Boguski 1998; Kumar and Subramanian 2002; Hardison et al. 2003) and are shown in black.

suggested previously from studies of the *cytb* gene (Kimura and Ozawa 2002) and also found in this study.

The inconsistency in divergence dates and fossils could be driven by undetected (and uncorrected) rate variation between mysticetes and odontocetes in the nuclear genome. Although our study detected no rate variation between odontocetes and mysticetes for autosomal genes, the low genetic variability across introns reduced the power of branch-length tests to detect departures from rate constancy (Bromham et al. 2000). Detection of rate variation in such a slowly evolving species will require a much larger sample of genomic DNA, although rate differences will always be difficult to detect, given the slow substitution rate estimated for the nuclear genome. A second possibility is that the early Oligocene calibration used to constrain the odontocete–mysticete divergence is an underestimate of true divergence time. A reasonable possibility is that the genetic divergence predates the acquisition of morphologically defining characters for the clade (and therefore their identification in the fossil record) and therefore that the calibration date is too recent. However, mitochondrial estimates based on earlier fossil calibrations also predicted dates ranging 34–36 Ma for this divergence (Sasaki et al. 2005; Nikaido et al. 2006).

The MYST data sets (balaenid constraint only, node B) supported an early Miocene divergence time for the balaenopterids, which greatly precedes (~8 My) appearance of their fossils in the late Miocene but is consistent with estimates from mitogenomes (Sasaki et al. 2005). This is in contrast to the CET data sets (node C constraint), which supported a mid-Miocene divergence time for the balaenopterids but which could also be subject to biases due to rate variation between odontocetes and mysticetes. If the early divergence time for the balaenopterids is correct, it implies a phylogenetic “fuse,” during which the crown group of balaenopterid whales is absent from the fossil record. The absence of crown balaenopterids in the fossil record at this time is surprising, as the Miocene period is relatively dense in other whale species, such as cetotheres. However, this may be due to difficulties in morphological identification of stem-group balaenopterids during this period rather than a true absence (see Sasaki et al. 2005). The alternative possibility that the balaenid–balaenopterid divergence is more recent awaits a more formal palaeontological description of the early balaenid fossil noted in Fordyce (2002).

Constraint of the MYST data sets to both balaenid and balaenopterid divergence times (nodes A and B) produced a slight conflict in the data sets, which was consistent over all loci. The posterior distributions of divergence dates around the constraint nodes A and B were skewed toward one another, suggesting that the genetic distance corresponding to the period of time between the divergence of the balaenids within the mysticetes (node B), and the subsequent radiation of balaenopterids (node A), is shorter than has been imposed by the constraints. This suggests that either the balaenopterid divergence (8–16 Ma) or the balaenid–balaenopterid divergence (28 Ma) is more recent than has been assumed based on the fossil record.

Noting the caveats discussed above, the balaenid-only (node B) constraint is the only fossil calibration to provide divergence times that are consistent with the fossil record

(i.e., predating fossil appearances rather than lagging them) and with the pattern of branch lengths across the tree. For this reason, we conclude that this scenario represents the best fit to the available data and provides the most likely genomic and mitochondrial substitution rates for mysticetes. Based on the results of our preferred calibration, we would predict that future fossil discoveries will extend the fossil record for balaenopterids back to the middle Miocene.

Gene Trees and Species Trees

There was a surprising lack of agreement between the phylogenetic reconstruction for the mitochondrial and nuclear loci used in this study. The mitogenomes strongly supported a sister-taxa relationship for humpbacks and fin whales, consistent with previous mitogenomic studies (Árnason et al. 2004; Sasaki et al. 2005). In contrast, the combined autosomal loci in our study supported a sister-taxa relationship between fin whales and blue whales. This was consistent with short interspersed nuclear element (SINE) loci (Nikaido et al. 2006) and the “Common Cetacean” long satellite repeat by Árnason et al. (1992), which both provided support for this sister-taxa relationship.

One possible explanation is that a rapid crown group radiation immediately following the evolution of “lunge-feeding” and the balaenopterid form could have led to incomplete lineage sorting of autosomal genes and corresponding differences in genetic histories among independent loci. This possibility has been raised by Nikaido et al. (2006) as a likely explanation for the observed inconsistencies in SINE loci and by Deméré et al. (2008) who identified an indel character in the *ENAM* gene uniting minke whales and balaenids.

A second explanation for the conflict is that hybridization among these species has eroded their phylogenetic signature through introgression and/or recombination. It is known that blue and fin whales can produce fertile hybrids (e.g., Spilliaert et al. 1991; Berubé and Aguilar 1998), a phenomenon which has not been confirmed among humpbacks and any other balaenopterid species (although a suspected blue/humpback hybrid has been photographed in French Polynesia; Poole M, unpublished data). The fitness of such hybrids, and their potential contribution to the descendent gene pool, is unknown. However, there is potential for hybrids to have caused considerable introgression among balaenopterids as they all have the same karyotype ($2n = 44$) and thus have full gamete and chromosomal compatibility (Árnason 1972, 1974). Under the scenario of rapid radiation, nonrecombining mitochondria would be expected to provide more consistent estimate of mysticete genetic history, although lineage sorting could still be unrepresentative of organismal phylogeny. Hybrid introgression could also obscure true interspecies relationships inferred from either mitochondrial or nuclear DNA. Consistent with this hypothesis, synapomorphic characters in individual autosomal loci supported multiple hypotheses of relationships. However, this could also be an effect of weak phylogenetic signal because individual loci are both slowly evolving and relatively short in length. The inclusion of further mysticete taxa is anticipated to provide a more robust picture of the evolutionary history of the balaenopterids,

and it is hoped that the work presented here can provide a useful framework for more detailed exploration of such questions.

Conclusions

1. Substitution rates estimated in a phylogenetic framework over multiple autosomal introns and all mitochondrial protein-coding genes confirm an approximately 10-fold slower rate of evolution within the mysticetes relative to most mammals. The slow rates are consistent for both mitochondrial and nuclear genes, although new control region rate estimates are higher than previously reported.
2. The slowdown is most consistent with an influence of longevity or generation time as well as body mass/metabolic rate. Despite similar longevity and generation times, humans exhibit a 5× higher genomic substitution rate, consistent with the hypothesis of body mass–metabolic rate differences.
3. Systematic relationships estimated from nuclear introns using multiple alleles to represent each species are only weakly consistent with mtDNA, perhaps reflecting 1) rapid radiation or 2) interspecies hybridization/recombination.

Supplementary Material

Supplementary figures 1–2 and supplementary tables 1–9 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

Funding was provided by a Marsden Grant to CSB (01-UOA-070) from the New Zealand Royal Society and a grant from the Lenfest Program of the Pew Charitable Trust to S.R.P. L.M.G. was supported by grants from Universidad Nacional Autónoma de México and Consejo Nacional de Ciencia y Tecnología. We thank the South Pacific Whale Research Consortium for providing DNA samples and H. Rosenbaum for access to unpublished control region sequences. The authors thank K. Ruegg and R. Hamner for a review of previous drafts of the manuscript, E. Fordyce for comments on fossil calibration, A. Drummond for comments on tree priors, L. Lyons for assistance with primer design, and B. Congdon for genotyping of Actin alleles. We would also like to thank two anonymous reviewers and the Assistant Editor for their useful comments.

Literature Cited

Alter SE, Rynes E, Palumbi SR. 2007. DNA evidence for historic population size and past ecosystem impacts of gray whales. *Proc Natl Acad Sci USA*. 104:15162–15167.

Árnason A, Gullberg A, Widegren B. 1991. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. *J Mol Evol*. 33:556–568.

Árnason A, Gullberg A, Widegren B. 1993. Cetacean mitochondrial DNA control region: sequences of all extant baleen whales and two sperm whale species. *Mol Biol Evol*. 10:960–970.

Árnason Ú. 1972. The role of chromosomal rearrangements in mammalian speciation with special reference to Cetacea and Pinnipeda. *Hereditas*. 70:113–118.

Árnason Ú. 1974. Comparative chromosome studies in Cetacea. *Hereditas*. 77:1–36.

Árnason Ú, Grétarsdóttir S, Widegren B. 1992. Mysticete (Baleen whale) relationships based on the sequence of the common cetacean DNA satellite. *Mol Biol Evol*. 9:1018–1028.

Árnason Ú, Gullberg A, Grétarsdóttir S, Ursing B, Janke A. 2000. The mitochondrial genome of the sperm whale and a new molecular reference for estimating Eutherian divergence dates. *J Mol Evol*. 50:569–578.

Árnason U, Gullberg A, Janke A. 2004. Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene*. 333:27–34.

Baker CS, Lento GM, Cipriano F, Palumbi SR. 2000. Predicted decline of protected whales based on molecular genetic monitoring of Japanese and Korean markets. *Proc Roy Soc Lond Ser B Biol Sci*. 267:1191–1199.

Baker CS, Medrano-González L. 2002. Worldwide distribution and diversity of humpback whale mitochondrial DNA lineages. In: Pfeiffer CJ, editor. *Molecular and cell biology of marine mammals*. Malabar (FL): Krieger Publishing Company. p. 84–99.

Baker CS, Medrano-Gonzalez L, Calambokidis J, Perry A, Pichler F, Rosenbaum H, Straley JM, Urban-Ramirez J, Yamaguchi M, Von Ziegler O. 1998. Population structure of nuclear and mitochondrial DNA variation among humpback whales in the North Pacific. *Mol Ecol*. 7:695–707.

Baker CS, Perry A, Bannister JL, et al. (13 co-authors). 1993. Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. *Proc Natl Acad Sci USA*. 90:8239–8243.

Baker CS, Slade RW, Bannister JL, Abernethy RB, Weinrich MT, Lien J, Urban J, Corkeron P, Calambokidis J, Vasquez O. 1994. Hierarchical structure of mitochondrial DNA gene flow among humpback whales *Megaptera novaeangliae*, worldwide. *Mol Ecol*. 3:313–327.

Barja G, Herrero A. 2000. Oxidative damage to mitochondrial DNA is inversely related to maximum life span in the heart and brain of mammals. *FASEB J*. 14:312–318.

Berubé M, Aguilar A. 1998. A new hybrid between a blue whale, *Balaenoptera musculus*, and a fin whale, *B-physalus*: frequency and implications of hybridization. *Mar Mammal Sci*. 14:82–98.

Bromham L, Penny D, Rambaut A, Hendy MD. 2000. The power of relative rates tests depends on the data. *J Mol Evol*. 50:296–301.

Bromham L, Rambaut A, Harvey PH. 1996. Determinants of rate variation in mammalian DNA sequence evolution. *J Mol Evol*. 43:610–621.

Bulmer M, Wolfe KH, Sharp PM. 1991. Synonymous nucleotide substitution rates in mammalian genes: implications for the molecular clock and the relationship of mammalian orders. *Proc Natl Acad Sci USA*. 88:5974–5978.

Caballero S, Jackson JA, Mignucci-Giannoni AA, Barrios-Garrido H, Beltran-Pedrerros S, Montiel-Villalobos MA, Robertson KM, Baker CS. 2008. Molecular systematics of South American dolphins *Sotalia*: sister taxa determination and phylogenetic relationships, with insights into a multi-locus phylogeny of the Delphinidae. *Mol Phylogenet Evol*. 46:252–268.

Carraher CJF. 2004. *Comparative mitogenomics of the Southern hemisphere dolphin genus Cephalorhynchus*. Auckland

- (New Zealand): School of Biological Sciences, University of Auckland. p. 161.
- Cooper A, Fortey R. 1998. Evolutionary explosions and the phylogenetic fuse. *Trends Ecol Evol.* 13:151–156.
- Croteau DL, Bohr VA. 1997. Repair of oxidative damage to nuclear and mitochondrial DNA in mammalian cells. *J Biol Chem.* 272:25409–25412.
- Croteau DL, Stierum RH, Bohr VA. 1999. Mitochondrial DNA repair pathways. *Mut Res.* 434:137–148.
- Deméré TA, Berta A, McGowen MR. 2005. The taxonomic and evolutionary history of fossil and modern balaenopterid mysticetes. *J Mamm Evol.* 12:99–143.
- Deméré TA, McGowen MR, Berta A, Gatesy J. 2008. Morphological and molecular evidence for a stepwise evolutionary transition from teeth to baleen in mysticete whales. *Syst Biol.* 57:15–37.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenies and dating with confidence. *PLoS Biol.* 4:699–710.
- Drummond AJ, Rambaut A. 2007. BEAST v1.4.7. Oxford: University of Oxford.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian inference of past population dynamics from molecular sequences. *Mol Biol Evol.* 22:1185–1192.
- Emerson BC. 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 SYW. 2008. A Bayesian evaluation of human mitochondrial substitution rates. *Am J Hum Genet.* 82:895–902.
- Engel MH, Fagundes NJR, Rosenbaum HC, Leslie MS, Ott PH, Schmitt R, Secchi E, Dalla Rosa L, Bonatto SL. 2008. Mitochondrial DNA diversity of the Southwestern Atlantic humpback whale (*Megaptera novaeangliae*) breeding area off Brazil, and the potential connections to Antarctic feeding areas. *Conserv Genet.* 9:1253–1262.
- Ewing B, Green P. 1998. Basecalling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8:186–194.
- Ewing B, Hillier L, Wendl M, Green P. 1998. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8:175–185.
- Farris JS. 1989. The retention index and homoplasy excess. *Syst Zool.* 38:406–407.
- Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol.* 128:415–423.
- Fordyce RE. 1989. Origins and evolution of Antarctic marine mammals. *Geol Soc Lond Spec Publ.* 47:269–281.
- Fordyce RE. 2002. Oligocene origins of skim-feeding right whales: a small archaic balaenid from New Zealand. *J Vertebr Paleontol.* 22:54a.
- Fordyce RE, Barnes LG. 1994. The evolutionary history of whales and dolphins. *Annu Rev Earth Planet Sci.* 22:419–455.
- Friesen VL, Congdon BC, Walsh HE, Birt TP. 1997. Intron variation in marbled murrelets detected using analyses of single-stranded conformational polymorphisms. *Mol Ecol.* 6:1047–1058.
- Gillooly JF, Allen AP, West GB, Brown JH. 2005. The rate of DNA evolution: effects of body size and temperature on the molecular clock. *Proc Natl Acad Sci USA.* 102:140–145.
- Gradstein FM, Ogg JG. 2004. Geologic Time Scale 2004 - why, how, and where next!. *Lethaia.* 37:175–181.
- Hardison RC, Roskin KM, Yang S, et al. (17 co-authors). 2003. Covariation in frequencies of substitution, deletion, transposition, and recombination during Eutherian evolution. *Genome Res.* 13:13–26.
- Harman D. 1957. Aging: a theory based on free radical and radiation chemistry. *J Gerontol.* 2:298–300.
- Hasegawa M, Kishino K, Yano T. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 22:160–174.
- Hoelzel AR, Hancock JM, Dover GA. 1991. Evolution of the cetacean mitochondrial D-loop region. *Mol Biol Evol.* 8:475–493.
- Hurvich CM, Tsai C-L. 1989. Regression and time series model selection in small samples. *Biometrika.* 76:297–307.
- Jablonski D, Bottjer DJ. 1990. The origin and diversification of major groups: environmental patterns and macroevolutionary lags. Pp. 17–58 In: Taylor PD, Larwood GP, editor. *Major evolutionary radiations.* Clarendon Press: Oxford.
- Kimura T, Ozawa T. 2002. Rates of mitochondrial DNA evolution are slower in mysticete relative to Odontocete Cetaceans. In: Pfeiffer CJ, editor. *Molecular and cell biology of marine mammals.* Malabar (FL): Krieger Publishing Company. p. 111–117.
- Kumar S, Subramanian S. 2002. Mutation rates in mammalian genomes. *Proc Natl Acad Sci, USA.* 99:803–808.
- Lambertsen RH. 1987. A biopsy system for large whales and its use for cytogenetics. *J Mamm.* 68:443–445.
- Lasken RS, Egholm M. 2003. Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. *Trends Biotechnol.* 21:531–535.
- Li W-H. 1997. *Molecular evolution.* Sunderland (MA): Sinauer Associates.
- Li WH, Tanimura M, Sharp PM. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J Mol Evol.* 25:330–342.
- Li WH, Yi S, Makova K. 2002. Male-driven evolution. *Curr Opin Genet Dev.* 12:650–656.
- Liu P, Qian L, Sung JS, de Souza-Pinto NC, Zheng L, Bogenhagen DF, Bohr VA, Wilson DM, Shen B, Dernple B. 2008. Removal of oxidative DNA damage via FEN1-dependent long-patch base excision repair in human cell mitochondria. *Mol Cell Biol.* 28:4975–4987.
- Lyons LA, Laughlin TF, Copeland NG, Jenkins NA, Womack JE, O'Brien SJ. 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nat Genet.* 15:47–56.
- Maddison DR, Maddison WP. 2005. *MacClade 4.* Sunderland (MA): Sinauer Associates, Inc.
- Makalowski W, Boguski M. 1998. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. *Proc Natl Acad Sci USA.* 95:9407–9412.
- Martin AP, Palumbi S. 1993. Body size, metabolic rate, generation time and the molecular clock. *Proc Natl Acad Sci USA.* 90:4087–4091.
- Matsumura S, Forster P. 2008. Generation time and effective population size in Polar Eskimos. *Proc Roy Soc Lond Ser B Biol Sci.* 275:1501–1508.
- McVean GT, Hurst LD. 1997. Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. *Nature.* 386:388–392.
- Milinkovitch MC. 1995. Molecular phylogeny of cetaceans prompts revision of morphological transformations. *Trends Ecol Evol.* 10:328–334.
- Milinkovitch MC, Bérubé M, Palsbøll PJ. 1998. Cetaceans are highly specialized artiodactyls. In: Thewissen JGM, editor. *The emergence of whales: evolutionary patterns in the origin of Cetacea.* New York: Plenum.
- Mitchell ED. 1989. A new cetacean from the Late Eocene La Meseta formation, Seymour Island, Antarctic Peninsula. *Can J Fish Aquat Sci.* 46:2219–2235.

- Nabholz B, Glémin S, Galtier N. 2008. Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Mol Biol Evol.* 25:120–130.
- Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics.* 156:297–304.
- Nikaido M, Hamilton H, Makino H, Sasaki T, Takahashi K, Goto M, Kanda N, Pastene L, Okada N. 2006. Baleen whale phylogeny and a past extensive radiation event revealed by SINE insertion analysis. *Mol Biol Evol.* 23: 866–873.
- Nikaido M, Matsuno F, Hamilton H, Brownell RL Jr., Cao Y, Ding W, Zuoyan Z, Shedlock AM, Fordyce RE, Hasegawa M, Okada N. 2001. Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. *Proc Natl Acad Sci USA.* 98:7384–7389.
- Nikolaev SI, Montoya-Burgos JI, Popadin K, Parand L, Margulles EH, National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program, Antonarakis SE. 2007. Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. *Proc Natl Acad Sci USA.* 104:20443–20448.
- Nishida S, Goto M, Pastene L, Kanda N, Koike H. 2007. Phylogenetic relationships among cetaceans revealed by Y-chromosome sequences. *Zool Sci.* 24:723–732.
- Olavarría C, Baker CS, Garrigue C, et al. (19 co-authors). 2007. Population structure of South Pacific humpback whales and the origin of the eastern Polynesian breeding grounds. *Mar Ecol Prog Ser.* 330:257–268.
- Palumbi SR. 1995. Nucleic acids II: the polymerase chain reaction. In: Hillis D, Moritz C, Mable BK, editors. *Molecular systematics.* Sunderland (MA): Sinauer Associates, Inc.
- Palumbi SR, Baker CS. 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Mol Biol Evol.* 11:426–435.
- Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics.* 14:817–818.
- Rambaut A, Drummond AJ. 2007. Tracer: MCMC trace analysis tool v1.4. Oxford: University of Oxford.
- Rohland N, Malaspina A, Pollack JL, Slatkin M, Matheus P, Hofreiter M. 2007. Proboscidean mitogenomics: chronology and mode of elephant evolution using *Mastodon* as outgroup. *PLoS Biol.* 5:1663–1671.
- Rooney AP, Honeycutt RL, Derr JN. 2001. Historical population size change of bowhead whales inferred from DNA sequence polymorphism data. *Evolution.* 55:1678–1685.
- Rosenbaum HC, Pomilla CC, Leslie MC, et al. (21 co-authors). *PLoS ONE*, in press. Mitochondrial DNA diversity and population structure of humpback whales from their breeding areas in the south Atlantic and Indian Oceans (breeding regions A, B, C and X).
- Rosenbaum HC, Weinrich MT, Stoleson SA, Gibbs JP, Baker CS, DeSalle R. 2002. The effect of differential reproductive success on population genetic structure: correlations of life history with matriline in humpback whales of the Gulf of Maine. *J Hered.* 93:389–399.
- Rychel AL, Reeder TW, Berta A. 2004. Phylogeny of mysticete whales based on mitochondrial and nuclear data. *Mol Phylogenet Evol.* 32:892–901.
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science.* 239:487–491.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual.* Cold Spring Harbor (NY): Cold Spring Harbor Laboratory.
- Sasaki T, Nikaido M, Hamilton H, et al. (10 co-authors). 2005. Mitochondrial phylogenetics and evolution of mysticete whales. *Syst Biol.* 54:77–90.
- Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Amer J Hum Genet.* 78:629–644.
- Schlötterer C, Amos B, Tautz D. 1991. Conservation of polymorphic simple sequence loci in cetacean species. *Nature.* 354:63–65.
- Spilliaert R, Vikingsson G, Árnason Ú, Palsdóttir A, Sigurjonsson J, Árnason A. 1991. Species hybridization between a female blue whale (*Balaenoptera musculus*) and a male fin whale (*B. physalus*): Molecular and morphological documentation. *J Hered.* 82:269–274.
- Swofford DL. 2002. PAUP: phylogenetic analysis using parsimony, 4.0b.10. Sunderland (MA): Sinauer Associates.
- Takezaki N, Rzhetsky A, Nei M. 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol Biol Evol.* 12:823–833.
- Taylor BL, Chivers SJ, Larese J, Perrin WF. 2007. Generation length and percent mature estimates for IUCN assessments of cetaceans. La Jolla (CA): National Marine Fisheries Service, Southwest Fisheries Science Center. p. 1–24.
- Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol.* 15:1647–1657.
- Uhen MD. 1998. Middle to late Eocene Basilosaurines and Dorudontines. In: Thewissen JGM, editor. *The emergence of whales.* New York: Plenum Press. p. 29–61.
- Uhen MD. 1999. New species of protocetid Archaeocete whale, *Eocetus wardii* (Mammalia: Cetacea) from the middle Eocene of North Carolina. *J Paleontol.* 73:512–528.
- Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. *Science.* 253:1503–1507.
- West GB, Brown JH, Enquist BJ. 1997. A general model for the origin of allometric scaling laws in biology. *Science.* 276:122–126.
- Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Mol Biol Evol.* 14:717–724.

Koichiro Tamura, Associate Editor

Accepted July 4, 2009