

Mitochondrial Phylogenetics and Evolution of Mysticete Whales

TAKESHI SASAKI,^{1,2} MASATO NIKAIIDO,¹ HEALY HAMILTON,³ MUTSUO GOTO,⁴ HIDEHIRO KATO,⁵
NAOHISA KANDA,⁴ LUIS A. PASTENE,¹ YING CAO,^{6,7} R. EWAN FORDYCE,⁸ MASAMI HASEGAWA,^{6,7}
AND NORIHIRO OKADA^{1,2}

¹Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku Yokohama, Kanagawa 226-8504, Japan;
E-mail: nokada@bio.titech.ac.jp (N.O.)

²Department of Evolutionary Biology and Biodiversity, National Institute for Basic Biology, Myodaiji, Okazaki, Japan

³California Academy of Sciences, Golden Gate Park, San Francisco, California 94118, USA; and Ecosystem Sciences Division, I SPAL,
University of California, Berkeley California 94720, USA

⁴The Institute of Cetacean Research, 4-5 Toyomi-cho, Chuo-ku, Tokyo 104-0055, Japan

⁵National Research Institute of Far Seas Fisheries, Cetacean Population Biology Section, 5-7-1 Orido, Shimizu, Shizuoka 424-8633, Japan

⁶Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan

⁷Department of Biosystems Science, Graduate University for Advanced Studies, Shonan Village, Hayama, Kanagawa 240-0193, Japan

⁸Department of Geology, University of Otago, P.O. Box 56, Dunedin, New Zealand

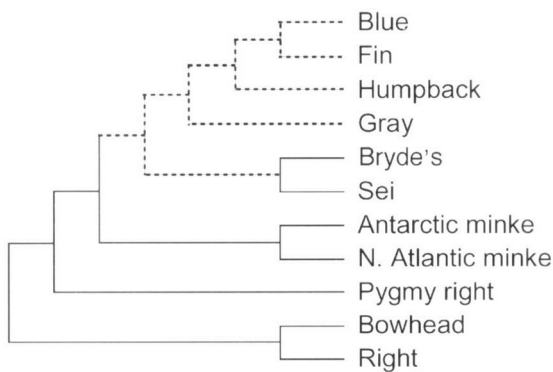
Abstract. The phylogenetic relationships among baleen whales (Order: Cetacea) remain uncertain despite extensive research in cetacean molecular phylogenetics and a potential morphological sample size of over 2 million animals harvested. Questions remain regarding the number of species and the monophyly of genera, as well as higher order relationships. Here, we approach mysticete phylogeny with complete mitochondrial genome sequence analysis. We determined complete mtDNA sequences of 10 extant Mysticeti species, inferred their phylogenetic relationships, and estimated node divergence times. The mtDNA sequence analysis concurs with previous molecular studies in the ordering of the principal branches, with Balaenidae (right whales) as sister to all other mysticetes base, followed by Neobalaenidae (pygmy right whale), Eschrichtiidae (gray whale), and finally Balaenopteridae (rorquals + humpback whale). The mtDNA analysis further suggests that four lineages exist within the clade of Eschrichtiidae + Balaenopteridae, including a sister relationship between the humpback and fin whales, and a monophyletic group formed by the blue, sei, and Bryde's whales, each of which represents a newly recognized phylogenetic relationship in Mysticeti. We also estimated the divergence times of all extant mysticete species, accounting for evolutionary rate heterogeneity among lineages. When the mtDNA divergence estimates are compared with the mysticete fossil record, several lineages have molecular divergence estimates strikingly older than indicated by paleontological data. We suggest this discrepancy reflects both a large amount of ancestral polymorphism and long generation times of ancestral baleen whale populations. [Ancestral polymorphism; baleen whale; evolution; mitochondrial DNA; molecular clock; phylogeny.]

Baleen whales, the largest of all living animals, are marine mammals that filter-feed on tiny prey. The 12 to 13 living species comprise the extant Mysticeti, one of the two suborders of the Cetacea. Many baleen whales were commercially harvested for more than two centuries, allowing significant advances in alpha taxonomy (description and preliminary classification; Mayr, 1969; LeDuc and Dizon, 2002), but also pushing some species close to extinction. There are currently four recognized families of mysticetes: the right whales (Balaenidae; three to four species), the pygmy right whale (Neobalaenidae, one species), the gray whale (Eschrichtiidae, one species), and the rorquals and humpback whale (Balaenopteridae; seven to eight species). With their large body size, oceanic distribution, and highly derived morphological characters, the baleen whales represent a challenge to cladistic analysis. Despite the wealth of scientific data represented by the harvested animals, few clear advances were made in understanding evolutionary relationships. Yet, a phylogeny of Mysticeti is the foundation for further understanding of their biology, ecology, conservation, and management.

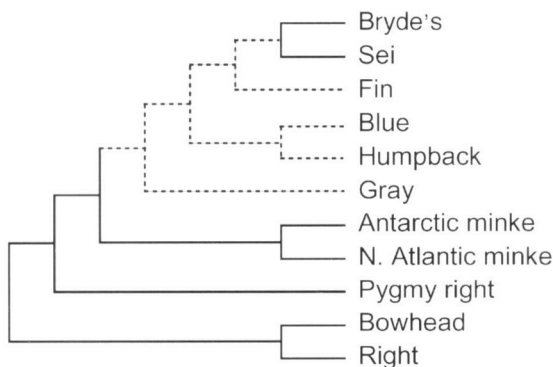
Previous studies of mysticete phylogeny have used both morphological and molecular approaches with varied success. Morphological analyses of modern and fossil material have generated diverse phylogenies with few common patterns (Barnes and McLeod, 1984; McLeod

et al., 1993; Geisler and Luo, 1996; Bisconti, 2000; Lindow, 2002; Sanders and Barnes, 2002; Geisler and Sanders, 2003). Although many conflicting arrangements of the families Balaenidae, Neobalaenidae, Eschrichtiidae, and Balaenopteridae have been proposed, there is general agreement that balaenids are the basal lineage of extant mysticetes. It is also widely accepted that the balaenopterids include two subfamilies, Balaenopterinae and Megapterinae, in recognition of the distinctness of the humpback whale, *Megaptera*. Consistently problematic issues are the position of the monotypic Neobalaenidae and Eschrichtiidae, and the relationships within Balaenopteridae. Overall, morphology has provided few insights into phylogenetic relationships, in part a reflection of the logistic difficulties in obtaining and comparing adequate sample sizes for mysticete whales (LeDuc and Dizon, 2002).

Investigation of mysticete phylogeny was an early target of molecular genetic techniques. Analysis of a common cetacean DNA satellite confirmed the early divergence of the Balaenidae, but removed the morphologically anomalous Neobalaenidae from their long-held association with this group, suggesting instead a closer association with the Eschrichtiidae and Balaenopteridae (Arnason and Best, 1991; Arnason et al., 1992). Analysis of mitochondrial control region sequences indicated balaenopterid paraphyly (Arnason et al., 1993;



A. Based on mtDNA complete control region sequences (Arnason et al. 1993)



B. Based on mtDNA complete cytochrome b sequences (Arnason and Gullberg 1994)

FIGURE 1. Previous molecular phylogenetic analyses of baleen whale based on partial mitochondrial DNA sequences (A and B). Dotted line indicates ambiguous phylogenetic relationships when these analyses are compared.

Fig. 1A). Cytochrome *b* sequence analysis also supported balaenopterid paraphyly, but suggested a different arrangement of relationships within the Eschrichtiidae and Balaenopteridae (Fig. 1B; Arnason and Gullberg, 1994). Thus, both morphological and molecular data have consistently failed to produce a well-supported phylogeny of mysticete whales, with relationships among the Balaenopteridae and Eschrichtiidae especially problematic. The mysticete mtDNA sequences presented here are the first major contribution to molecular phylogenetic analysis of mysticete relationships since Arnason and Gullberg (1994), while providing a vastly increased character set for phylogenetic analysis.

As a genetic marker, animal mitochondrial DNA (mtDNA) has distinct characteristics, such as relatively fast evolutionary rate (Brown et al., 1979), lack of recombination (Olivio et al., 1983), and maternal inheritance. Since the advent of polymerase chain reaction (PCR), phylogenetic analysis of partial mtDNA sequences has elucidated many branching patterns in the tree of life. Current advances in high-throughput DNA sequencing have spurred a robust new field of whole-mitochondrion phylogenetics (e.g., Grande et al., 2002; Nardi et al.,

2003). Although higher-order mammalian relationships are now routinely investigated with complete mtDNA sequence analyses (Nikaido et al., 2000, 2001a; Cao et al., 2000; Lin et al., 2002; Murata et al., 2003), there are few examples of beta-level studies, in which phylogenetic relationships of relatively close taxa are examined with complete mitochondrial sequences (Ingman et al., 2000; Kijas and Andersson, 2001).

This study has three main aims: (1) to determine and document the complete mtDNA sequences of ten Mysticeti species; (2) to clarify the uncertain phylogenetic relationships of all mysticete whale species by complete mtDNA sequence analysis; and (3) to use this extensive molecular data set to estimate divergence times of lineages leading to all major extant Mysticeti species, also comparing predictions with the fossil record. Previously, the divergence times for mysticete lineages were based mainly on fossil evidence, with only limited estimates based on early molecular studies (e.g., Wada and Numachi, 1991; Arnason and Gullberg, 1993). Integrating the mtDNA divergence estimates with the relevant mysticete fossil record provides further insights into the evolution of this clade.

MATERIALS AND METHODS

DNA Samples

The 10 baleen whale species for which we determined the complete mitochondrial genome sequence are shown in Table 1. One individual was sampled for each species. To confirm the species and its locality of each of four baleen whales, we performed phylogenetic analysis by neighbor-joining method (Saitou and Nei, 1987) (see Appendices 1 to 4, available at the Society of Systematic Biologists website, <http://systematicbiology.org>). They have distinct genetic populations, in which sufficient sequence data were accumulated and reported on the DNA data bank. Total genomic DNA was isolated from either liver or muscle tissue by phenol chloroform extraction (Sambrook et al., 1989) and stored at 4 °C. The mtDNA sequences for the blue whale *Balaenoptera musculus* and the fin whale *Balaenoptera physalus* were obtained from GenBank, accession numbers X72204 and X61145, respectively (Arnason and Gullberg, 1993; Arnason et al., 1991a).

LA-PCR of the Mitochondrial DNA Genome

Two sets of primers were designed for amplification of the complete mitochondrial genome in two fragments (Nikaido et al., 2001a). The amplicons were the "short" fragment, located from tRNA-Leu (CUN) to downstream of 16S rRNA, approximately 7000 bp in length, and the "long" fragment, starting downstream of 16S rRNA to tRNA-Leu (CUN), approximately 9000 bp in length. The short fragment primers were 5'-GGTCTTAGGAACC AAAAATTGGTGC AACTC-3' and 5'-CTCCGCCTCG AACTCAGATCACGTTAGGACT-3', and the long fragment primers were 5'-AGTCCTACGTGATCTGAGTTC AGACCGGAG-3' and 5'-GAGTTGCACCAATTTTTC GTTCCTAAGACC-3'. The long and accurate (LA)-PCR

TABLE 1. Specimen information for the 12 mysticete mt-genomes analyzed in this study.

Family	Scientific name	English name	Sampling location	Accession no.
Balaenopteridae	<i>Balaenoptera borealis</i>	Sei whale	Antarctic ocean	AP006470 (this work)
	<i>Balaenoptera brydei</i>	Bryde's whale	Northwest Pacific ocean	AP006469 (this work)
	<i>Balaenoptera acutorostrata</i>	North Atlantic minke whale	North Atlantic ocean	AP006468 (this work)
	<i>Balaenoptera bonaerensis</i>	Antarctic minke whale	Antarctic sea ocean	AP006466 (this work)
	<i>Balaenoptera physalus</i>	Fin whale	No geographic information reported	X61145 (Arnason and Gullberg, 1991)
	<i>Balaenoptera musculus</i>	Blue whale	No geographic information reported	X72204 (Arnason et al., 1993)
Eschrichtiidae	<i>Megaptera novaeangliae</i>	Humpback whale	Antarctic ocean	AP006467 (this work)
	<i>Eschrichtius robustus</i>	Gray whale	Eastern North Pacific (Southern California)	AP006471 (this work)
Balaenidae	<i>Balaena mysticetus</i>	Bowhead whale	Okhotsk Sea	AP006472 (this work)
	<i>Eubalaena australis</i>	Southern right whale	Antarctic ocean	AP006473 (this work)
	<i>Eubalaena japonica</i>	Northern right whale	Bering Sea	AP006474 (this work)
Neobalaenidae	<i>Caperea marginata</i>	Pygmy right whale	New Zealand	AP006475 (this work)

amplification profile consisted of 30 cycles of denaturation at 94 °C for 30 s, then annealing and extension at 68 °C for 15 min. The reaction mixture contained 2.5 units LA *Taq* polymerase (Takara), 1 × LA *Taq* buffer, 0.4 mM dNTPs, 2.5 mM MgCl₂, 10 pM primers, and 100 ng of genomic DNA, in a final volume 50 μL.

In order to determine the sequence for LA-PCR primer annealing, a separate PCR reaction was required. We amplified a 2113-bp fragment spanning from mid-ND4 to mid-ND5, and a second, 855-bp fragment spanning the middle of 16S rRNA to mid-ND1. The primer set for the ND4-ND5 fragment, which spans tRNA-Leu (CUN), was 5'-TGCAGCCGTAATAAACTTGG-3' and 5'-AGGGCTCAGGCGTTGGT-3'. The 16S rRNA-ND1 fragment was amplified with primers 5'-AACA GCGCAATCCTATTC-3' and 5'-AGGAGCCATTATTA GGAGT-3'. The PCR profile for amplification of these two fragments consisted of 30 cycles with denaturation at 94 °C for 30 s, annealing at 55 °C for 45 s, and extension at 72 °C for 1 min. The PCR mixture contained 2.5 units Ex *Taq* polymerase (Takara), 1 × Ex *Taq* buffer, 0.4 mM dNTPs, 10 pM primers, and 100 ng of genomic DNA in a final volume of 50 μL. Products of both LA-PCR and standard PCR were confirmed by electrophoresis in a 1.0% agarose gel (Takara) and stained with ethidium bromide for band characterization via ultraviolet transillumination.

Direct Sequencing and Primer Walking

Excess primers and nucleotides were removed from PCR products by treatment with exonuclease I (Exo I) and shrimp alkaline phosphatase (SAP). An SAP-Exo I mixture containing 1 unit of Exo I and 1 unit of SAP (USB Corporation) for 50 μL of PCR product was heated to 37 °C for 15 min and 85 °C for 15 min. Purified PCR products were used for direct cycle sequencing, with 25 cycles of denaturation at 96 °C for 30 s, annealing at 50 °C for 15 s, and extension at 60 °C for 1 min. The reaction mixture contained 4 μL BigDye ver. 3.0 terminator premix (Applied Biosystems), 5 pM sequence primer, and 5 μL purified LA-PCR products. Sequences were determined with an automated sequencer (Applied Biosystems, models

310 and 3100). Seventeen sets of universal primers for direct sequencing were derived from Nikaido et al. (2000). Sequences obtained from regions adjacent to these 17 primers were analyzed and a second set of primers was then designed. This "primer walking" procedure was repeated until the entire mitochondrial genome was sequenced.

Phylogenetic Analysis

The 12 proteins encoded in the same strand of mtDNA were prepared for phylogenetic analyses. Alignments of sequences were carefully checked by eye, and all positions with gaps or ambiguous alignment plus overlapping regions between ATP6 and ATP8 and between ND4 and ND4L were excluded. The total number of remaining codons is 3535. Furthermore, the small (12S) and the large (16S) mt-rRNA sequences were aligned manually by taking account of the secondary structure model (Cao et al., 1994), and again all positions with gaps or ambiguous alignment were excluded. The total number of remaining sites of 12S + 16S rRNA is 2416 bp.

The phylogenetic relationships by the neighbor-joining (NJ) and maximum parsimony (MP) analyses were constructed by 12 concatenated protein sequence data using MEGA version 2.1 program package (Kumar et al., 2001). In the NJ analysis, we chose the gamma distance (gamma shape parameter $\alpha = 0.5, 1.0, 2.0,$ and 3.0) for calculation of genetic distance and bootstrapped with 1000 replicates. MP tree was constructed by branch-and-bound search with 500 bootstrap resampling.

We evaluated the phylogenetic relationships by the ML method (Felsenstein, 1981; Kishino et al., 1990). We used the ProtML program in the MOLPHY package (version 2.3) (Adachi and Hasegawa, 1996a) and the TREE-PUZZLE program for quartet-puzzling (QP) analysis (Strimmer and Haeseler, 1996). We also used the CodeML program in the PAML package (version 3.14; Yang, 1997) for analysis of the amino acid sequences of the mt-proteins with the mtREV-F model (Adachi and Hasegawa, 1996b) and nucleotide sequences of the protein-encoding genes with the codon-substitution model (Goldman and Yang, 1994; Yang et al., 1998).

For the codon-substitution model, we used equal distance among 20 amino acids, and Grantham's (1974) and Miyata et al.'s (1979) distance both with linear and geometric formulae (Yang et al., 1998), and the best model was chosen with the Akaike information criterion (AIC), defined by

$$\text{AIC} = -2 \times (\log\text{-likelihood}) + 2 \\ \times (\text{number of parameters})$$

A model that minimizes AIC may be considered to be the most appropriate model (Akaike, 1974; Kishino and Hasegawa, 1990; Adachi et al., 1993). Furthermore, the BaseML program in PAML was used for the analyses of nucleotide sequences of rRNAs and of protein-encoding genes with the HKY+ Γ and the GTR+ Γ models (Hasegawa et al., 1985; Yang, 1994, 1996). In analyzing the protein genes with the BaseML, three codon positions were analyzed separately and then the whole protein-encoding nucleotide data set was evaluated by the TotalML program in MOLPHY.

In using the CodeML, BaseML, and TREE-PUZZLE programs, the discrete Γ distribution (with eight categories except for the codon-substitution model, where we used three categories because of the computational burden) for the site-heterogeneity (Yang, 1996) was adopted, and the shape parameter (α) of the Γ model was optimized. Bootstrap probabilities (BPs) were estimated by the REL method (Kishino et al., 1990) with 10,000 bootstrap resamplings. The REL method has been shown to be efficient in estimating BPs without performing ML estimation for each resampled data (Hasegawa and Kishino, 1994).

From preliminary analyses of the concatenated 12 mt-proteins by the NJ, MP, and TREE-PUZZLE methods, we identified some clades if supported with nearly 100% BPs or QP. Those well-supported clades that are noncontroversial were constrained in to reduce the number of candidate trees we analyzed with more computationally intensive approaches. Even so, the constrained searches were large to allow exhaustive analysis with a computationally intensive method, and we conducted an approximate likelihood analysis with the ProtML program for all the candidate trees. The most serious problem of the ML method when applied to data from many species is the explosively increasing number of possible trees. However, most of these trees are very bad and unpromising and we eliminated such trees by an approximate method. In estimating the branch lengths for each tree topology by the ML, we used the Newton-Raphson method, which is time consuming. The approximate likelihood option implemented in ProtML avoids this process and estimates an "approximate likelihood" from the initial values for the Newton-Raphson method given by the ordinary least-squares. We examined all the possible trees with the approximate likelihood method, excluded unpromising trees by this approximate criterion, and selected the best 10,000 trees for the full likelihood analy-

sis. It has been shown that there is a strong correlation between the approximate likelihood and the maximum likelihood, and that this is practically good method to reduce the computational burden (Adachi and Hasegawa, 1996a). Because even 10,000 trees were too many for the most sophisticated model, we further reduced the number of candidate trees by selecting the best trees among the 10,000, which have log-likelihood scores differing by less than 3 SEs from that of the highest likelihood tree with a simpler model.

Estimation of Branching Dates with a Relaxed Molecular Clock

From the sequences of mt-proteins, we estimated branching dates on the Mysticeti tree. However, the molecular clock does not hold for the mt-protein tree, as will be shown by the likelihood ratio test between the clock and nonclock model (Felsenstein, 1988). Therefore, the rate difference must be taken into account in estimating the dates. The Bayesian method of Thorne and colleagues (Thorne et al., 1998; Kishino et al., 2001) is useful for this purpose, as shown by Nikaido et al. (2001a, 2001b), Cao et al. (2000), and Hasegawa et al. (2003), who applied this method to cetacean SINE flanking sequences and mammalian mt-proteins. Here, we applied this method to the Mysticeti mt-proteins. By using cow as an outgroup, we calibrated the clock by choosing 55 ± 3 Mya (± 1 SD) for the hippopotamus/whales separation (McKenna and Bell, 1997; Bajpai and Gingerich, 1998; Waddell et al., 1999). We also set the Mysticeti/Odontoceti separation older than 34.2 Mya from the fossil evidence that the oldest reported fossil Mysticeti is the archaic toothed *Ulanocetus denticrenatus*, dated at about 34.2 Mya (Fordyce, 1989; Mitchell, 1989; Dingle and Lavelle, 1998).

RESULTS AND DISCUSSION

Features of the Mysticete Mitochondrial Genome

The mitochondrial genomes of the 10 mysticete species determined in the present study are in accordance with the general features of the mammalian mtDNA genome, consisting of 13 protein-coding genes, 2 rRNAs, and 22 tRNAs. Because the ND6 gene was coded on the strand opposite to the other 12 protein-coding genes, we excluded it from the phylogenetic analysis. The protein-coding genes start with the codons ATG or ATA, and stop with codons TAA, TAG, or AGA, as usual. For 10 of the 12 baleen species, the ND4L start codon is CAC. Only the humpback whale and fin whale (Arnason et al., 1991a) share ATG as the ND4L start codon, supporting the novel hypothesis of a sister relationship between these two species (see below).

Estimation of the Phylogenetic Tree Using Mt-Genome Data

Figure 2 shows a QP tree of the concatenated amino acid sequences of 12 mt-proteins. This tree indicates that the Antarctic/North Atlantic minke whales clade, the sei/Bryde's whales clade, and northern/southern right whales clade were confirmed all with 100% QP support

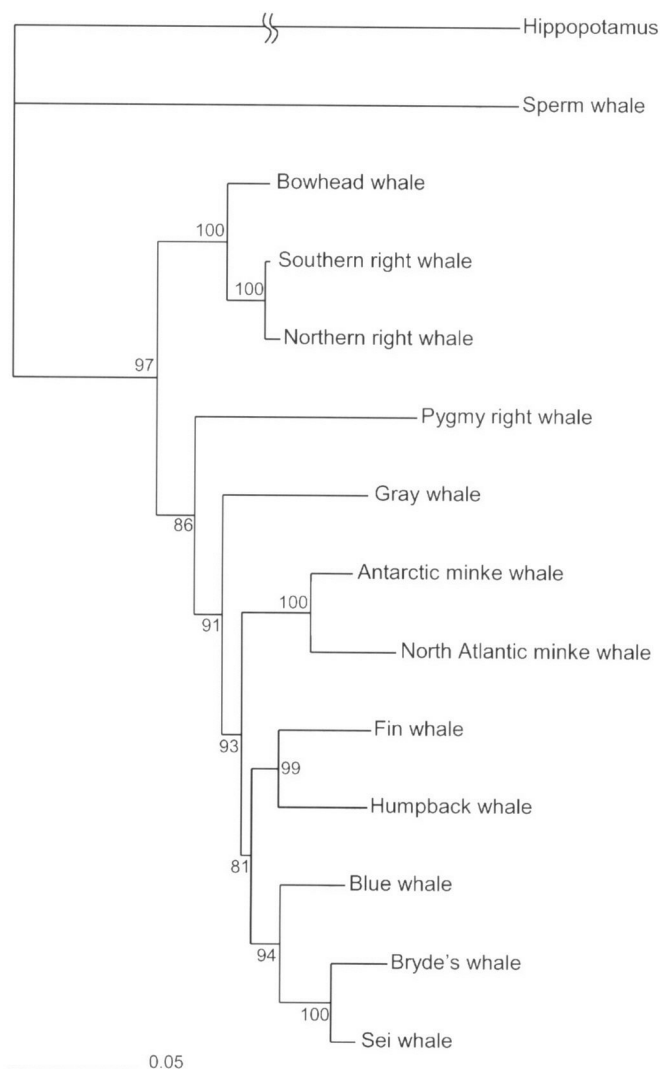


FIGURE 2. A quartet-puzzling tree (Strimmer and Haeseler, 1996) of the concatenated amino acid sequences of 12 mt-proteins with mtREV-F+ Γ model. The horizontal length of each branch is proportional to the estimated number of amino acid substitutions. Numbers indicate QP support values. Hippopotamus and sperm whale were used as outgroups.

These clades were also supported with 100% BP from the NJ using four kinds of gamma shape parameter (α) and MP analyses except 91% BP for the clades of the two minkes from MP (data not shown), and these clades were constrained in the subsequent maximum likelihood (ML) analyses. With hippopotamus and sperm whale as outgroups (Gatesy et al., 1996; Ursing and Arnason, 1998; Nikaido et al., 1999; Arnason et al., 2000), the number of possible trees among minke, fin, humpback, sei/Bryde's, blue, gray, pygmy right, right, and bowhead whales is 2,027,025. These trees were examined by using the approximate likelihood option of ProtML for the concatenated mt-proteins, and the best 10,000 trees were selected for full likelihood analyses by the CodeML program with the mtREV-F+ Γ model and by the BaseML program,

with the HKY+ Γ and GTR+ Γ models distinguishing among three codon positions (assign different parameters to different codon positions).

We think the amino acid sequence analysis with the CodeML is better than the nucleotide sequence analysis with the BaseML in the phylogenetic inference from the protein-encoding genes if synonymous substitutions are not very informative, because the latter does not take account of the correlation among different positions in a codon. The merit of nucleotide sequence analysis is that it can take account of synonymous substitutions that might be informative in the timescale of evolution we are analyzing in this work. Therefore, we also applied the codon-substitution model implemented in the CodeML program to the protein-encoding nucleotide sequences. Because the program is too slow to be applied to all 10,000 trees, we chose the best 939 trees among the 10,000, which have log-likelihood scores differing by less than 3 SEs from that of the highest likelihood tree by the amino acid analysis and were provided for an analysis with the codon-substitution model.

Figure 3 shows the ML tree of the concatenated mt-proteins with the mtREV-F+ Γ model and BPs for the proteins, rRNAs, and the total of proteins + rRNAs estimated by the REL method applied to the 10,000 candidate trees. Figure 4 shows the ML tree of the concatenated protein-encoding genes with the codon-substitution+ Γ model and BPs estimated by the REL method applied to the 939 candidate trees. In order to check whether the set of 939 trees provided for the detailed analyses well represents candidate trees, CodeML analyses of the amino acid sequences were performed for the 939 trees and BPs were shown in Figure 4. It turned out that these estimates do not differ from those shown in Figure 3 estimated from the 10,000 candidate trees. This justifies our choosing of only 939 trees for the most sophisticated analysis. The codon-substitution model used in Figure 4 is based on geometric Grantham's (1974) distance among 20 amino acids with the vertebrate mitochondrial code table. This distance gives the best AIC and therefore represents the most appropriate model examined in Table 2. In Table 2, the conventional models of independent nucleotide substitution, which distinguishes among different codon positions, are compared with the codon-substitution models by using AIC. The best model of independent substitution (GTR+ Γ) is much worse than any of the codon-substitution models, even the equal-distance model which does not take account of physicochemical distances among amino acids and the worst among the codon-substitution models. This demonstrates the importance of incorporating the code table in analyzing protein-encoding nucleotide sequences, and even the most sophisticated model is only a poor approximation if it is not incorporated (Cao and Hasegawa, in preparation).

We performed phylogenetic analyses by using mt-genome data with careful modeling of nucleotide and amino acid substitutions during evolution. Although some branching orders within Balaenopteridae still remain unresolved, our analyses gave a clear picture of

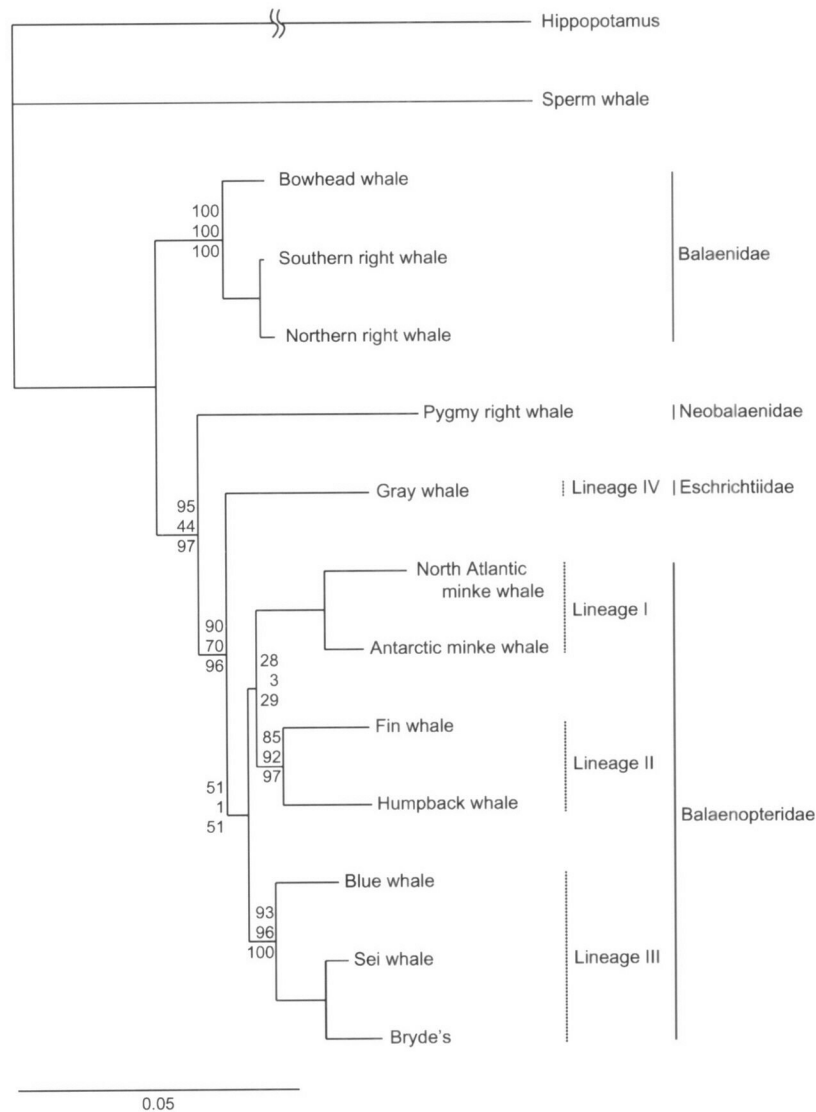


FIGURE 3. An ML tree of the concatenated amino acid sequences of 12 mt-proteins with mtREV-F+ Γ model (Adachi and Hasegawa, 1996b; Yang, 1996). The horizontal length of each branch is proportional to the estimated number of amino acid substitutions. Numbers indicate percent bootstrap probabilities for the proteins (top), 12S + 16S rRNAs (middle), and the total of proteins + rRNAs (bottom) estimated by the RELL method (Hasegawa and Kishino, 1994; Kishino et al., 1990) applied to the 10,000 candidate trees with 10,000 bootstrap replications.

the phylogeny of baleen whales which could not be attained by previous studies.

Phylogenetic Relationships of Baleen Whales

The phylogenetic relationships in this study differ markedly from previous hypotheses of mysticete relationships in the pattern of species groupings, but corroborate earlier molecular studies in the broad arrangement of mysticete families. The two families containing more than a single species, namely Balaenidae and Balaenopteridae, are monophyletic, confirming earlier morphological studies that recognized these groups as clades.

The position of right whales (family Balaenidae) as the most basal clade of Mysticeti was supported with

95%, 97%, and 100% BI¹ from the proteins, proteins + rRNAs in Figure 3, and protein-encoding codons in Figure 4, respectively, and is unsurprising in light of past studies (Arnason et al., 1992, 1993; McLeod et al., 1993; Adegoke et al., 1993; Arnason and Gullberg, 1994, 1996; Milinkovitch et al., 1994). The data indicate a clear distinction between the basal right whales and a clade comprised of all remaining species. Immediately crownward from the right whales is the pygmy right whale, *Caperea marginata*, which represents the only living species of the family Neobalaenidae. This species is quite disparate from other mysticetes in terms of structure and biology. Although some morphological features have allied *Caperea* with the right whales (McLeod et al., 1993; Bisconti 2000), the mt-genome analysis places the pygmy right whale as sister to Eschrichtiidae + Balaenopteridae. The

TABLE 2. Comparison of models used in analyzing the mt-protein encoding genes for the tree shown in Figure 4.

Model	<i>p</i>	λ	AIC	κ	<i>a</i>	<i>b</i>	<i>dN/dS</i>	α
Codon-substitution model								
Equal distance	87	-44,033.4	88,240.8	14.9			0.025	1.35
Geometric								
Grantham's distance	88	-43,830.1	87,836.2	16.3	0.068	3.56	0.025	1.31
Miyata et al.'s distance	88	-43,842.5	87,861.0	16.0	0.059	3.24	0.025	1.32
Linear								
Grantham's distance	88	-43,896.3	87,968.6	15.2	0.037	1.00	0.025	1.34
Miyata et al.'s distance	88	-43,890.1	87,956.2	15.2	0.037	1.00	0.025	1.34
Different parameters for each codon positions								
HKY+ Γ	90	-44,824.7	89,829.4					
GTR+ Γ	102	-44,618.4	89,440.8					

Note. Grantham's and Miyata et al.'s distances refer to physicochemical distances among 20 amino acids defined in Grantham (1974) and Miyata et al. (1979), respectively. *p* is the number of parameters in the model including the 25 branch lengths in the tree, and λ is log-likelihood. Parameters *a* and *b* are defined in Eqs. (11) and (12) in Yang et al. (1998), κ is the transition/transversion ratio, *dN/dS* is nonsynonymous/synonymous rate ratio, and α is the shape parameter of the Γ model for site heterogeneity. Details of the codon substitution models were described in Yang et al. (1998). In each codon-substitution model, codon frequencies with $60 - 1 = 59$ free parameters are used. The model with different parameters for each codon positions using the HKY+ Γ model (Hasegawa et al., 1985; Yang, 1996) estimates three nucleotide frequencies and two parameters for κ and α in addition to the 25 branch lengths for each codon positions, and therefore the total number of parameters is 90. The GTR+ Γ model (Yang, 1994, 1996) differs from the HKY+ Γ model in estimating five GTR rates instead of κ , and therefore the total number of parameters is 102.

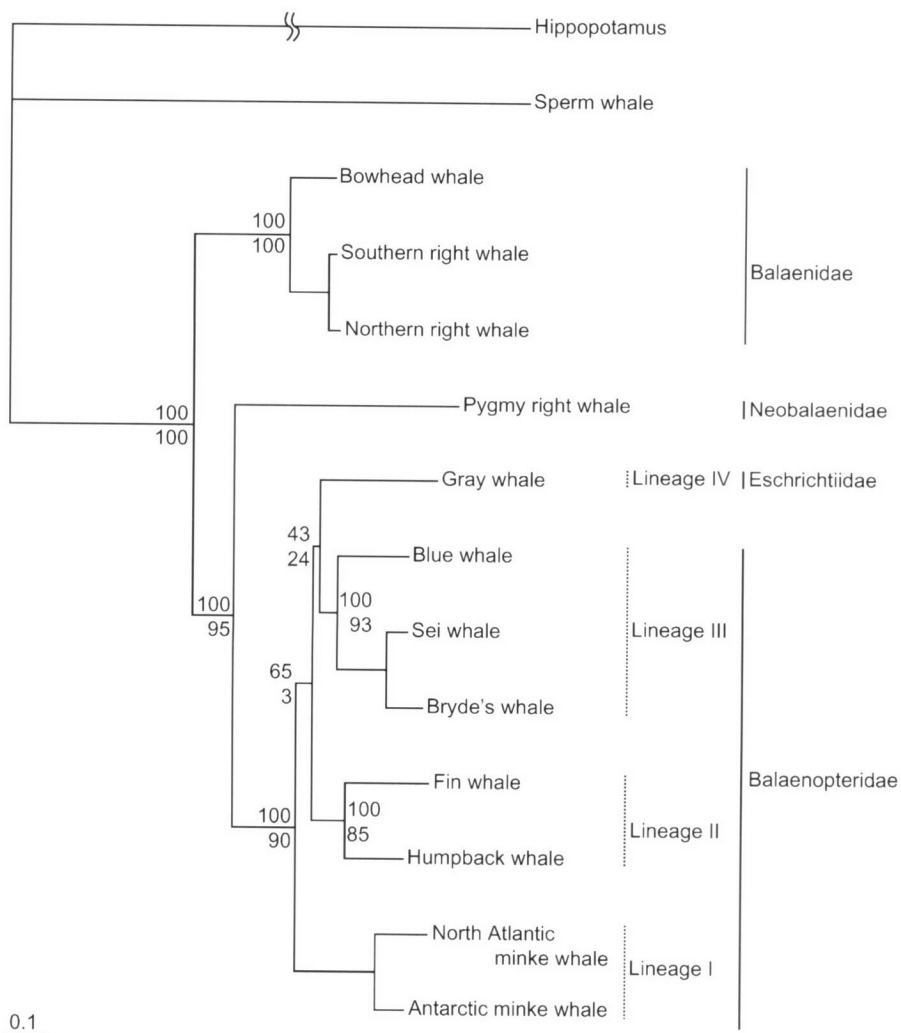


FIGURE 4. An ML tree of the concatenated nucleotide sequences of 12 mt-protein-encoding genes with the codon-substitution+ Γ model (Yang, 1996; Yang et al., 1998). The horizontal length of each branch is proportional to the estimated number of nucleotide substitutions. Numbers indicate percent bootstrap probabilities for the codons (top) and the amino acids (bottom) estimated by the RELL method applied to the 939 candidate trees with 10,000 bootstrap replications.

monophyly of this Neobalaenidae + Eschrichtiidae + Balaenopteridae clade is well supported by the mtDNA analysis and is also supported by a shared DNA satellite (Arnason and Best, 1991; Arnason et al., 1992; Adegoke et al., 1993). However, a recent exhaustive morphological analysis of extant and fossil cetaceans found multiple synapomorphies suggesting the traditional arrangement of a Neobalaenidae + Balaenidae sister relationship (Geisler and Sanders, 2003). Thus, the morphological and molecular data sets continue to conflict with respect to the position of the enigmatic pygmy right whale. Evidence from the nuclear genome regarding the phylogenetic affinities of *Caperea* would help distinguish between these two contrasting hypotheses.

The greatest degree of uncertainty in mysticete relationships lies within the clade of Eschrichtiidae + Balaenopteridae. Many alternative arrangements have been proposed within the Balaenopteridae, and questions remain regarding the position of the gray whale and its taxonomic status as a monotypic family. Our mt-genome analysis strongly supports the monophyly of the Eschrichtiidae + Balaenopteridae clade (Figs. 2 to 4). Furthermore, the mt-protein results identify four principal lineages among the Eschrichtiidae + Balaenopteridae: lineage I (the two minke whales), lineage II (the fin and humpback whales), lineage III (the sei, Bryde's, and blue whales), and lineage IV (the gray whale; Fig. 3). Both lineages II and III are recognized for the first time in this study. The proposed sister relationship between the fin and humpback whales (lineage II) is well supported, as measured by BPs of 85%, 97%, and 100% from the proteins, proteins + rRNAs (Fig. 3), and codons (Fig. 4). Even greater support is found for lineage III, grouping the blue whale and sei + Bryde's whales with BPs of 93%, 100%, and 100% from the proteins, proteins + rRNAs (Fig. 3), and codons in (Fig. 4). The fin/blue whale clade was never recovered during 10,000 bootstrap replications from the analyses of proteins + rRNAs and codons and was rejected. However, the mt-genome analysis fails to resolve the relationships among the four lineages, as indicated by the polytomy in Figure 5.

Relatively distant phylogenetic position of minke whales among the Eschrichtiidae + Balaenopteridae was suggested in previous molecular phylogenetic analyses, which used representative species of baleen whales (Arnason et al., 1993; Arnason and Gullberg, 1994). The mt-genome results also confirmed that lineage I, namely minke whales clade, had unique evolutionary history, which can be regarded as an independent lineage in the four principal lineages among the Eschrichtiidae + Balaenopteridae clade.

The humpback whale is the only living species of *Megaptera*. *Megaptera* differs markedly from species of *Balaenoptera* in habits and morphology (Clapham and Mead, 1989), and thus has been placed in its own subfamily Megapterinae, sister clade to the subfamily Balaenopterinae, together comprising the family Balaenopteridae (Rice, 1998). Although its position within the clade of Eschrichtiidae + Balaenopteridae is uncontroversial (Messenger and McGuire, 1998), previous

mtDNA studies contradict the sister arrangement of Megapterinae + Balaenopterinae. Analysis of mtDNA control region sequences placed the humpback as the sister group to blue + fin whales, excluding sei, Bryde's, and minke whales (Arnason et al., 1993; Fig. 1A) whereas mtDNA cytochrome *b* sequences suggested a humpback + blue whale sister relationship with a clade of (sei + Bryde's) + fin whales (Arnason and Gullberg, 1994; Fig. 1B). Although both studies implied a paraphyletic Balaenopterinae, these hypotheses for *Megaptera* relationships lacked strong statistical support. The novel sister relationship of humpback + fin whale, indicated by analysis of mt-genomes, is supported by high BPs irrespective of the models used in the analyses. The association of humpback and fin whale indicates the paraphyly of the Balaenopterinae; in turn, a separate subfamily may not be justified for *Megaptera*.

Very strong statistical support is presented for lineage III, the blue whale as sister to sei + Bryde's whales. This arrangement contrasts sharply with the close relationship between blue and fin whales inferred by natural hybridizations (Arnason et al., 1991b; Berube and Aguilar, 1998).

Lineage IV is composed solely of the gray whale *Eschrichtius*. Morphological analyses have placed the gray whale variously close to the Balaenopteridae, the Balaenidae, or as a sister taxon to Balaenopteridae + Balaenidae (McLeod et al., 1993; Geisler and Luo, 1996; Bisconti, 2000; Lindow, 2002). Previous mitochondrial sequence and nuclear DNA satellite analyses (Arnason et al., 1992, 1993; Adegoke et al., 1993; Arnason and Gullberg, 1994, 1996) suggest the gray whale (Eschrichtiidae) lies within, or is sister taxon to, the Balaenopteridae (Fig. 1A and B), again indicating its paraphyly. In the ML tree of mt-proteins and mt-proteins + mt-rRNAs, gray whale is sister to Balaenopteridae, but the BP for the monophyly of Balaenopteridae is only 51% (Fig. 3), and its monophyly is no longer the ML relationship by the codon-substitution model only with 13% BP (data not shown). Examining data from the nuclear genome will be an important independent source of molecular evidence for the position of *Eschrichtius*.

Divergence Times

In order to check whether or not the clock model holds for the mt-protein sequences, the likelihood ratio test was carried out as follows: twice the log-likelihood difference between the nonclock and clock models for the tree topology in Figure 5 is $2\Delta\lambda = 2 \times [-17, 570.1 - (-17, 661.0)] = 181.8$, much greater than the critical value $\chi_{0.01}^2 = 32.9$ with degrees of freedom = 12 (difference of the number of parameters between the two models), and the clock model is thus rejected. Therefore, we used Thorne and Kishino's method (Thorne et al., 1998; Kishino et al., 2001), which takes account of the violation of the clock in estimating divergence times. Another merit of the method is that it can incorporate fossil evidence via constraints on node times.

We used 12 mt-protein sequences to estimate divergence times of lineages leading to different species

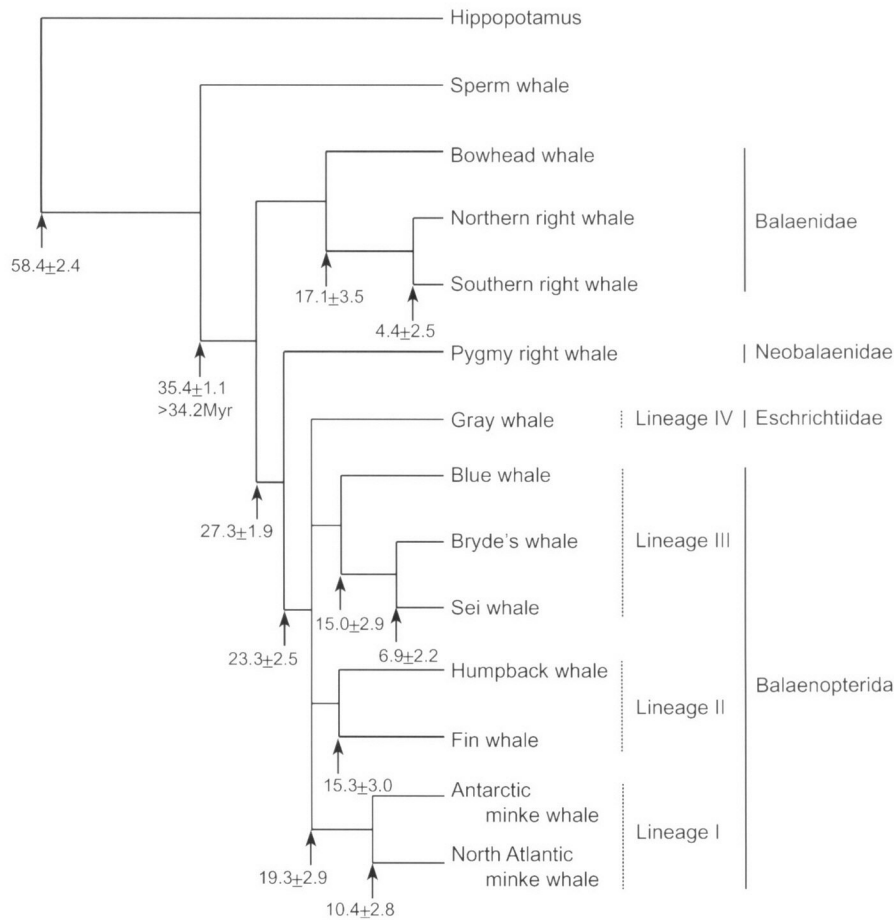


FIGURE 5. Estimates of branching dates (in Mya) in the Mysticeti evolution by using the mtDNA 12-protein concatenated sequences.

of Mysticeti (Fig. 5). In estimating divergence time, branches among lineages I to IV were treated as a polytomy, because those BPs were relatively low in the ML tree. When comparing molecular divergence times with the fossil record, crown-group and stem-group terminology helps to clarify major taxonomic groups of mysticetes (Craske and Jefferies, 1989). The crown-group is a clade of living species, together with the most recent common ancestor of those living species and all of its descendants. The stem-group comprises those species more closely related to the crown-group than to any other clade.

The first divergence of extant mysticete lineages, separating right whales from other extant mysticetes, is estimated at 27.3 ± 1.9 Mya (late Oligocene). This accords closely with the oldest reported fossil stem-balaenid at about 28 Mya (Fordyce, 2002; Fig. 6, stem \dagger sb). The long interval spanning much of the late Oligocene and the Miocene, about 28 to 5–6 Mya, is a major gap in the history of right whales, with only *Morenocetus parvus* (early Miocene, 20 to 22 Mya; M.Á. Cozzuol, personal communication) and some fragmentary specimens reported. The fossil record is unrevealing about the inferred molecular divergence of the two extant genera *Balaena* and *Eubalaena* at about 17 Mya. However, Bisconti (2000;

Fig. 5b) implied that *Balaena* and *Eubalaena* diverged before *Morenocetus*, that is, before 20 to 22 Mya. Because *Balaena* and *Eubalaena* are phenetically similar in structure and indeed were long regarded as congeneric, such an old divergence time implies extremely slow rates of morphological change. Alternatively, *Morenocetus* is notably archaic in form relative to *Balaena* and *Eubalaena*, and perhaps diverged earlier. Two alternatives are shown in Figure 6.

The Neobalaenidae divergence is predicted at 23.3 ± 2.5 Mya, near the Oligocene/Miocene boundary, but no fossil pygmy right whales older than Quaternary (0 to 2 Mya) have been reported (Fordyce and Muizon, 2001). Similarly, the suggested early Miocene divergence time of Eschrichtiidae, at 19.3 ± 2.9 Mya, contrasts dramatically with a 0.5 Mya Pleistocene date for the oldest reliably identified fossil gray whale (Barnes and McLeod, 1984). Thus, the molecular results imply substantial ghost lineages (Weishampel, 1996) for these two groups. This suggests the marked structural disparity between the pygmy right whale, the gray whale, and other living baleen whales is not a reflection of geologically recent origins associated with rapid evolutionary/developmental change. Rather, the disparity reflects ancient origins and long gradual structural divergence between these

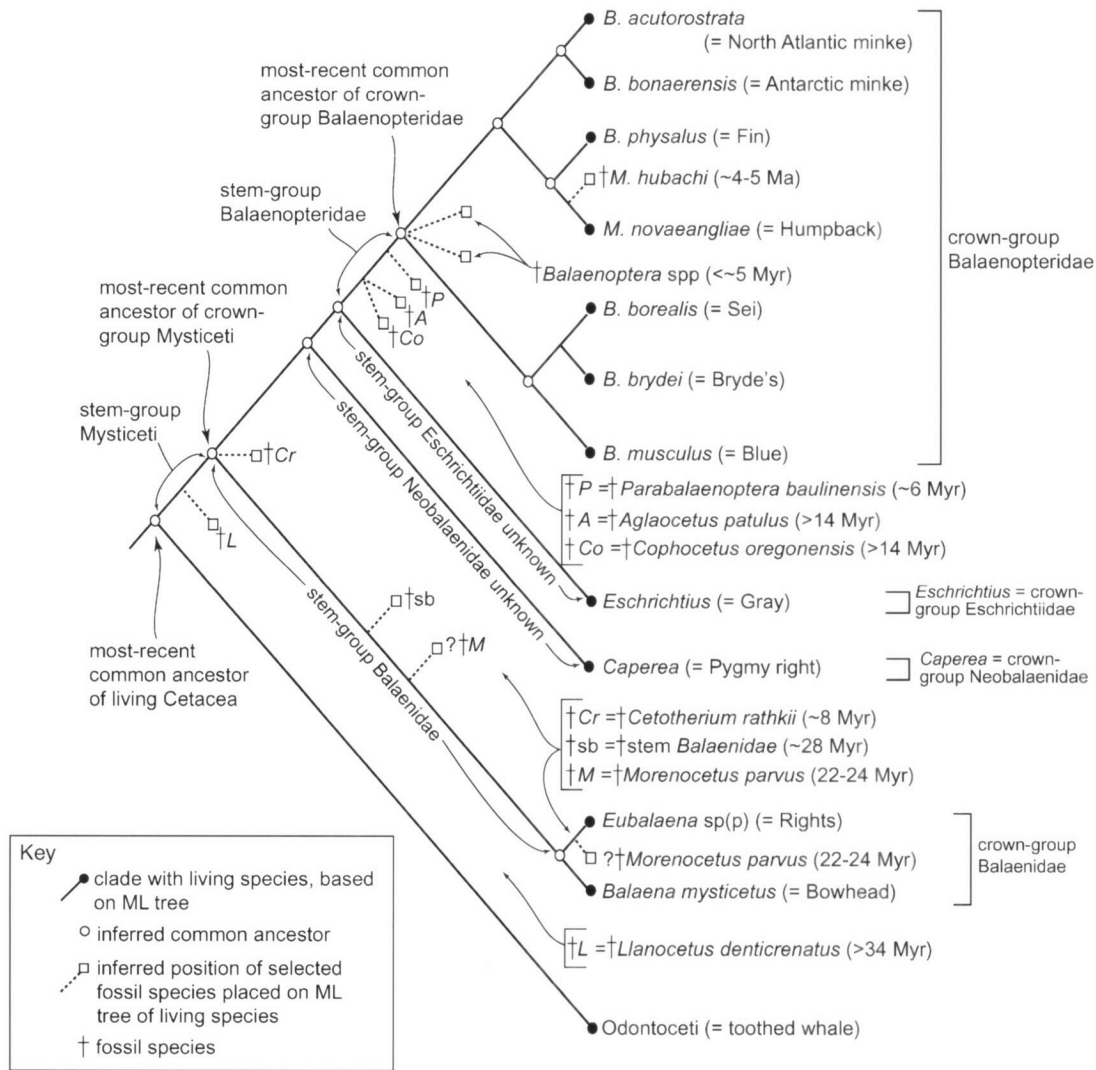


FIGURE 6. Composite tree for modern and selected fossil Mysticeti, using the ML tree of Figure 3 as a framework. Fossils, marked by a dagger †, are placed using published literature or best current knowledge; see text for fossil dates. Crown-groups and selected stem-groups are indicated, using the style of Craske and Jefferies (1989). Alternative positions are shown for the fossil right whale species *Morenocetus parvus*.

groups and other mysticetes. Two factors may account for the difference between predicted ancient molecular divergence times and the short fossil record. Stem-Neobalaenidae and stem-Eschrichtiidae may indeed reside in fossil collections already, but may not be identified because synapomorphies with their crown-species are not recognized. Alternatively, stem-Neobalaenidae and stem-Eschrichtiidae perhaps occupied habitats poorly sampled in the fossil record, as has been suggested for the gray whale (Arnason and Best, 1991).

Discussion of balaenopterid origins is hampered because crown-group and stem-group concepts have not been used widely for the Balaenopteridae. Reliably dated fossil rorquals (*Balaenoptera*) and humpbacks (*Megaptera*) are known from the Pliocene, <5 Mya (Dathe, 1983; Demere, 1986). Late Miocene dates for these genera, 7 to 12 Mya (Barnes, 1977), are less certain because the fossils are incomplete. Early supposed balaenopterids such

as *Parabalaenoptera baulinensis* (Zeigler et al., 1997) differ from their putative ancestors, the Miocene mysticetes traditionally termed "cetotheres," in changes in jaw muscle origins that suggest refined gulp-feeding habits. Recent studies suggest that the Cetotheriidae in the cladistic sense is not clearly related to the Balaenopteridae (Fordyce and Muizon, 2001), and a cautious approach to relationships is prudent (see *Cetotherium rathkii* in Fig. 6). However, "cetotheres" in the traditional grade sense include apparent stem-Balaenopteridae (Fig. 6), for example, late early or early middle Miocene (>14 Mya) species of *Aglaocetus* and *Cophocetus* (see Kimura and Ozawa, 2002: fig. 13b). Thus, the molecular date of 19.3 ± 2.9 Mya for balaenopterid divergence from other living mysticetes has reasonable accord with the fossil record of stem-balaenopterids.

The pattern of lineages I, II, and III, when taken with molecular divergence times, predicts the origin of

crown-*Balaenoptera*, as well as crown-Balaenopteridae, at 19 Mya. Molecular divergence times of lineages between fin and humpback, and between blue and sei + Bryde's, are early middle Miocene (15 Mya). Previously, divergence times predicted by molecular studies were 5 Mya for fin and blue (Arnason and Gullberg, 1993) and 7 Mya for the divergence of the minke whale lineage from other balaenopterids (Wada and Numachi, 1991). The divergence time predicted by the present study is older than that predicted using the mtDNA partial region or by allozyme analysis, but there is some accord with predictions of Nikaido et al. (2001b) using nuclear DNA sequences. According to Nikaido et al. (2001b), minke whales lineage diverged from a fin + humpback lineage at 20 Mya, with a divergence time for fin/humpback predicted at 17 Mya. In summary, the complete mtDNA results suggest that 10 out of 12 lineages leading to extant Mysticeti species appeared between early Miocene (23 Mya) and middle Miocene (10 Mya).

An inferred origin for *Balaenoptera* at 19 Mya implies that fossil rorquals of essentially modern appearance should be found well before the late Miocene, but this is clearly not the case. The huge global collections of Miocene baleen whales lack well-identified fossils of *Balaenoptera* or *Megaptera* older than late, indeed latest, Miocene. Further, the collections do include many mysticetes that were similar in size and probably habit to modern rorquals, suggesting that the lack of early fossil *Balaenoptera* is real, and not an environmental/preservational bias. Reliably identified species of *Balaenoptera* and *Megaptera* only become common in the Pliocene, <5 Mya (Dathe, 1983; Demere, 1986), and it is likely that *Balaenoptera* and the crown-group Balaenopteridae originated then, about the Miocene/Pliocene boundary.

Toward a Mysticete Phylogeny

The most recent common ancestor of mysticete whales existed in the very late Oligocene. The fossil and molecular records concord that balaenid whales diverged soon thereafter, near the Oligocene/Miocene boundary, with right and bowhead whales as the extant representatives of this basal mysticete lineage. Molecular data strongly indicate the next branch in mysticete evolution is the Neobalaenidae, of which the disparate pygmy right whale, *Caperea*, is the monotypic living representative. The unique morphology of *Caperea* combined with an mtDNA divergence estimate of 23 Mya for the Neobalaenidae suggests a long independent evolutionary history. However, described fossil specimens are so far restricted to the Quaternary, implying the existence of undescribed or undiscovered neobalaenid fossils.

The molecular data suggest the remaining extant genera, *Eschrichtius*, *Megaptera*, and *Balaenoptera* (lineages I to IV), shared a unique common ancestor in the late early Miocene, but cannot clearly discern the splitting pattern within this clade. The preponderance of morphological and molecular evidence suggests that *Eschrichtius* (lin-

age IV) is the sister taxon to *Megaptera* + *Balaenoptera* (the latter of which is likely paraphyletic, see below). Like *Caperea*, the extant monotypic *Eschrichtius* has only a limited, late Pleistocene fossil record, whereas mt-protein-based node divergence estimates indicate a long independent history for this lineage and its morphology is highly distinct from its probable sister clade. Increased efforts to identify fossil Eschrichtiids and the analysis of nuclear phylogenetic markers could further clarify the evolutionary history of this clade.

Based on the relatively rich fossil record of stem-balaenopterids, molecular and paleontological data broadly concur with a late early Miocene origin for the Balaenopteridae lineage. However, the most significant conflict in evolutionary timing between the fossil and molecular data is encountered among the balaenopterid crown group. Molecular divergence estimates for lineages I, II, and III predict the appearance of fossil *Balaenoptera* or *Megaptera* in the middle Miocene, a time frame that the dense Miocene baleen whale record strongly refutes.

In addition to problematic divergence times, the phylogenetic relationships among *Megaptera* + *Balaenoptera* (lineages I, II, and III) remain unresolved. The sheer diversity of published phylogenetic hypotheses for this clade attests to the challenge of its resolution; almost every possible arrangement has been proposed. Our results provide further evidence that minke whales (lineage I) represent a relatively unique branch within *Balaenoptera*, an insight that does not readily emerge from comparative morphology. The mt-genome analysis suggests yet another novel set of phylogenetic relationships among the blue, Bryde's, fin, humpback, and sei whales (lineages II and III); however, unlike many previous studies, the nodes for these lineages are supported by high BPs. If the mitochondrial phylogeny proposed here is upheld by nuclear data, it would indicate the genus *Balaenoptera* is paraphyletic, and the subfamily designations of Megapterinae and Balaenopterinae, invoked to reflect the morphological distinctiveness of the humpback whale as compared to rorquals, are unwarranted. In turn, this has significant implications for rates of morphological and behavioral character evolution in this clade.

There are several possible explanations for the substantial conflict between the molecular and paleontological divergence time estimates for balaenopterid evolution. Molecular evolution may be accelerated in the balaenopterid lineage, or significant rate variation could exist among extant balaenopterid species. Recently, Kimura and Ozawa (2002) analyzed rates of mtDNA evolution in cetaceans, concluding that mysticete mtDNA evolves at least twice as slowly as odontocete DNA. In our calculations of divergence time estimates, we used statistical models that attempt to account for molecular evolutionary rate variation among different lineages. Thus we consider mutational rate variation is unlikely to account for discordant results.

Certain population genetic histories can cause considerable departures between divergence time estimates inferred from individual genetic markers and the

population bifurcation events that generate the actual species trees. If the ancestral baleenopterid population exhibited a large degree of mitochondrial polymorphism, the splitting of mtDNA haplotypes (the timing of which we are estimating) could significantly predate the population splitting event that gave rise to baleenopterid species. The influence of ancestral polymorphism in generating gene tree versus species tree divergence time discrepancies would be considerable if effective population size (N_e) is very large and generation times are long (Nei, 1987). Baleen whales are large-bodied, long-lived animals, widely distributed throughout their oceanic habitat. N_e of large, vagile animals living in oceanic environments may be very large, due to the relatively limited opportunities for geographic subdivision. We suggest that molecular divergence time estimates for baleenopterid whales have been strongly influenced by the combined effects of ancestral polymorphism, large ancestral N_e , and long generation times.

In addition to a rapid pace of speciation, these same three interacting forces may also account for the consistent lack of phylogenetic resolution in the Eschrichtiidae + Balaenopteridae clade. It is well established that cladograms inferred from single molecules are particularly susceptible to complications of ancestral polymorphisms and incomplete lineage sorting (Brower et al., 1996). However, because N_e for mtDNA is smaller than that of the nuclear genome, mtDNA should have a significantly higher chance of accurately tracking short internodes (Moore, 1995). To date, no nuclear phylogeny of mysticetes has been published. If ancestral polymorphism and incomplete lineage sorting have been important forces shaping the Eschrichtiidae + Balaenopteridae genomes, some inconsistencies in branching order would be an expected result of a nuclear gene phylogeny. A substantial amount of comparative nuclear data will likely be required for further resolution of mysticete phylogeny.

ACKNOWLEDGMENTS

This work was supported by a Grant-in-Aid to N.O. and M.H. from the Ministry of Education, Science, Sports and Culture of Japan and from JSPS. We thank A. E. Dizon, J. Mead, and K. Robertson for permissions and research support from the Marine Mammal Tissue Archive, National Marine Fisheries Service Protected Resources Division.

REFERENCES

- Adachi, J., Y. Cao, and M. Hasegawa. 1993. Tempo and mode of mitochondrial DNA evolution in vertebrates at the amino acid sequence level: Rapid evolution in warm-blooded vertebrates. *J. Mol. Evol.* 36:270–281.
- Adachi, J., and M. Hasegawa. 1996a. MOLPHY: Programs for molecular phylogenetics ver. 2.3. Computer Science Monographs, No. 28. Institute of Statistical Mathematics, Tokyo.
- Adachi, J., and M. Hasegawa. 1996b. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol.* 42:459–468.
- Adegoke, J. A., U. Arnason, and B. Widegren. 1993. Sequence organization and evolution, in all extant whalebone whales, of a DNA satellite with terminal chromosome localization. *Chromosoma* 102:382–388.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Contr.* AC-19:716–723.
- Arnason, U., and P. Best. 1991. Phylogenetic relationships within the Mysticeti (whalebone whale) based upon studies of highly repetitive DNA in all extant species. *Hereditas* 114:263–269.
- Arnason, U., S. Gretarsdottir, and B. Widegren. 1992. Mysticete (baleen whale) relationships based upon the sequence of the control or cetacean DNA satellite. *Mol. Biol. Evol.* 9:1018–1028.
- Arnason, U., and A. Gullberg. 1993. Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature. *J. Mol. Evol.* 37:312–322.
- Arnason, U., and A. Gullberg. 1994. Relationship of baleen whales established by cytochrome *b* gene sequence comparison. *Nature* 367:726–728.
- Arnason, U., and A. Gullberg. 1996. Cytochrome *b* nucleotide sequences and the identification of five primary lineages of extant cetaceans. *Mol. Biol. Evol.* 13:407–417.
- Arnason, U., A. Gullberg, S. Gretarsdottir, B. Ursing, and A. Janke. 2000. The mitochondrial genome of the sperm whale and a new molecular reference for estimating eutherian divergence dates. *J. Mol. Evol.* 50:569–578.
- Arnason, U., A. Gullberg, and B. Widegren. 1991a. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. *J. Mol. Evol.* 33:556–568.
- Arnason, U., A. Gullberg, and B. Widegren. 1993. Cetacea mitochondrial DNA control region: Sequences of all extant baleen whale and two sperm whale species. *Mol. Biol. Evol.* 10:960–970.
- Arnason, U., R. Spilliaert, A. Palsdottir, and A. Arnason. 1991b. Molecular identification of hybrids between the two largest whale species—the blue whale (*Balaenoptera musculus*) and the fin whale (*B. physalus*). *Hereditas* 115:183–189.
- Bajpai, S., and P. D. Gingerich. 1998. A new Eocene archaeocete (*Mammalia*, Cetacea) from India and the time of origin of whales. *Proc. Natl. Acad. Sci. USA* 95:15404–15468.
- Barnes, L. G. 1977. Outline of eastern north Pacific fossil cetacean assemblage. *Syst. Zool.* 25:321–343.
- Barnes, L. G., and S. A. McLeod. 1984. The fossil record and phyletic relationships of gray whales. Pages 3–32 in *The gray whale* (M. T. Jones, S. J. Swartz, and S. Leatherwood, eds.). Academic Press, New York.
- Berube, M., and A. Aguilar. 1998. A new hybrid between a blue whale (*Balaenoptera musculus*) and a fin whale, *B. physalus*. Frequency and implications of hybridization. *Mar. Mamm. Sci.* 14:82–98.
- Bisconti, M. 2000. New description, character analysis, and preliminary phyletic assessment of two Balaenidae skulls from the Italic Pliocene. *Palaontogr. Ital.* 87:37–66.
- Brower, A. V. Z., R. DeSalle, and A. Volger. 1996. Gene trees, species trees, and systematics: A cladistic perspective. *Ann. Rev. Ecol. Syst.* 27:423–450.
- Brown, W. M., M. George, and A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 76:767–771.
- Cao, Y., J. Adachi, and M. Hasegawa. 1994. Eutherian phylogeny as inferred from mitochondrial DNA sequence data. *Jpn. J. Genet.* 69:455–472.
- Cao, Y., M. Fujiwara, M. Nisaido, N. Okada, and M. Hasegawa. 2000. Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. *Gene* 259:149–158.
- Clapham, P. L., and J. G. Mead. 1989. *Megaptera novaeangliae*. *Mamm. Species* 604:1–9.
- Craske, A. J., and R. P. S. Jefferies. 1989. A new mitrate from the upper Ordovician of Norway and a new approach to subdividing a clasion. *Palaontology* 32:69–99.
- Dathe, E. 1983. *Megaptera hubachi* sp. n., ein fossiler Bartenwald aus marinen Sandsteinschichten des tieferen Pliozäns Chiles. *Z. Geol. Wissen.* 11:813–848.
- Demere, T. A. 1986. The fossil whale, *Balaenoptera davidsonii* (Cope, 1872), with a review of other Neogene species of *Falangeria* (Cetacea: Mysticeti). *Mar. Mamm. Sci.* 2:277–298.
- Dingle, R. V., and M. Lavelle. 1998. Antarctic peninsular cryosphere—early Oligocene (c. 30 Ma) in tiation and a revised glacial chronology. *J. Geol. Soc.* 155:433–437.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17:368–376.

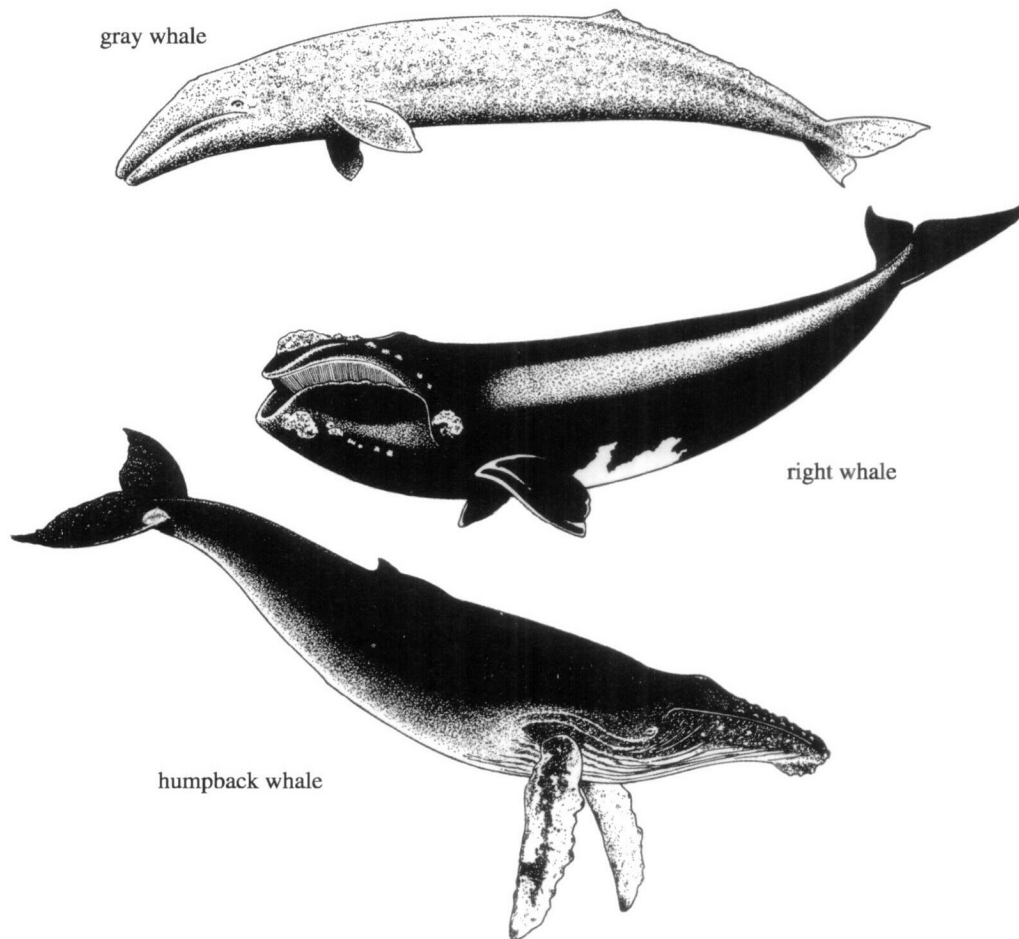
- Felsenstein, J. 1988. Phylogenies from molecular sequences: Inference and reliability. *Annu. Rev. Genet.* 22:521–565.
- Fordyce, R. E. 2002. Oligocene origins of skin-feeding right whales: A small archaic balaenid from New Zealand. *J. Vertebr. Paleontol.* 22 (3, supplement): 54a.
- Fordyce, R. E. 1989. Origins and evolution of Antarctic marine mammals. *Geol. Soc. Spec. Publ.* 47:269–281.
- Fordyce, R. E., and C. de Muizon. 2001. Evolutionary history of cetaceans: A review. Pages 169–233 in *Secondary Adaptation of Tetrapods to Life in Water* (J.-M. Mazin and V. de Buffrenil, eds.). Verlag Dr. Friedrich Pfeil, Munchen, Germany.
- Gatesy, L. C., Hayashi, M., A. Cronin, and P. Arcander. 1996. Evidence from milk casein genes that cetaceans are close relatives of hippopotamid artiodactyls. *Mol. Biol. Evol.* 13:954–963.
- Geisler, J. H., and Z. Luo. 1996. The petrosal and inner ear of *Herpetocetus* sp. (Mammalia: Cetacea) and their implications for the phylogeny and hearing of archaic mysticetes. *J. Paleontol.* 70:1045–1066.
- Geisler, J. H., and A. Sanders. 2003. Morphological evidence for the phylogeny of Cetacea. *J. Mamm. Evol.* 10:23–129.
- Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11:725–736.
- Grande, C., J. Templado, J. L. Cervera, and R. Zardoya. 2002. The complete mitochondrial genome of the nudibranch *Roboastra europaea* (Mollusca: Gastropoda) Supports the monophyly of opisthobranchs. *Mol. Biol. Evol.* 19:1672–1685.
- Grantham, R. 1974. Amino acid difference formula to help explain protein evolution. *Science* 185:862–864.
- Hasegawa, M., and H. Kishino. 1994. Accuracies of the simple methods for estimating the bootstrap probability of a maximum likelihood tree. *Mol. Biol. Evol.* 11:142–145.
- Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–174.
- Hasegawa, M., J. L. Thorne, and H. Kishino. 2003. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet. Syst.* 78:267–283.
- Ingman, M., H. Kaessmann, S. Pääbo, and U. Gyllenstein. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713.
- Kijas, J. M. H., and L. Andersson. 2001. A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. *J. Mol. Evol.* 52:302–308.
- Kimura, E., and I. Ozawa. 2002. A new cetothere (Cetacea: Mysticeti) from the early Miocene of Japan. *J. Vertebr. Paleontol.* 22:684–702.
- Kishino, H., and M. Hasegawa. 1990. Converting distance to time: An application to human evolution. *Methods Enzymol.* 183:550–570.
- Kishino, H., T. Miyata, and M. Hasegawa. 1990. Maximum likelihood inference of protein phylogeny, and the origin of chloroplasts. *J. Mol. Evol.* 31:151–160.
- Kishino, H., J. L. Thorne, and W. J. Bruno. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol. Biol. Evol.* 18:352–361.
- Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245.
- LeDuc, R., and A. E. Dizon. 2002. Reconstructing the orquials phylogeny: With comments on the use of molecular and morphological data for systematic study. Pages 100–110 in *Molecular and cell biology of marine mammals* (C. J. Pfeiffer and P. E. Nachtigall, eds.). Krieger.
- Lin, Y.-H., P. A. McLenachan, A. R. Gore, M. J. Phillips, R. Ota, M. D. Hendy, and D. Penny. 2002. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. *Mol. Biol. Evol.* 19:2060–2070.
- Lindow, B. E. K. 2002. Bardehvalernes indbyrdes slægtskabsforhold—en foreløbig analyse [The internal relationships of the baleen whales—a preliminary analysis]. Pages 12–19 in *Resumé* (hæfte). Hvaldag 2002, 24 September 2002 (B. E. K. Lindow, eds.). Mtsønderjyllands Museum, Gram.
- Mayr, E. 1969. Principles of systematic zoology. McGraw-Hill, New York.
- McKenna, M. C., and S. K. Bell. 1997. Classification of mammals: Above the species level. Columbia University Press, New York.
- McLeod, S. A., F. C. Whitmore, Jr. and L. G. Barnes. 1993. Evolutionary relationships and classification. Pages 45–70 in *The Bowhead Whale* (J. I. Burns, J. E. Montague, and C. J. Cowles, eds.). Society of Marine Mammals Special Publication No. 2. Society of Marine Mammals, Lawrence, Kansas.
- Messenger, S. J., and J. McGuire. 1998. Morphology, molecules, and phylogenetics of cetaceans. *Syst. Biol.* 47:90–124.
- Milinkovitch, M. C., A. Meyer, and J. R. Powell. 1994. Phylogeny of all major groups of cetaceans based on DNA sequences from three mitochondrial genes. *Mol. Biol. Evol.* 11:939–948.
- Mitchell, E. D. 1989. A new cetacean from the late Eocene La Mesta Formation, Seymour Island, Antarctic Peninsula. *Can. J. Fish. Aquat. Sci.* 46:2219–2235.
- Miyata, T., S. Miyazawa, and I. Yasunaga. 1979. Two types of amino acid substitutions in protein evolution. *J. Mol. Evol.* 12:219–236.
- Moore, W. S. 1995. Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49:718–726.
- Murata, Y., M. Nikaido, I. Sasaki, Y. Cao, Y. Fukumoto, M. Hasegawa, and N. Okada. 2003. Afrotherian phylogeny as inferred from complete mitochondrial genomes. *Mol. Phylogenet. Evol.* 28:253–260.
- Nardi, E., G. Spinsanti, I. E. Boore, V. Carapelli, R. Dallai, and E. Frati. 2003. Hexapod origins: Monophyletic or paraphyletic? *Science* 299:1887–1889.
- Nei, M. 1987. Molecular evolution genetics. Columbia University Press, New York.
- Nikaido, M., K. Kawai, Y. Cao, M. Harada, S. Tomita, N. Okada, and M. Hasegawa. 2001a. Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *J. Mol. Evol.* 53:508–516.
- Nikaido, M., M. Harada, Y. Cao, M. Hasegawa, and N. Okada. 2000. Monophyletic origin of the order Chiroptera and its phylogenetic position among Mammalia, as inferred from the complete sequence of the mitochondrial DNA of a Japanese megabat, the ryukyu flying fox (*Pteropus dassumallus*). *J. Mol. Evol.* 51:318–328.
- Nikaido, M., E. Matsuno, H. Hamilton, R. J. Brownell Jr., Y. Cao, W. Ding, Z. Zuoyan, A. M. Shedlock, R. E. Fordyce, M. Hasegawa, and N. Okada. 2001b. 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.
- Nikaido, M., A. P. Rooney, and N. Okada. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: Hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. USA* 96:10261–10266.
- Olivio, P. D., M. J. Van De Walle, P. I. Tapis, and W. W. Hauswirth. 1983. Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. *Nature* 306:400–402.
- Rice, D. W. 1998. Marine mammals of the world: Systematics and distribution. Society of Marine Mammals, Lawrence, Kansas.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: Laboratory manual, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sanders, A. E., and L. G. Barnes. 2002. Paleontology of the late Oligocene Ashley and Chandler bridge formations of South Carolina. 3: Eomysticetidae, a new family of primitive mysticetes (Mammalia: Cetacea). *Smithsonian Contrib. Paleobiol.* 93:313–356.
- Strimmer, K., and A. von Haeseler. 1996. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13:964–969.
- Thorne, J., H. Kishino, and J. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15:1647–1657.
- Ursing, B. M., and U. Arnason. 1998. Analyses of mitochondrial genomes strongly support a hippopotamus-whale clade. *Proc. R. Soc. Lond. B. Biol. Sci.* 265:2251–2255.
- Wada, S., and K. Numachi. 1991. Allozyme analyses of genetic differentiation among the populations and species of the *Balaenoptera*. *Rep. Int. Whal. Commn. Spec. Issue* 13:126–154.

- Waddell, P. J., Y. Cao, M. Hasegawa, and D. P. Mindell. 1999. Assessing the Cretaceous superordinal divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an extended statistical framework. *Syst. Biol.* 48:119–137.
- Weishampel, D. B. 1996. Fossils, phylogeny, and discovery: A cladistic study of the history of tree topologies and ghost lineage durations. *J. Vertebr. Paleontol.* 16:191–197.
- Yang, Z. 1994. Estimating the pattern of nucleotide substitution. *J. Mol. Evol.* 39:105–111.
- Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. *TREE* 11:367–372.
- Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *CABIOS* 13:555–556.
- Yang, Z., R. Nielsen, and M. Hasegawa. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol. Biol. Evol.* 15:1600–1611.
- Zeigler, C. V., G. I. Chan, and T. G. Barnes. 1997. A new late Miocene baleenopterid whale (Cetacea: Mysticeti), *Parabalaenoptera baltica* (new genus and species), from the Santa Cruz Mudstone, Point Reyes Peninsula, California. *Proc. Calif. Acad. Sci.* 50:115–133.

First submitted 20 January 2004; in final form 1 April 2004.

Final acceptance 4 August 2004.

Associate Editor: Jack Sullivan



gray whale

right whale

humpback whale