Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 41 (2006) 12–27
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Phylogenetic relationships among the baleen whales based on maternally and paternally inherited characters

Leila T. Hatch [a,*], Erik B. Dopman [b], Richard G. Harrison [a]

[a] *Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA*
[b] *Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA*

Received 28 January 2005; revised 31 March 2006; accepted 21 May 2006
Available online 26 May 2006

Abstract

Phylogenetic relationships in the Cetacean suborder Mysticeti (baleen whales) have recently been the focus of increased attention. Here, we examine the evolutionary history of this group by comparing genealogies derived from Y chromosome and mitochondrial DNA sequences. We generated topologies based on paternally and maternally inherited characters for males from nine baleen whale species, including representatives of three families (Balaenidae, Eschrichtiidae, and Balaenopteridae) and four genera (*Balaena*, *Eschrichtius*, *Balaenoptera*, and *Megaptera*). Divergence among species was fifteen times greater for mtDNA than for Y-specific DNA. Both mtDNA and yDNA topologies revealed the family Balaenopteridae to be paraphyletic, but this relationship was neither strongly supported nor consistent across phylogenetic analysis methodologies. Humpback and fin whales, representing different genera, were reciprocally monophyletic sister species according to mtDNA. Although the monophyly of fin whales decayed for yDNA, a close relationship between fin and humpback whales was retained in yDNA trees. The paraphyly of fin whales and the long branch leading to humpback whales for the yDNA marker may suggest life history differences between these species. Specifically, male humpback whales showed higher than average divergence from other baleen whales at yDNA, although not at mtDNA, suggesting a potential for smaller effective population sizes among male humpbacks on an evolutionary timescale. The observation that those species that have been found to hybridize in nature (blue/fin and blue/humpback) do not reveal evidence for paraphyly for either maternal or paternal markers suggests that introgressive hybridization has not historically been extensive and thus may not represent a substantial source of phylogenetic error for Mysticeti.

Keywords: Phylogeny; Y chromosome; Baleen whales

1. Introduction

Evolutionary relationships at all levels within the order Cetacea (whales, dolphins, and porpoises) have been controversial (reviewed in Perrin and Reeves, 2004). Although the suborder Mysticeti (baleen whales) is generally accepted as a monophyletic group, relationships among some species within this clade remain unresolved in spite of a recent surge in the collection of molecular and morphological data (Rosenbaum et al., 2000; Gatesy et al., 2002; Nishida et al., 2003; Wada et al., 2003; Rychel et al., 2004).

Mysticetes are comprised of four families: the Balaenidae (the right and bowhead whales), the Neobalaenidae (the pygmy right whale), the Eschrichtiidae (the gray whale), and the Balaenopteridae (the humpback whale, the fin whale, the blue whale, the minke whales, the Bryde's whales, and the sei whale). Members of the family Balaenidae were initially characterized as the most basal of the mysticetes based on morphology and early molecular evidence from allozymes and the distribution of satellite DNA (Wada and Numachi, 1991; Árnason et al., 1992). More recent molecular data have supported bowhead and right whales as basal mysticetes, but have also led to recognition of three rather than two species within the right whale

* Corresponding author. Present address: Gerry E. Studds Stellwagen Bank National Marine Sanctuary, US National Oceanic and Atmospheric Administration, 175 Edward Foster Road, Scituate, MA 02066, USA. Fax: +1 781 545 8036.
E-mail address: Leila.Hatch@noaa.gov (L.T. Hatch).

1055-7903/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2006.05.023

genus (*Eubalaena*) (Rosenbaum et al., 2000). That Neobalaenidae, a monotypic family, are closely related to Balaenidae was resolved in early studies and remains uncontroversial. Phylogenetic relationships of the gray whale (family Eschrichtiidae), and the rorquals (family Balaenopteridae), however, are less clear. Phylogenies based on the mitochondrial cytochrome *b* gene and the mitochondrial control region nest the gray whale within the rorquals, consistently supporting the family Balaenopteridae as paraphyletic (Árnason et al., 1993; Árnason and Gullberg, 1994). Consensus trees that integrate morphological evidence, however, place gray whales outside the Balaenopteridae, due, presumably, to the large number of morphological differences associated with the evolution of this species' unique bottom-sucking feeding behavior (Messenger and McGuire, 1998).

Until recently, molecular phylogenetic analyses of cetaceans have relied primarily on the mitochondrial genome, which, because it does not recombine and is maternally inherited in mammals, represents a single, female-mediated, estimate of evolutionary relationships among species. Relationships within and among the two genera within the family Balaenopteridae, the short-flippered fin, blue, sei, Bryde's and minke whales (genus *Balaenoptera*) and the long-flippered humpback whale (genus *Megaptera*) vary among mtDNA gene genealogies. For example, cytochrome *b* sequences indicate humpback and blue whales as sister taxa (Árnason and Gullberg, 1994), mtDNA control region sequences reveal fin and blue whales as sister taxa (Árnason et al., 1993), and ND4 sequences suggest humpback and fin whales as sister taxa (Rychel et al., 2004). Relationships within the family Balaenopteridae are of particular interest in light of known instances of hybridization among rorqual species. Five blue-fin hybrid whale specimens have been described (Árnason et al., 1991; Spilliaert et al., 1991; Bérubé and Aguilar, 1998), as well as one humpback-blue hybrid (M. Poole, personal communication). Whether introgression occurs among rorqual species remains unknown, but is of great importance given their endangered status (Perry et al., 1999).

In an attempt to address inconsistencies between cetacean phylogenies, Messenger and McGuire (1998) compared a large data set of morphological characters to mitochondrial 12S, 16S, and cytochrome *b* sequence data. However, this study sampled only two rorqual species and was unable to resolve relationships within the family Balaenopteridae. Likewise, Gatesy et al.'s (2002) supertree analysis based on multiple morphological characters as well as nuclear and mitochondrial sequence data did not include all rorqual species, but found fin and humpback whales to be sister taxa to the exclusion of blue and minke whales. Recently, Rychel et al. (2004) compared a nuclear gene genealogy (α-lactalbumin) to genealogies based on two regions of the mitochondrial genome (ND4 and cytochrome *b*). Combined analysis again showed fin and humpback whales to be sister taxa, with gray whales nested in the Balaenopteridae and minkes forming a basal clade within that family. However, when analyzed alone, nuclear data supported a more basal position for fin whales, not minke whale species, within the Balaenopteridae. Rychel et al. (2004) proposed that recent mtDNA introgression into the fin whale lineage could explain the different positions of fin whales in genealogies derived from independent markers.

Nishida et al. (2003) published a gene geneology for cetaceans based on sequence data from the sex-determining region (SRY) of the Y chromosome. The Y chromosome (yDNA) is paternally inherited, and the majority of the chromosome, termed "male-specific" (Skaletsky et al., 2003) does not recombine. Using 750 base pairs of yDNA, Nishida et al. (2003) resolved relationships among a subset of six mysticete species. They found a sister relationship between fin and humpback whales, with minke whales basal among the four rorqual species sampled, and right whales basal within the mysticetes. Gray whales, sei whales, and Bryde's whales were not included in the Nishida et al. (2003) study, and therefore the paraphyly of the Balaenopteridae was not addressed.

We have recently developed additional markers on the Y chromosome of the fin whale (Hatch, 2004). These new markers represent previously uncharacterized non-coding regions of the Y chromosome. Here, we present sequences from nine mysticete species and one odontocete outgroup for two anonymous yDNA loci totaling 1040 base pairs. In addition, partial mitochondrial control region sequences are generated for the same males, to allow direct comparison between paternally and maternally inherited estimates of baleen whale phylogeny. We re-examine relationships within the mysticete clade, with particular attention to the placement of fin and gray whales, and discuss hypotheses accounting for differences among the published phylogenies for this group.

2. Methods

2.1. DNA extraction and sexing

We received tissue samples from wild free-swimming or stranded whales from multiple individual collectors and institutions. Samples were either tissues (skin and/or blubber) preserved in 20% dimethyl sulfoxide (DMSO) or aliquots of pre-extracted DNA in TE buffer (10 mM Tris–Cl, pH 8.4, and 0.5 mM EDTA). Tissues from 45 fin whales, three blue whales, three Northern minke whales, four "small type" Bryde's whales, four sei whales, four gray whales, three bowhead whales, and one sperm whale were the starting material for our study.

Genomic DNA was extracted using Qiaquick DNA Extraction Kits (Qiagen, Inc.) and eluted in dilute TE buffer (10 mM Tris–Cl and 0.5 mM EDTA). The gender of all individuals sampled was determined by amplification of a Y-specific region (SRY) and an X/Y homologous gene (ZFX/Y) using primers and cycling conditions described in Palsbøll et al., 1992 and Bérubé and Palsbøll, 1996; 10 μl PCRs contained buffer (20 mM Tris–Cl, pH 8.4, and 50 mM

KCl), 0.2 mM each dNTP, 1 mM $MgCl_2$, 1 μM each primer and 0.5 U *Taq* (Platinum® *Taq* DNA polymerase, 5 U/μl). Amplifications were performed in Hybaid® Omni Gene or Hybaid® PCR Express (Hybaid, Inc.) thermal cyclers. Following sex-typing, 45 male whales, including 28 male fin whales, two males from each of six additional baleen whale species, and a single male sperm whale were available for analysis (Table 1). Fin whale samples were collected from individuals in the central and eastern North Pacific Ocean and the western North Atlantic Ocean. Humpback and sei whale samples were collected from North Pacific and North Atlantic populations. Samples for all other species were collected solely from North Pacific populations. Females from each species were used as negative controls in verifying the male-specificity of PCR products.

2.2. Data collection and alignment

The 45 male samples were PCR amplified and sequenced for three "anonymous" loci developed on the fin whale Y chromosome; ylocus2, ylocus10, and ylocus13 (Hatch, 2004). Primers and PCR conditions for ylocus10 and ylocus13 were as presented in Hatch (2004). Ylocus2 failed to amplify consistently in sperm, bowhead, sei, and Bryde's whales using conditions developed for fin whales. In response, we used "touchdown" PCR (modified from

Table 1
Tissue samples from male baleen whales used in this study

Identification		Taxonomy			
Source	LablD#	Family	*Genus*	Species	Common
SWFSC[a]	4631	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	4767	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	7832	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	7835	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	8644	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	10744	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	14336	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	15897	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	23640	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	24798	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	25397	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	25399	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	25400	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	25403	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	1862	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	2821	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	5819	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	5820	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	5822	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	5824	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	6249	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
SWFSC	6254	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
M. Berube	GM950003	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
M. Berube	GM950027	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
NEFSC[b]	DE020744	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
NEFSC	DE020745	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
NEFSC	DE020746	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
NEFSC	DE020753	Balaenopteridae	*Balaenoptera*	*physalus*	Fin
S. Cerchio	00S151	Balaenopteridae	*Megaptera*	*novaeangliae*	Humpback
S. Cerchio	99C027	Balaenopteridae	*Megaptera*	*novaeangliae*	Humpback
SWFSC	2313	Balaenopteridae	*Balaenoptera*	*acuturostrata*	Minke
SWFSC	5318	Balaenopteridae	*Balaenoptera*	*acuturostrata*	Minke
SWFSC	3999	Balaenopteridae	*Balaenoptera*	*borealis*	Sei
NEFSC	82	Balaenopteridae	*Balaenoptera*	*borealis*	Sei
SWFSC	7619	Balaenopteridae	*Balaenoptera*	*musculus*	Blue
SWFSC	7620	Balaenopteridae	*Balaenoptera*	*musculus*	Blue
SWFSC	15911	Balaenopteridae	*Balaenoptera*	*edeni*	Bryde's
SWFSC	15912	Balaenopteridae	*Balaenoptera*	*edeni*	Bryde's
SWFSC	546	Eschrichtiidae	*Eschrictius*	*robustus*	Grey
SWFSC	1287	Eschrichtiidae	*Eschrictius*	*robustus*	Grey
SWFSC	6970	Balaenidae	*Balaena*	*mysticetus*	Bowhead
SWFSC	6971	Balaenidae	*Balaena*	*mysticetus*	Bowhead
SWFSC	75	Physeteridae	*Physeter*	*macrocephalus*	Sperm

[a] US National Marine Fisheries Service's Southwest Fisheries Science Center (NOAA).
[b] US National Marine Fisheries Service's Northeast Fisheries Science Center (NOAA).

Sambrook and and Russel, 2001, p. 8.112) for these taxa, with annealing temperatures starting at 64 °C and gradually lowered to 55 °C. Bands of expected size were gel extracted using Qiaquick Gel Extraction Kit protocol (Qiagen, Inc.), and the purified template was PCR amplified for sequencing.

We designed primers to amplify ~416 base pairs of the fin whale mitochondrial control region (15,851–16,267, Accession #X72204, Árnason et al., 1993). Mitochondrial DNA (mtDNA) PCR was performed in 10 μl total volumes containing 1× buffer (20 mM Tris–HCl, pH 8.4, and 50 mM KCl), 0.2 mM each dNTP, 1 mM $MgCl_2$, 1 μM each primer and *Taq* (Platinum® *Taq* DNA polymerase, 5 U/μl). MtDNA was also amplified using touchdown PCR. Annealing temperatures were lowered gradually from 62 to 53 °C over the course of 16 cycles. The final 19 cycles annealed at 53 °C, followed by a 10 min extension at 72 °C. All sequences were determined using ABI PRISM® 377 DNA Sequencer technology (Applied Biosystems, Foster City, CA, USA).

Sequences were compared and aligned using Lasergene Navigator software (DNAStar, Inc., Madison, WI, USA). Ylocus10 and ylocus13 showed few indels among sequences, and alignment was easily verified by eye. However, alignment of ylocus2 sequences proved more difficult, and preliminary results suggested that not all sequences were orthologs. Studies of human and ape Y chromosomes have shown the mammalian Y to be highly repetitive in nature (Skaletsky et al., 2003; Rozen et al., 2003). Additionally, the majority of sequence in the non-recombining region of the human Y chromosome was found to show high similarity to X chromosome-specific and autosomal sequences (Skaletsky et al., 2003). These studies underscore the importance of strict orthology criteria in estimating higher-level relationships among species based on yDNA. Therefore, ylocus2 was removed from further analyses. MtDNA control region sequences were aligned using the ClustalW algorithm (Chenna et al., 2003), with gap opening and closing costs set to defaults (15, 6.66). Varying default settings produced alignments that differed only in non-informative positions.

2.3. Phylogenetic analyses

We employed ModelTest 3.04 (Posada and Crandall, 1998) to determine the model of DNA substitution that best fit each data partition (ylocus10 and ylocus13, both individually and together, and mtDNA). Estimates of base frequencies, the ratio of transitions to transversions, the distribution of substitution rates (γ) and the proportion of invariant sites for each sequence were used to inform decisions about data combinability and phylogenetic analysis. Pairwise sequence divergences for yDNA (combined) and mtDNA, both uncorrected and corrected for within-species divergence, were calculated among all whales, including the sperm whale outgroup, using PAUP* (version 4.0b10, Swofford, 2003) and Arlequin v2.001 (Schneider et al., 2000).

PAUP* 4.0b10 was used to estimate the phylogenetic signal at the Y loci and mtDNA using the g1 statistic for skewness of tree distributions estimated from 100,000 random trees and published critical values (Hillis and Huelsenbeck, 1992). Partition homogeneity tests among loci (ylocus10, ylocus13, and mtDNA) and among genomes (yDNA versus mtDNA) were used to test whether data could be combined.

Y-specific DNA, mtDNA and a combined mtDNA and yDNA dataset were analyzed using maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI). In all analyses, the few gaps were coded as missing data. Unweighted MP and ML searches for yDNA, mtDNA and the combined dataset were performed in PAUP* 4.0b10 using the heuristic search option, with 100 random sequence-addition replicates and tree-bisection-reconnection (TBR) branch swapping. Results from ModelTest 3.04 were used to define the model of nucleotide substitution, the level of among site rate variation, and presence/absence of invariant sites for ML and BI analyses. Original branch lengths for ML analyses were obtained using the Rogers–Swofford approximation method (Rogers and Swofford, 1998), and branch lengths were optimized using the one-dimensional Newton–Raphson method with pass limit and delta settings at defaults (20, 10^{-6}). BI analyses of yDNA, mtDNA, and combined yDNA and mtDNA were conducted with MrBayes 3.0b3 (Huelsenbeck and Ronquist, 2001). Random starting trees were used, and analyses were run for 10 million generations, sampling the Markov chains every 1000 generations. The first 5000 trees were discarded. Multiple BI analyses were completed to ensure that searches were not uncovering local optima. All trees were rooted using sequences from a male sperm whale (*Physeter macrocephalus*). Although three methods for estimating phylogeny were employed, trees resulting from ML and BI analyses were preferred because they allowed us to further parameterize the models of molecular evolution underlying each dataset. Additionally, ML and BI methods are more robust to biases that can lead MP methods to estimate incorrect topologies (reviewed in Swofford et al., 2001).

We assessed relative nodal support for MP and ML consensus topologies using nonparametric bootstrap analysis (Felsenstein, 1985) and decay indices (Bremer, 1994). Under both MP and ML criteria, 100 pseudoreplicates were conducted with ML settings again determined by ModelTest 3.04. For MP and ML bootstrapping, values $\geqslant 70\%$ were taken to indicate strong support for a clade (Hillis and Bull, 1993), while posterior probability values $\geqslant 95\%$ were taken to indicate strong support under BI criteria (Rannal and Yang, 1996). Decay indices (Bremer support) and partitioned decay indices (partitioned Bremer support) were calculated using TreeRot (version 2, Sorenson, 1999). Support was partitioned to reflect the relative support at each node from yDNA versus mtDNA (Baker et al., 1998), and to explore conflicts between the topologies derived from the two genomes.

Hypotheses of species and family-level monophyly within balaenopterids were tested by comparing consensus topologies to alternative topologies generated using MacClade version 3.08a (Maddison and Maddison, 1992). The shortest trees were used as consensus MP topologies, while trees with the most positive log likelihood values were used as ML consensus topologies. ML tree comparisons utilized branch length information. Because preliminary phylogenetic analyses showed differences between ylocus10 and ylocus13 with respect to the relationship among sampled fin and humpback whales, statistical support for alternative hypotheses for fin/humpback arrangements were tested using ylocus10. Alternative baleen whale topologies were compared for ylocus10, yDNA, mtDNA, and combined data, under MP and ML criteria. Finally, consensus topologies from MP and ML analyses were compared to each other. Templeton's (1983) Wilcoxon signed-rank was used to assess the statistical significance of topology differences under MP; one-tailed SH tests (Shimodaira and Hasegawa, 1999) were used for ML (with likelihood settings based on ModelTest results). All tree congruence testing was implemented in PAUP*4.0b10, assuming $\alpha = 0.05$ for statistical significance based on 10,000 pseudoreplicates.

3. Results

3.1. Sequence alignment and properties

We generated 1049 bp of yDNA sequence and 416 bp of mtDNA sequence for analysis. Unique sequences for ylocus10 and ylocus13 were submitted to GenBank and were assigned the following accession numbers: ylocus10 AY822119–AY822136, and ylocus13 AY822152–AY822164. Unique sequences for ylocus2 (removed from analysis) were assigned accession numbers AY822137–AY822151. Unique mtDNA control region sequences were assigned accession numbers AY822087–AY822115.

Alignments of variable sites among sampled baleen whale species and the sperm whale outgroup for ylocus10 and ylocus13 are presented in Table 2. Among the 42 sampled baleen whales (excluding sperm whale outgroup), ylocus10 (403 bp) contained 44 variable sites, 38 of which were parsimony–informative, while ylocus13 (646 bp) contained 30 variable sites, 19 of which were informative. Therefore, the sampled yDNA included 74 sites that were polymorphic among baleen whale species (69 substitutions and 5 indels), 57 of which were parsimony-informative in the combined data set. Alignment of variable sites among sampled baleen whale species and the sperm whale outgroup for mitochondrial control region DNA is available as Supplemental Information (1). 190 sites within the sequenced portion of the mitochondrial control region were variable among the sampled baleen whales (158 substitutions and 32 indels), 152 of which were parsimony–informative.

Percent sequence divergence (corrected for divergence within species) ranged from 0.2% (between sei and small-type Bryde's whales) to 5.9% (between humpback and sperm whales) for yDNA, while mtDNA showed divergence on average 15 times higher, with minimum divergence 5.9% (again, between sei and small-type Bryde's whales) and maximum 32.8% (again, between humpback and sperm whales) (Table 3). In both data sets, pairwise comparisons that included humpback whales showed high sequence divergence relative to other species' comparisons.

All three loci contained significant ($P < 0.01$) phylogenetic signal according to skewness statistics. For ylocus10, ModelTest chose as most likely a model incorporating six rates of substitution with a gamma distribution of rates (GTR+G). The same model, but incorporating a proportion of the sites as invariant (GTR+I+G), was the most likely model of evolution for ylocus13. Estimates of transitions/transversions (ti/tv) (ylocus10 ti/tv = 0.76, ylocus13 ti/tv = 1.16), and rate distribution parameters (ylocus10 $\gamma = 0.14$, ylocus13 $\gamma = 0.15$) were similar for the two loci. When both Y-specific loci were analyzed together, ModelTest chose the GTR+I+G model as most likely, with ti/tv = 0.89, $I = 0.93$, $\gamma = 0.01$. A simpler model (HKY+G) was found to explain patterns of substitution among baleen whales at the mitochondrial control region. The ratio of transitions to transversions at mtDNA was higher than estimated for yDNA, while the distribution of substitution rates among sites was less variable for mtDNA than for yDNA (mtDNA ti/tv = 2.56, $I = 0.54$, $\gamma = 0.4588$).

3.2. Data combinability

Based on initial data exploration, we combined the two Y loci into a single data set. Ylocus10 and ylocus13 showed no significant differences in best-fit models of substitution, and preliminary analyses of each locus individually showed no conflict of signal, thus providing no reason to suspect that these loci differ in phylogenetic signal. Likewise, partition homogeneity tests fail to reject ylocus10 and ylocus13 as reflecting heterogeneous phylogenetic signal (P-value = 0.15), thus supporting the combined analysis of these two loci. These results are expected; the entire Y chromosome is inherited as a single unit, so different Y chromosome loci are not independent estimates of the mysticete's evolutionary history and should retrieve the same phylogenetic relationships.

Partition homogeneity tests also failed to detect heterogeneity among mtDNA and yDNA (P-value = 0.54). However, given our interest in comparing genealogies derived from paternally and maternally inherited loci, these loci were treated separately as well as combined in phylogenetic analysis and hypothesis testing. Given differences in the models of evolution and parameters mediating substitution between yDNA and mtDNA, lack of statistically significant heterogeneity may reflect limited power to discriminate due to low levels of yDNA variation.

Table 2
Alignment of variable sites among sampled baleen whales and sperm whale for the two Y chromosome loci

	Y chromosome locus 10																																		
	21	29	34	44	57	64	67	76	90	94	118	126	127	133	142	151	153	155	156	157	161	162	168	172	177	219	213	215	231	253	258	262	275	276	283	288
Fin_Common	C	C	C	G	T	T	T	A	G	T	G	G	A	0	G	C	A	A	G	T	T	T	G	C	G	A	G	T	C	C	T	A	T	T	C	G
Fin_GulfCalifornia	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Fin_SENPacific	—	—	—	—	—	—	—	—	—	—	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Fin_SENPacific	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Fin_NWNPacific	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	T	—	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Fin_WAlaska	—	A	—	C	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	T	—	—	—	—	—	—	—
Fin_Alaska	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	C	—	—	—	—	—	—	—	—	—	—	T	—	—	—	—	—	—	—
Humpback_NPacific	T	—	A	—	C	—	G	C	A	C	—	A	G	A	T	—	C	—	—	G	G	A	—	—	—	G	A	C	—	—	G	C	C	—	G	T
Humpback_NAtlantic	T	—	A	—	C	—	G	C	A	C	—	A	G	A	T	—	C	—	—	G	G	A	—	—	—	G	A	C	—	—	G	C	C	—	G	T
Grey_NPacific	—	—	—	C	—	C	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	T	T	—	—	—	T	—	—	—	—	—	—	—
Minke NPacific1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	A	—	—	—	—	T	T	—	—	—	T	—	—	—	—	—	—	—
Minke_NPacific2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	A	—	—	—	—	T	T	—	—	—	T	—	—	—	C	—	—	T
Blue_Antarctic	—	A	—	C	—	—	—	—	—	—	—	—	—	—	—	—	—	—	A	—	—	—	—	T	T	—	—	—	T	—	—	—	—	—	—	—
Sei_Northern	—	A	—	C	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	T	T	—	—	—	T	—	—	—	—	—	—	—
Brydes_NPacific1	—	A	—	C	—	C	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	T	T	—	—	—	T	—	—	—	—	—	—	—
Brydes_NPacific2	—	A	—	C	—	C	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	T	T	—	—	—	T	T	—	—	—	—	—	—
Bowhead_Alaska	—	A	—	C	—	C	—	—	—	—	—	—	—	—	—	—	—	—	A	—	—	—	—	T	T	—	—	—	T	T	—	—	—	—	—	—
Sperm_NPacitic	—	A	—	C	—	C	—	—	—	—	—	—	—	—	—	—	—	—	A	—	—	—	C	T	T	—	—	C	T	T	—	—	—	C	—	—

	Y chromosome locus 10								Y chromosome locus 13																												
	298	321	324	357	359	366	367	371	429	459	458	460	522	544	574	610	611	616	626	647	656	662	675	687	691	731	741	747	777	788	795	800	818	827	830	853
Fin_Common	T	T	G—	A	A	T	T	T	C	G	A	A	G	A	A	C	C	A—	A	G	C	T	G	G	C	C	C	A	T	T	T	A	0	T	T	C
Fin_GulfCalifornia	—	—	—	—	—	—	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—
Fin_SENPacific	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Fin_SENPacific	—	—	—	—	—	—	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Fin_NWNPacific	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Fin_WAlaska	—	—	—	—	—	—	—	—	—	—	—	—	A	—	—	—	G	—	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Fin_Alaska	—	—	—	—	—	—	—	—	—	C	—	—	A	—	—	—	G	G	—	A	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Humpback_NPacific	C	—	T	C	—	G	—	—	—	C	—	—	A	—	—	T	—	G	0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Humpback_NAtlantic	C	—	T	C	—	G	—	—	—	C	—	—	A	—	—	T	—	G	0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	A	—
Grey_NPacific	—	—	—	—	—	—	—	—	—	C	C	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	G	A	—
Minke_NPacific1	—	—	—	—	—	—	—	—	—	C	C	T	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	T	—	—	A	—
Minke_NPacific2	—	A	T	C	T	G	C	—	—	C	C	T	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	T	—	—	A	—
Blue_Antarctic	—	—	—	—	—	—	—	—	—	C	C	—	A	—	—	—	—	—	—	—	—	—	—	—	—	T	—	—	G	—	—	—	—	—	A	—
Sei_Northern	—	—	—	—	—	—	—	—	—	C	C	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	C	—	—	—	A	—
Brydes_NPacific1	—	—	—	—	—	—	—	—	—	C	C	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	C	—	—	—	A	—
Brydes_NPacific2	—	—	—	—	—	—	—	—	—	C	C	—	A	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	C	—	—	—	A	—
Bowhead_Alaska	—	—	—	—	—	—	—	—	—	C	C	—	A	G	T	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	A	T
Sperm_NPacific	—	—	—	—	—	—	—	—	—	C	C	—	A	—	—	—	—	—	—	—	A	—	—	A	A	T	T	G	—	C	—	—	A	—	A	—

Numbers along the top correspond to the position of the variable site in the combined ylocus10 and ylocus13 data set. The vertical line denotes separation between ylocus10 and ylocus13. The horizontal line denotes separation between fin whale haplotypes and haplotypes from additional whale species. (—) denotes the presence of the same nucleotide as seen in the Fin_Common haplotype. (0) denotes the presence of a single nucleotide gap.

Table 3
Pairwise sequence divergences among all species included in this study for yDNA (A) and mtDNA (B)

	N	Fin	Blue	Humpback	Grey	Minke	Brydes	Sei	Bowhead	Sperm
(A)										
Fin	28	**0.0017**	0.0123	0.0391	0.0125	0.0144	0.0129	0.0106	0.0161	0.0284
Blue	2	0.0114	**0.0000**	0.0433	0.0076	0.0105	0.0062	0.0038	0.0095	0.0200
Humpback	2	0.0377	0.0428	**0.0010**	0.0433	0.0413	0.0437	0.0413	0.0490	0.0594
Grey	2	0.0117	0.0076	0.0428	**0.0000**	0.0124	0.0062	0.0057	0.0114	0.0238
Minke	2	0.0087	0.0057	0.0361	0.0076	**0.0095**	0.0128	0.0105	0.0152	0.0276
Brydes	2	0.0116	0.0057	0.0428	0.0057	0.0076	**0.0010**	0.0024	0.0090	0.0214
Sei	2	0.0097	0.0038	0.0409	0.0057	0.0057	0.0019	**0.0000**	0.0095	0.0219
Bowhead	2	0.0152	0.0095	0.0485	0.0114	0.0105	0.0086	0.0095	**0.0000**	0.0181
Sperm	1	0.0276	0.0200	0.0589	0.0238	0.0228	0.0209	0.0219	0.0181	**0.0000**
(B)										
Fin	28	**0.0132**	0.1180	0.1380	0.1572	0.1394	0.1548	0.1661	0.2065	0.2882
Blue	2	0.1041	**0.0144**	0.1695	0.1226	0.1430	0.1683	0.1707	0.2067	0.2632
Humpback	2	0.1159	0.1466	**0.0313**	0.1899	0.2043	0.2163	0.2219	0.2404	0.3438
Grey	2	0.1483	0.1130	0.1719	**0.0048**	0.1442	0.1791	0.1815	0.2151	0.3101
Minke	2	0.1291	0.1322	0.1851	0.1382	**0.0072**	0.1947	0.1827	0.2175	0.3065
Brydes	2	0.1351	0.1478	0.1875	0.1635	0.1779	**0.0264**	0.0841	0.2500	0.3137
Sei	2	0.1476	0.1514	0.1942	0.1671	0.1671	0.0589	**0.0240**	0.2608	0.2957
Bowhead	2	0.1988	0.1983	0.2236	0.2115	0.2127	0.2356	0.2476	**0.0024**	0.2608
Sperm	1	0.2815	0.2560	0.3281	0.3077	0.3029	0.3005	0.2837	0.2596	**0.0000**

Above diagonal: Average percent pairwise differences between species. Diagonal elements: Average percent pairwise differences within species. Below diagonal: Average percent pairwise differences between species corrected for within species differences. N = number of males sampled per species.

3.3. Phylogenetic analyses

Topologies representing the best estimates of phylogeny for mtDNA (A) and yDNA (B) are compared under MP (Fig. 1) and ML (Fig. 2) criteria. MP and ML phylogenies for the combined dataset are presented in Fig. 3(A and B). For all three datasets (mtDNA, yDNA, and combined), ML and BI analyses resulted in identical topologies. Thus, BI posterior probabilities are presented on nodes in ML trees. All trees recovered the traditionally accepted basal position of the *Balaenidae* (bowhead whale) and showed strong support for the monophyly of blue, Bryde's, humpback, gray, and bowhead species. Most trees also showed strong support for the monophyly of minke and sei whales. In general, nodal support for yDNA topologies was consistently weaker than for mtDNA topologies. However, some notable patterns recovered by yDNA showed moderate posterior probability and bootstrap support.

ML and BI analyses of Y chromosome DNA showed low-level support for paraphyly in the family Balaenopteridae, with blue whales basal to gray whales and minke whales sister to a humpback-fin clade (Fig. 2B). ML and BI showed moderate support for fin whales as paraphyletic with respect to humpback whales, and for sei and Bryde's whales as sister taxa (Fig. 2B). None of these relationships, however, were strongly supported in the MP analysis (Fig. 1B). The MP analysis retained 34 trees, and all phylogenetic analyses detected single topological islands. Patterns of divergence at yDNA were driven mainly by ylocus10. Fin whales sampled in the Gulf of Alaska were highly divergent from all other North Pacific and North Atlantic-sampled fin whales at ylocus10. Additionally, humpback whales (including samples from both North Atlantic and North Pacific populations) were highly divergent from all other sampled baleen whales at this locus, including fin whales, and were on a "long branch" in the ylocus10 topology.

All mtDNA topologies also showed low to moderate bootstrap and posterior probability support for Balaenopteridae as paraphyletic; however, this relationship was created by minke whales (rather than blue whales) taking a basal position relative to gray whales (Figs. 1A and 2A). All topologies showed strong support for fin whales as monophyletic and sister to humpback whales, and for sei and Bryde's whales as sister taxa (Figs. 1A and 2A). The MP search for mtDNA retained 316 trees. Both MP and ML analysis detected two tree islands. Although under both methodologies the secondary tree was found much less often (~1:9), it was not significantly different from the primary tree (Templeton's test, P-value = 0.221; one-tailed SH test, P-value = 0.403). The primary and secondary trees varied in their resolution of minke, blue and gray whale positions. The strict consensus MP tree presented in Fig. 1A represents this lack of consistency between the two tree islands as a trichotomy for blue, gray and sei/Bryde's whales. The ML tree (Fig. 2A) placed blue whales sister to sei/Bryde's whales and placed gray whales more basally. Low bootstrap support at some nodes reflects the lack of consistency between the two equally-likely trees.

In combined analyses of both data sets (Fig. 3), mtDNA contributed almost three times as many informative characters as yDNA. Therefore, in most cases the resulting topologies reflected stronger support for relationships derived from mtDNA. MP analysis retained 5920 trees and

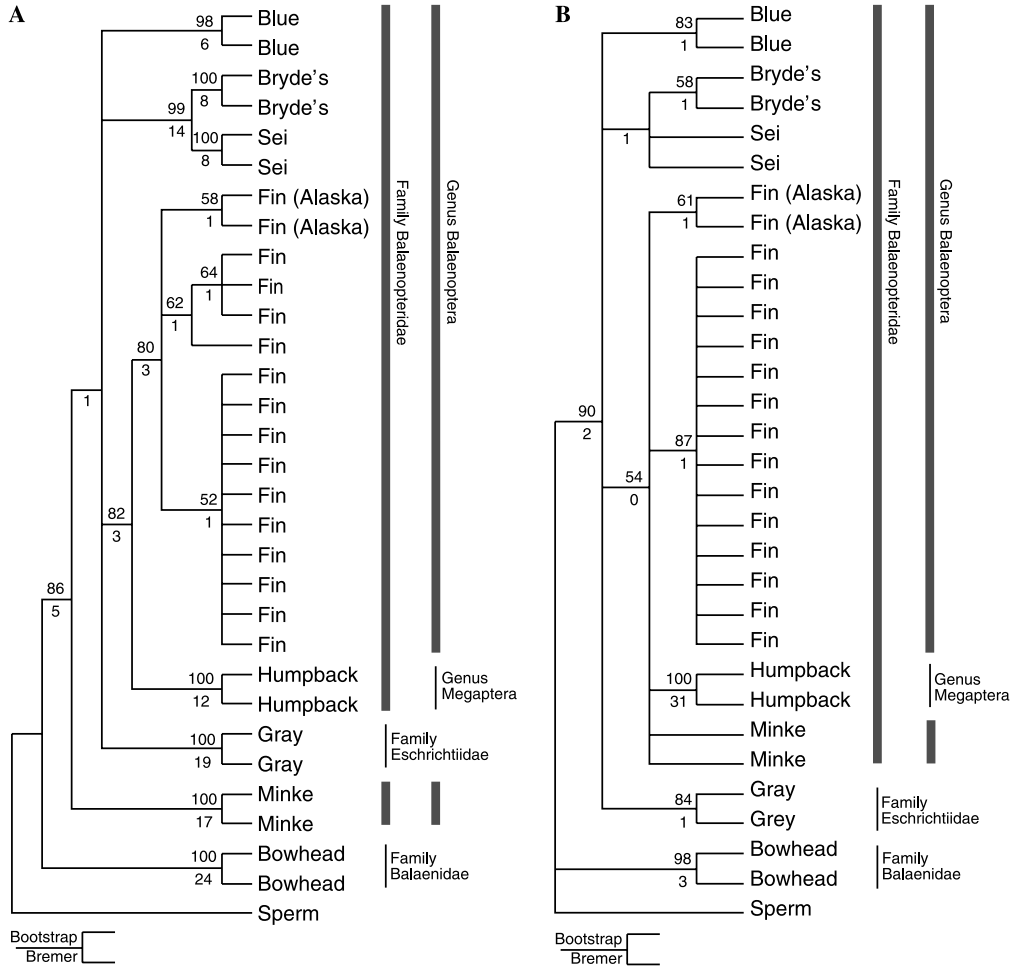

Fig. 1. Strict consensus maximum parsimony cladograms for sampled baleen whale species based on mtDNA (A) and yDNA (B). All genealogies are rooted with sequences from a single male sperm whale. Bootstrap support $\geqslant 50\%$ and Bremer support values are presented above and below nodes, as indicated in the legends.

detected a single topological island, but failed to provide any resolution among blue, gray, sei/Bryde's and minke whales (Fig. 3A). ML analysis found two trees, with the secondary tree less likely, but not significantly so (one-tailed SH test, P-value = 0.475). Primary and secondary ML trees for the combined dataset differed in the placement of minke and blue whales. The primary ML tree (Fig. 3B) retained minkes in their mitochondrially-derived position, basal among the rorquals, and placed gray, blue and sei/Bryde's whales in sister clade to fin/humpback whales. BI analysis resulted in an identical topology. The secondary ML tree for the combined dataset was more reflective of Y chromosome-derived relationships, placing minke whales as sister to the fin/humpback whale clade and blue whales as the most basal rorqual species. Loss of resolution through lack of congruence between yDNA and mtDNA in the combined analysis is further corroborated by partitioned Bremer support values. Negative decay indices indicate a lack of support in yDNA for nodes largely derived from mtDNA. For example, in the MP topology (Fig. 3A) fin whale monophyly had negative Bremer support for yDNA, and positive Bremer support from mtDNA. In contrast, Bremer values indicated strong support for sister relationships between both humpback and fin whales and sei and Bryde's whales.

3.4. Tree congruence testing

Templeton's signed-rank tests and SH tests yielded similar results in comparisons among MP and ML consensus topologies for ylocus10, combined yDNA, and mtDNA and in comparisons of the topologies with hypotheses of monophyly of Balaenopteridae and/or fin whales (Tables 4 A, B). When ylocus10 was analyzed alone, Templeton's test found that a tree in which all sampled fin whales were monophyletic was significantly longer than the consensus topology, in which humpback whales are nested within fin whales. The corresponding SH test found fin whale monophyly to be marginally less likely. When ylocus13 was included in a combined yDNA data set, both tests found differences between fin whale paraphyly and monophyly to be not significant. All mitochondrial topologies in which fin whales were paraphyletic were rejected as significantly longer than the consensus topology in Templeton's tests,

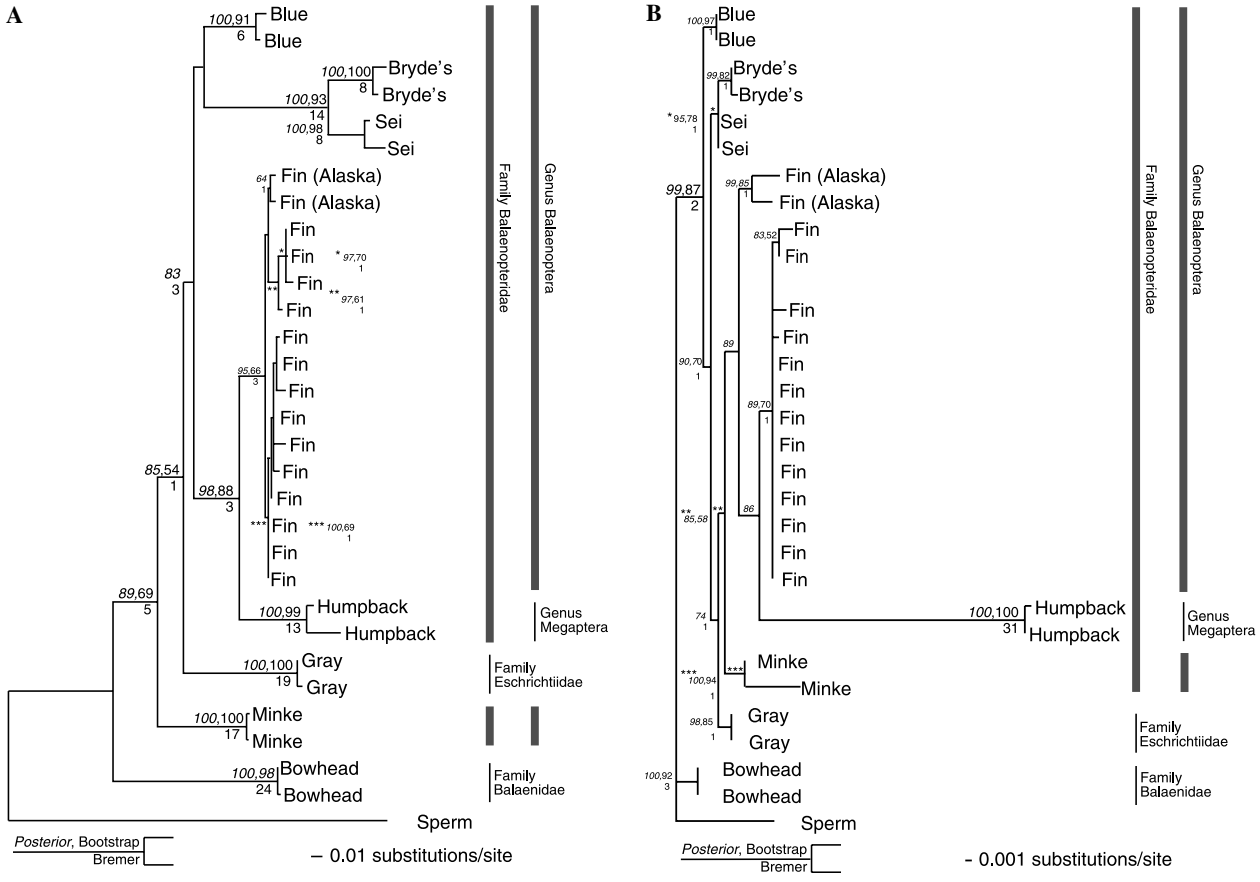

Fig. 2. Maximum likelihood phylograms for sampled baleen whale species based on mtDNA (A) and yDNA (B). All genealogies are rooted with sequences from a single male sperm whale. Bayesian posterior probabilities $\geqslant 50\%$, bootstrap support $\geqslant 50\%$, and Bremer support values are presented above and below nodes, as indicated in the legends.

while two of three were significantly less likely in SH tests. Neither Templeton's nor SH tests revealed significant differences between trees in which rorquals were paraphyletic versus monophyletic, all other conditions remaining the same. This result is further indication of the weak support for this relationship in the mtDNA, yDNA and combined datasets.

MP and ML consensus topologies for mtDNA were found to differ significantly according to Templeton's test (P-value = 0.008), but not according to the SH test (P-value = 0.07). For yDNA, consensus topologies inferred under the two different phylogenetic analysis methods were not significantly different according to either test. Despite differences in topologies resulting from different analysis methods, both yDNA consensus topologies were found to be significantly longer or less likely when compared to mitochondrial consensus topologies, and visa versa. Thus, incongruence in topologies based on the maternally and paternally inherited DNA was significant, supporting separate analysis of Y-specific and mitochondrial data sets. Loss of resolution though combined analysis of yDNA and mtDNA is further supported in tree congruence tests, as none of the tested alternatives were found to be significantly different from the MP and ML consensus trees.

4. Discussion

4.1. Y chromosomes and baleen whale divergence

Among the 45 males sampled here, representing nine baleen whale species, sequence divergence based on Y chromosome sequences was only about 7% of the divergence based on the mitochondrial control region. Estimates of diversity within species and divergence among species depend on rates of mutation and the influences of selection and recombination, as well as changes in effective population sizes (N_e) (including variation due to unequal sex ratios, non-random mating and/or sex-biased dispersal) and population structure. Mutation rates for mammalian mitochondrial DNA are several times higher than mutation rates for autosomal nuclear loci (Ballard and Whitlock, 2004; and reviewed in Avise, 2004). Thus mitochondrial divergence is expected to exceed divergence at Y-specific loci, although the magnitude of the difference varies among taxa and among loci (Ballard and Whitlock, 2004). Different selective histories for Y chromosome versus mitochondrial DNA and differences in N_e (e.g. smaller male N_e due to differences in male and female mating behavior) also influence the observed ratio of yDNA and mtDNA divergence.

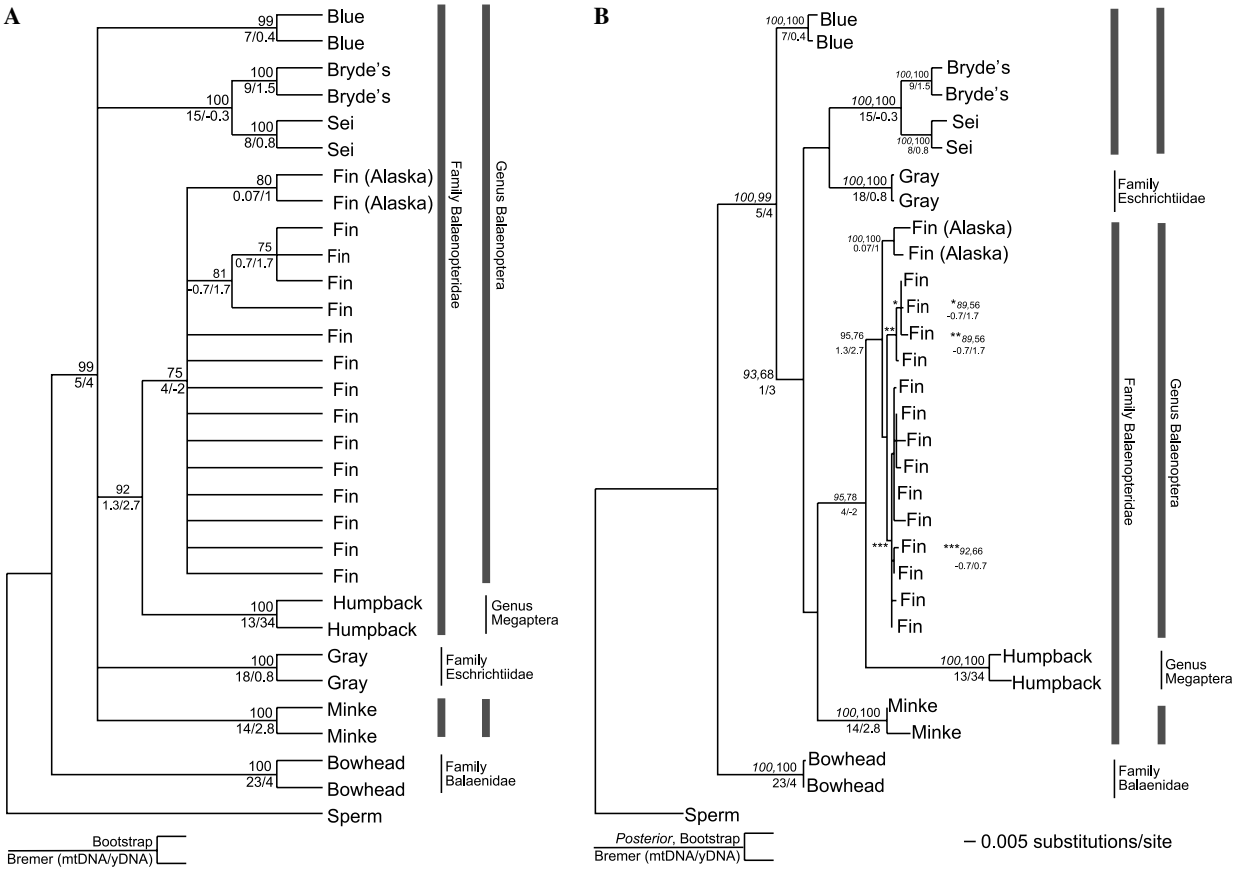

Fig. 3. Strict consensus maximum parsimony cladogram (A) and maximum likelihood phylogram (B) for sampled baleen whale species based on a combined mtDNA and yDNA dataset. All genealogies are rooted with sequences from a single male sperm whale. Bayesian posterior probabilities ⩾75%, bootstrap support ⩾50%, and Bremer support values are presented above and below nodes, as appropriate and as indicated in the legends.

For example, the high ratio of yDNA divergence to mtDNA divergence for humpback whales (Table 3A and B) may reflect this species' highly gregarious mating behavior, in which males are believed to engage in both acoustic and physical competition for female mates (Clapham, 1996). If such a mating system results in a hierarchical, dominance-based structure in which reproductive success among males is highly variable, the effective population size among male humpbacks will be lower than under random mating conditions. Smaller effective population size and greater population substructure for males than females could result in increased rates of evolution for Y chromosome loci. As diversity at Y-specific sequences within humpback whales has not been characterized, the relative importance of demography/behavior versus selection or mutation in explaining the divergence of humpback yDNA requires further investigation.

In assessing divergence among cetacean species using 750 base pairs surrounding the Y-specific sex-determining region (SRY), Nishida et al. (2003) found no polymorphism within species, but moderately high divergence among species, and therefore argued for SRY's utility in resolving species'-level relationships. Recently, a survey of nucleotide diversity within multiple mammalian species targeting Y chromosome-specific gene introns suggested that selective-sweeps on sex-limited chromosomes have resulted in generally low levels of Y chromosome variation within mammals (Ellegren, 2003; Hellborg and Ellegren, 2004). Hellborg and Ellegren (2004) did not present estimates of divergence among the six sampled species, but intraspecific diversity ranged from zero to four substitutions in 0.7–3.5 kb of yDNA. In fin whales (L. Hatch and R. Harrison, in preparation), levels of Y diversity were higher than corresponding values reported for other mammals by Hellborg and Ellegren (2004). Furthermore, divergence among cetacean species at the anonymous Y loci ranged from 0.2 to 5.9%, values similar to estimates of yDNA divergence among primate species.

The average ratio of yDNA divergence to mtDNA divergence seen here among baleen whales (1:14) falls within the range of values reported for comparisons among primate species. Comparison between chimpanzee and human Y loci have estimated sequence divergence between these species to range between 1.4 and 2.3% (Rozen et al., 2003), comparable to divergence estimates based on autosomal nuclear DNA (Chen and Li, 2001; Britten, 2002). MtDNA divergence between chimpanzee and humans was estimated in early studies to be 9.6% (Gibbons, 1990); therefore, divergence between humans and chimpanzee based on yDNA is 14–24% of mtDNA divergence. The nucleotide diversity of the

Table 4
(A) Results of maximum likelihood (A) and maximum parsimony (B) tree congruence testing

Partition	Topology		SH test	
	Are fin whales paraphyletic with respect to humpback whales?	Are rorquals paraphyletic with respect to grey whales?	Ln	*P*-value
(A)				
yDNA locus 10	Yes*	Yes*	908.55995	
	No	Yes*	928.98014	0.0646
yDNA	Yes*	Yes*	2140.5243	
	No	Yes*	2146.4318	0.4848
	No	No	2152.6747	0.2339
	Yes*	No	2146.6996	0.4573
	mtDNA consensus topology		2183.1209	**0.028**
mtDNA	No*	Yes*	2313.1629	
	No*	No	2319.6032	0.1337
	Yes	Yes*	2325.3804	0.0934
	Yes	No	2331.8639	**0.0263**
	yDNA consensus topology		2474.9148	**0.0013**
combined data	No*	Yes*	4919.2214	
	No*	No	4929.1735	0.2695
	Yes	No	4927.062	0.3421
	Yes	Yes*	4937.5305	0.0692

Partition	Topology		Templeton			
	Are fin whales paraphyletic with respect to humpback whales?	Are rorquals paraphyletic with respect to grey whales?	Length	*N*	*z*	*P*-value
(B)						
yDNA locus 10	Yes*	Yes*	55			
	No	Yes*	64	6	−2.2514	**0.0244**
yDNA	Yes*	Yes*	101			
	No	Yes*	104	1	−1	0.3173
	No	No	103	5	−1.3416	0.1797
	Yes*	No	110	4	−1	0.3173
	mtDNA consensus topology		108	8	−2.1106	**0.0348**
mtDNA	No*	Yes*	362			
	No*	No	371	10	−1.2649	0.2059
	Yes	Yes*	367	14	−2.1828	**0.029**
	Yes	No	399	5	−2.2361	**0.0253**
	yDNA consensus topology		411	24	−2.9133	**0.0034**
combined data	No*	Yes*	488			
	No*	No	492	16	−1	0.3173
	Yes	No	496	25	−1.5119	0.1306
	Yes	Yes*	493	11	−1.5076	0.1317

(A) Results of maximum likelihood tree congruence testing. SH test (Shimodaira and Hasegawa 1999) one-tailed scores are reported, with bold indicating *P*-values < 0.05. (B) Results of maximum parsimony tree congruence testing. Templeton's test (1993) (Wilcoxon signed-ranks) test scores are reported. *, topology consistent with consensus tree.

human Y chromosome has been estimated to be ~20% of that found in human autosomes (Shen et al., 2000; The International SNP Map Working Group, 2001). Stone et al. (2002) found chimpanzees and bonobos to be considerably more diverse when compared to humans based on male-specific loci, as has long been known to be true for mtDNA (Ferris et al., 1981). However, Stone et al. (2002) estimated yDNA sequence divergence between chimpanzees and bonobos to be only ~0.25%; a small fraction of the divergence estimated from mtDNA comparison (~13%). Thus, for these close relatives the disparity between divergence estimates based on the two genomes is even greater than we found in this study of baleen whales. Tosi et al. (2003) found the proportion of variable sites at Y chromosome loci among species in the genus *Macaca* to be only 10% of the proportion for mtDNA.

4.2. Y chromosomes and baleen whale phylogeny

Fig. 4 presents a schematic of eight hypotheses for relationships among baleen whale species, including mtDNA and yDNA phylogenies from this study, as well as three mtDNA phylogenies and three nuclear/morphological phylogenies taken from other published studies. We chose to include cladograms from our ML analyses in this comparison figure, as these topologies were also found in BI analysis. The trees in Fig. 4 were reproduced without support values or branch lengths that may have appeared in their original publications; thus, this figure cannot be used to infer which topology represents the best estimate for the mysticete clade nor which relationships within the clade have been the most commonly resolved among studies. Fig. 4 is presented purely to aid the reader in keeping track

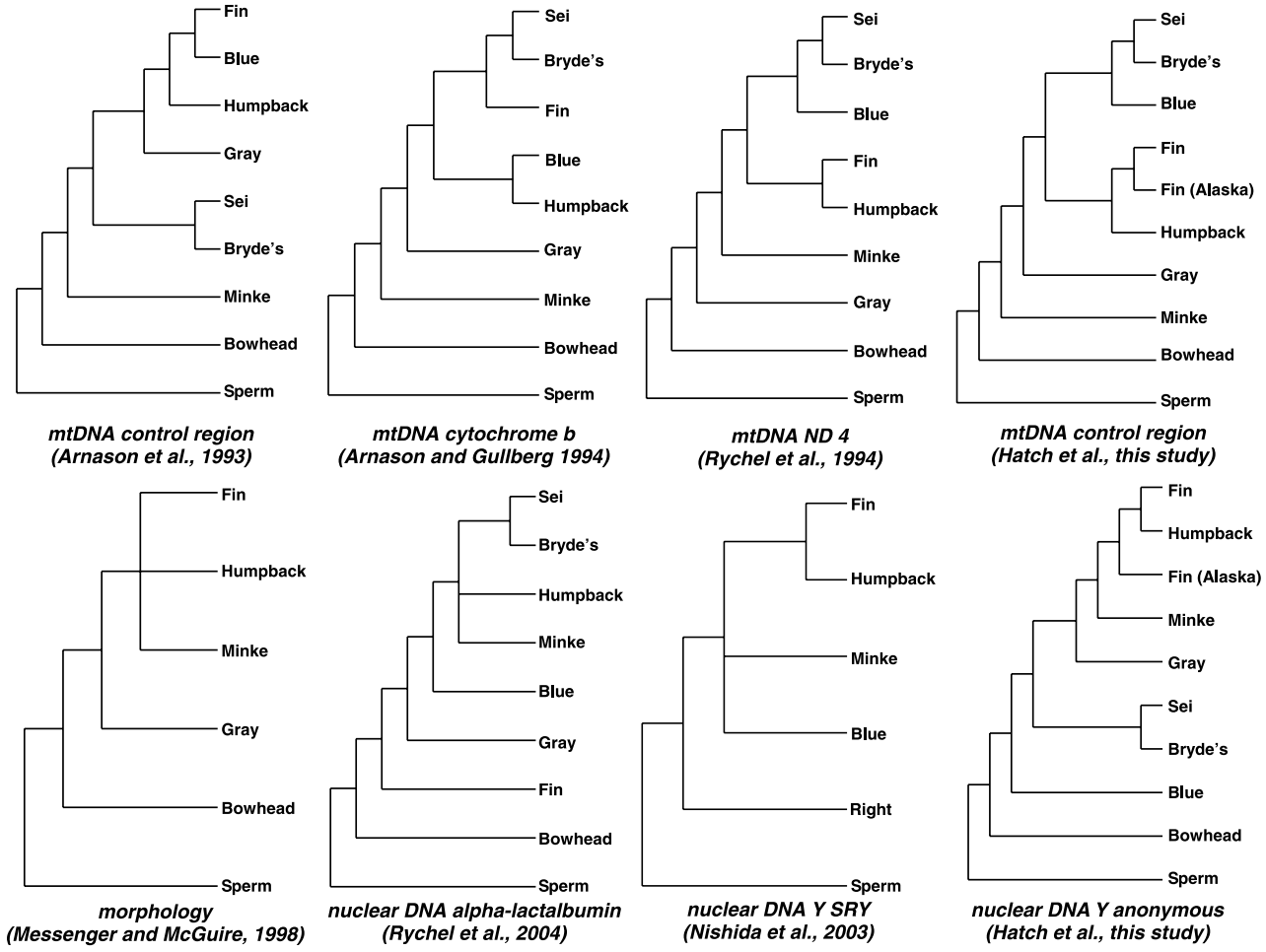

Fig. 4. Comparison of hypothetical relationships among baleen whale species reproduced from previous studies and this study (references cited in figure), including four mitochondrial gene genealogies (top) and four nuclear/morphological gene genealogies (bottom).

of the results of multiple estimates of phylogeny for this clade as we discuss differences between studies.

Anonymous Y chromosome loci showed low-level support under ML criteria for gray whales nested within a paraphyletic family Balaenopteridae, and moderate support for humpback whales nested within a paraphyletic fin whale clade. The first result is in agreement with the placement of gray whales in the majority of published baleen whale genealogies, most of which are based on mitochondrial DNA, but contradicts evidence based on morphological characters (Fig. 4). MP analysis of yDNA and tree-comparison tests between MP and ML topologies with monophyletic and paraphyletic rorquals indicated that yDNA did not significantly support Balaenopteridae paraphyly. The second result (fin whale paraphyly) is due to divergence at a single Y-specific locus of fin whales sampled in the Gulf of Alaska. Although fin whale paraphyly is unique to topologies based on this Y locus, a close relationship between fin and humpback whales was also found in the yDNA analysis of Nishida et al. (2003) (Fig. 4), two of the four mtDNA topologies (Fig. 4), and Gatesy et al.'s (2002) supertree analysis of multiple molecular and morphological characters (not shown). Y chromosome data revealed some support for minke whales as sister to a humpback-fin clade, a relationship also revealed in Gatesy et al.'s (2002) consensus analysis but not seen in any other published genealogy (Fig. 4). Sei and Bryde's whales were closely related in all of the phylogenetic analyses in which these taxa were included (Fig. 4). Finally, the nuclear genealogy of Rychel et al. (2004) placed six fin whale samples basal to all other samples from both the family Balaenopteridae (the rorquals) and the family Eschrichtiidae (the gray whale). In contrast, our study, which included 28 fin whales from two ocean basins, did not find fin whales to be basal among rorqual species.

Our mitochondrial topologies, like previously published studies of control region, ND4, and cytochrome *b* sequences, placed minke whales in a basal position among rorquals, and nested the gray whale within the rorqual family (Fig. 4). However, multiple trees were detected for both MP and ML analyses of mtDNA, and inconsistencies among results are reflected in the lack of resolution and/or support for blue, gray and minke whale positions in both MP and ML topologies. In addition, tree-compar-

ison tests between MP and ML topologies with monophyletic and paraphyletic rorquals indicated that mtDNA support for Balaenopteridae paraphyly was not significant.

Mysticete genealogies based on different regions of the mitochondrial genome (assumed to represent only a single estimate of the evolutionary relationships among species) have been found to vary substantially in their topologies (Fig. 4). For example, we found strong mtDNA control region support for a sister relationship between humpback and fin whales, while an earlier phylogeny based on mtDNA control region (Árnason et al., 1993) suggested that blue whales and fin whales are sister taxa (although bootstrap support for this relationship was <50%). To determine whether the relationships we see are somehow unique to our dataset, we added 11 whale sequences (two fin whales, two blue whales and single humpback, minke, gray, sei, Bryde's, bowhead and sperm whales) represented in GenBank by whole mitochondrial genome sequences, to our mtDNA control region dataset. Phylogenetic analyses restricted to this 416 bp segment of the control region, but including these additional samples, continued to support a sister relationship between fin and humpback whales (bootstrap value 77%; data not shown). Lack of congruence among different estimates of phylogeny based on the same locus has been found previously for this group (e.g., α-lactalbumin sequences from Bérubé and Aguilar, 1998 and Rychel et al., 2004).

We found baleen whale gene phylogenies based on maternally and paternally inherited DNA to be significantly different from each other (Table 4), although, in our analyses, the two markers agreed as to a basal position for the Balaenidae (represented by the bowhead whale, *Balaena mysticetus*) and close relationships (shared genetic variation) between both fin and humpback whales and sei and Bryde's whales. ML genealogies based on yDNA showed some support for minke whales as more deeply nested among the rorquals than was seen in mitochondrial genealogies from this study (control region) and past studies (control region and cytochrome *b* data: Árnason et al., 1993; Árnason and Gullberg, 1994; see Fig. 4). A more deeply nested position for minke whales was also apparent in the nuclear gene genealogy of Rychel et al. (2004) compared to those based on mtDNA (Fig. 4). Incongruence between nuclear and Y-based phylogenies suggests locus-specific effects (e.g., different evolutionary rates) or sex-biases in introgression among species. However if nodes with low support in the MP consensus trees for mtDNA and yDNA are collapsed, minke and gray whales become members of the same polytomy; thus, individual hypotheses for relationships among these species are not considered well supported.

Although signal heterogeneity between yDNA and mtDNA partitions was not statistically significant, patterns and rates of substitution led to different choices of models of evolution for Y-specific and mitochondrial sequences in hierarchical likelihood testing (Posada and Crandall, 1998). Consensus topologies produced by the two data sets were significantly different from each other. Bull et al. (1993) suggest that combining partitions with different rates of change may lead to lowered chances of recovering the correct phylogeny. In this study, topologies based on yDNA were shallower but showed strong support at nodes that were in conflict with longer-branched mtDNA-based topologies. However, as mtDNA contained more phylogenetically informative sites, signal derived from yDNA was overwhelmed in combined analysis, and significant information from the Y loci was lost. This result suggests that generating an accurate estimate of Mysticeti relationships may necessitate sampling multiple independent (autosomal nuclear) loci, both to generate enough informative sites, and to accurately characterize variation among topologies that may reflect the lack of congruence seen here in male versus female derived characters.

4.3. Y chromosomes and baleen whale introgression

Hybridization between humpback whales and blue whales and between fin whales and blue whales has been documented (Árnason et al., 1991; Spilliaert et al., 1991; Bérubé and Aguilar, 1998; M. Poole personal communication), but thus far there is no direct evidence for hybridization between humpback whales and fin whales. Evidence regarding the fertility of documented hybrids remains inconclusive. Although examination of male hybrid blue-fin whales suggested reproductive impairment, one female hybrid was carrying a backcrossed fetus, the viability of which was unknown (Bérubé and Aguilar, 1998).

In this study, fin whales were paraphyletic with respect to humpback whales based on a single segment of Y-specific DNA. At the same Y locus, humpback whales were highly divergent from all sampled baleen whales, and thus were on a "long branch" in yDNA topologies. Fin whale paraphyly was not supported by either mtDNA or additional Y-specific sequence. Allele-specific paraphyly is common in gene trees, and may reflect interspecific hybridization and/or incomplete lineage sorting (Harrison, 1998; Funk and Omland, 2003). In this study, paraphyly for this Y-specific segment suggests incomplete lineage sorting among fin whales due to regional diversity (i.e. Gulf of Alaska) captured by our more comprehensive sampling. Alternative explanations for paraphyly at ylocus10 appear unlikely. Accelerated rates of change for humpback whales could not have distorted the topology via long-branch attraction, as only humpback whales occupy a long branch in our topologies. Likewise, there is no evidence for recent introgression between fin and humpback whales, given the substantial divergence of Y haplotypes between the two species.

Therefore, our results support a more recent divergence between humpback and fin whales than between

humpbacks and blue and/or blue and fin whales, as evidenced by incomplete sorting of Y chromosome lineages between these two species. This hypothesis is not supported by phenetic evidence, however, as fin whales and humpbacks show less similarity in coloration and body proportions than fin whales and the remaining rorqual species. Instead, humpback whales show similarly "accelerated" rates of phenotypic change when compared to other baleen whale species, both in morphology (i.e. flipper size) and acoustic behavior (i.e. highly derived male song).

4.4. Introgression and defining baleen whale species

Understanding relationships between protected baleen whale species, as well as among protected populations, is critical to efforts to manage their conservation effectively (Perrin and Reeves, 2004). Recently, new specimens, reexamination of molecular data, and collection of data from new molecular markers (Rosenbaum et al., 2000; Wada et al., 2003) have prompted cetacean systematists to discuss increasing the number of recognized baleen whale species from ten to fifteen (Perrin and Reeves, 2004). At least one researcher has questioned whether these new species represent previously misidentified or unknown taxonomic diversity or recent hybridization (Rychel et al., 2004). Willis et al. (2004) suggest that although difficulty in detecting wild cetacean hybrids in the marine environment has led researchers to assume hybridization events are rare, karyological uniformity among cetaceans may allow whale species to hybridize more readily than other mammals. As increasing instances of hybridization among cetacean species are documented (see review in Bérubé and Aguilar, 1998; Kingston and Rosel, 2004; Willis et al., 2004; D. Duffield unpublished data), determining whether rates of detection or rates of occurrence are driving this trend is generating interest, not only among cetacean researchers, but among the managers responsible for the recovery of these endangered species.

Evidence from molecular markers, such as sharing of haplotypes and/or species paraphyly, is often used to support hypotheses of contemporary hybridization events (reviewed in Funk and Omland, 2003). However, to assess whether introgression played a role in generating phylogenic relationships reflected by a single locus, researchers must first look for corroborating evidence from multiple independent loci. Inconsistencies among multiple gene genealogies are then assessed in light of differences in the evolutionary processes mediating rates of change at sampled loci.

Although managers have cautioned against relying solely on mtDNA in estimating baleen whale molecular phylogeny, until recently, few nuclear markers have shown both low levels of homoplasy and yet sufficient variation to resolve relationships within this clade. Observed differences in estimates of diversity and divergence within and among baleen whale species, both across sampled species and across sampled loci, underscores the importance of using multiple independent markers in characterizing species' level relationships within this clade. In addition, male and female cetaceans, as is true for other mammalian species, are expected to be under different selective pressures for mate choice (Greenwood, 1980). Comparing genealogical patterns among maternally and paternally inherited loci may help researchers to better understand whether differences among sexes influence the origin and maintenance of species boundaries.

Acknowledgments

This research was supported by a National Science Foundation Graduate Research Fellowship, a National Science Foundation Doctoral Dissertation Improvement Grant (Award Number IBN-0308617), and an Andrew W. Mellon Student Research Grant awarded to Leila Hatch. Genetic samples were generously provided by the US National Marine Fisheries Service (US NMFS)'s Southwest Fisheries Science Center (SWFSC) Marine Mammal Tissue Archive, with permission from the Smithsonian Institution (SI) and the International Whaling Commission (IWC). Additional genetic samples were provided by the US NMFS's Northeast Fisheries Science Center (NEFSC), Salvatore Cerchio, Martine Bérubé, and Daniel DenDanto. We thank Andrew Dizon, Susan Chivers, and Kelly Robertson at SWFSC, James Mead and Charley Potter at SI, Greg Donovan at the IWC, and Phillip Clapham at NEFSC for their assistance in provided samples.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2006.05.023.

References

Avise, J., 2004. Molecular Markers, Natural History and Evolution. Sinauer Associates, Sunderland, MA.

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

Árnason, Ú., Gullberg, A., 1994. Relationships of baleen whales established by cytochrome *b* gene sequence comparison. Nature 367, 726–728.

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, 960–970.

Árnason, Ú., Spilliaert, R., Palsdottir, A., Árnason, A., 1991. 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.

Baker, R.H., Yu, X., DeSalle, R., 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. Mol. Phylogenet. Evol. 9, 427–436.

Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13, 729–744.

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. Mam. Sci. 14, 82–98.

Bérubé, M., Palsbøll, P.J., 1996. Identification of sex in Cetaceans by multiplexing with three ZFX and ZFY specific primers. Mol. Ecol. 5, 283–287.

Bremer, K., 1994. Branch support and tree stability. Cladistics 10, 295–304.

Britten, R.J., 2002. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. Proc. Natl. Acad. Sci. USA 99, 13633–13635.

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, 384–397.

Chen, F., Li, W., 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am. J. Hum. Genet. 68, 444–456.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T., Higgins, D., Thompson, J., 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31, 3497–3500.

Clapham, P.J., 1996. The social and reproductive biology of humpback whales: an ecological perspective. Mammal Rev. 26, 27–49.

Ellegren, H., 2003. Levels of polymorphism on the sex-limited chromosome: a clue to Y from W? BioEssays 25, 163–167.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Ferris, S.D., Brown, W.M., Davidson, W.S., Wilson, A.C., 1981. Extensive polymorphism in the mitochondrial DNA of apes. Proc. Natl. Acad. Sci. USA 78, 6319–6323.

Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Ann. Rev. Ecol. Syst. 34, 397–423.

Gatesy, J., Matthee, C., DeSalle, R., Hayashi, C., 2002. Resolution of a supertree/supermatrix paradox. Syst. Biol. 51, 652–664.

Gibbons, A., 1990. Our chimp cousins get that much closer. Science 250, 376.

Greenwood, P.J., 1980. Mating systems, philopatry and dispersal in birds and mammals. Anim. Behav. 28, 1140–1162.

Harrison, R.G., 1998. Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. In: Howard, D.J., Berlocher, S.H. (Eds.), Endless Forms: Species and Speciation. Oxford University Press, New York, NY, pp. 19–31.

Hatch, L.T., 2004. Male genes and male songs: Integrating genetic and acoustic data in defining fin whale, *Balaenoptera physalus*, management units (Ph.D. dissertation). Ithaca, NY: Cornell University.

Hellborg, L., Ellegren, H., 2004. Low levels of nucleotide diversity in mammalian Y chromosomes. Mol. Biol. Evol. 21, 158–163.

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., Huelsenbeck, J.P., 1992. Signal, noise, and reliability in molecular phylogenetic analysis. J. Hered. 83, 189–195.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

The International SNP Map Working Group, 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 407, 900–903.

Kingston, S.E., Rosel, P.E., 2004. Genetic differentiation among recently diverged delphinid taxa determined using AFLP markers. J. Hered. 95, 1–10.

Maddison, W.P., Maddison, D.R., 1992. MacClade, Version 3.08a: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, MA.

Messenger, S.L., McGuire, J.A., 1998. Morphology, molecules, and the phylogenetics of cetaceans. Syst. Biol. 47, 90–124.

Nishida, S., Pastene, L.A., Goto, M., Hiroko, K., 2003. SRY gene structure and phylogeny in the cetacean species. Mamm. Study [Japan] 28, 57–66.

Palsbøll, P.J., Vader, A., Bakke, I., Raafat El-Gewely, M., 1992. Determination of gender in cetaceans by the polymerase chain reaction. Can. J. Zool. 70, 2166–2170.

Perrin, W.F., Reeves, R.R., 2004. Appendix 5: report of the working group on species- and subspecies-level taxonomy. In: Reeves, R.R., Perrin, W.F., Taylor, B.L., Baker, C.S., Mesnick, S.L. (Eds.), Report of the Workshop on Shortcomings of Cetacean Taxonomy in Relation to Needs of Conservation and Management, April 30–May 2, La Jolla, California. NOAA technical memorandum NMFS NOAA-TM-NMFS-SWFSC #363, pp. 25–60.

Perry, S.L., DeMaster, D.P., Silber, G.K., 1999. The great whales: history and status of six species listed as endangered under the US Endangered Species Act of 1973. Mar. Fish. Rev. [Special Issue] 61, 1–74.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rannal, B., Yang, Z., 1996. Probability distribution of molecular evolution trees: a new method of phylogenetic inference. J. Mol. Evol. 43, 304–311.

Rogers, J.S., Swofford, D.L., 1998. A fast method for approximating maximum likelihoods of phylogenetic trees from nucleotide sequences. Syst. Biol. 47, 77–89.

Rosenbaum, H.C., et al. (19 co-authors). 2000. World-wide genetic differentiation of *Eubalaena*: questioning the number of right whale species. Mol. Ecol. 9, 1793–1802.

Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Page, D.C., 2003. Abundant gene conversion between arms of palidromes in human and ape Y chromosomes. Nature 423, 873–876.

Rychel, A.L., Reeder, T.W., Berta, A., 2004. Phylogeny of mysticete whales based on mitochondrial and nuclear data. Mol. Phylogenet. Evol. 32, 892–901.

Sambrook, J., Russel, D.W., 2001. Molecular Cloning, a Laboratory Manual (third ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Volume 2.

Schneider, S., Roessli, D., Excoffier, L., 2000. Arlequin ver. 2.000: A Software for Population Genetics Analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Shen, P., Wang, F., Underhill, P.A. et al. (13 co-authors), 2000. Population genetic implications from sequence variation in the human genome. Genome Res. 12, 1350–1356.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., et al. (37 co-authors), 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423, 825–837.

Sorenson, M.D., 1999. TreeRot, Version 2. Boston University, Boston, MA.

Spilliaert, R., Vikingsson, G., Árnason, Ú., Palsdottir, A., Sigurjonsson, J., Árnason, A., 1991. Species hybridization between a female blue whale (Balaenoptera musculus) and a male fin whale (*B. physalus*): molecular and morphological documentation. J. Hered. 82, 269–274.

Stone, A.C., Griffiths, R.C., Zegura, S.L., Hammer, M.F., 2002. High levels of Y-chromosome nucleotide diversity in the genus Pan. Proc. Natl. Acad. Sci. USA 99, 43–48.

Swofford, D.L., 2003. PAUP* Phylogenetic Analysis using Parsimony (* and other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Swofford, D.L., Waddell, P.J., Huelsenbeck, J.P., Foster, P.G., Lewis, P.O., Rogers, J.S., 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst. Biol. 50, 525–539.

Templeton, A.R., 1983. Phylogenetic inference form restriction endonuclease cleavage site maps with particular reference to the humans and apes. Evolution 37, 221–244.

Tosi, A.J., Morales, J.C., Melnick, D.J., 2003. Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of Macaque monkeys. Evolution 57, 1419–1435.

Wada, S., Numachi, K., 1991. Allozyme analyses of genetic differentiation among the populations and species of the *Balaenoptera*. Rep. Int. Whal. Commis. 13 [special issue], 125–154.

Wada, S., Oishi, M., Yamada, T.K., 2003. A newly discovered species of living baleen whale. Nature 426, 278–281.

Willis, P.M., Crespi, B.J., Dill, L.M., Baird, R.W., Hanson, M.B., 2004. Natural hybridization between Dall's porpoises (*Phocoenoides dalli*) and harbour porpoises (*Phocoena phocoena*). Can. J. Zool. 82, 828–834..