

# Phylogeny of mysticete whales based on mitochondrial and nuclear data

Amanda L. Rychel,<sup>a,\*</sup> Tod W. Reeder,<sup>a</sup> and Annalisa Berta<sup>a</sup>

<sup>a</sup> Department of Biology, San Diego State University, San Diego, CA 92182, USA

Received 25 September 2003; revised 4 February 2004

Available online 27 April 2004

## Abstract

Mysticetes or baleen whales are comprised of four groups: Eschrichtiidae, Neobalaenidae, Balaenidae, and Balaenopteridae. Various phylogenetic hypotheses among these four groups have been proposed. Previous studies have not satisfactorily determined relationships among the four groups with a high degree of confidence. The objective of this study is to determine the relationships among the mysticete whales. Mitochondrial and nuclear DNA were sequenced for phylogenetic analysis. Most species relationships determined using these data were well resolved and congruent. Balaenidae is the most basal group and Neobalaenidae is the second most basal and sister group to the balaenopterid–eschrichtiid clade. In this phylogenetic study, the resolution of Eschrichtiidae with two main lineages of Balaenopteridae was problematic. Some data partitions placed this group within the balaenopterids, and other partitions placed it as a sister taxon to the balaenopterids. An additive likelihood approach was used to determine the most optimal trees. Although it was not found in the combined phylogenetic analyses, the “best” tree found under the additive likelihood approach was one with a monophyletic Balaenopteridae.

© 2004 Elsevier Inc. All rights reserved.

**Keywords:** Baleen whales; Phylogenetics; ND 4;  $\alpha$ -Lactalbumin; Combined data

## 1. Introduction

Mysticetes (baleen whales) are among the largest and best known mammals on Earth. They belong to a monophyletic group that is diagnosed by three unequivocal synapomorphies (Geisler and Sanders, 2003): zygomatic process of maxilla bears a steep face that clearly separates the rostrum from the antorbital process, long postorbital process, and wide and bulbous basioccipital crest. Mysticetes currently number 12 to 15 species, with the majority of them (seven to nine) (Wada et al., 2003) belonging to the Balaenopteridae. The gray whale (*Eschrichtius robustus*) is the sole member of the Eschrichtiidae and the Neobalaenidae also is represented by a single species, the pygmy right whale (*Caperea marginata*). The Balaenidae is the fourth line-

age of baleen whales and is composed of three to four species [two to three species of right whales (Rosenbaum et al., 2000) and the bowhead whale]. Cetacean relatives nearest to the mysticetes are the odontocetes (toothed whales), and the nearest extant non-marine relative to the whales is the hippopotamus, based on molecular data (Gatesy, 1997; Shimamura et al., 1997) and morphology (Geisler and Uhen, 2003). In fact, recent molecular studies have found that Cetacea is nested within Artiodactyla (even-toed ungulates), with this new more inclusive clade now named Cetartiodactyla (Gatesy, 1997). While the systematics of cetaceans in general has attracted much attention, the phylogenetic relationships among mysticetes have been largely ignored.

The phylogenetic relationships among baleen whales have been inferred in two previous molecular systematic studies. These studies employed the mitochondrial cytochrome *b* gene (Árnason and Gullberg, 1994) and the mitochondrial control region (Árnason et al., 1993). Neither study included as many species as this study nor did they yield well-resolved relationships between the

\* Corresponding author. Fax: 1-206-543-3041.

E-mail address: [arychel@u.washington.edu](mailto:arychel@u.washington.edu) (A.L. Rychel).

<sup>1</sup> Present address: Department of Biology, University of Washington, Seattle, WA 98195, USA.

four major mysticete groups (i.e., Neobalaenidae, Balaenidae, Balaenopteridae, and Eschrichtiidae). Similar phylogenetic conclusions resulted from both studies: (1) Balaenidae was the sister taxon to the remaining mysticetes, with Neobalaenidae being sister to a Balaenopteridae + Eschrichtiidae clade, (2) minke whales formed a separate and distinct clade within the Balaenopteridae, and (3) Balaenopteridae was rendered paraphyletic by the inclusion of Eschrichtiidae.

Resolving the interrelationships among the four major groups of mysticetes is the main goal of this study. Phylogenetic relationships are inferred among mysticete whales using the following mitochondrial and nuclear gene regions: *cyt b*, ND4L, ND4, and a primarily intronic section of  $\alpha$ -lactalbumin. The phylogenetic placement of gray whale is of particular interest, given its surprising position in previous molecular phylogenetic studies.

## 2. Materials and methods

### 2.1. DNA extraction, amplification, and sequencing

One representative from each species was sequenced and included in the phylogenetic analysis (Table 1). Samples were obtained either in the form of tissue (skin) or previously extracted DNA. DNA was extracted using the phenol–chloroform protocol of Hillis et al. (1996). Primers used in this study are given in Table 2. ND4L (297 bp) and ND4 (1378 bp) were amplified in one piece as part of a larger fragment of the mitochondrial genome (~1850 bp) that also included flanking tRNAs (i.e., His-tRNA and Ser-tRNA). A second portion of the mitochondrial genome was amplified (~1250 bp) which included *cyt b* (1140 bp) and the flanking Thr-tRNA. This second fragment was only amplified for the taxon (i.e., *Eubalaena glacialis*) not included in Arnason and Gullberg (1994). The mostly intronic portion of the nuclear  $\alpha$ -lactalbumin gene ( $\alpha$ -la; ~1150 bp) was amplified in two sections (~600 bp each).

PCR is typically performed in 25  $\mu$ l reactions using 50–200 ng total DNA, 10 mM Tris (pH 8.8), 50 mM KCl, 2 mM MgCl<sub>2</sub>, 0.1% Triton X-100, 0.12 mM each dNTP, 0.16  $\mu$ M each primer, and 1.25 U of *Taq*. Cycling parameters are noted in Table 2. PCR products are purified using polyethylene glycol precipitation (20% PEG 8000, 2.5 M NaCl) and cycle sequenced using the BigDye Terminator Sequencing Reaction Kit (Applied Biosystems). Sequencing reactions are run on an ABI 377 automated sequencer. The same primers were used for PCR and cycle sequencing, but internal cycle sequencing primers were designed as needed (Table 2). Sequencher 3.0 was used to edit and assemble sequences. Because of conserved codon reading frames, the mitochondrial protein coding gene sequences (ND4L, ND4, and *cyt b*) were easily aligned by eye (flanking tRNAs were not

included in the alignment). The  $\alpha$ -lactalbumin intron sequences were aligned with CLUSTAL X (Jeanmougin et al., 1998). Alignment of the  $\alpha$ -lactalbumin sequences was stable to varying gap opening and closing costs, both higher and lower than the default (15, 6.66). Settings of 10,5 and 5,10 produced identical alignments as the default values, and settings of 0,10, 10,0, 0,5, and 5,0 produced alignments that varied only in non-informative positions. The default setting alignment was used and no nucleotide positions were excluded from the subsequent phylogenetic analyses (Gatesy et al., 1993).

### 2.2. Phylogenetic analysis

Phylogenetic analyses were performed on the mitochondrial and nuclear data separately and on a combined data set using PAUP\* v4b10 (Swofford, 1999). Initial maximum parsimony (MP) analyses (TBR branch swapping, 100 random addition heuristic tree search) were conducted to generate a pool of trees with which to begin maximum likelihood (ML) analyses. In all cases, no more than six trees were found with the initial MP searches. When only one tree was found, it was used to estimate ML model parameters and used as the starting tree in the ML tree search. However, if more than one tree resulted from the parsimony tree search, one of the trees was chosen at random. Since ML is robust to changes in model parameter estimates (Yang, 1994) it was not critical which MP tree was chosen to begin the analysis.

Modeltest (Posada and Crandall, 1998) was used to determine the best model of sequence evolution. For each analysis, a MP tree was chosen as the basis for Modeltest, estimating the parameters to the best model found, and beginning a heuristic ML tree search. Once a ML tree was found, model testing, parameter estimation, and tree searching were repeated. This process was continued until the tree topology stabilized on the same topology for two successive rounds (Wilgenbusch and de Queiroz, 2000). The resulting trees were rooted with the cow and hippopotamus gene sequences.

Bayesian phylogenetic analyses were performed on the mitochondrial and nuclear data using MrBayes 3.0 (Huelsenbeck and Ronquist, 2001). The Bayesian analyses are launched with random starting trees and run for  $2.0 \times 10^6$  generations (sampling the Markov chains at intervals of 100 generations) using the same models as previously determined by Modeltest. Likelihood scores of sample points were plotted against generation time to determine when the Bayesian analyses had reached stationarity. Sample points generated before reaching stationarity were discarded as “burn-in” samples. Analyses were run twice for each data set to ensure the Bayesian analyses were not trapped on local optima. Apparent stationarity levels from these duplicate analyses were compared for convergence on the same posterior

Table 1  
Taxon list and classification

	GenBank Accession Nos.		
	cyt <i>b</i>	ND4L/ND4	$\alpha$ -Lactalbumin
<b>Ingroup</b>			
Mysticeti			
<i>Balaenidae</i>			
<i>Balaena mysticetus</i> —bowhead whale	X75588 <sup>a</sup>	AY398630	AJ007809
<i>Eubalaena australis</i> —Southern right whale	X75587 <sup>a</sup>	AY398627	AY398660
<i>Eubalaena glacialis</i> —Northern right whale	AY398662	AY398626	AY398647
<i>Balaenopteridae</i>			
<i>Balaenoptera acutorostrata</i> —Northern minke whale	X75753 <sup>a</sup>	AY398635	AY398649
<i>Balaenoptera bonaerensis</i> —Antarctic minke whale	X75581 <sup>a</sup>	AY398633	AY398645
<i>Balaenoptera borealis</i> —sei whale	X75582 <sup>a</sup>	AY398632	AY398658
<i>Balaenoptera edeni</i> —Bryde's whale	X75583 <sup>a</sup>	Y398631	AY398659
<i>Balaenoptera musculus</i> —blue whale	NC_001601 <sup>a</sup>	NC_001601 <sup>a</sup>	AY398655
<i>Balaenoptera physalus</i> —fin whale	NC_001321 <sup>a</sup>	NC_001321 <sup>a</sup>	AY398652
<i>Megaptera novaeangliae</i> —humpback whale	X75584 <sup>a</sup>	AY398624	AY398651
<i>Eschrichtiidae</i>			
<i>Eschrichtius robustus</i> —gray whale	X75585 <sup>a</sup>	AY398625	AF304099 <sup>a</sup>
<i>Neobalaenidae</i>			
<i>Caperea marginata</i> —pygmy right whale	X75586 <sup>a</sup>	AY398629	AY398657
<b>Outgroup</b>			
Odontoceti			
<i>Monodontidae</i>			
<i>Delphinapterus leucas</i> —beluga whale	X92531 <sup>a</sup>	AY398628	AF228409 <sup>a</sup>
<i>Ziphiidae</i>			
<i>Ziphius cavirostris</i> —Cuvier's beaked whale	AF304075 <sup>a</sup>	AY398623	AF228412 <sup>a</sup>
<i>Physeteridae</i>			
<i>Physeter macrocephalus</i> —sperm whale	AJ277029 <sup>a</sup>	AJ277029 <sup>a</sup>	AF304098 <sup>a</sup>
Artiodactyla			
<i>Bovidae</i>			
<i>Bos taurus</i> —cow	NC_001567 <sup>a</sup>	NC_001567 <sup>a</sup>	X06366 <sup>a</sup>
<i>Hippopotamidae</i>			
<i>Hippopotamus amphibius</i> —hippopotamus	NC_000889 <sup>a</sup>	NC_000889 <sup>a</sup>	AJ007813 <sup>a</sup>

<sup>a</sup> Downloaded from GenBank for this study.

Table 2  
Primers and PCR conditions

Gene	Fragment	Primer name and sequence (5'–3')	PCR conditions
$\alpha$ -Lactalbumin (Milinkovitch et al., 1998)	1	LacI.R ctcactgtcacaggagatgt	94 °C 5 min, 94 °C 30 s, 50 °C 30 s, 72 °C 30 s × 40, 72 °C 5 min
	2	LacII.F ccaaaatgatgtcctttgtc LacIII.F2 gggctctgtaccgtatttcata LacIV.R gactcaccagtaggtaattc	
cyt <i>b</i> (Árnason and Gullberg, 1996)	1	cyt <i>b</i> tRNA-glu cgagatctgaaaaaacatcgttg	94 °C 5 min, 94 °C 30 s, 54 °C 30 s, 72 °C 30 s × 40, 72 °C 5 min
	2	cyt <i>b</i> 772R gggtrtagttrtckgggtctcc cyt <i>b</i> 400F ccctgaggacaaatcatt cyt <i>b</i> tRNA-thr ggaattcatctctccggtttacaagac	
ND4L and ND4 (designed for this study)	Whole fragment	ND4 tRNA-arg gtttaarayaaaayaartgatttcgac	94 °C 5 min, 94 °C 30 s, 54 °C 30 s, 72 °C 30 s × 40, 72 °C 5 min
		ND4 tRNA-leu acttttatttggagttgcaccaat	
ND4 and ND4L (designed for this study)	Internal sequencing	ND4 29F tactaataccctracctgaytatcaa	
		ND4 371F ctacaytaatccccacccttat	
		ND4 661F gtagaagcccyatycagg	
		ND4 464R ttgagttgrttaaggaggtag ND4 794R gatccarytagttraraahaggaa	

probability distributions (Huelsenbeck and Bollback, 2001). The percentage of samples (pooled for a given data set) recovering any particular clade represents that clade's posterior probability.

While MP, ML, and Bayesian methods were employed, trees resulting from maximum likelihood and Bayesian analyses were preferred because models employed by these methods allow character weights to be chosen objectively (Felsenstein, 1981) and many aspects of molecular evolution (e.g., among site variation, unequal base frequencies) to be incorporated.

2.3. Testing competing tree hypotheses

Maximum likelihood non-parametric bootstrap support values were estimated (100 pseudoreplicates) for

trees found in each analysis. Clades with bootstrap values of  $\geq 70\%$  were considered strongly supported (Hillis and Bull, 1993). Clades with posterior probabilities of  $\geq 0.95$  were considered strongly supported because they are true probabilities of clades under the assumed models (Rannala and Yang, 1996). Any conflicting well-supported clades in the separate analyses were used to identify areas of data incongruence. Tree incongruence was also assessed using the Shimodaira–Hasegawa (S–H) test (Shimodaira and Hasegawa, 1999). Pairwise comparisons were made between trees and their data sets that contained the most conflict in topology. Significant *P* values ( $P < 0.05$ ) resulting from the comparisons highlighted incongruence for a given data set and its model parameter values for any one topology against an alternate topology. Unfortunately, we were

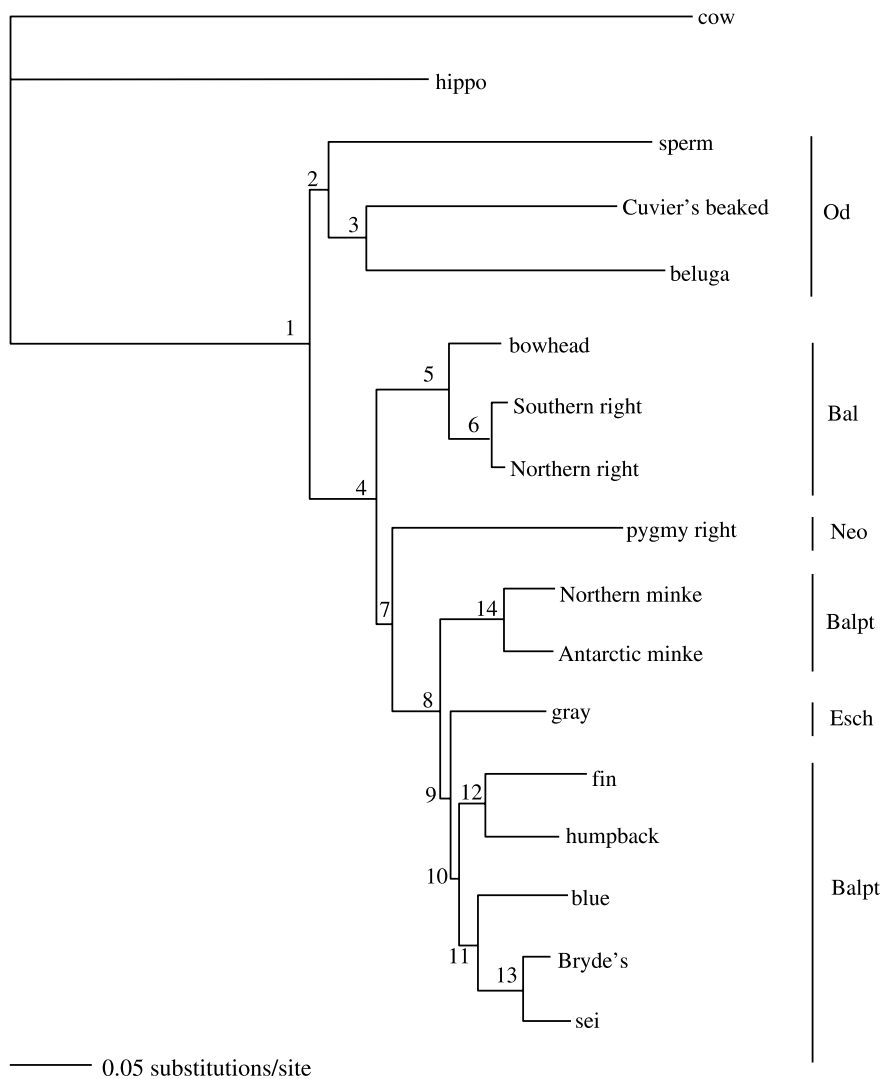


Fig. 1. Combined analysis phylogeny. The same topology resulted from the ML and Bayesian analyses of mtDNA alone as well as mt and nuDNA combined. Abbreviations: Od, Odontoceti; Bal, Balaenidae; Neo, Neobalaenidae; Esch, Eschrichtiidae; and Balpt, Balaenopteridae. Nodes are labeled to identify bootstrap and posterior probability values from each analysis reported in Table 3. GTR + I +  $\Gamma$  model used in mtDNA ML analysis and in mt and nuDNA combined ML analysis.

unable to perform maximum likelihood tree searches using a heterogeneous model analysis. However, the additive nature of the likelihood score allows for the summation of likelihood scores from individual data partitions using models and parameters specific for each partition on competing tree topologies (Wilgenbusch and de Queiroz, 2000). Using this method we have the ability to uncover hidden support for trees not necessarily found in a one model (one parameter set) combined data tree search.

### 3. Results

#### 3.1. Phylogenetic trees

##### 3.1.1. MtDNA

The phylogeny resulting from the ML and Bayesian analyses of the mitochondrial data (826 variable positions; 551 parsimony informative positions) is shown in Fig. 1. In general, most nodes of this tree were strongly supported (bootstrap [BS]  $\geq 70\%$  and posterior probabilities [PP]  $\geq 0.95$ ; Table 3). The monophyly of the Balaenidae (bowhead and right whales; node 5) was strongly supported, as was its placement as the sister taxon to the clade containing the remaining mysticetes (i.e., neobalaenids, eschrichtiids, and balaenopterids). Within this more exclusive clade, the pygmy right whale (*C. marginata*; Neobalaenidae) was strongly placed as the sister taxon to the remaining mysticetes. One of the weakest supported relationships involved the placement of the gray whale (*E. robustus*; Eschrichtiidae) within the Balaenopteridae (node 9; BS = 58, PP = 0.82); thus rendering the balaenopterids paraphyletic.

##### 3.1.2. NuDNA

The phylogeny inferred from the ML and Bayesian analyses of the  $\alpha$ -lactalbumin data (36 variable posi-

tions; 27 parsimony informative positions) was not as well resolved (Fig. 2) as the phylogeny resulting from the mtDNA, with a trichotomy at the base of the mysticete clade and another trichotomy within Balaenopteridae. As in the mtDNA phylogeny, the Balaenopteridae is paraphyletic with respect to the gray whale. Whereas the mtDNA and nuDNA are congruent in supporting balaenopterid paraphyly, the inferred species relationships within the balaenopterid–eschrichtiid clade are incongruent between these two sources of data. The nuDNA strongly places the fin whale (*Balaenoptera physalus*) as the sister species of the remaining taxa of this clade whereas the mtDNA strongly places the fin whale with the humpback (*Megaptera novaeangliae*).

##### 3.1.3. Combined DNA

The phylogenetic hypothesis based on the combined mt and nuDNA was identical to the phylogeny inferred from the separate ML and Bayesian analyses of the mtDNA. Also, as in the mtDNA analyses, all mysticete relationships are strongly supported except for the basal relationships within the balaenopterid–eschrichtiid clade (node 8). Whereas there is strong support for this major clade containing the gray whale and balaenopterids, balaenopterid paraphyly is not well supported because of the weak placement of the gray whale (also evident in the separate analysis of the mtDNA). This is in sharp contrast to the separate analysis of the nuDNA which strongly supported the paraphyly of Balaenopteridae (Fig. 2).

#### 3.2. Additive likelihood and S–H test

Additive likelihood scores were compiled in two sets. The first set was the sum of likelihood scores from the mtDNA codon position data partitions plus  $\alpha$ -lactalbumin. The second set was the sum of the likelihood scores from the gene data partitions plus  $\alpha$ -lactalbumin.

Table 3  
Maximum likelihood bootstrap support (BS) and Bayesian posterior probabilities (PP)

Node	Mt data BS	Mt data + nu data BS	Mt data PP	Mt data + nu data PP
1	100	100	100	100
2	66	97	59	100
3	90	100	100	100
4	100	100	100	100
5	100	100	100	100
6	100	100	100	100
7	67	71	97	97
8	99	100	100	100
9	58	<50	82	90
10	69	<50	93	89
11	100	89	100	100
12	79	76	100	100
13	100	100	100	100
14	100	100	100	100

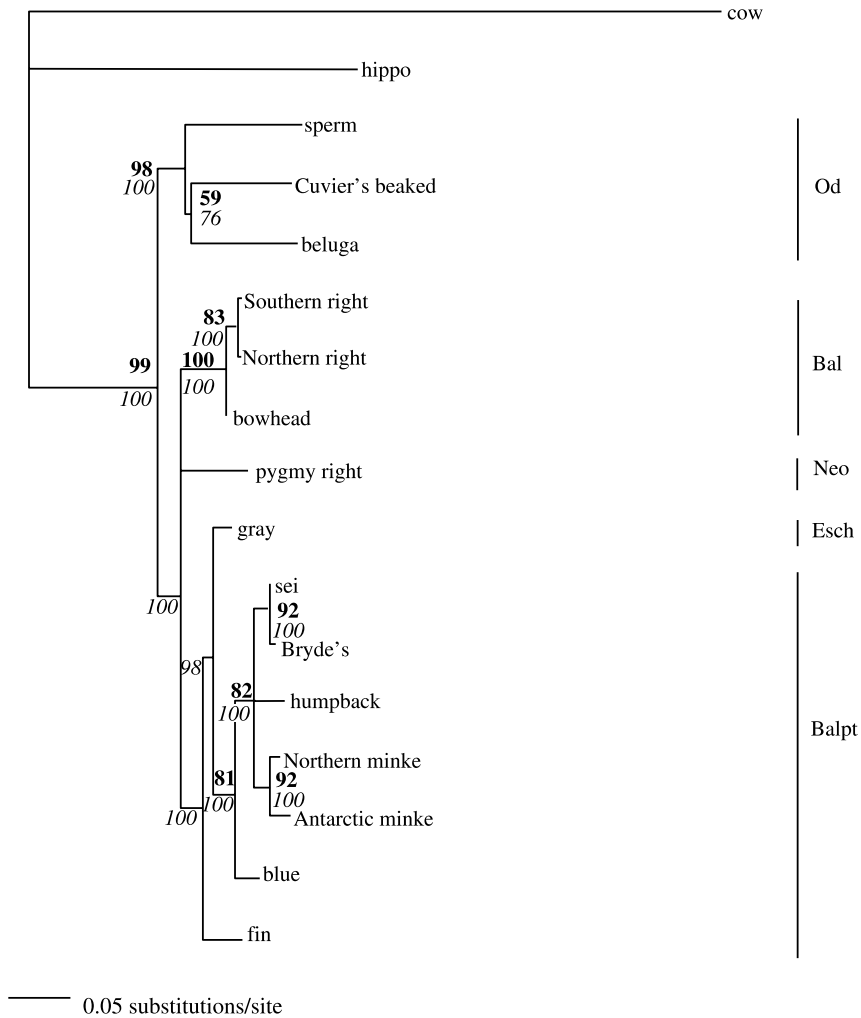


Fig. 2.  $\alpha$ -Lactalbumin phylogeny. Maximum likelihood bootstrap  $\geq 50$  shown in bold, Bayesian posterior probabilities  $\geq 50$  shown in italics. The Bayesian tree supported the position of the pygmy right whale between the Bal and Balpt-Esch clades with a posterior probability of 50. Abbreviations: Od, Odontoceti; Neo, Neobalaenidae; Bal, Balaenidae; Esch, Eschrichtiidae; and Balpt, Balaenopteridae. ML model used HKY +  $\Gamma$ .

Four competing topologies were selected for this analysis. Since the species relationships among the Balaenidae and the Balaenopteridae varied little between any analyses conducted, only the positions of the pygmy right whale (Neobalaenidae) and the gray whale (Eschrichtiidae) varied in the competing topologies. The first tree (Fig. 3A) was one of the two trees found in a separate analysis of only the ND4 gene. This tree had a monophyletic Balaenopteridae and pygmy right whale diverging off the stem of the tree between the balaenids and the gray whale. The second tree (Fig. 3B) was the combined data tree with balaenopterids paraphyletic and with the pygmy right whale in the same position as the first tree. The third tree (Fig. 3C) was the  $\alpha$ -lactalbumin tree, and the fourth tree (Fig. 3D) was the cytochrome *b* tree (paraphyletic balaenopterids and the pygmy right whale sister to the balaenids). On each of these four competing topologies, the likelihood scores were calculated, using partition specific models and es-

timated parameters. Then the scores found for each partition were added, resulting in an additive likelihood score for each competing topology (Table 4). The topology with the best overall likelihood score was considered the best estimate of mysticete phylogeny given these data sets. The best scores generated for both partitioning schemes were for the topology in which the balaenopterids were monophyletic with the gray whale as a sister taxon and the pygmy right whale diverging off the stem between the balaenids and the gray whale (Fig. 3A).

S–H tests were conducted between data sets and their corresponding trees that were most in conflict. The combined data (and original estimated model parameters) were compared to the best alternative hypothesis tree found with the additive likelihood scores (monophyletic balaenopterids). This comparison was not significantly different ( $P = 0.280$ ), indicating this alternate hypothesis is not a statistically less favorable explana-

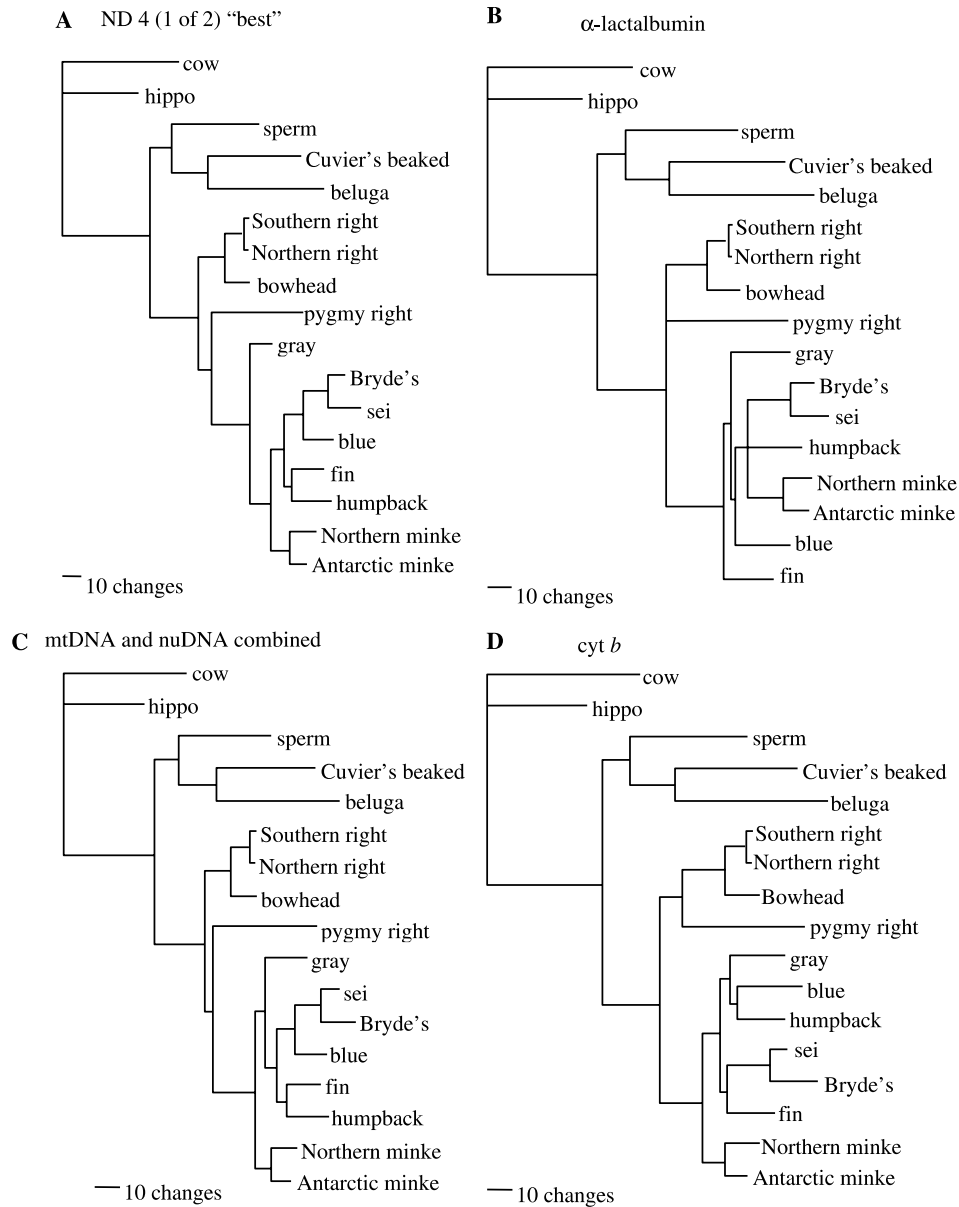


Fig. 3. Comparison trees for additive likelihood.

Table 4  
Additive likelihood scores

Partition	"Best"	Combo	$\alpha$ -La tree	<i>cyt b</i>	Standard error
ND4	<b>8075.69871</b>	<b>8075.69871</b>	8107.71421	8108.46506	4.675855337
ND4L	1813.82109	<b>1813.79173</b>	1820.48812	1820.10531	0.937614013
<i>cyt b</i>	6459.4547	6457.31723	6470.46279	<b>6455.84654</b>	1.657321325
$\alpha$ -La	3374.04901	3388.15553	<b>3336.40883</b>	3398.4489	6.786850184
Total	<b>19723.0235</b>	19734.9632	19735.074	19782.8658	
Difference		11.93969	12.05044	59.8423	
1pos	<b>4259.80692</b>	<b>4259.80692</b>	4278.3473	4261.75101	2.001289275
2pos	<b>2212.95769</b>	2215.08688	2221.64098	2218.57769	1.180790446
3pos	8972.70028	<b>8970.08487</b>	9006.73757	9006.73757	3.943234315
$\alpha$ -La	3496.33894	3509.65068	<b>3458.63706</b>	3509.78709	6.786850184
Total	<b>18941.80383</b>	18954.62935	18965.36291	18996.85336	
Difference		13.44377	14.54136	67.41782	

The sum of data partitions favors the tree with balaenopterids monophyletic ("best"). Total scores were lower than the combined data analysis score ( $-\ln$  likelihood: 20208.16567). Best scores for each partition or total are in bold.

tion for the outcome of the combined data. The last two S–H tests were reciprocal comparisons between the mtDNA and nuDNA data and optimal trees. Both comparisons were significant ( $P = 0.001$  for mtDNA data against the nuDNA tree and  $P = 0.034$  for nuDNA data against the mtDNA tree). Thus, the optimal nuDNA tree is a statistically worse explanation of the mtDNA data than the optimal mtDNA tree and vice versa, indicating a general incongruence between these data sets.

## 4. Discussion

### 4.1. Data partitions and combined analysis

Based on morphology, cetologists have always assumed a monophyletic Balaenopteridae (Geisler and Sanders, 2003; McLeod et al., 1993). The type and number of ventral grooves (numerous and extensive), the degree of arching of the rostrum (nearly flat), and the type of telescoping (extreme interdigitation of rostral and cranial bones excluding parietal and frontal from vertex) are all very similar among balaenopterids. The gray whale has very different states for these characters when compared to other baleen whales i.e., ventral grooves confined to throat region, moderately arched rostrum, and moderate interdigitation of rostral and cranial bones with parietal and frontal exposed on vertex (Deméré and Berta, unpublished). If the gray whale were included within the balaenopterids, then all of the morphological characters that unite the balaenopterids to the exclusion of the gray whale would have to had reversed in the gray whale lineage. Our preferred tree presented in the paper of a monophyletic Balaenopteridae with the gray whale as sister presents a more parsimonious explanation of the evolution of this group.

By combining data into a single analysis, a better estimate of the true phylogeny may be obtained and/or overall clade support may be improved. However, this is based on the assumption that the several data partitions are not strongly incongruent. In the debate over whether to combine data or not, both possibilities have been shown to occur. Bull et al. (1993) caution against combining data partitions if heterogeneity is known to exist between them. In fact, combining partitions with different rates of change may lead to lowered chances of recovering the correct phylogeny. Using a differential weighting scheme (Chippindale and Wiens, 1994) is one method to improve on combining character sets with different rates of evolution, only we do not often know what values to weight different character partitions, and this approach limits the method of phylogenetic inference to parsimony. Even though combining data will continue to be a contentious issue, an argument for combining data, even between data sets with different

phylogenetic histories can be made (Wiens, 1998). In simulations, Wiens (1998) demonstrates that localized areas of conflict between data sets may not disrupt overall analyses, and in areas of data congruence, combining data strengthens the overall accuracy of the analysis. As long as the data are first analyzed separately and areas of incongruence are identified, the combined analysis may still be the preferred result, although those clades strongly in conflict should be cautiously interpreted and treated as unresolved.

The bulk of the data collected for this study is from protein coding regions of the mitochondrial genome. These data share the same evolutionary history since they are physically linked. Furthermore, they should be evolving in a relatively similar fashion since they both code for functional proteins and are not structural. The  $\alpha$ -lactalbumin data were collected in order to have an additional independent data set. The S–H analysis indicated strong incongruities between the mt and nuDNA data sets. Comparing the bootstrap values and posterior probabilities between the two trees allows us to pinpoint the locations and taxa involved with the incongruence. Much of this incongruence lies in the placement of the gray and fin whales. Overall, the topologies between the mtDNA and nuDNA trees are not strikingly different, although in the  $\alpha$ -lactalbumin (nuDNA) tree, the position of the pygmy right whale was unresolved and the basal position of the fin whale had never been seen before in any other analysis (Fig. 2). This data set in general exhibited very little homoplasy (C.I. = 0.880) yet the samples of fin whale sequenced shared four characters with non-balaenopterid whales where no other balaenopterid shared these characters. These four characters, shown in Table 5, are responsible for positioning the fin whale basal to the gray whale. Only one sequence of fin whale  $\alpha$ -lactalbumin is known from the literature (Bérubé and Aguilar, 1998), and it did not have any of these anomalous characters. For this study six fin whale individuals from populations in the Eastern and Western Pacific and the Atlantic were sequenced to confirm these four characters. Since these sequences were generated for this study, they are the ones preferred in the analysis. A possible explanation for the difference between the sequence from the literature and ours is introgression may have occurred in the fin whale lineage. Recently, a new balaenopterid species was described, *Balaenoptera omurai* (Wada et al., 2003). Specimens of this new whale were not available at the time the present study was conducted, hence it was not included. However, we can speculate on where it may be placed had it been included. Likely, *B. omurai* would be a sister taxon to the sei + Bryde's whale clade in a mitochondrial analysis. However, in light of the fact that fin and blue whales are known to hybridize, it is possible that the mitochondrial lineage of *B. omurai* is the result of a hybridization event between a fin or blue whale and

Table 5  
 $\alpha$ -Lactalbumin matrix indicating (with shading) homoplastic characters for the fin whale (Bphy)

Balpt	Bphy2364	TCATCTGTTTACTCTTTTATTACATTTATTACCTATCTCTCCTTTCTCC	CATTGTCTGAT	TTT
	Bphy2413	.....	.....	.....
	Bphy10743	.....	.....	.....
	Bphy4767	.....	.....	.....
	Bphy26295	.....	.....	.....
	Bphy761	.....	.....	.....
	Bphyref	.....	C.....C.....	.....T.....T.....
	finXblue	.....	C.....C.....	.....T.....TK.....
	Bmuscref	.....	C.....C.....	.....T.....TG.....
	Bmusc	.....	C.....C.....	.....T.....TG.....
	Bbor	.....	C.....C.....	.....T.....TG.....
	Bedeni	.....	C.....C.....	.....T.....TG.....
	Mnov	.....	C.....C.....	.....T.....TG.....
	Bacutonp	.....	C.....C.....	.....T.....TG.....
	Bacutona	.....	C.....C.....	.....T.....TG.....
	Bbon	.....	C.....C.....	.....T.....TG.....
Neo	Cmar	.....	.....	.....G.....
Esch	Erobust	.....	.....	.....
	Eglacial	.....	A.....A.....	.....C.....T.....A.....
Bal	Bmyst	.....	A.....	.....C.....T.....G.....
	Ejap	.....	A.....A.....	.....C.....T.....A.....
	Eaust	.....	A.....A.....	.....C.....T.....A.....
Od	Zcav	C.....	.....	.....C.....G.....
	Pmac	.....	.....	.....C.....A.....
	Dleu	.....	.....	.....T.....TG.....

Bphyref, Bmuscref, and finXblue (hybrid) sequences were taken from Bérubé and Aguilar (1998). Abbreviations: Balpt, Balaenopteridae; Neo, Neobalaenidae; Esch, Eschrichtiidae; Bal, Balaenidae; and Od, Odontoceti (outgroup). Bphy GenBank accession numbers from this study are AY398648, AY398650, AY398652, AY398653, AY398654, and AY398656.

a sei or Bryde's whale. Nuclear markers will be required in order to test this hypothesis, and as such it would be difficult to predict where in the  $\alpha$ -lactalbumin analysis, *B. omurai* would be placed.

Since most of the incongruence between the mtDNA and nuDNA data sets resides in the position of the gray and fin whales, it is interesting that the posterior probability for the placement of the gray whale actually increased (82–90) upon the inclusion of  $\alpha$ -lactalbumin in the combined analysis. The posterior probability values increasing with the inclusion of  $\alpha$ -lactalbumin maybe due to the fact that Bayesian methods may be more susceptible to model misspecification (Buckley, 2002; Shimodaira, 2001).

Two problems that could potentially disrupt the strength of a combined analysis are model misspecification and data set heterogeneity. One way to explore potential data heterogeneity is to use a parametric bootstrap approach (Hillis and Bull, 1993). The parametric bootstrap was not employed in this study because it is not computationally feasible. The S–H tests employed in this study did find significant differences between all tree topologies compared which indicates the strong possibility of data heterogeneity playing a major role in this study. Model misspecification is more difficult to assess. The model parameters implemented in the combined analysis do not perfectly fit any one of the data partitions. Another method that avoids the pitfalls of model misspecification is an additive likelihood ap-

proach that uses partition specific models and parameters. In this method, the combination of data partitions supports a topology not found as the best tree in the combined analysis (Table 4). Most likely, this result is influenced by the  $\alpha$ -lactalbumin data that was possibly overwhelmed or mis-modeled in the combined data analysis.

### Acknowledgments

This project was conducted in conjunction with the MS thesis work of Amanda L. Rychel at San Diego State University. Without the generous support of Andy Dizon and Kelly Robertson at the Southwest Fisheries Science Center, this project would not have been possible, since they provided us with the majority of the whale tissues or DNA needed. Steve Donnellan at the South Australian Museum and Úlfur Árnason at the University of Lund, Sweden made other DNA and tissue loans essential to the completion of this project. We also thank two anonymous reviewers for providing helpful insights and comments on the manuscript. Funding for this project was provided by an RSCA grant from San Diego State University to A.B. and SDSU graduate student funds to A.L.R. NSF funding (DEB-9707428) to T. R. indirectly contributed to the completion of this project.

## References

- Árnason, Ú., Gullberg, A., 1994. Relationship of baleen whales established by cytochrome *b* gene sequence comparison. *Nature* 367, 726–728.
- Árnason, Ú., Gullberg, A., 1996. Cytochrome *b* nucleotide sequences and the identification of five primary lineages of extant cetaceans. *Mol. Biol. Evol.* 13 (2), 407–417.
- Árnason, Ú., 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 (5), 960–970.
- Bérubé, 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 (1), 82–98.
- Buckley, T.R., 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. *Syst. Biol.* 51 (3), 509–523.
- Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* 42 (3), 384–397.
- Chippindale, P.T., Wiens, J.J., 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. *Syst. Biol.* 43 (2), 278–287.
- Felsenstein, J., 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *J. Linn. Soc.* 16, 183–196.
- Gatesy, J., 1997. More DNA support for a Cetacea/Hippopotamidae clade: the blood-clotting protein gene  $\gamma$ -fibrinogen. *Mol. Biol. Evol.* 14 (5), 537–543.
- Gatesy, J., DeSalle, R., Wheeler, W., 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. *Mol. Phylogenet. Evol.* 2, 152–157.
- Geisler, J.H., Sanders, A.E., 2003. Morphological evidence for the phylogeny of the Cetacea. *J. Mam. Evol.* 10 (1/2), 23–129.
- Geisler, J.H., Uhen, M., 2003. Support for a close relationship between hippos and whales. *J. Vertebr. Paleontol.* 23 (4), 991–996.
- Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42, 182–192.
- Hillis, D.M., Moritz, C., Mable, B.K., 1996. *Molecular Systematics*. Sinauer Associates, Sunderland, MA.
- Huelsenbeck, J.P., Bollback, J.P., 2001. Empirical and hierarchical Bayesian estimation of ancestral states. *Syst. Biol.* 50 (3), 351–366.
- Huelsenbeck, J.P., Ronquist, F.R., 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17, 751–755.
- Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., Gibson, T.J., 1998. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* 23, 403–405.
- McLeod, S.A., Whitmore Jr., F.C., Barnes, L.G., 1993. Evolutionary relationships and classification. In: Burns, J.J., Montague, J.J., Cowles, C.J. (Eds.), *The Bowhead Whale*. Special Publ. 2. Society for Marine Mammalogy, Lawrence, pp. 45–70.
- Milinkovitch, M.C., Bérubé, M., Palsbøll, P.J., 1998. Cetaceans are highly derived artiodactyls. In: Thewissen, J.G.M. (Ed.), *The Emergence of Whales*. Plenum Press, New York, pp. 113–131.
- Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14 (9), 817–818.
- Rannala, B., Yang, Z., 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol.* 43, 304–311.
- Rosenbaum, H.C., Brownell Jr., R.L., Brown, M.W., Schaeff, C., Portway, V., White, B.N., Malik, S., Pastene, L.A., Patenaude, N.J., Baker, C.S., Goto, M., Best, P.B., Clapham, P.J., Hamilton, P., Moore, M., Payne, R., Rowntree, V., Tynan, C.T., Bannister, J.L., DeSalle, R., 2000. World-wide genetic differentiation of *Eubalaena*: questioning the number of right whale species. *Mol. Ecol.* 9 (11), 1793–1802.
- Shimamura, M., Yasua, H., Ohshimia, K., Abe, H., Kato, H., Kishiro, T., Gotos, M., Munechikai, I., Okada, N., 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* 388, 666–670.
- Shimodaira, H., 2001. Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic tree selection. *Comm. Stat. A Theory Methods* 30, 1751–1772.
- Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16 (8), 1114–1116.
- Swofford, D.L., 1999. PAUP\*: Phylogenetic Analysis Using Parsimony (\* and other methods). Version 4. Sinauer Associates, Sunderland, MA.
- Wada, S., Oishi, M., Yamada, T.K., 2003. A newly discovered species of living baleen whale. *Nature* 426, 278–281.
- Wiens, J.J., 1998. Combining data sets with different phylogenetic histories. *Syst. Biol.* 47, 568–581.
- Wilgenbusch, J., de Queiroz, K., 2000. Phylogenetic relationships among the phrynosomatid sand lizards inferred from mitochondrial DNA sequences generated by heterogeneous evolutionary processes. *Syst. Biol.* 49 (3), 592–612.
- Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* 39, 306–314.