Molecular Phylogenetics and Evolution 53 (2009) 891–906

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Divergence date estimation and a comprehensive molecular tree of extant cetaceans

Michael R. McGowen *, Michelle Spaulding [1], John Gatesy

Department of Biology, University of California Riverside, Riverside, CA 92521, USA

ARTICLE INFO

Article history:
Received 26 February 2009
Revised 4 August 2009
Accepted 14 August 2009
Available online 21 August 2009

Keywords:
Cetacea
Odontoceti
Mysticeti
Phylogeny
Supermatrix

ABSTRACT

Cetaceans are remarkable among mammals for their numerous adaptations to an entirely aquatic existence, yet many aspects of their phylogeny remain unresolved. Here we merged 37 new sequences from the nuclear genes *RAG1* and *PRM1* with most published molecular data for the group (45 nuclear loci, transposons, mitochondrial genomes), and generated a supermatrix consisting of 42,335 characters. The great majority of these data have never been combined. Model-based analyses of the supermatrix produced a solid, consistent phylogenetic hypothesis for 87 cetacean species. Bayesian analyses corroborated odontocete (toothed whale) monophyly, stabilized basal odontocete relationships, and completely resolved branching events within Mysticeti (baleen whales) as well as the problematic speciose clade Delphinidae (oceanic dolphins). Only limited conflicts relative to maximum likelihood results were recorded, and discrepancies found in parsimony trees were very weakly supported. We utilized the Bayesian supermatrix tree to estimate divergence dates among lineages using relaxed-clock methods. Divergence estimates revealed rapid branching of basal odontocete lineages near the Eocene–Oligocene boundary, the antiquity of river dolphin lineages, a Late Miocene radiation of balaenopteroid mysticetes, and a recent rapid radiation of Delphinidae beginning ~10 million years ago. Our comprehensive, time-calibrated tree provides a powerful evolutionary tool for broad-scale comparative studies of Cetacea.

1. Introduction

Cetaceans (whales and dolphins) are the most speciose living group of aquatic mammals, comprising 87–89 extant species in 14 families. They are ecomorphologically diverse, ranging in adult length from approximately 1.45 m (*Phocoena sinus*, vaquita) to 33 m (*Balaenoptera musculus*, blue whale), and inhabit every ocean basin and many large river systems (Jefferson et al., 2008). Traditionally, extant cetaceans are divided into two distinct groups based on their morphology: Mysticeti (baleen whales) and Odontoceti (toothed whales, dolphins, and porpoises). Both groups have significantly diverged from terrestrial mammals via a suite of characteristic adaptations for the aquatic environment. Extant mysticetes possess baleen, a keratinous sieve that continuously grows from their palate and is used to filter food from the water. Odontocetes use echolocation to detect prey, which they then capture via suction or tooth-assisted predation (Fordyce and de Muizon, 2001). Due to their unique morphology and relative diversity, cetaceans pose many challenging questions to fields as diverse as neurobiology (Marino et al., 2007), behavior (Connor, 2007), bioacoustics (May-Collado et al., 2007), physiology (Tyack et al., 2006), biomechanics (Fish, 1998; Woodward et al., 2006), molecular evolution (Levenson and Dizon, 2003; Iwanami et al., 2006; Deméré et al., 2008), and paleontology (Geisler and Sanders, 2003; Bisconti, 2005; Geisler and Uhen, 2003; Fordyce, 2009).

The secondarily aquatic nature of whales and dolphins has inspired numerous phylogenetic studies and prompted debate concerning their origins among terrestrial mammals (reviewed in Gatesy and O'Leary, 2001; O'Leary and Gatesy, 2008). No less attention has been given to elucidating relationships among the major groups of cetaceans. The monophyly of Cetacea is well-supported by multiple molecular and morphological analyses (Messenger and McGuire, 1998; Gatesy et al., 1999a; Nikaido et al., 2001a; Geisler and Sanders, 2003; Árnason et al., 2004; Agnarsson and May-Collado, 2008; O'Leary and Gatesy, 2008). However, many relationships within Cetacea remain less certain, despite multiple efforts to resolve discrete portions of the phylogeny using a diverse array of systematic markers (mitochondrial [mt] DNA [Árnason et al., 1991a, 2000, 2004; Árnason and Gullberg, 1993, 1994, 1996; Milinkovitch et al., 1993, 1994, 1996; Rosel et al., 1995; Montgelard et al., 1997; LeDuc et al., 1999; Cassens et al., 2000; Hamilton et al., 2001; Dalebout et al., 2002, 2003, 2004; Rychel et al, 2004; Sasaki et al., 2005, 2006; Yan et al., 2005; Caballero et al., 2007; McGowen et al., 2008; Xiong et al., 2009], exons, introns, pseudogenes [references in Table 1], transposons [Nikaido et al. 2001a, 2001b, 2006, 2007], and morphology [Heyning,

* Corresponding author. Address: Department of Biology, University of California Riverside, Spieth Hall, 900 University Ave., Riverside, CA 92521, USA.
E-mail address: mmcgo002@student.ucr.edu (M.R. McGowen).
[1] Present addresses: Department of Earth and Environmental Sciences, Columbia University, NY 10027, USA. Division of Paleontology, American Museum of Natural History, NY 10024, USA.

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.08.018

Table 1
List and description of nuclear genes included in this study.

Gene	Protein	Length (bp)	Type	References
ACTA2	Alpha-2-actin	1235	Intron	Alter et al. (2007), Caballero et al. (2007, 2008), Dalebout et al. (2004), Hare et al. (2002), Harlin-Cognato and Honeycutt (2006), Palumbi and Baker (1994), Palumbi et al. (2001)
AMBN	Ameloblastin	607	Exon (Pseudogene in some taxa)	Deméré et al. (2008), McGowen et al. (2008)
AMEL	Amelogenin	435	Exon (Pseudogene in some taxa)	Gatesy et al. (in prep)
ATP7A	Copper-transporting ATPase-1	677	Exon	Deméré et al. (2008); Murphy et al. (2001)
BDNF	Brain-derived neurotrophic factor	581	Exon	Deméré et al. (2008), Murphy et al. (2001)
BTN1A1	Butyrophilin 1A1	672	Intron	Alter et al. (2007), Hare et al. (2002), McGowen et al. (2008)
CAT	Catalase	513	Exon; Intron	Caballero et al. (2007, 2008)
CHRNA1	Alpha-1 nicotinic cholinergic receptor	360	Exon; Intron	Caballero et al. (2007, 2008)
CSN2	Beta casein	439	Exon	Gatesy (1998), O'Leary and Gatesy (2008)
DBY	DEAD-box protein 3, Y-linked	397	Exon; Intron	Caballero et al. (2007, 2008)
DMP1	Dentin matrix protein 1	517	Exon	Deméré et al. (2008)
ENAM	Enamelin	536	Exon (Pseudogene in some taxa)	Deméré et al. (2008)
ESD	Esterase D	828	Exon; Intron	Alter et al. (2007)
FGG	Gamma fibrinogen	677	Intron	Alter et al. (2007)
G6PD	Glucose-6-phosphate dehydrogenase	308	Exon; Intron	Caballero et al. (2007, 2008)
GBA	Beta glucosidase (glucocerebrosidase)	308	Exon; Intron	Caballero et al. (2007, 2008)
IFN	Interferon	337	Exon	Caballero et al. (2007, 2008)
IRBP	Inter-retinoid binding protein	1268	Exon	Cassens et al. (2000, Gatesy et al., 1999a, Smith et al., 1996, Waddell et al. (2000
KITLG	KIT ligand	582	Exon; Intron	Deméré et al. (2008)
LALBA	Alpha-lactalbumin	1171	Exon; Intron	Caballero et al. (2007, 2008), Cassens et al. (2000), Milinkovitch et al. (1998), Rychel et al. (2004), Waddell et al. (2000)
MB	Myoglobin	459	Exon derived from mRNA	Iwanami et al. (2006) and references therein
OPN1SW	Short-wave-sensitive opsin	695	Exon; Intron (Pseudogene)	Levenson and Dizon (2003), Newman and Robinson (2005)
PKDREJ	Polycystic kidney disease and REJ homolog	207	Exon	Deméré et al. (2008)
PLP	Proteolipid protein 1	899	Exon; Intron	Alter et al. (2007)
PRM1	Protamine P1	449	Exon; Intron; Flanking regions	Deméré et al. (2008), Gatesy (1998), McGowen et al. (2008), O'Leary and Gatesy (2008); this study
RAG1	Recombination activating gene 1	811	Exon	Deméré et al. (2008), McGowen et al. (2008), O'Leary and Gatesy (2008); this study
SINEs	SINE flanking regions (11 loci)	2790	Anonymous	Nikaido et al. (2001b)
SMCY	Smcy homolog	449	Intron	Caballero et al. (2007, 2008)
SPTBN1	Non-erythrocytic beta spectrin 1	937	Intron	Matthee et al. (2001), McGowen et al. (2008)
SRY	Sex-determining region Y	2085	Exon; Flanking regions	Nishida et al. (2003, 2007)
STAT5	Signal transducer and activator of transciption factor 5A	801	Exon; Intron	Deméré et al. (2008), Matthee et al. (2001)
TBX4	T-box 4 transcription factor	1362	Exon	Onbe et al. (2007)
TF	Transferrin	1577	Exon; Intron	Cassens et al. (2000)
vWF	Von Willebrand factor	1237	Exon	Cassens et al. (2000, Gatesy et al., 1999a, Waddell et al. (2000
WT1	Wilms tumor 1	526	Exon; Intron	Alter et al. (2007)

1997; Messenger and McGuire, 1998; Geisler and Sanders, 2003; Deméré et al., 2008]). For example, several early phylogenetic hypotheses, separately derived from mt genes and the nuclear (nu) gene *IRBP*, suggested that Mysticeti was nested within Odontoceti (Milinkovitch et al., 1993, 1994, 1996; Smith et al., 1996), contrary to abundant morphological evidence (Heyning, 1997; Messenger and McGuire, 1998; Geisler and Sanders, 2003; O'Leary and Gatesy, 2008), and recent analyses of complete mt genomes yielded low support for a monophyletic Odontoceti (Árnason et al., 2004; Yan et al., 2005). Alternatively, individual nu genes (Nishida et al., 2007), concatenations of nu genes (Gatesy, 1998) and SINE insertion data (Nikaido et al., 2001b, 2007) displayed consistent support for odontocete monophyly. In addition to this conflict, previous phylogenetic hypotheses disagree concerning the branching sequence of the three most basal odontocete clades (Physeteroidea, Ziphiidae, Platanistidae), the exact position of the now extinct Chinese river dolphin *Lipotes vexillifer*, and the relationships of species within the delphinid, ziphiid, and balaenopteroid radiations.

Although a large number of molecular characters have been published for Cetacea, much of this evidence is scattered throughout the group and remains to be combined and evaluated in a comprehensive supermatrix analysis encompassing every extant species. The supermatrix approach to systematics summarizes the strongest hierarchical signals in the character data and permits emergence of phylogenetic support that may not be apparent when analyzing different data sets independently (Barrett et al., 1991); the relative weight of character support for conflicting relationships can be directly assessed in a single phylogenetic test that arbitrates the evidence from distinct data partitions. This contrasts with consensus or supertree methods, in which trees derived from separate loci are used to generate summary topologies (Bininda-Emonds, 2004).

In the absence of a cohesive species-level phylogeny, there is also a lack of divergence estimates for most cetacean speciation events. A handful of analyses have attempted to date deep splits within Odontoceti (Cassens et al., 2000; Nikaido et al., 2001b; Árnason et al., 2004; Xiong et al., 2009), Mysticeti (Sasaki et al., 2005, 2006), and more exclusive clades (*Mesoplodon* spp., Dalebout et al., 2008), but most odontocete studies have included only one species per family. Therefore, divergence date estimates in many clades remain largely unexplored. This is especially true of oceanic

dolphins (Delphinidae, 36 spp.), a rapid radiation that accounts for over 40% of extant cetacean diversity; the timing of most phylogenetic splits within this group has not been investigated using molecular data.

Here we combined most published molecular data for Cetacea, including both DNA sequences and transposon insertion events, to produce a comprehensive species-level tree for 87 out of 89 extant species of Cetacea. New sequences from two nu genes (*PRM1* and *RAG1*) were incorporated, creating a substantial supermatrix of 42,335 characters from 45 nu loci, mt genomes, and transposon insertion events. We compared this combined approach to other molecular analyses using one or few loci and to a recent matrix representation with parsimony (MRP) supertree (Price et al., 2005). Lastly, we used a relaxed-clock Bayesian approach to yield the first divergence date estimates encompassing nearly all extant cetacean species, and interpreted these dates in the context of major Cenozoic geologic and climatic events.

2. Methods

2.1. Taxa sampled and sequencing of RAG1 and PRM1

DNA samples were provided by P. Morin, A. Dizon and K. Robertson (SWFSC: Southwest Fisheries Science Center, NOAA, La Jolla, CA), G. Amato (NYZS: New York Zoological Society), P. Palsbøll (Stockholm University), and M. Milinkovitch (Free University of Brussels). Donating institutions (or persons) and sample reference numbers (if applicable) are listed after each species below. For this study, new taxa sampled for *RAG1* included: *Tursiops truncatus* (SWFSC Z38274), *Stenella attenuata* (SWFSC Z592), *Lissodelphis borealis* (SWFSC Z176), *Feresa attenuata* (SWFSC Z3944 [Mote Marine Lab]), *Globicephala macrorhynchus* (SWFSC Z39091, Z537 [Smithsonian Institution, Division of Mammals]), *Grampus griseus* (SWFSC Z483 [Smithsonian Institution, Division of Mammals]), *Phocoenoides dalli* (SWFSC Z711), *Neophocaena phocaenoides* (SWFSC Z984), *Delphinapterus leucas* (NYZS), *Mesoplodon peruvianus* (Milinkovitch), *Tasmacetus shepherdi* (SWFSC Z4971 [Smithsonian Institution, Division of Mammals]), *Platanista minor* (SWFSC Z15224 [Gill Braulik, World Wildlife Fund]), *Kogia sima* (SWFSC Z23604), *Eubalaena australis* (SWFSC Z18928 [South Australian Museum]), and *Eubalaena glacialis* (SWFSC Z13086 [Northeast Fisheries Science Center, Stranding Network]). All of the above taxa except *S. attenuata and D. leucas* were also sampled for *PRM1*, plus the following additional species: *Hippopotamus amphibius* (NYZS), *Inia geoffrensis* (SWFSC Z505 [Smithsonian Institution, Division of Mammals]), *Pontoporia blainvillei* (SWFSC Z7349), *Leucopleurus acutus* (SWFSC Z7842), *Monodon monoceros* (Palsbøll), *Mesoplodon bidens* (SWFSC Z3859), *Mesoplodon grayi* (SWFSC Z6997), *Berardius bairdii* (SWFSC Z4963), and *Kogia breviceps* (SWFSC Z10119 [Mote Marine Lab]).

RAG1 sequences were PCR amplified and sequenced using primers and conditions described in Deméré et al. (2008). *PRM1* sequences were PCR amplified and sequenced using primers from Queralt et al. (1995) and Deméré et al. (2008) with conditions described in Deméré et al. (2008). All sequences were deposited in Genbank (Accession Nos. GQ368508–GQ368546).

2.2. Compilation of supermatrix

Database searches were conducted in Genbank using scientific names of species, higher-level taxa, gene name, and gene symbol. There were no molecular data available for *Sousa teuszii*, and none identified as *Orcaella heinsohni*, although it is likely that some sequences attributed to *O. brevirostris* are *O. heinsohni* due to the very recent elevation of *O. heinsohni* to species status (Beasley et al., 2005). Sequences listed as "unpublished" were not included in this study. Due to the date of compilation, sequences published after August 2008 also were not included in this study. We selected four cetartiodactylan outgroup taxa (*Hippopotamus amphibius* [hippo], *Choeropsis liberiensis* [pygmy hippo], *Bos taurus* [cow], *Sus scrofa* [pig]). *Hippopotamus* and *Choeropsis* are members of Hippopotamidae, the extant sister group of Cetacea (Irwin and Árnason, 1994; Gatesy, 1997, 1998; Montgelard et al., 1997; Gatesy et al., 1999a,b; Ursing et al., 2000; Geisler and Uhen, 2003; O'Leary and Gatesy, 2008). *Bos taurus* and *Sus scrofa* are domestic animals within the more distantly related clades of Ruminantia and Suina, respectively; *Bos taurus* has a published genome available through NCBI. At least one cetacean sequence per gene was used to search the *Bos* genome for orthologs. Genes that lacked a published sequence from any of the four outgroup taxa were excluded. For genes with multiple sequences available per species, only one sequence was chosen to represent the taxon. Genes that were represented by four or less sequences from different cetacean species, or that showed no phylogenetically informative variation within the ingroup were not included in the supermatrix.

Mt genomes from 25 cetaceans and three outgroup species were available on Genbank as of August 2008 (Árnason et al., 1991a, 2000, 2004; Árnason and Gullberg, 1993; Sasaki et al., 2005, 2006; Yan et al., 2005). Twelve protein-coding genes and two rRNA genes were downloaded separately from each mt genome. These sequences were combined with an extensive database of cetacean mt genes from other species: *MT-CYTB* (cytochrome *b*, Árnason and Gullberg, 1994, 1996; Rosel et al., 1995; Milinkovitch et al., 1996; LeDuc et al., 1999; Cassens et al., 2000; Hamilton et al., 2001; Dalebout et al., 2002, 2003, 2004; Caballero et al., 2007), *MT-ND4/ND4L* (NADH dehydrogenase 4/4L, Rychel et al., 2004), *MT-COX2* (cytochrome *c* oxidase II, McGowen et al., 2008), 12S rRNA and 16S rRNA (Milinkovitch et al., 1993, 1994; Montgelard et al., 1997; Cassens et al., 2000; Hamilton et al., 2001). Every operational taxonomic unit (OTU) in the study was at the least represented by an *MT-CYTB* sequence. In addition to mt sequences, 45 nu DNA fragments were included in the supermatrix; nu fragments are listed in Table 1. Missing data in the combined data set are summarized in Fig. 1.

All genes were aligned individually using Clustal X (Thompson et al., 1994) with a gap-opening penalty of 10 and gap-extension penalty of 1. The Clustal output was inspected by eye and some adjacent gaps were consolidated using Seqapp 1.9a (Gilbert, 1992). Indels (insertions or deletions) were coded for each gene in SeqState (Müller, 2005) using the simple gap-coding method of Simmons and Ochoterena (2000). Transposon insertion events from Nikaido et al. (2001a, 2001b, 2006, 2007) also were added to the matrix. The total data set consisted of 41,540 nucleotide characters (13,535 mt, 28,005 nu), 694 gap characters, and 101 transposon insertion events. The final supermatrix was submitted to TreeBase.

2.3. Phylogenetic analysis

Markov chain Monte Carlo (MCMC) Bayesian analyses were conducted using default parameters and four simultaneous chains (three "cold", one "heated") in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The combined data set was analyzed using separate partitions for mt and nu nucleotide data. Both gaps and transposons were treated as separate partitions, and each used the binary model of character evolution (Ronquist et al., 2005). For DNA sequence alignments, MrModeltest 2.2 (Nylander, 2004) was employed to choose optimal models for each partition according to the AIC (Posada and Buckley, 2004; mt: GTR + I + G; nu: GTR + I + G). Two concurrent runs of 15,000,000 generations were employed for each analysis with trees sampled every 1000

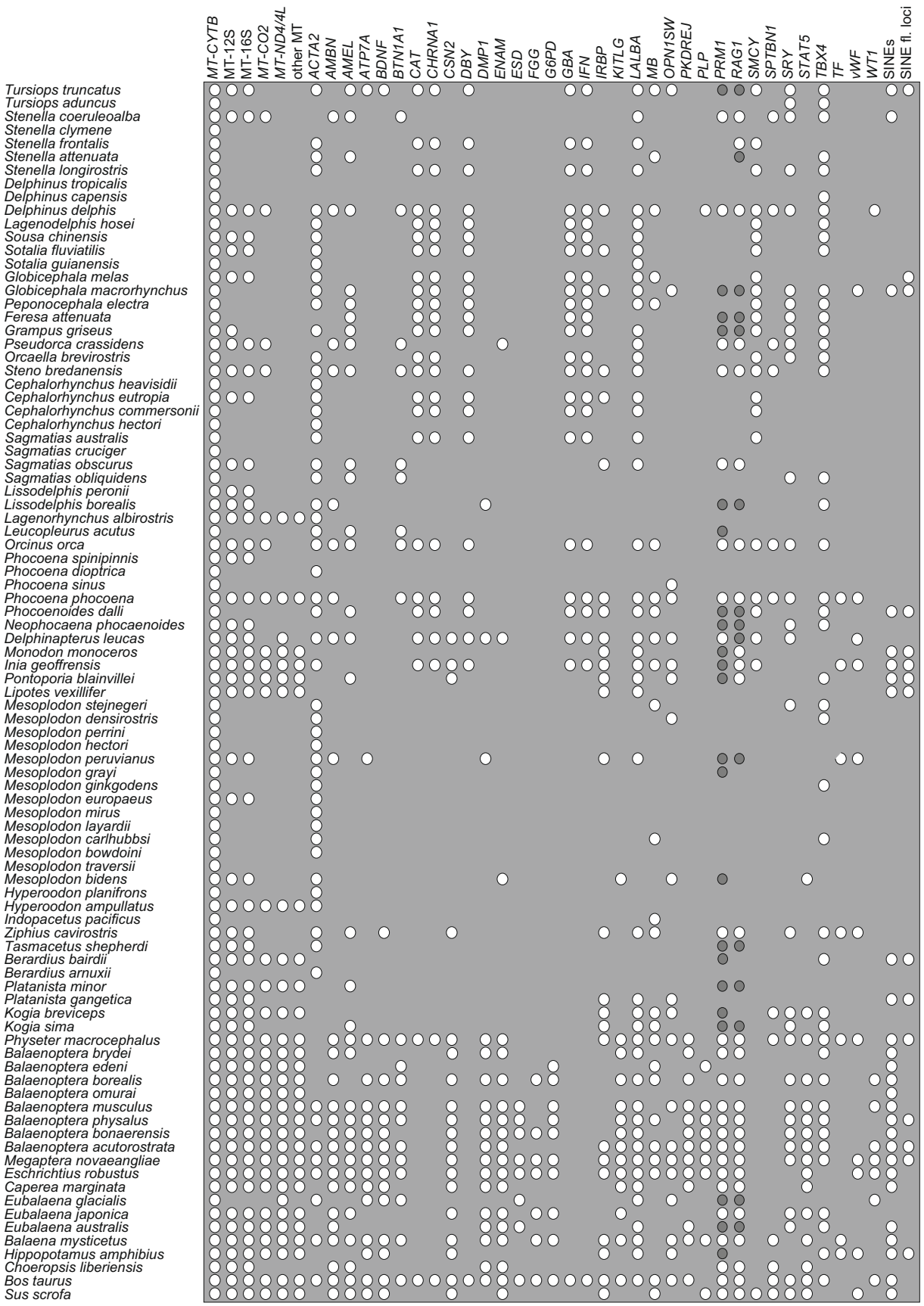

Fig. 1. Data sets (top) versus species (left) included in the supermatrix analysis. White dots symbolize the presence of some or all of a data set for a particular species. Dark gray dots represent data that are new to this study.

generations to assess convergence. The stationarity of the likelihood scores of sampled trees was evaluated in Tracer v1.04 (Rambaut and Drummond, 2007), and the first 10% of trees were discarded as burn-in. The standard deviation of split frequencies between runs also was evaluated to establish that concurrent runs had converged. A 50% majority-rule consensus of post burn-in trees from both runs was constructed to summarize posterior probabilities for each clade. A second Bayesian analysis was executed in which gap characters were excluded; analysis parameters were as in the initial Bayesian analysis described above.

Maximum likelihood (ML) analyses were conducted using RAxML 7.0.4 (Stamatakis, 2006) via the Cyberinfrastructure for Phylogenetic Research (CIPRES) Portal (www.phylo.org). Gaps and transposons were excluded, as RAxML cannot implement binary models of character evolution. RAxML can incorporate separate models of nucleotide substitution for individual data partitions; here, we used two strategies: (1) two partitions: mt and nu, and (2) 119 partitions: each codon position for each protein-coding gene, 12S rDNA, 16S rDNA, and a final partition for non-coding nu DNA. The search for the optimal likelihood tree was initiated using "GTRMIX." This option first searches for trees via the GTRCAT algorithm, which incorporates rate heterogeneity with low computational cost (Stamatakis, 2006). Then, GTRMIX compares likelihood scores for trees under a GTR + G model and chooses the optimal tree. To assess support, 100 ML bootstrap replicates were run in RAxML using the CIPRES interface (Stamatakis et al., 2008).

Maximum parsimony (MP) analyses were implemented in PAUP* 4.0b10 (Swofford, 2002) with 1000 random stepwise-addition replicates and tree bisection reconnection (TBR) branch-swapping. All nucleotide, gap, and transposon characters were given equal weight, and internal branches were collapsed if the minimum length of an internode was zero ("amb-minus" option in PAUP*). Support for nodes was evaluated by non-parametric bootstrapping using 1000 pseudoreplicates with 10 random-addition sequence replicates (Felsenstein, 1985) in PAUP*, and Bremer support scores (Bremer, 1994) for selected nodes were estimated using 20 random stepwise-addition replicates in PAUP* with TreeRot v.3 (Sorenson and Franzosa, 2007). An additional MP analysis that incorporated morphological data from Geisler and Sanders (2003) was executed. Three hundred and four phenotypic characters coded for 19 taxa were merged with the molecular data in the supermatrix; tree search and assessments of support were performed as described above.

2.4. Divergence time estimation

Divergence times were estimated using *MT-CYTB* only, because *MT-CYTB* is the only gene that has been sampled for every OTU in this analysis. A relaxed-clock MCMC approach using the uncorrelated log-normal model was implemented in BEAST v1.4.8 (Drummond et al., 2006). Conditional log-normal minimum priors for age estimates were established for the Cetacea–Hippopotamidae split (54 million years [Ma], log-normal distribution with mean of 55.0 and standard deviation [SD] of 1.0; based on the oldest stem cetacean *Himalayacetus* [Bajpai and Gingerich, 1998]), crown Cetacea (34.2 Ma, 35.2 mean, 1.0 SD; based on the oldest mysticete fossil *Llanocetus* [Mitchell, 1989; Steeman, 2007]), crown Mysticeti (28 Ma, 29.0 mean, 1.0 SD; based on an unnamed balaenid from New Zealand [Fordyce, 2002]), and the Phocoenidae–Monodontidae split (8.75 Ma, 9.75 mean, 1.0 SD; based on the oldest phocoenid, *Salumiphocaena stocktoni* [Barnes, 1985]). Divergence times also were estimated with the log-normal minimum prior of the Cetacea–Hippopotamidae split set at 50 Ma (51.0 mean, 1.0 SD), the age of the earliest confirmed stem cetacean (*Pakicetus*) that has been integrated into phylogenetic analyses (Geisler and Uhen, 2003). A GTR + I + G model was applied to three partitions corresponding to the first, second, and third codon positions. Base frequencies, rate heterogeneity, and substitution models were unlinked across all partitions. A Yule process prior was implemented, assuming a constant rate of speciation. Analyses were performed both with minimal constraints (only calibration nodes and Odontoceti) and with all nodes fixed to relationships represented in the Bayesian majority-rule consensus tree derived from our supermatrix. Use of the Bayesian consensus tree in this context seems justified, because this topology was supported by combined analysis of multiple, independent molecular markers, and represents a more robust estimate of phylogeny relative to an analysis of *MT-CYTB* alone. Furthermore, only one node supported by Bayesian analysis of *MT-CYTB* with PP $\geqslant$ 0.95 conflicted with our Bayesian supermatrix. Each clock analysis was executed for 30 million generations with parameters sampled every 1000 generations. Results were examined using Tracer v1.04 (Rambaut and Drummond, 2007) to evaluate stationarity, and the first three million generations were discarded as burn-in.

3. Results

3.1. Supermatrix phylogeny

Overall there was a large disparity in the number of species sampled for each gene, ranging from seven (*TF* and *FGG*) to 91 (*MT-CYTB*) (Fig. 1). Twenty-three species were represented by less than four data sets, and four of these species included only *MT-CYTB*. Ziphiidae (beaked whales) had low coverage, with 11 out of 21 (52.4%) species having less than four data sets in the supermatrix. In comparison, Mysticeti (baleen whales) was well-sampled with only one species characterized by less than 10 data sets (Fig. 1).

Topologies derived from the two Bayesian runs were nearly identical; the inclusion or exclusion of gap characters had little effect on the resulting topologies and support scores (Fig. 2). Furthermore, the two ML analyses were highly congruent with the Bayesian consensus trees, weakly conflicting at only six or seven nodes (Fig. 2). The ML trees ($-\ln L$ [119 partitions] = 208,680.348504; $-\ln L$ [two partitions] = 223,274.327263) also were strongly concordant with each other (Fig. S1). All model-based phylogenetic analyses recovered cetacean monophyly (posterior probability [PP] = 1.0; ML bootstrap [BS] = 100%), odontocete monophyly (PP = 1.0; ML BS = 100%), and supported Physeteroidea (Physeteridae + Kogiidae) as the most basal clade of odontocetes (PP = 1.0; ML BS = 100%; Fig. 2). Other suprafamilial groupings that were supported by Bayesian PP = 1.0 and ML BS = 100% included: Mysticeti, Balaenopteroidea (Balaenopteridae + Eschrichtiidae), Balaenopteroidea + Neobalaenidae, Physeteroidea, Inioidea (Iniidae + Pontoporiidae), Delphinida (Inioidea + Lipotidae + Delphinoidea), Delphinoidea (Delphinidae + Phocoenidae + Monodontidae), and Phocoenidae + Monodontidae (Figs. 2 and S1). The position of Platanistidae as sister to Ziphiidae + Delphinida was resolved but with more moderate support scores (PP = 0.850–0.88; ML BS = 87–90%), and Lipotidae was placed as sister to Inioidea in all Bayesian and ML analyses (PP = 0.69–0.88; ML BS = 55–62%).

Cetacean families (except Balaenopteridae) and delphinid subfamilies with more than one species were overwhelmingly supported by Bayesian and ML analyses (PP = 1.0; ML BS $\geqslant$ 97%). All relationships within Mysticeti were strongly supported by Bayesian PP = 1.0, and the ML analysis with two partitions agreed with these results (ML BS = 83–100%). However, the ML analysis with 119 partitions positioned *Eschrichtius* (Eschrichtiidae) in a more basal position within Balaenopteridae (Fig. S1), and grouped *Megaptera* and *Balaenoptera physalus* with a cluster of five *Balaenoptera* species (ML BS = 64%). Within Delphinidae, both *Orcaella*

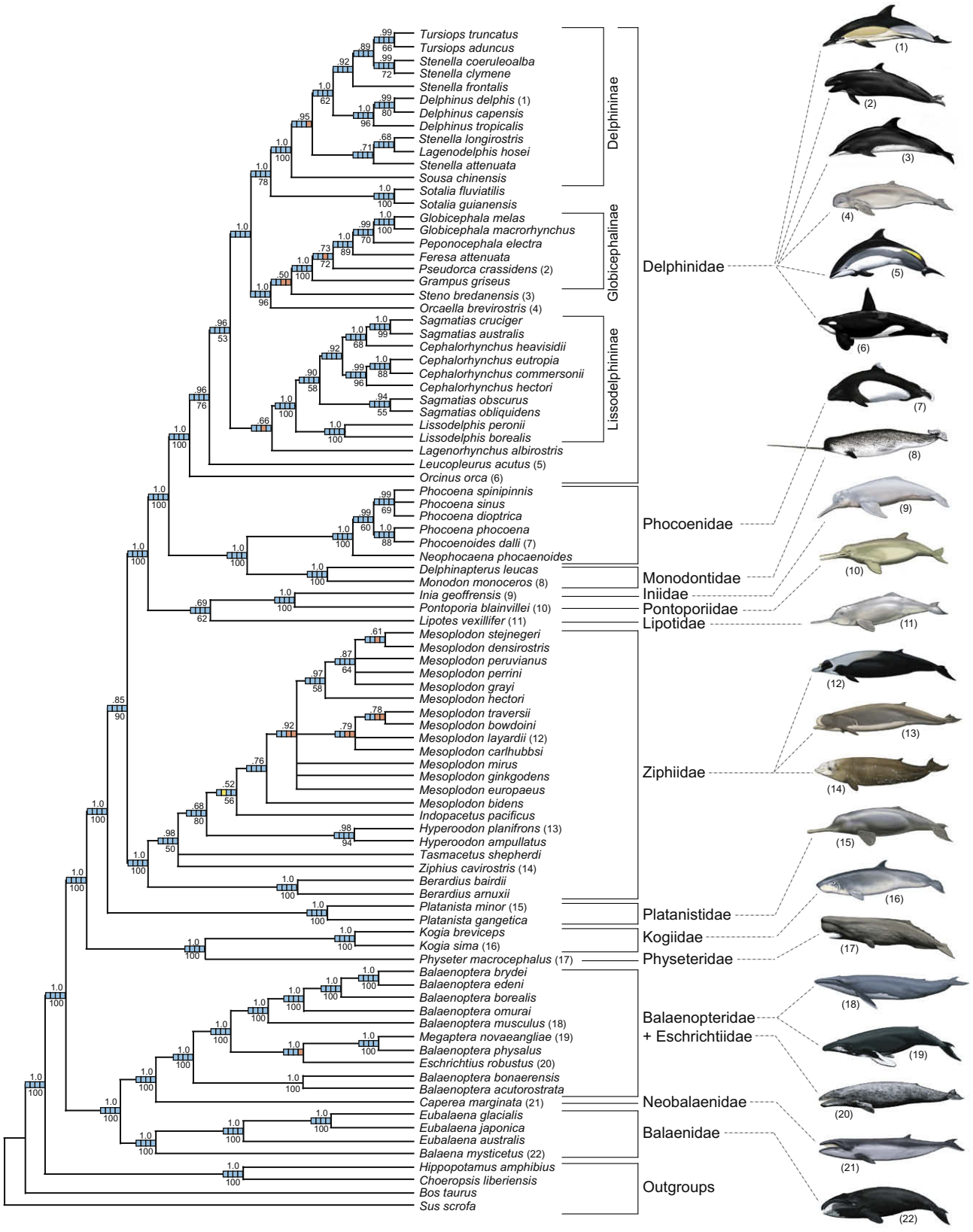

Fig. 2. Phylogenetic relationships favored by model-based analyses of the cetacean molecular supermatrix. The topology derived from Bayesian analysis of nucleotides, gap characters, and transposon insertions is shown. The four squares at each internode indicate congruence (blue), conflict (red), or lack of resolution (yellow) in the four model-based analyses of the supermatrix data set: Bayesian analysis of nucleotides, gaps, and transposons (leftmost square), Bayesian analysis of nucleotides and transposons (2nd from left), maximum likelihood (ML) analysis of nucleotides with two model partitions (2nd from right), and ML analysis of nucleotides with 119 model partitions (rightmost square). Bayesian clade posterior probabilities $\geqslant 0.50$ for the analysis of nucleotides, gaps, and transposons are above internodes. ML bootstrap support scores $\geqslant 50\%$ for the 119 partition analysis of nucleotides are below internodes. Cetacean families and three subfamilies of Delphinidae are delimited by brackets to the right of the tree along with representative members.

and *Steno* were placed with Globicephalinae (PP = 1.0; ML BS = 94–96%), *Sotalia* was sister to Delphininae (PP = 1.0; ML BS = 63–78%), and *Orcinus orca* and *Leucopleurus acutus* represented early diverging delphinid lineages (PP = 0.95–0.96; ML BS < 50–76%). In all model-based phylogenetic analyses, five genera (*Stenella, Sagmatias*, *Cephalorhynchus*, *Phocoena*, *Balaenoptera*) were not monophyletic (Figs. 2 and S1).

The two MP analyses of the supermatrix, all molecular data (266 optimal trees, 38107 steps, CI = 0.470, RI = 0.539) and all molecular plus morphological data (528 optimal trees, 39229

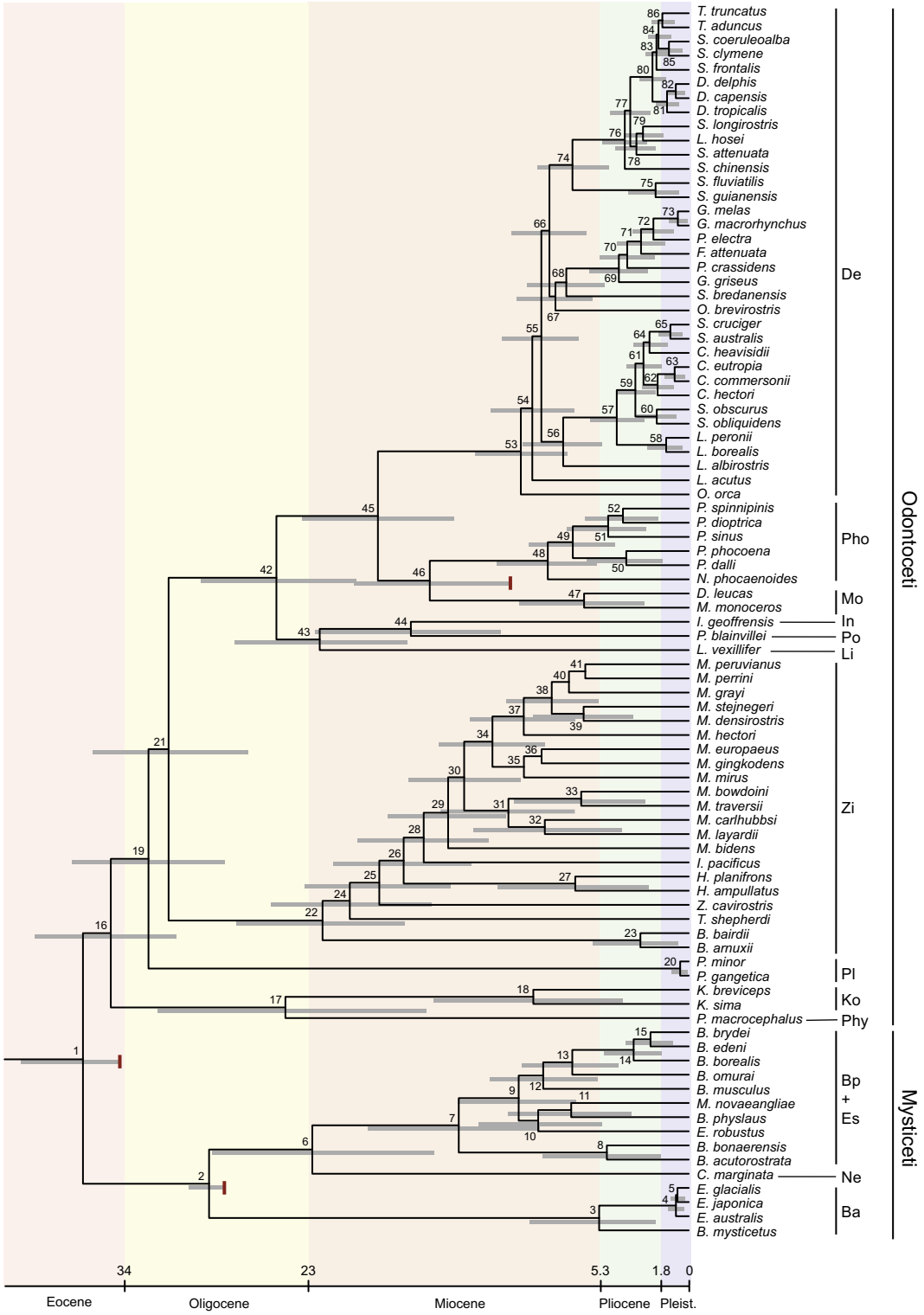

Fig. 3. Time-calibrated phylogenetic tree of Cetacea derived from BEAST. Numbers at each node correspond to divergence date estimates listed in Table 2. Horizontal gray bars correspond to the 95% highest posterior density (HPD) of each node. Vertical red bars show the positions of minimum calibration ages for three nodes within Cetacea used in this analysis. Nodes that do not have 95% HPDs are those with less than 50% clade posterior probabilities in the BEAST analysis. Cenozoic epochs are shaded by color. Family names are abbreviated to the right of the tree (De, Delphinidae; Pho, Phocoenidae; Mo, Monodontidae; In, Iniidae; Po, Pontoporiidae; Li, Lipotidae; Zi, Ziphiidae; Pl, Platanistidae; Ko, Kogiidae; Phy, Physeteridae; Bp + Es, Balaenopteroidea [Balaenopteridae + Eschrichtiidae]; Ne, Neobalaenidae; Ba, Balaenidae).

Table 2
Divergence date estimates for nodes in Fig. 3. Means and 95% highest posterior distributions (HPD) are shown for this study along with dates from five previous studies. "S" refers to the analysis constrained to nodes present in the Bayesian supermatrix tree; "cytb" refers to the analysis with minimal constraints. Asterisks symbolize HPDs that were unavailable for nodes with less than 50% clade posterior probabilities. Numbers in parentheses listed in the column "Cassens et al. (2000)" correspond to approximate error bars. All dates are reported in millions of years (Ma). Numbers in bold indicate nodes where fossil calibrations were integrated into the molecular clock analysis.

Node	Mean (S)	95% HPD (S)	Mean (cytb)	95% HPD (cytb)	Cassens et al. (2000)	Nikaido et al. (2001b)	Árnason et al. (2004)	Sasaki et al. (2006)	Xiong et al. (2009)
1 Cetacea	**36.36**	**34.24–40.14**	**36.10**	**34.26–39.20**	48 (40–52)	**32.3 ± 5.1**	**35.0**	**35.5 ± 1.1**	**33.74 (33.52–35.43)**
2 Mysticeti	**28.79**	**28.03–30.07**	**28.72**	**28.43–29.84**	31 (26–36)		20.9	27.6 ± 1.8	
3 Balaenidae	5.38	2.06–9.60	5.05	2.19–8.51				17.9 ± 3.5	
4 *Eubalaena*	0.77	0.34–1.28	1.39	0.49–2.52				4.6 ± 2.5	
5	0.68	0.30–1.15							
6 Balaenopteroidea + *Caperea*	22.59	15.35–28.64					17.6	23.9 ± 2.4	
7 Balaenopteroidea	13.80	8.99–19.32	14.25	9.09–20.18		19.8 ± 4.7	14.7	20.1 ± 2.7	
8	4.92	1.72–8.83	5.03	1.74–8.86				11.1 ± 2.7	
9	10.21	6.82–13.89	10.38	7.14–19.00			10.3		
10	9.04	5.30–12.69			20 (16–22)				
11	7.06	3.49–10.90	7.37	3.70–11.20		17.2 ± 4.5		15.8 ± 2.8	2.49 (2.46–9.20)
12	8.74	5.50–11.99	8.19	5.35–11.39				17.0 ± 2.7	
13	6.99	4.29–10.06	6.61	4.05–9.32					
14	3.31	1.69–5.13	3.19	1.66–4.87				8.8 ± 2.2	
15	2.30	1.02–3.85	2.24	0.99–3.58				6.3 ± 2.2	
16 Odontoceti	34.69	30.82–39.31	34.45	30.85–38.34	47 (39–51)	30.0 ± 5.0	32.1 ± 1.7		31.29 (25.35–33.20)
17 Physeteroidea	24.21	15.83–31.93	23.88	16.19–31.09	37 (32–42)		24.5		
18 Kogiidae	9.33	4.03–15.38	9.47	4.65–15.43	21 (17–25)				
19	32.43	27.92–37.07	32.26	28.02–36.45	45 (39–49)	28.9 ± 4.9	30.5 ± 1.3		28.77 (23.62–31.68)
20 *Platanista*	0.52	0.12–1.09	0.51	0.14–1.02					
21 Delphinida + Ziphiidae	31.22	26.52––35.82				28.2 ± 4.9			27.95
22 Ziphiidae	21.98	17.11–27.23	22.06	16.47–27.74			19.2		
23 *Berardius*	2.90	0.68–5.81	2.53	0.70–4.70					
24	20.34	15.51–25.14							
25	18.57	14.40–23.00							
26	17.11	13.13–21.42							
27 *Hyperoodon*	6.81	2.44–11.53	5.84	2.37–9.59					
28	15.91	12.11–19.96		*					
29 *Mesoplodon*	14.44	11.03–18.13	10.72						
30	13.48	10.13–16.86							
31	10.83	6.90–14.96		*					
32	8.64	4.00–12.80	8.24	*					
33	6.46	2.67–10.52	4.07						
34	11.78	8.75–14.60							
35	9.89	*		*					
36	8.83		8.84						
37	9.90	6.89–13.21							
38	8.23	5.50–11.05	8.62	6.39–13.3					
39	6.31	3.42–9.41							
40	7.19	*							

41	6.21	*						
42 Delphinida	24.75	20.02–29.35	24.83	20.1–29.44	34 (29–37)	25.0 ± 4.8	22.4 ± 1.5	20.78 (18.17–24.82)
43 Inioidea + *Lipotes*	22.15	16.93–27.30	22.09	16.96–27.38		21.5 ± 4.6		20.70 (17.18–24.01)
44 Inioidea	16.68	11.35–22.51	16.97	11.53–22.31	24 (20–27)	19.9 ± 4.4	18.4	11.95 (11.25–16.88)
45 Delphinoidea	18.66	14.14–23.26	18.53	13.99–23.42	25 (22–28)	19.8 ± 4.0	16.0	14.21 (12.36–17.32)
46 Phocoenidae + Monodontidae	**15.54**	**10.82–20.12**	**15.36**	**10.73–19.92**	20 (17–23)	13.1 ± 4.0	12.8	11.10 (10.51–11.59)
47 Monodontidae	6.28	2.73–10.23	6.22	2.96–9.95				
48 Phocoenidae	8.46	5.60–11.60	7.98	5.16–10.93				
49	6.96	4.51–9.68						
50	3.75	1.65–6.17	3.81	1.78–6.24				
51	4.84	2.59–7.37	5.16	2.75–7.89				
52	3.95	1.80–5.50	4.35	1.89–7.04				
53 Delphinidae	10.08	7.34–12.88	9.72	6.98–12.65	**12 (11–13)**			
54	9.39	6.95–11.98						
55	8.85	6.67–11.24						7.21 (4.61–9.46)
56	7.54	5.26–10.01						
57 Lissodelphininae	4.32	2.74–5.98	4.25	2.78–5.97				
58 *Lissodelphis*	1.36	0.44–2.52	1.33	0.47–2.31				
59	3.20	2.06–4.44	3.12	2.04–4.28	5 (4–6)			
60	1.92	0.80–3.16	1.83	0.77–2.98				
61	2.72	1.69–3.80	2.72	1.78–3.30				
62	1.87	0.96–2.85						
63	0.85	0.28–1.51	0.82	0.27–1.43				
64	2.36	1.35–3.41						
65	1.11	0.44–1.89	1.22	0.47–2.08				
66	8.36	6.23–10.70						
67	8.00	5.85–10.37						
68	7.35	5.11–9.74						
69 Globicephalinae	4.20	2.50–6.02	4.70	2.78–6.76				
70	3.69	2.14–5.42						
71	2.87	1.50–4.38	2.91	1.63–4.37				
72	2.12	0.97–3.46	2.24	1.09–3.53				
73 *Globicephala*	0.66	0.16–1.26	0.67	0.16–1.28				
74	6.98	4.88–9.17						
75 *Sotalia*	1.99	0.63–3.67	1.90	0.69–3.44				
76 Delphininae	3.84	2.57–5.23	4.12	2.58–5.68				2.35 (1.77–3.53)
77	3.51	2.38–4.79						
78	3.14	2.03–4.47						
79	2.75	1.60–3.97						
80	2.18	1.46–3.02	1.80	1.12–2.52				1.45 (1.04–2.26)
81 *Delphinus*	1.31	0.64–2.04	1.05	0.57–1.62				
82	0.78	0.27–1.34	0.66	0.27–1.10				
83	1.94	1.24–2.67						
84	1.89	1.13–2.50						
85	1.10	0.44–1.96	1.03	0.38–1.66				
86 *Tursiops*	1.59	0.93–2.29						

steps, CI = 0.470, RI = 0.538), were broadly consistent with the model-based analyses that more directly accounted for rate heterogeneity and base frequency differences among molecular data partitions (Figs. S2 and S3). Overall, only three clades resolved in the 50% majority-rule bootstrap tree for the molecular MP analysis (Fig. S2) conflicted with clades supported by the Bayesian consensus tree (Fig. 2). These groups had low MP BS scores of 52%, 52%, and 56%. Other conflicts in the molecular MP analysis were characterized by BS scores ⩽50% and Bremer supports of only 1 or 2. When morphological data (Geisler and Sanders, 2003) were included in MP analysis, more groups were resolved in the 50% majority-rule bootstrap consensus (Fig. S3). Odontoceti was recovered (MP BS = 60%) in contrast to the MP result for molecular data alone (Fig. S2), Platanistidae was positioned in a weakly supported alliance with Ziphiidae (MP BS = 53%), and *Caperea* grouped with Balaenopteroidea (MP BS = 59%). Only four clades with MP BS >50% conflicted with the Bayesian consensus (Fig. 2); these groups were characterized by only marginal support (MP BS = 53–59%). Other conflicting clades in the MP analysis that included morphology had BS scores <50% and Bremer support scores of only 1 or 2.

3.2. Divergence date estimation

Mean divergence date estimates did not vary substantially whether the date of 50 or 54 Ma was used as a minimum age for the most recent common ancestor (MRCA) of Hippopotamidae and Cetacea; results using the 54 Ma date are presented here. For the analysis using the combined Bayesian tree and separate models for each codon position, BEAST produced a maximum clade credibility tree (Fig. 3) with a mean negative log-likelihood ($-\ln L$) of 17,825.64 (17,805.62–17,845.79, 95% highest posterior density [HPD]), after discarding the first 3 million generations as burn-in. The mean coefficient of variation was 0.51 (0.40–0.62 HPD), indicating significant rate heterogeneity in this data set. Covariance of ancestor and descendant branches was not significantly different from zero (mean 0.04, −0.11–0.19 HPD), implying little if any autocorrelation of rates among lineages. In comparison, the analysis with minimal topological constraints produced a maximum clade credibility tree with mean $-\ln L$ = 17,780.34 (17,758.59–17,802.42 HPD), mean coefficient of variation = 0.47 (0.37–0.57 HPD) and covariance also not significantly different from zero (0.01, −0.11–0.21 HPD).

All estimated dates for nodes in Fig. 3 are shown in Table 2, compared with estimates from the analysis based primarily on the *MT-CYTB* topology, as well as dates from other studies (Cassens et al., 2000; Nikaido et al., 2001b; Árnason et al., 2004; Sasaki et al., 2006; Xiong et al., 2009). There were limited differences between divergence times in the two analyses executed here. Ages at all concordant nodes except for two (*Mesoplodon* = node 29; *Eubalaena* = node 4) are contained within the HPDs of both analyses (Table 2). Due to this similarity, divergence dates discussed below refer to those derived from the analysis with topological constraints from the Bayesian supermatrix analysis.

The split between Mysticeti and Odontoceti was estimated to have occurred in the Late Eocene, shortly before the appearance of the first documented fossil mysticete *Llanocetus* (∼34.2 Ma), a minimum prior used in this analysis (Fig. 3). The mean age estimates of further branching events within Cetacea ranged from 34.69 to 0.52 Ma (Table 2). Within ∼5 Ma of the MRCA of all extant cetaceans, successive branching events had given birth to Mysticeti and four extant odontocete lineages (Physeteroidea, Platanistidae, Ziphiidae, Delphinida). Diversification of Mysticeti in the Miocene produced the *Caperea* lineage and the balaenopteroid radiation (Fig. 3). All lineages leading to modern odontocete families were present by the Middle Miocene. Branching events within the speciose clade Ziphiidae took place over a wide time frame from the Early Miocene to the Pliocene. In contrast, Delphinidae underwent a relatively recent radiation of ∼36 species beginning at 10.08 Ma, with most divergences in the Plio-Pleistocene (Fig. 3).

4. Discussion

4.1. Comparison of supermatrix with other methods

A comprehensive and well-supported phylogenetic hypothesis is the goal of modern systematics and paramount to any comparative biological study (Harvey and Pagel, 1991). Here we presented a phylogenetic hypothesis for Cetacea that includes most, if not all, molecular markers known for more than four species as of August 2008. By merging diverse data from multiple studies, we were able to simultaneously assess the relative weight of evidence for conflicting relationships in different data sets by concatenated analyses of all the data sets combined. Bayesian and ML methods yielded a well-resolved, consistently supported estimate of species relationships (Fig. 2), even in the presence of substantial missing data (Fig. 1). Parsimony analyses of the molecular database were highly congruent with the model-based analyses, and conflicts generally were restricted to very weakly supported nodes with Bremer supports of 1 or 2 and MP BS scores of <60% (Figs. S2 and S3). Therefore, for simplicity, we refer primarily to the Bayesian topology, which is well-corroborated by ML analyses (Fig. 2), in interpretations of our results below.

The *MT-CYTB* topology of Agnarsson and May-Collado (2008) is the only DNA sequence-based study for which there was comparable taxonomic coverage to our study. Relative to the *MT-CYTB* tree, the supermatrix topology (Fig. 2) showed more resolution (in the 50% majority-rule consensus tree), higher support scores for most shared nodes, and also resolved several clades with strong support scores that conflicted with the *MT-CYTB* analysis. Thirty-two out of 85 nodes in the *MT-CYTB* tree (37.6%) disagreed with our Bayesian consensus tree (77 nodes total, Fig. 2); all but three conflicting nodes were concentrated in the delphinid, ziphiid, and balaenopteroid radiations. Most discordant relationships in the *MT-CYTB* tree were not strongly supported (only one node had PP ⩾ 0.95); by contrast, 28 nodes in our Bayesian consensus tree conflicted with the *MT-CYTB* analysis. Fourteen were well-supported by Bayesian PP ⩾ 0.95 and nine nodes by ML BS ⩾ 70%. Of the clades in agreement between *MT-CYTB* (Agnarsson and May-Collado, 2008) and the supermatrix topology (Fig. 2), 14 had higher PP scores in the supermatrix analysis (11 of these increased >25%), three nodes had lower PP scores in the supermatrix, and the remainder stayed the same at 0.99 or 1.0.

Our study also diverged in topology from the MRP supertree of Price et al. (2005), perhaps due to the inclusion of more data and the interaction of characters from separate data sets in our supermatrix searches. Despite the inclusion of a comparable number of species, resolution in the supertree analysis was much lower (40 versus 77 nodes), with limited resolution within Delphinidae and no resolution within Ziphiidae. Eight out of 40 nodes (20%) in the MRP supertree disagreed with our Bayesian consensus tree (Fig. 2); all but three nodes concerned relationships within Delphinidae.

4.2. Phylogeny of Mysticeti

Mysticeti represented the most-resolved and well-supported region of all trees (Fig. 2; S1–S3), possibly due to less missing data relative to other major clades (Fig. 1). In agreement with previous molecular hypotheses, there was a basal split between Balaenidae and the remaining mysticetes, and Neobalaenidae clustered with Balaenopteroidea (Árnason and Gullberg, 1994; Rychel et al.,

2004; Sasaki et al., 2005, 2006; Nikaido et al., 2006; Deméré et al., 2008). In contrast to our study and most published molecular analyses, morphological evidence strongly places Neobalaenidae in a clade with Balaenidae, in part due to characters related to skim filter-feeding (Bouetel and de Muizon, 2006; Bisconti, 2007; Steeman, 2007). Relationships within Balaenopteroidea have been more difficult to resolve using molecular data, but there is a consensus that four major lineages exist: *Eschrichtius*, a *Balaenoptera borealis* group, *B. acutorostrata* + *B. bonaerensis*, and *Megaptera* + *B. physalus* (Rychel et al., 2004; Sasaki et al., 2005, 2006; Nikaido et al., 2007; Deméré et al., 2008). Several researchers have suggested that the base of Balaenopteroidea represents a rapid radiation that has been obscured by incomplete lineage sorting (Sasaki et al., 2005; Nikaido et al., 2006), and some species within this group have been known to hybridize in nature (Árnason et al., 1991b). Here, all analyses firmly placed the *B. acutorostrata* + *B. bonaerensis* group as basal. Bayesian analyses and ML analysis with two partitions allied *Eschrichtius* with *Megaptera* + *B. physalus*. This controversial grouping conflicted with MP topologies (Figs. S2 and S3) and the ML analysis with 119 partitions (Fig. S1), which instead placed *E. robustus* sister to all balaenopteroids except minke whales (*B. acutorostrata* + *B. bonaerensis*). The possible paraphyly of Balaenopteridae, due to the nesting of *Eschrichtius* (Eschrichtiidae) among balaenopterids, also has been noted in previous molecular studies (Árnason; Gullberg, 1994; Rychel et al., 2004 and Sasaki et al., 2005; Hatch et al., 2006), but balaenopterids are defined by multiple, unique, morphological characters associated with their complex method of engulfment filter-feeding (Deméré et al., 2008). Our supermatrix results uniformly imply that *Eschrichtius* lost this specialized suite of anatomical traits and acquired new features utilized in suction assisted filter-feeding. Some previous molecular analyses, however, did not conflict with a monophyletic Balaenopteridae (Rychel et al., 2004; Sasaki et al., 2005), and the inclusion of morphology and fossils with molecular data repositioned *Eschrichtius* outside balaenopterids in a recently published supermatrix analysis of mysticetes (Deméré et al., 2008).

4.3. Phylogeny of Odontoceti

Contrary to some previous studies using one or few loci, combined ML and Bayesian molecular analyses presented robust support for odontocete monophyly, despite the merging of data sources that exhibited strongly conflicting signals when analyzed separately. For example, mt 12S rDNA, mt 16S rDNA, and nu *IRBP* supported odontocete paraphyly (Milinkovitch et al., 1993, 1994, 1996; Smith et al., 1996), and nu *SRY*, concatenations of nu genes, and SINE transposons supported monophyly (Gatesy, 1998; Nikaido et al., 2007; Nishida et al., 2007; Deméré et al., 2008). The combination of all data in model-based analyses provided consistent support for short internodes at the base of Cetacea and strong placement for the cetacean root (Fig. 2). The combined molecular signal for odontocete monophyly was compatible with overwhelming morphological evidence, such as loss of the olfactory nerve, maxilla overlapping supraorbital process of the frontal, single external narial opening, hypertrophied melon, enlarged maxillofacial muscles, and facial asymmetry, among others (Heyning, 1997; Messenger and McGuire, 1998; Geisler and Sanders, 2003). The latter three features are involved in the production of sound in underwater echolocation, a unique sensory innovation uniting all living and fossil odontocetes (Fordyce and de Muizon, 2001).

Physeteroidea (sperm whales) has been allied to Mysticeti in some molecular analyses (Milinkovitch et al., 1993, 1994, 1996; Smith et al., 1996), and to Ziphiidae in selected morphological analyses (de Muizon, 1988; Geisler and Sanders, 2003). Our Bayesian and ML trees agreed with research that placed Physeteroidea as sister to all other extant odontocetes (Heyning, 1997; Messenger and McGuire, 1998; Cassens et al., 2000; Hamilton et al., 2001; Nishida et al., 2003, 2007; Árnason et al., 2004; Yan et al., 2005; Agnarsson and May-Collado, 2008; Xiong et al., 2009). All extant odontocetes except Physeteroidea are united by the restructuring of their nasal complex, including nasal passage confluence and the presence of blowhole ligaments, premaxillary sacs, and modified proximal sacs (Heyning, 1997).

In ML and Bayesian analyses, Ziphiidae (beaked whales) and Platanistidae (South Asian river dolphins) were positioned as successive sister taxa to Delphinida (Fig. 2). Mt genome and morphological data have placed Platanistidae as sister to Delphinida, but with low support (Heyning, 1997; Messenger and McGuire, 1998; Árnason et al., 2004; Yan et al., 2005; but see Xiong et al., 2009). The inclusion here of nu sequence and SINE insertion data corroborated previous analyses using SINE transposons (Nikaido et al., 2001b) that grouped Ziphiidae with Delphinida to the exclusion of Platanistidae and Physeteroidea. In fact, all odontocete clades previously supported by transposon insertion events (Nikaido et al., 2001a,b, 2007) were replicated in our ML and Bayesian topologies (Fig. 2). Importantly, our ML supermatrix analyses (Fig. S1) included no transposon characters, but the ML results based on DNA sequences alone corroborated the transposon insertion data.

As there were few molecular data available for a large number of ziphiids at the time of compilation (Fig. 1), resolution within this family, especially among *Mesoplodon* species, was incomplete and inconsistent (Fig. 2). Our Bayesian and ML analyses supported *Berardius* as the sister taxon to all remaining ziphiids, conflicting with the *MT-CYTB* tree of Agnarsson and May-Collado (2008), which placed *Tasmacetus* as basal. Morphological analysis of fossil and extant ziphiids corroborated that the *Berardius* lineage was first to diverge from other beaked whales (Bianucci et al., 2007). The supermatrix analysis nested *Tasmacetus shepherdi* [the only ziphiid with multiple functional teeth in both jaws and in both sexes (Mead, 2002)], among species that have severely reduced dentitions. Most male ziphiids have one or two pairs of teeth on the dentary used in aggressive behavior, but females lack functional teeth (Mead, 2002). Based on the supermatrix results, it is unclear whether dentition was reduced once at the base of the ziphiid clade and regained in *Tasmacetus shepherdi*, or was reduced separately in *Berardius* and in other ziphiids. A very recent phylogenetic investigation of *Mesoplodon* ziphiids using nu introns was completed after compilation and execution of analyses presented here (Dalebout et al., 2008). This study (Dalebout et al., 2008) largely agreed with that of our Bayesian topology (Fig. 2), providing solid support for a six species *M. hectori* clade, a three species *M. layardii* clade (Dalebout et al., 2008 did not include *M.traversii*), and *M. bidens* as the sister group to the remaining members of the genus.

Higher-level relationships within Delphinida varied in congruence with published phylogenetic hypotheses. Inioidea (Iniidae [Amazon river dolphin] + Pontoporiidae [La Plata dolphin]) and Delphinoidea (Monodontidae [beluga and narwhal] + Phocoenidae [porpoises] + Delphinidae [oceanic dolphins]) have been supported by many systematic studies of Cetacea (Messenger and McGuire, 1998; Cassens et al., 2000; Hamilton et al., 2001; Nikaido et al., 2001b; Árnason et al., 2004). *Lipotes vexillifer* (Chinese river dolphin), the sole member of Lipotidae, has been more difficult to place, possibly due to rapid splitting events at the base of Delphinida. Some molecular analyses positioned Lipotidae as the sister taxon to all other delphinidans (Cassens et al., 2000; Hamilton et al., 2001). In contrast, analysis of mt genomes supported Lipotidae + Inioidea (Yan et al., 2005; Xiong et al., 2009), a result weakly supported here (Fig. 2). There was no compelling supermatrix support for the monophyly of all river dolphins (Inioidea +

Lipotidae + Platanistidae), which has been supported by cladistic analysis of morphological characters from living and extinct species (Geisler and Sanders, 2003). Several morphological and molecular analyses have grouped Phocoenidae with Delphinidae (Milinkovitch et al., 1993; Heyning, 1997; Hamilton et al., 2001; Geisler and Sanders, 2003), while other molecular analyses have supported the monophyly of Phocoenidae + Monodontidae (Waddell et al., 2000; Cassens et al., 2000; Hamilton et al., 2001; Árnason et al., 2004; Agnarsson and May-Collado, 2008; Xiong et al., 2009). Our compilation of the conflicting evidence robustly placed Monodontidae and Phocoenidae in a clade (Fig. 2). Relationships within Phocoenidae conformed to previous analyses of mt DNA (Rosel et al., 1995; Agnarsson and May-Collado, 2008), likely due to the paucity of nu DNA data for 50% of phocoenid species (Fig. 1).

With the combination of evidence from mt and nu sources, our study presented the most-resolved and well-supported phylogenetic hypothesis for delphinid cetaceans to date. Many molecular analyses were unable to resolve or provide solid support for the branching events at the base of Delphinidae (LeDuc et al., 1999; May-Collado and Agnarsson, 2006; Agnarsson and May-Collado, 2008; Caballero et al., 2008). Here *Orcinus orca* (killer whale), the largest delphinid, and *Leucopleurus acutus* (Atlantic white-sided dolphin) were basal delphinids in all model-based analyses (Fig. 2). A recent analysis based on olfactory receptor genes also supported *Orcinus* as an early branching delphinid lineage, but taxon sampling was much more limited in that study (McGowen et al., 2008). We found neither support for a close relationship between *Orcinus* and *Orcaella*, as in analyses of *MT-CYTB* (LeDuc et al., 1999; May-Collado and Agnarsson, 2006; Agnarsson and May-Collado, 2008), nor support for the inclusion of *Orcinus* within Globicephalinae as suggested by some morphologists (Fraser and Purves, 1960; Mead, 1975; de Muizon, 1988; Perrin, 1989). However, there was high Bayesian PP for a close relationship between Globicephalinae and Delphininae to the exclusion of Lissodelphininae. Globicephalinae (blunt-headed dolphins), Delphininae (bottle-nose-type dolphins), and Lissodelphininae (piebald dolphins) were each monophyletic as in all previous taxonomically comprehensive molecular studies (LeDuc et al., 1999; May-Collado and Agnarsson, 2006; Agnarsson and May-Collado, 2008; Caballero et al., 2008).

In addition, this study robustly supported the alliance of *Steno* and *Orcaella* with Globicephalinae, which also has been favored by separate analyses of some nu genes (Onbe et al., 2007; Caballero et al., 2008; McGowen et al., 2008), and disagreed with mt DNA evidence placing *Steno* in a clade with delphinines (LeDuc et al., 1999; Agnarsson and May-Collado, 2008; McGowen et al., 2008). We found no support for the monophyly of the former constituent species of *Lagenorhynchus* (*Lagenorhynchus albirostris* + *Leucopleurus acutus* + *Sagmatias* spp.) and followed the provisional recommendation of LeDuc et al., 1999 to refer the appropriate species to new genera. Like Harlin-Cognato and Honeycutt (2006), we were unable to confirm the monophyly of *Sagmatias*, and further studies will be needed to discern the interrelationships of *Sagmatias* and *Cephalorhynchus* species.

In the Bayesian consensus (Fig. 2), we recovered *Sotalia* as sister to Delphininae, and *Sousa* as the most basal delphinine, in agreement with Caballero et al. (2008). In contrast, *Sotalia* was allied with *Steno* in a previous molecular analysis (Agnarsson and May-Collado, 2008), and some morphologists have suggested the alliance of *Sotalia* or both *Sotalia* and *Sousa* with *Steno* (de Muizon, 1988; Mead, 1975; Perrin, 1989). Many other relationships within Delphininae have been difficult to resolve due to fairly recent origins and short internodes; the genera *Delphinus*, *Lagenodelphis*, *Stenella*, and *Tursiops* have been hypothesized to represent a closely related species flock (Kingston and Rosel, 2004). We found robust support for the monophyly of *Delphinus*, and contrary to other molecular analyses (LeDuc et al., 1999; Agnarsson and May-Collado, 2008; Xiong et al., 2009), we discovered some emergent but weak support for a monophyletic *Tursiops*. Due to the polyphyly of *Stenella* in our study, we agree with the proposal of LeDuc et al. (1999) to synonymize *Tursiops*, *Stenella*, and *Lagenodelphis* with *Delphinus*; however, more study should be focused on this complex species group to corroborate these findings, as sampling of nu loci across Delphinidae is presently inconsistent (Fig. 1).

4.4. Timing and tempo of cetacean evolution

According to the fossil record, crown cetaceans (Mysticeti + Odontoceti) probably arose near the Eocene–Oligocene boundary (~34 Ma) (Fordyce, 2003). Their appearance and subsequent diversification has been hypothesized to coincide with decreased ocean temperatures and increased upwelling of nutrient resources as a result of rapid Antarctic glaciation and the opening of the Southern Ocean (Fordyce, 2003; Berger, 2007). Here, estimates derived from calibration using the earliest known crown cetacean (*Llanocetus*, ~34.2) create a Late Eocene window (34.24–40.14 Ma) for the origination of the crown cetacean MRCA (Fig. 3; Table 2). Although slightly earlier, this time frame overlapped in range with results from previous molecular studies (Nikaido et al., 2001b; Sasaki et al., 2005, 2006; Xiong et al., 2009), but conflicted with the earlier estimates of Cassens et al. (2000) (Table 2). Consistent with our results of a Late Eocene origin for crown Cetacea, Fordyce (2009) noted the presence of unnamed Late Eocene odontocetes from Washington State, USA, which would push the known fossil record of Odontoceti into the Eocene. In addition, a recent morphological analysis suggested that *Llanocetus*, although a stem mysticete, might not be the most basal fossil, implying mysticete diversification prior to 34.2 Ma (Steeman, 2007).

By our estimate, crown mysticetes originated in the Early Oligocene (Fig. 3), shortly before the appearance of the first balaenid ~28 Ma (Fordyce, 2002). This is within ~1 Ma of some previous molecular estimates for this node (Sasaki et al., 2005, 2006), but contrasted with both the slightly older date of Cassens et al. (2000) and the younger date of Árnason et al. (2004) (Table 2). Stem mysticetes such as toothed aetiocetids also existed during this time, but were extinct by the end of the Oligocene (Fordyce and de Muizon, 2001).

Within Balaenidae, our estimates for the split between *Balaena* and *Eubalaena* and between *Eubalaena australis* and other *Eubalaena* spp. were much younger than those of Sasaki et al. (2005, 2006) (Table 2). Few balaenid fossils have been found in deposits that predate the Pliocene, and many Pliocene balaenids can be assigned to the *Balaena* or *Eubalaena* lineages (Bisconti, 2005). The phylogenetic analysis of Bisconti (2005) weakly placed the Early Miocene fossil *Morenocetus parvus* (~20–22 Ma) as more closely related to *Eubalaena* than *Balaena*, in direct conflict with the much younger age for the divergence between *Balaena* and *Eubalaena* presented here. However, others regarded *Morenocetus* to be a stem balaenid due to primitive characteristics (Sasaki et al., 2005). Our mean estimate for the split between Neobalaenidae and Balaenopteroidea (22.59 Ma) agrees with Sasaki et al. (2005), but is slightly older than Árnason et al. (2004) (Table 2); currently, no fossil record exists for Neobalaenidae.

Dates for divergences within Balaenopteroidea were much younger than those of Sasaki et al. (2005, 2006), which suggests an Early Miocene origin for the group (~19 Ma; Table 2). Estimates presented here were more in line with the fossil record, where the appearance of modern balaenopteroids around 10–12 Ma was coeval with Middle Miocene cooling events (Whitmore, 1994). The ages of *Megaptera* and *Eschrichtius,* two morphologically distinct genera nested within *Balaenoptera,* were consistent with recent

phylogenetic analyses that placed Pliocene taxa with these lineages (Bisconti, 2007; Deméré et al., 2008). Some phylogenetic analyses also positioned various members of the fossil grade "Cetotheriidae," including *Cetotherium rathkii* and *Herpetocetus* spp. (Middle Miocene to Pliocene age), as more closely related to *Eschrichtius*, but excluded *Eschrichtius* from Balaenopteridae (Bisconti, 2007; Steeman, 2007). This presents a slightly older origin for the *Eschrichtius* lineage than our molecular estimates. It is unclear where other members of "Cetotheriidae" fit into the overall mysticete tree, as many recent phylogenetic analyses variously placed them as stem mysticetes (Geisler and Sanders, 2003; Bouetel and de Muizon, 2006; Deméré et al., 2008) or both stem and crown balaenopteroids (Bisconti, 2007; Steeman, 2007).

According to our estimates, divergence of the odontocete lineages Physeteroidea, Platanistidae, Ziphiidae, and Delphinida occurred primarily in the Early Oligocene (Fig. 3). The upper bounds of estimates presented here are slightly older than other studies, but largely overlapped in range at shared nodes (Nikaido et al. 2001b; Árnason et al., 2004; Xiong et al., 2009; Table 2). Our clock results disagreed strongly with Cassens et al. (2000), which positioned major odontocete splits deep in the Eocene (Table 2). Due to substantial global erosion of Early Oligocene marine deposits, cetacean fossils from this time period are rare. There is little evidence for the presence of modern odontocete clades in the Early Oligocene, although very primitive stem odontocetes existed during this time (Fordyce, 2003). Early representatives of Physeteroidea (*Ferecetotherium*), Platanistoidea (*Waipatia* and others), and possibly Delphinida (the fossil grade "Kentriodontidae") were present in the Late Oligocene, demonstrating that these clades were established by this time (Barnes et al., 1985; Fordyce, 2003, 2009). However, the morphological analysis of Geisler and Sanders (2003) grouped most crown odontocetes in a clade exclusive of many Oligocene fossil taxa, implying a much more recent age for crown odontocete diversification than that estimated here.

The split between Physeteridae and Kogiidae represents the age estimate with the most uncertainty (Fig. 3). Our mean age, 24.21 Ma, agrees well with Árnason et al., 2004, but is much younger than Cassens et al. (2000) (Table 2). Fossils attributed to either family by phylogenetic analysis were present by the Middle Miocene (~16–11.6 Ma; Bianucci and Landini, 2006), but it is not clear whether the Late Oligocene *Ferecetotherium* is a stem or crown physeteroid (Fordyce, 2009). The latter position would pull the known age of Physeteridae back to ~23 Ma, a date that is close to molecular estimates presented here.

Our age estimate for the ziphiid MRCA is consistent with another analysis that included *Berardius* and *Hyperoodon* (Árnason et al., 2004; Table 2). After the divergence of *Berardius*, intergeneric splits within Ziphiidae generally occurred in the Early to Middle Miocene according to our analysis (Fig. 3). Previous paleontological work has shown that multiple fossils of Middle Miocene to Early Pliocene age are nested throughout the extant ziphiid tree, corroborating our estimates that lineages leading to modern genera were present at this time (Lambert, 2005; Lambert and Louwye, 2006; Bianucci et al., 2007). Divergence times within the genus *Mesoplodon* were slightly younger than a recent study using Bayesian methods and an estimate of *Mesoplodon* root age of 15–25 Ma (Dalebout et al., 2008). The gradual cladogenesis of extant ziphiid lineages took place over a wide time frame beginning in the Early Miocene (Fig. 3).

As indicated by our estimates, the river dolphin clade that includes Iniidae, Lipotidae, and Pontoporiidae diverged from delphinoids in the Late Oligocene, and origination of the three families took place in the Early to Middle Miocene (Fig. 3). This conflicts with dates from Cassens et al. (2000) but is consistent with some other molecular dating analyses (Nikaido et al., 2001b; Árnason et al., 2004; Xiong et al., 2009). However, our mean date for the Pontoporiidae–Iniidae split is slightly younger than Nikaido et al. (2001b) and Árnason et al. (2004), and older than Xiong et al. (2009) (Table 2). Early fossils that can be attributed to these families first appeared in the Middle to Late Miocene (reviewed in Hamilton et al., 2001), implying extensive ghost lineages.

Delphinoidea originated in the Early Miocene, and eventually gave rise to phocoenids, monodontids, and delphinids (Fig. 3); our estimate was within the range of some earlier assessments (Nikaido et al., 2001b; Árnason et al., 2004; Xiong et al., 2009), and younger than Cassens et al. (2000) (Table 2). Our date was consistent with the age of fossils attributed to the extinct "Kentriodontidae", a grade usually included within Delphinoidea or the more inclusive Delphinida. "Kentriodontids" range in age from Late Oligocene to Late Miocene and are in need of further study to establish definitive relationships with modern delphinoid clades (Fordyce and de Muizon, 2001). Within Delphinoidea, the split between Phocoenidae and Monodontidae occurred much earlier than the oldest representative of either family, the Late Miocene phocoenid *Salumiphocaena stocktoni*; our estimate was also slightly earlier than other analyses (Nikaido et al., 2001b; Árnason et al., 2004; Xiong et al., 2009). Speciation events within Phocoenidae were much older than those predicted by Rosel et al. (1995).

The origin of modern delphinids (oceanic dolphins), the largest family of extant cetaceans, occurred in the Late Miocene (~10 Ma; Fig. 3). This was followed by a succession of speciation events during the next ~3 Ma that established all recognized modern subfamilies (Fig. 3). Subsequent diversification in Lissodelphininae (4.32 Ma), Globicephalinae (4.20 Ma), and Delphininae (3.84 Ma) began almost simultaneously during the Early Pliocene and continued through the Pleistocene (Fig. 3; Table 2). Intriguingly, contrary to extant diversity, the fossil record of Delphinidae is limited, but dolphin fossils, some from modern lineages, are present in Pliocene sediments (Bianucci, 1996; Fordyce et al., 2002). Paleontologists have noted that the earliest delphinids are approximately 10–12 Ma in age, a date that agrees with our estimates; however, these specimens remain undescribed (Fordyce and de Muizon, 2001; Barnes, 2002).

The rapidity of speciation in delphinid dolphins is remarkable considering their relatively large size, mobility, long generation time, sympatric association, and the near global distribution of some species (LeDuc, 2002). Delphinid species vary widely in diet, head and feeding morphology, coloration, body size, social structure, and habitat (LeDuc, 2002). Some have proposed that their explosive rise may have driven the competitive exclusion of other delphinoids and extinct groups, due to the evolution of more refined echolocation abilities (Fordyce and de Muizon, 2001; LeDuc, 2002). Paleontologists also have noted that pulses of cooling in the Middle-Late Miocene and Pliocene could have played a role in diversification (Whitmore, 1994; Fordyce and de Muizon, 2001); both of these events correspond with two distinct pulses of speciation within Delphinidae (Fig. 3). Further investigation of correlations among morphological and ecological features, as well as diversification patterns will be needed to corroborate or refute these hypotheses.

5. Conclusion

We produced a comprehensive, time-calibrated phylogenetic hypothesis for extant Cetacea from the integration of multiple sources of molecular data, including 45 nu loci, mt genomes, and transposon insertion events. ML and Bayesian analyses of the combined data (Fig. 2) yielded greater resolution and support than previously published single locus studies and a recent MRP supertree analysis with similar taxonomic coverage. Divergence date estimates suggest a basal diversification of Cetacea at the

Eocene–Oligocene boundary, the antiquity of river dolphin clades, and the recent explosive radiation of delphinid dolphins beginning in the Late Miocene. Our supermatrix tree should prove to be a valuable resource for future comparative studies of cetacean evolution.

Acknowledgments

Funding was provided by NSF grants to JG (DEB-0743724, DEB-0640313, EF-0629860, DEB-0213171). We thank Carl Buell for providing illustrations of cetacean species, as well as P. Morin, K. Robertson, and R. LeDuc (Southwest Fisheries Science Center, NOAA, La Jolla, CA, USA), the South Australian Museum, Northeast Fisheries Science Center Stranding Network, Smithsonian Institution Division of Mammals, Gill Braulik (World Wildlife Fund), Mote Marine Laboratory (Sarasota, FL, USA), G. Amato (New York Zoological Society, Bronx, NY, USA), P. Palsbøll (Stockholm University, Sweden), and M. Milinkovitch (Free University of Brussels, Belgium) for providing DNA samples. N. Ayoub, J. Starrett, J. Geisler, L. May-Collado, I. Agnarsson, and an anonymous reviewer provided helpful comments on previous drafts of this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2009.08.018.

References

Agnarsson, I., May-Collado, L.J., 2008. The phylogeny of Cetartiodactyla: the importance of dense taxon sampling, missing data, and the remarkable promise of cytochrome b to provide reliable species-level phylogenies. Mol. Phylogenet. Evol. 48, 964–985.

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

Árnason, U., Gullberg, A., 1993. Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature. J. Mol. Evol. 37, 312–322.

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

Árnason, U., Gullberg, A., 1996. Cytochrome *b* nucleotide sequences and the identification of five primary lineages of extant cetaceans. Mol. Biol. Evol. 13, 407–417.

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

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

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

Árnason, U., Spillaert, R., Palsdottir, A., Árnason, A., 1991b. Molecular identification of hybrids between the two largest whale species, the blue whale (*Balaenoptera musculus*) and the fin whale (*B. physalus*). Hereditas 115, 183–189.

Bajpai, S., Gingerich, P.D., 1998. A new Eocene archaeocete (Mammalia, Cetacea) from India and the time of origin of whales. Proc. Natl. Acad. Sci. USA 95, 15464–15468.

Barnes, L.G., 2002. Delphinoids, evolution of the modern families. In: Perrin, W., Würsig, B., Thewissen, J.G.M. (Eds.), Encyclopedia of Marine Mammals. Academic Press, San Diego, pp. 314–316.

Barnes, L.G., Domning, D.P., Ray, C.E., 1985. Status of studies on fossil marine mammals. Mar. Mammal Sci. 1, 15–53.

Barrett, M., Donoghue, M.J., Sober, E., 1991. Against consensus. Syst. Zool. 40, 486–493.

Beasley, I., Robertson, K.M., Arnold, P., 2005. Description of a new dolphin, the Australian snubfin dolphin *Orcaella heinsohni* sp. N. (Cetacea, Delphinidae). Mar. Mammal Sci. 21, 365–400.

Berger, W.H., 2007. Cenozoic cooling, Antarctic nutrient pump, and the evolution of whales. Deep Sea Res. II 54, 2399–2421.

Bianucci, G., 1996. The Odontoceti (Mammalia, Cetacea) from Italian Pliocene. Systematics and phylogenesis of Delphinidae. Palaeontogr. Italia 83, 73–167.

Bianucci, G., Landini, W., 2006. Killer sperm whale: a new basal physeteroid (Mammalia, Cetacea) from the Late Miocene of Italy. Zool. J. Linn. Soc. 148, 103–131.

Bianucci, G., Lambert, O., Post, K., 2007. A high diversity in fossil beaked whales (Mammalia, Odontoceti, Ziphiidae) recovered by trawling from the sea floor off South Africa. Geodiversitas 29, 561–618.

Bininda-Emonds, O.R.P., 2004. The evolution of supertrees. Trends Ecol. Evol. 19, 315–322.

Bisconti, M., 2005. Skull morphology and phylogenetic relationships of a new diminutive balaenid from the lower Pliocene of Belgium. Palaeontology 48, 793–816.

Bisconti, M., 2007. A new basal balaenopterid whale from the Pliocene of Northern Italy. Palaeontology 50, 1103–1122.

Bouetel, V., de Muzion, C., 2006. The anatomy and relationships of *Piscobalaena nana* (Cetacea, Mysticeti), a Cetotheriidae s.s. from the early Pliocene of Peru. Geodiversitas 28, 319–395.

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

Caballero, S., Jackson, J., Mignucci-Giannoni, A.A., Barrios-Garrido, H., Beltrán-Pedreros, S., Montiel-Villalobos, M.G., Robertson, K.M., Baker, C.S., 2008. Molecular systematics of South American dolphins *Sotalia*: sister taxa determination and phylogenetic relationships, with insights into a multilocus phylogeny of the Delphinidae. Mol. Phylogenet. Evol. 46, 252–268.

Caballero, S., Trujillo, F., Vianna, J.A., Barrios-Garrido, H., Montiel, M.G., Beltrán-Pedreros, S., Marmontel, M., Santos, M.C., Rossi-Santos, M., Santos, F.R., Baker, C.S., 2007. Taxonomic status of the genus *Sotalia*: species level ranking for "tucuxi" (*Sotalia fluviatilis*) and "costero" (*Sotalia guianensis*) dolphins. Mar. Mammal Sci. 23, 358–386.

Cassens, I., Vicario, S., Waddell, V.G., Balchowsky, H., Van Belle, D., Ding, W., Fan, C., Lal Mohan, R.S., Simões-Lopes, P.C., Bastida, R., Meyer, A., Stanhope, M.J., Milinkovitch, M.C., 2000. Independent adaptation to riverine habitats allowed survival of ancient cetacean lineages. Proc. Natl. Acad. Sci. USA 97, 11343–11347.

Connor, R.C., 2007. Dolphin social intelligence: complex alliance relationships in bottlenose dolphins and a consideration of selective environments for extreme brain size evolution in mammals. Phil. Trans. R. Soc. B 362, 587–602.

Dalebout, M.L., Baker, C.S., Anderson, R.C., Best, P.B., Cockcroft, V.G., Hinsz, H.L., Peddemors, V., Pitman, R.L., 2003. Appearance, distribution, and genetic distinctiveness of Longman's beaked whale, *Indopacetus pacificus*. Mar. Mammal Sci. 19, 421–461.

Dalebout, M.L., Baker, C.S., Mead, J.G., Cockcroft, V.G., Yamada, T.K., 2004. A comprehensive and validated molecular taxonomy of beaked whales, Family Ziphiidae. J. Hered. 95, 459–473.

Dalebout, M.L., Mead, J.G., Baker, C.S., Baker, A.N., van Helden, A.L., 2002. A new species of beaked whale *Mesoplodon perrini* sp. N. (Cetacea: Ziphiidae) discovered through phylogenetic analyses of mitochondrial DNA sequences. Mar. Mammal Sci. 18, 577–608.

Dalebout, M.L., Steel, D., Baker, C.S., 2008. Phylogeny of the beaked whale genus *Mesoplodon* (Ziphiidae: Cetacea) revealed by nuclear introns: implications for the evolution of male tusks. Syst. Biol. 57, 857–875.

Deméré, T.A., McGowen, M.R., Berta, A., Gatesy, J., 2008. Morphological and molecular evidence for a stepwise evolutionary transition from teeth to baleen in mysticete whales. Syst. Biol. 57, 15–37.

de Muizon, C., 1988. Les relations phylogenetiques des Delphinida (Cetacea, Mammalia). Ann. Paleontol. 74, 159–227.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e5.

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

Fish, F.E., 1998. Comparative kinematics and hydrodynamics of odontocete cetaceans: morphological and ecological correlates with swimming performance. J. Exp. Biol. 201, 2867–2877.

Fordyce, R.E., 2002. Oligocene origins of skim-feeding right whales: a small archaic balaenid from New Zealand. J. Vert. Paleont. 22 (Suppl. To 3), 54A.

Fordyce, R.E., 2003. Cetacean evolution and Eocene–Oligocene oceans revisited. In: Prothero, D.R., Ivany, L.C., Nesbitt, E.A. (Eds.), From Greenhouse to Icehouse: the Marine Eocene–Oligocene Transition. Columbia University Press, New York, pp. 154–170.

Fordyce, R.E., 2009. Cetacean fossil record. In: Perrin, W., Würsig, B., Thewissen, J.G.M. (Eds.), Encyclopedia of Marine Mammals, second ed. Academic Press, San Diego, pp. 201–207.

Fordyce, R.E., de Muizon, C., 2001. Evolutionary history of cetaceans: a review. In: Mazin, J.M., de Buffrenil, V. (Eds.), Secondary Adaptation of Tetrapods to Life in Water. Verlag Dr. Friedrich Pfeil, Munich, pp. 169–233.

Fordyce, R.E., Quilty, P.G., Daniels, J., 2002. *Australodelphis mirus*, a bizarre new toothless ziphiid-like fossil dolphin (Cetacea: Delphinidae) from the Pliocene of Vestfold Hills, east Antarctica. Antarct. Sci. 14, 37–54.

Fraser, F.C., Purves, P.E., 1960. Hearing in cetaceans. Evolution of the accessory air sacs and the structure and function of the outer and middle ear in recent cetaceans. Bull. Br. Mus. Nat. Hist. 7, 1–140.

Gatesy, J., 1997. More DNA support for a Cetacea/Hippopotamidae clade: the blood-clotting protein gene γ-fibrinogen. Mol. Biol. Evol. 14, 537–543.

Gatesy, J., 1998. Molecular evidence for the phylogenetic affinities of Cetacea. In: Thewissen, J. (Ed.), The Emergence of Whales. Plenum, New York, pp. 63–111.

Gatesy, J., Milinkovitch, M., Waddell, V., Stanhope, M., 1999a. Stability of cladistic relationships between Cetacea and higher-level artiodactyl taxa. Syst. Biol. 48, 6–20.

Gatesy, J., O'Grady, P., Baker, R.H., 1999b. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. Cladistics 15, 271–313.

Gatesy, J., O'Leary, M.A., 2001. Deciphering whale origins with molecules and fossils. Trends Ecol. Evol. 16, 562–570.
Geisler, J.H., Sanders, A.E., 2003. Morphological evidence for the phylogeny of Cetacea. J. Mammal. Evol. 10, 23–129.
Geisler, J.H., Uhen, M.D., 2003. Morphological support for a close relationship between hippos and whales. J. Vert. Paleontol. 23, 991–996.
Gilbert, D.G., 1992. SeqApp Version 1.9a. Indiana University, Bloomington.
Hamilton, H., Caballero, S., Collins, A.G., Brownell Jr., R.L., 2001. Evolution of river dolphins. Proc. R. Soc. B 268, 549–556.
Hare, M.P., Cipriano, F., Palumbi, S.R., 2002. Genetic evidence on the demography of speciation in allopatric dolphin species. Evolution 56, 804–816.
Harlin-Cognato, A.D., Honeycutt, R.L., 2006. Multi-locus phylogeny of dolphins in the subfamily Lissodelphininae: character synergy improves phylogenetic resolution. BMC Evol. Biol. 6, 87.
Harvey, P.H., Pagel, M.D., 1991. The Comparative Method in Evolutionary Biology. Oxford University Press, Oxford.
Hatch, L.T., Dopman, E.B., Harrison, R.G., 2006. Phylogenetic relationships among the baleen whales based on maternally and paternally inherited characters. Mol. Phylogenet. Evol. 41, 12–27.
Heyning, J.E., 1997. Sperm whale phylogeny revisited: analysis of the morphological evidence. Mar. Mammal Sci. 13, 596–613.
Irwin, D.M., Árnason, U., 1994. Cytochrome *b* gene of marine mammals: phylogeny and evolution. J. Mammal. Evol. 2, 37–55.
Iwanami, K., Mita, H., Yamamoto, Y., Fujise, Y., Yamada, T., Suzuki, T., 2006. CDNA-derived amino acid sequences of myoglobins from nine species of whales and dolphins. Comp. Biochem. Physiol. B 145, 249–256.
Jefferson, T.A., Webber, M.A., Pitman, R.L., 2008. Marine Mammals of the World: a Comprehensive Guide to their Identification. Academic Press, San Diego.
Kingston, S.E., Rosel, P.E., 2004. Genetic differentiation among recently diverged delphinid taxa determined using AFLP markers. J. Hered. 95, 1–10.
Lambert, O., 2005. Systematics and phylogeny of the fossil beaked whales *Ziphirostrum* du Bus, 1868 and *Choneziphius* Duvernoy, 1851 (Mammalia, Cetacea, Odontoceti), from the Neogene of Antwerp (North of Belgium). Geodiversitas 27, 443–497.
Lambert, O., Louwye, S., 2006. *Archaeoziphius microglenoideus*, a new primitive beaked whale (Mammalia, Cetacea, Odontoceti) from the Middle Miocene of Belgium. J. Vert. Paleontol. 26, 182–191.
LeDuc, R.G., 2002. Delphinids, overview. In: Perrin, W., Würsig, B., Thewissen, J.G.M. (Eds.), Encyclopedia of Marine Mammals. Academic Press, San Diego, pp. 310–314.
LeDuc, R.G., Perrin, W.F., Dizon, A.E., 1999. Phylogenetic relationships among delphinid cetaceans based on full cytochrome *b* sequences. Mar. Mammal Sci. 15, 619–648.
Levenson, D.H., Dizon, A., 2003. Genetic evidence for the ancestral loss of short-wavelength-sensitive cone pigments in mysticete and odontocete cetaceans. Proc. R. Soc. Lond. B 270, 673–679.
Marino, L., Connor, R.C., Fordyce, R.E., Herman, L.M., Hof, P.R., Lefebvre, L., Lusseau, D., McCowan, B., Nimchinsky, E.A., Pack, A.A., Rendell, L., Reidenberg, J.S., Reiss, D., Uhen, M.D., Van der Gucht, E., Whitehead, H., 2007. Cetaceans have complex brains for complex cognition. PLoS Biol. 5, e139.
Matthee, C.A., Burzlaff, J.D., Taylor, J.F., Davis, S.K., 2001. Mining the mammalian genome for artiodactyl systematics. Syst. Biol. 50, 367–390.
May-Collado, L., Agnarsson, I., 2006. Cytochrome *b* and Bayesian inference of whale phylogeny. Mol. Phylogenet. Evol. 38, 344–354.
May-Collado, L.J., Agnarsson, I., Wartzok, D., 2007. Reexamining the relationship between body size and tonal signals frequency in whales: a comparative approach using a novel phylogeny. Mar. Mammal Sci. 23, 524–552.
McGowen, M.R., Clark, C., Gatesy, J., 2008. The vestigial olfactory receptor subgenome of odontocete whales: phylogenetic congruence between gene-tree reconciliation and supermatrix methods. Syst. Biol. 57, 574–590.
Mead, J.G., 1975. Anatomy of the external nasal passages and facial complex in the Delphinidae (Mammalia: Cetacea). Smithsonian Contrib. Zool. 207, 1–72.
Mead, J.G., 2002. Beaked whales, overview. In: Perrin, W., Würsig, B., Thewissen, J.G.M. (Eds.), Encyclopedia of Marine Mammals. Academic Press, San Diego, pp. 81–84.
Messenger, S.L., McGuire, J.A., 1998. Morphology, molecules, and the phylogenetics of cetaceans. Syst. Biol. 47, 90–124.
Milinkovitch, M.C., Bérubé, M., Palsbøll, P.J., 1998. Cetaceans are highly derived artiodactyls. In: Thewissen, J. (Ed.), The Emergence of Whales. Plenum, New York, pp. 113–161.
Milinkovitch, M.C., LeDuc, R.G., Adachi, J., Farnir, F., Georges, M., Hasegawa, M., 1996. Effects of character weighting and species sampling on phylogeny reconstruction: a case study based on DNA sequence data in cetaceans. Genetics 144, 1817–1833.
Milinkovitch, M.C., Meyer, A., Powell, J.R., 1994. Phylogeny of all major groups of cetaceans based on DNA sequences from three mitochondrial genes. Mol. Biol. Evol. 11, 939–948.
Milinkovitch, M.C., Ortí, G., Meyer, A., 1993. Revised phylogeny of whales suggested by mitochondrial ribosomal DNA sequences. Nature 361, 346–348.
Mitchell, E.D., 1989. A new cetacean from the late Eocene La Meseta formation, Seymour Island, Antarctic peninsula. Can. J. Fish. Aquat. Sci. 46, 2219–2235.
Montgelard, C., Catzeflis, F.M., Douzery, E., 1997. Phylogenetic relationships of artiodactyls and cetaceans as deduced from the comparison of cytochrome *b* and 12S rRNA mitochondrial sequences. Mol. Biol. Evol. 14, 550–559.
Müller, K., 2005. SeqState - primer design and sequence statistics for phylogenetic DNA data sets. Appl. Bioinformatics 4, 65–69.
Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A., O'Brien, S.J., 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409, 614–618.
Newman, L.A., Robinson, P.R., 2005. Cone visual pigments of aquatic mammals. Vis. Neurosci. 22, 873–879.
Nikaido, M., Hamilton, H., Makino, H., Sasaki, T., Takahashi, K., Goto, M., Kanda, N., Pastene, L.A., Okada, N., 2006. Baleen whale phylogeny and a past extensive radiation event revealed by SINE insertion analysis. Mol. Biol. Evol. 23, 866–873.
Nikaido, M., Matsuno, F., Abe, H., Shimamura, M., Hamilton, H., Matsubayashi, H., Okada, N., 2001a. Evolution of CHR-2 SINEs in cetartiodactyl genomes: possible evidence for the monophyletic origin of toothed whales. Mammal. Genome 12, 909–915.
Nikaido, M., Matsuno, F., Hamilton, H., Brownell Jr., R.L., Cao, Y., Ding, W., Zuoyan, Z., Shedlock, A.M., Fordyce, R.E., Hasegawa, M., Okada, N., 2001b. Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. Proc. Natl. Acad. Sci. USA 98, 7384–7389.
Nikaido, M., Piskurek, O., Okada, N., 2007. Toothed whale monophyly reassessed by SINE insertion analysis: the absence of lineage sorting effects suggests a small population of a common ancestral species. Mol. Phylogenet. Evol. 43, 216–224.
Nishida, S., Pastene, L.A., Goto, M., Koike, H., 2003. SRY gene structure and phylogeny in the cetacean species. Mammal Study 28, 57–66.
Nishida, S., Goto, M., Pastene, L.A., Kanda, N., Koike, H., 2007. Phylogenetic relationships among cetaceans revealed by Y-chromosome sequences. Zool. Sci. 24, 723–732.
Nylander, J.A.A., 2004. MrModeltest v2. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University.
O'Leary, M.A., Gatesy, J., 2008. Impact of increased character sampling on the phylogeny of Cetartiodactyla (Mammalia): combined analysis including fossils. Cladistics 24, 397–442.
Onbe, K., Nishida, S., Sone, E., Kanda, N., Goto, M., Pastene, L.A., Tanabe, S., Koike, H., 2007. Sequence variation in the Tbx4 gene in marine mammals. Zool. Sci. 24, 449–464.
Palumbi, S.R., Baker, C.S., 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol. Biol. Evol. 11, 426–435.
Palumbi, S.R., Cipriano, F., Hare, M.P., 2001. Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. Evolution 55, 859–868.
Perrin, W.F., 1989. Dolphins, porpoises, and whales. An action plan for the conservation of biological diversity: 1988–1992, second ed. IUCN, Gland, Switzerland.
Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.
Price, S.A., Bininda-Emonds, O.R.P., Gittleman, J.L., 2005. A complete phylogeny of the whales, dolphins and even-toed hoofed mammals (Cetartiodactyla). Biol. Rev. 80, 1–29.
Queralt, R., Adroer, R., Oliva, R., Winkfein, R.J., Retief, J.D., Dixon, G.H., 1995. Evolution of protamine P1 genes in mammals. J. Mol. Evol. 40, 601–607.
Rambaut A., Drummond, A.J., 2007. Tracer v1.4. Distributed by the Authors. Available from: <http://beast.bio.ed.ac.uk/Trace>.
Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Ronquist, F., Huelsenbeck, J.P., van der Mark, P., 2005. MrBayes 3.1 Manual Draft 5/26/2005. Distributed by the Authors. Available from: <http://mrbayes.csit.fsu.edu/manual.php>.
Rosel, P.E., Haygood, M.G., Perrin, W.F., 1995. Phylogenetic relationships among the true porpoises (Cetacea: Phocoenidae). Mol. Phylogenet. Evol. 4, 463–474.
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.
Sasaki, T., Nikaido, M., Hamilton, H., Goto, M., Kato, H., Kanda, N., Pastene, L.A., Cao, Y., Fordyce, R.E., Hasegawa, M., Okada, N., 2005. Mitochondrial phylogenetics and the evolution of mysticete whales. Syst. Biol. 54, 77–90.
Sasaki, T., Nikaido, M., Wada, S., Yamada, T.K., Cao, Y., Hasegawa, M., Okada, N., 2006. *Balaenoptera omurai* is a newly discovered baleen whale that represents an ancient evolutionary lineage. Mol. Phylogenet. Evol. 41, 40–52.
Smith, M.R., Shivji, M.S., Waddell, V.G., Stanhope, M.J., 1996. Phylogenetic evidence from the IRBP gene for the paraphyly of toothed whales, with mixed support for Cetacea as a suborder of Artiodactyla. Mol. Biol. Evol. 13, 918–922.
Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49, 369–381.
Sorenson, M.D., Franzosa, E.A., 2007. TreeRot, Version 3, Boston University, Boston, MA.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57, 758–771.
Steeman, M.E., 2007. Cladistic analysis and a revised classification of fossil and recent mysticetes. Zool. J. Linn. Soc. 150, 875–894.
Swofford, D.L., 2002. PAUP* Phylogenetic Analysis using Parsimony (*and other methods), Version 4.0b10. Sinauer Associates, Sunderland, MA.
Thompson, J.D., Higgins, D.G., Gibson, T.G., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence

weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
Tyack, P.L., Johnson, M., Aguilar Soto, N., Sturlese, A., Madsen, P.T., 2006. Extreme diving of beaked whales. J. Exp. Biol. 209, 4238–4253.
Ursing, B.M., Slack, K.E., Árnason, U., 2000. Subordinal artiodactyl relationships in the light of phylogenetic analysis of 12 mitochondrial protein-coding genes. Zool. Scripta 29, 83–88.
Waddell, V.G., Milinkovitch, M.C., Bérubé, M., Stanhope, M.J., 2000. Examination of the Delphinoidea trichotomy: congruent evidence from three nuclear loci indicates that porpoises (Phocoenidae) share a more recent common ancestry with white whales (Monodontidae) than they do with true dolphins (Delphinidae). Mol. Phylogenet. Evol. 15, 314–318.
Whitmore Jr., F.C., 1994. Neogene climate change and the emergence of the modern whale fauna of the North Atlantic Ocean. In: Berta, A., Démére, T. (Eds.), Contributions in Marine Mammal Paleontology Honoring Frank C. Whitmore, Jr. Proceedings of the San Diego Society of Natural History, vol. 29. pp. 223–228.
Woodward, B.L., Winn, J.P., Fish, F.E., 2006. Morphological specializations of baleen whales associated with hydrodynamic performance and ecological niche. J. Morphol. 267, 1284–1294.
Xiong, Y., Brandley, M.C., Xu, S., Zhou, K., Yang, G., 2009. Seven new mitochondrial genomes and a time-calibrated phylogeny of whales. BMC Evol. Biol. 9, 20.
Yan, J., Zhou, K., Yang, G., 2005. Molecular phylogenetics of 'river dolphins' and the baiji mitochondrial genome. Mol. Phylogenet. Evol. 37, 743–750.