Phylogenomic Resolution of the Cetacean Tree of Life Using Target Sequence Capture

MICHAEL R. McGOWEN1,2, GEORGIA TSAGKOGEORGIA1, SANDRA ÁLVARÉZ-CARRÉTERO1, MARIO DOS REIS1, MONIKA STRÜBEG1, ROBERT DEAVILLE3, PAUL D. JEPSON3, SIMON JARMAN4, ANDREA POLANOWSKI5, PHILIP A. MORIN6, AND STEPHEN J. ROSSITER1, *  

1 School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK; 2Department ofVertebrate Zoology, Smithsonian Museum ofNatural History, 10th & Constitution Ave. NW, Washington DC 20560, USA; 3Institute of Zoology, Zoological Society of London, Outer Circle, London NW1 4RY, UK; 4School of Biological Sciences, University of Western Australia, 35 Stirling Highway, Kensington TAS 7050, Australia; and 5Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 8901 La Jolla Shores Dr., La Jolla CA 92037 USA  

*Correspondence to be sent to: School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK;  
E-mail: s.j.rossiter@qmul.ac.uk.

Received 15 November 2018; revisions returned 2 October 2019; accepted 6 October 2019  
Associate Editor: Hernán López-Fernández

Abstract.—The evolution of cetaceans, from their early transition to an aquatic lifestyle to their subsequent diversification, has been the subject of numerous studies. However, although the higher-level relationships among cetacean families have been largely settled, several aspects of the systematics within these groups remain unresolved. Problematic clades include the oceanic dolphins (37 spp.), which have experienced a recent rapid radiation, and the beaked whales (22 spp.), which have not been investigated in detail using nuclear loci. The combined application of high-throughput sequencing with techniques that target specific genomic sequences provide a powerful means of rapidly generating large volumes of orthologous sequence data for use in phylogenomic studies. To elucidate the phylogenetic relationships within the Cetacea, we combined sequence capture with Illumina sequencing to generate data for ~3200 protein-coding genes for 68 cetacean species and their close relatives including the pygmy hippopotamus. By combining data from >38,000 exons with existing sequences from 11 cetacean and seven outgroup taxa, we produced the first comprehensive comparative genomic data set for cetaceans, spanning 6,522,596 aligned base pairs (bp) and 89 taxa. Phylogenetic trees reconstructed with maximum likelihood and Bayesian inference of concatenated loci, as well as with coalescence analyses of individual gene trees, produced mostly concordant and well-supported trees. Our results completely resolve the relationships among beaked whales as well as the contentious relationships among oceanic dolphins, especially the problematic subfamily Delphininae. We carried out Bayesian estimation of species divergence times using MCMCTree and compared our complete data set to a subset of clocklike genes. Analyses using the complete data set consistently showed less variance in divergence times than the reduced data set.

In addition, integration of new fossils (e.g., Mystacodon selenensis) indicates that the diversification of Crown Cetacea began before the Late Eocene and the divergence of Crown Delphininae as early as the Middle Miocene. [Cetaceans; phylogenomics; Delphininae; Ziphiidae; dolphins; whales.]

Cetaceans (whales, dolphins, and porpoises) have undergone the most dramatic morphological transformation of all mammals, having originated from a clade of terrestrial even-toed ungulates >50 Ma (Gatesy and O’Leary 2003). The origin and evolution of cetaceans have emerged as a textbook case for macroevolution and are arguably one of the best examples of morphological transition in the fossil record (Thewissen et al. 2009). Numerous remarkable fossils from the Eocene (56–34 Ma) have documented this seemingly insurmountable transition from land to sea, detailing such adaptations as the reduction of the hind limbs, reconfiguration of the spine, movement of the nostrils posteriorly, and development of underwater hearing (Berta et al. 2015; Marx et al. 2016).

After their transition to the sea, cetaceans further diversified into two groups with unique adaptations. Toothed whales (Odontoceti) acquired echolocation to hunt using ultrasonic pulses and a highly specialized inner ear, whereas baleen whales (Mysticeti) lost their teeth and evolved a novel keratinous material for filtering aggregate prey (Gatesy et al. 2013). Modern extant cetacean number 89 recognized species, including 75 odontocetes and 14 mysticetes. These species have achieved a cosmopolitan distribution, living in tropical, temperate, and polar marine waters with some species exclusively inhabiting estuaries and river systems (Jefferson et al. 2015; Society for Marine Mammalogy Committee on Taxonomy 2017). Many cetaceans also possess other distinctive specializations, including reduced olfactory and gustatory capacity, the ability to see in dim light, large brains, enormous body size, extended longevity, complex social behavior, osmoregulatory innovations, and respiratory and circulatory systems for extended dives, all of which have made them supremely adapted to their aquatic environment (Gatesy et al. 2013; McGowen et al. 2014; Berta et al. 2015).

Although the evolution of cetaceans from an even-toed “ungulate” ancestor is well understood, there are aspects of their systematics that have proven more challenging. This is particularly the case for relationships within cetacean families, some of which remain problematic (Hamilton et al. 2001; McGowen et al. 2009; Steeman et al. 2009; Chen et al. 2011; Geisler et al. 2011; Zhou et al. 2011; Hassanin et al. 2012). For example, the most speciose cetacean family, Delphinidae (oceanic dolphins, ~35 species), has been especially difficult to resolve despite
recent attempts, likely due to this group’s recent rapid radiation (Leduc et al. 1999; Nishida et al. 2007; Caballero et al. 2008; McGown et al. 2009; McGown 2011; Amaral et al. 2012; Perrin et al. 2013). Particular confusion surrounds the phylogenetic relationships among ~14 species of bottlenose-like dolphins (subfamily Delphininae); these radiated within ~5 myr by some estimates (e.g. McGown et al. 2009; Steeman et al. 2009; Slater et al. 2010) and little consensus exists among data sets, possibly due to incomplete lineage sorting, introgression, hybridization (either ancient or ongoing), and the slow mutation rate in cetaceans (Fig. 1; Kingston et al. 2009; McGown et al. 2009; McGown 2011; Amaral et al. 2012, 2014; Perrin et al. 2013).

Aside from oceanic dolphins, the relationships among taxa within two other speciose clades, the Balaenopteroidea (rorquals plus gray whale; at least 9 species) and the Ziphiidae (beaked whales; 22 species), have also been problematic to disentangle, with several conflicting internal nodes between studies (Nikaido et al. 2005; Sasaki et al. 2005; Nishida et al. 2007; Dalebout et al. 2008, 2014; Deméré et al. 2008; McGown et al. 2009; Hassanian et al. 2012). For example, multiple molecular analyses have revealed that the morphologically distinct gray whale (a benthic suction feeder and the sole member of the family Eschrichtiidae) is nested within the engulfment feeding rorquals of Balaenopteridae and relationships at the base of Balaenopteridae have varied between studies (Nikaido et al. 2005; Sasaki et al. 2005; Deméré et al. 2008; McGown et al. 2009; Steeman et al. 2009; Hassanian et al. 2012; Árnason et al. 2018). In addition, most data gathered for Ziphiidae, especially the genus Mesoplodon (i.e., mitochondrial [mt] genes, 2 nuclear loci), have not robustly resolved species-level relationships (e.g., Dalebout et al. 2008, 2014). There is a pressing need for a good understanding of cetacean systematics, especially in light of their status as highly protected species; smaller cetaceans, in particular, are under increasing threat, as evidenced by the recent extinction of the Chinese river dolphin (Lipotes vexillifer; Turvey et al. 2007) and the rapidly declining population of the vaquita porpoise, which may have <30 individuals left in the wild (Thomas et al. 2017).

The release of several cetacean genomes and transcriptomes in recent years has made it possible to detail the molecular differences between species, as well as identify variable regions or sites for use in population-level and phylogenetic studies (Gui et al. 2013; Yim et al. 2013; Zhou et al. 2013; Foote et al. 2015; Keane et al. 2015; Tsagkogeorga et al. 2015; Camden et al. 2016; Warren et al. 2017; Árnason et al. 2018; Zhou et al. 2018). In addition, new advances in state-of-the-art target sequence capture approaches underpinned by short-read high-throughput sequencing technologies means that huge volumes of genetic data (e.g., thousands of genetic markers per sample) are now obtainable from small amounts of starting material at lower cost (Mamanova et al. 2010; Gasc et al. 2016b). Such approaches offer unprecedented opportunities for studying the genomes of non-model organisms such as cetaceans and developing methods that can be used by researchers for a diverse array of non-model systems. Target sequence capture shows especially great prospects in phylogenomic studies to investigate the generation of multiple loci for large-scale systematic studies. Thus utilizing target capture of exons to sequence large numbers of loci has led to increased resolution of vertebrate clades both deep and shallow (McCormack et al. 2013; Bragg et al. 2015; Portik et al. 2016; Schott et al. 2017). However, with the increase in genomic data, reconstructing divergence dates using standard approaches is computationally intensive, and some researchers have called for the use of reduced data sets using clocklike genes (Smith et al. 2018).

To resolve uncertain relationships among cetacean lineages, we generated new sequence data for 391 protein-coding genes in 68 species of cetaceans, two hippopotamuses and three ruminants. By supplementing these data with available sequences from 38 taxa (11 cetacean, 7 outgroup) obtained from a combination of published genomes, transcriptomes and other data sets, our final alignments spanned 100 individuals from 77 cetacean and 12 outgroup species. We used more than 6.5 million bp of aligned sequence from 38,367 exons of 391 genes to construct a large-scale well-supported species tree of Cetacea using both concatenated and coalescence methods. Every node was well-resolved, including those within Ziphiidae and the problematic Delphinidae (Atlantic and Pacific dolphins). Our results resolve a long, debated over the contentious relationships among species within the subfamily Delphininae, which includes some of the most recognizable cetaceans, such as common dolphins (Delphinus) and bottlenose dolphins (Tursiops). Importantly, our large data set also allowed us to unravel the pattern of molecular rate variation in cetaceans, and thus obtain a precise species-level timetree of cetacean divergences using our complete data set.

**MATERIALS AND METHODS**

**Sample Description, DNA Extraction, and Library Construction**

We obtained tissue or DNA from national repositories for 68 species (77 total individuals) of cetaceans, two species of hippopotamuses, and three species of ruminants (Appendix Table A1). These DNAs were extracted using Qiagen DNeasy extraction kits (Qiagen UK Ltd., Manchester, UK). DNA quality was then evaluated using the Agilent Tape Station 2200, and ~100–200 ng per sample was sheared using a Covaris focused ultrasonicator to achieve ~200 bp fragments. Some degraded samples of <100 ng were not sheared due to their already fragmentary nature. After shearing, fragment size, quantity, and quality of the DNA were then determined using the Agilent 2100 Bioanalyzer. Illumina libraries were constructed for each sample using the NEBXNext Ultra and Ultra II DNA Library Prep Kits with NEBNext Multiplex Oligos (Dual Index
Table 1. Four representative recent phylogenetic analyses of Delphininae (a–d) showing the disagreement in relationships between studies. All numbers above nodes represent Bayesian PP. Multiple individuals for species in Kingston et al. (2009) have been condensed into single OTUs, but support values have been retained.

Design of Biotinylated RNA Baits

A list of 1:1 orthologous protein-coding genes for the Tursiops truncatus genome version turTru1 (as compared with protein-coding genes from Homo sapiens and other available laurasiatherians) was compiled using Ensembl v. 75. We included genes belonging to specific gene ontology (GO) categories based on genes of interest and added these to a larger subset of randomly selected genes. Our target loci covered a range of GO categories ranging from “regulation of centrosome cycle” to “lung development”. Official HGNC gene names were used to search the coding sequence (CDS) databases of two delphinid genomes, Tursiops truncatus (version Ttru_1.4) and Orcinus orca (version Oorc_1.1) on NCBI GenBank (Foote et al. 2015). The longest CDS for each gene, whether Tursiops or Orcinus, was downloaded. For some sequences, no delphinid sequence was available, and another cetacean CDS was used (Lipotes vexillifer, Physeter macrocephalus, Balaenaoptera acutorostrata; Zhou et al. 2013; Yim et al. 2013; Warren et al. 2017). This resulted in 10,271 individual CDS sequences with a total of 18,386,718 bp.
Biotinylated RNA baits (MYbaits) of 100 nucleotides in length were designed by MYCroarray (Ann Arbor, MI, USA; now Arbor Biosciences) using these 10,271 individual CDS sequences. Baits were evaluated via a MYCroarray in-house algorithm, and those with potential to cause cross-hybridization to multiple targets (based on the T. truncatus genome [version Tru_1.4] as a reference) were filtered using a relaxed “4” setting. We then initiated a pilot study of four cetaceans (Mesoplodon bidens, Lagenodelphis hosei, Caperea marginata, Stenella coeruleoalba), and the pygmy hippopotamus (Choeropsis liberiensis), to determine the success of target sequence capture before proceeding further.

After target sequence capture of these five species using the same protocols described below, we reduced the number of included genes to 3266 based on the success of capture of at least two species for a majority of exons of a particular gene. We constructed a new round of baits for these sequences using the same parameters. We then used these baits to capture DNA sequences from all 77 individuals, representing 68 species cetacean, both species of hippopotamuses, and three species of ruminants (Appendix Table A1).

**Target Sequence Capture and Sequencing**

Target sequence capture was performed following the protocol contained in MYBaits Manual version 3.1, in which biotinylated RNA baits were hybridized to individual sample libraries for ~20–24 h. Captured DNA was recovered using Streptavidin C1 magnetic beads (MyOne) and washed to remove any unhybridized fragments. Then all captured DNA was amplified and individual samples were pooled into two batches for sequencing. Each batch was paired-end sequenced by individual samples were pooled into two batches for sequencing. Each batch was paired-end sequenced by the “Bart’s and the London Genome Centre” of Queen Mary, University of London using the Illumina NextSeq platform with the high output mode and a read length of 150 bp. All raw reads were deposited in the Sequence Read Archive (SRA) of NCBI, BioProject PRJNA575269.

**Assembly of Reads and Identification of Contigs**

We assessed the quality of raw reads using FastQC version 0.11.5 (Babraham Bioinformatics), and raw reads were cleaned by removing adapters and low-quality bases using Trimmomatic 0.36 (Bolger et al. 2014). A total number of reads for each sample are shown in Supplementary Table S1 available on Dryad at https://doi.org/10.5061/dryad.jq40b.0f. Trimmed reads for each individual sample were de novo assembled separately using Trinity v2.2.0 with default settings (Grabherr et al. 2011). To identify each Trinity contig, we then conducted reciprocal blast searches (blastn; E-value cutoff of 10−8; retained only top blast hit) of each Trinity assembly using FASTA files with all exons drawn from the O. orca and T. truncatus genomes. Per species counts of contigs with a reciprocal blast hit are shown in Supplementary Table S1 available on Dryad. In addition, we also conducted reciprocal blast searches with FASTA files containing CDSs from existing cetacean genomes, partial genomes, or transcriptions including Balaenoptera acutorostrata, B. physalus, Megaptera novaeangliae, Balaena mysticetus, Physeter macrocephalus, Neophocaena phocaenoides, Lipotes vexillifer (Zhou et al. 2013; Yam et al. 2013; Keane et al. 2015; Tsagkogeorga et al. 2015; Warren et al. 2017), as well as outgroup genomes from Bos taurus, Oris aries, Panthelops hodgsonii, Sus scrofa, Vicugna pacos, Camelus bactrianus, and Equus caballus (Lindblad-Toh et al. 2011; Bactrian Camels et al. 2012; Groenen et al. 2012; Ge et al. 2013; Jiang et al. 2014). All contigs were then trimmed to the length of the desired exon. We then kept all exons (38,832) present in our original baits for further downstream analyses.

**Alignments**

Each individual exon was aligned separately using mafft version 7 (Katoh and Standley 2013) for a total of 38,832 exon alignments. These were then concatenated into complete gene alignments. To assess the efficacy of this process, alignments were then translated into amino acids to identify potential stop codons. Alignments with stop codons were excluded by eye. In some cases, insertions at the end of exon boundaries were introduced from blastn, and these were removed. In other cases, exons were missing from the original annotations of the O. orca and T. truncatus genomes; these exons were then introduced in order for the whole alignment to remain in the correct reading frame. In addition, we removed 65 genes (665 exons) from the overall data set if the gene was difficult to align or difficult to differentiate sequences from closely related paralogues. For the remaining 3191 genes, presence of premature stop codons and/or indels that were not multiples of three nucleotides were taken as potential evidence for the presence of a pseudogene and noted for further analysis.

We also added isolated sequences from NCBI GenBank for Platania gangetica (South Asian river dolphin) and Balaenoptera omurai (Omura’s whale), two species for which we did not have capture data and for which a whole genome is not available. This consisted of a total of 72 sequences for P. gangetica (57,770 bp) and 67 for B. omurai (57,886 bp). See Supplementary Table S2 available on Dryad for a list with accession numbers of these sequences and their publications.

**Phylogenomic Analyses**

We created two concatenated alignments, both with a total of 3191 genes (38,167 exons) and 6,527,596 bp: Dataset A and Dataset B. Dataset A contained sequences from P. gangetica and B. omurai, whereas Dataset B excluded these sequences. For both data sets,
we conducted three concatenated maximum likelihood analyses using RAxML v8.2 (Stamatakis 2014): (i) unpartitioned, (ii) 3191 partitions, one for each gene, and (iii) a partition scheme of 1573 partitions selected using PartitionFinder v2.1.1 (Lanfear et al. 2016). We performed each analysis using the GTRCAT model for every partition. Each analysis used default parameters in RAxML and support scores were generated using the rapid bootstrapping option with at least 1000 replicates. To confirm our findings using an alternative method, a Bayesian analysis of Dataset A was implemented in ExaBayes using default parameters and a GTR+G model of evolution (Aberer et al. 2014). Two unpartitioned analyses of Dataset A were conducted for 1,000,000 generations with two coupled chains instituted for each analysis and trees sampled every 500 generations. The initial 25% of runs were discarded as burn-in. Results of the Bayesian analyses were examined in Tracer v1.7 (Rambaut et al. 2018) to evaluate whether parameters, node ages, and likelihood values had converged. All RAxML and ExaBayes runs were implemented using the CIPRES Science Gateway v3.3 (Miller et al. 2010).

We also implemented a species tree analysis, which takes into account the potential discordance between individual gene trees and the underlying species tree due to incomplete lineage sorting. We first generated individual maximum likelihood gene trees in RAxML v8.2 for each of the 3191 genes using a GTRCAT model. Due to the computational constraints imposed by the large number of sequences present for P. gangetica and B. omurai, we excluded them from all gene tree analyses. We used a partitioned likelihood calculation allowing Bayesian divergence time inference of phylogenomic data sets (dos Reis and Rannala 2006). At this stage, we did not want to use the prior on node ages. These parameters lead to an approximate likelihood method that cannot be used for marginal likelihood calculation (dos Reis et al. 2018) and thus the computationally expensive exact method must be used. Therefore, to decide the best-fitting clock model, we carried out Bayesian model selection on smaller subsets of the data suitable for exact likelihood calculation: 1 randomly selected locus for subsets of 20, 40, and all 85 species; 5 randomly selected loci for a subset of 20 species; and 20 randomly selected loci for a subset of both 20 and 40 species. As in dos Reis et al. (2018), for analyses which used less than 85 species, we chose taxa from representative clades to reflect the true diversity of rate variation across taxa. Note that the sampling of genes was random and, based on inference theory, we had no reason to expect any biases in model selection.

Test runs of the program were carried out to ensure the convergence of the Markov Chain Monte Carlo (MCMC) chains and that enough likelihood samples had been collected for Bayes factors calculation. The birth–death process with $\lambda = 1$ (birth and death rates) and $\rho = 0.1$ (fraction of species sampled) was used to construct the prior on node ages. These parameters lead to an approximately uniform distribution on node ages (Yang and Rannala 2006). At this stage, we did not want to estimate divergence times but simply select the most
appropriate clock model given the data, thus the root age was fixed to 1. In MCMCtree, this may be done by using a narrow uniform distribution between 0.999 and 1.001. No other fossil calibrations were used at this step.

We used the HKY85+Γ5 substitution model (Hasegawa et al. 1985; HKY model accommodating among site rate heterogeneity using a gamma distribution with parameter was not smaller than 100 (Nascimento et al. 2018). We checked that the estimated sample size (ESS) for each node ages, and likelihood values had converged. We discarded as burn-in. Results were examined in Tracer v. 2017).

Our phylogenetic distance from the Tursiops and Orcinus genomes did not appear to affect the success of sequence capture, as 21.6 million reads were obtained for the ruminant Gazella ambi. For each sample, reads were assembled into Trinity contigs numbering from 11,156 (H. planifrons) to 375,798 (Mesoplodon caribbicus) with an average N50 of 310 bp. After reciprocal blasting of the contigs to the Orcinus genome using blastn, we recovered between 7428 (H. planifrons) and 31,888 exons (Stenella longirostris 28493), with an average of 28,324 exons per species (74% recovery of initial exons). Delphinids had a higher average of 30,106 exons (79% recovery). The five noncetaceans ranged from 21,259 to 26,179 exons, with an average of 24,085 exons (63% recovery).

**Phylogenetic Analysis**

The same tree topology was generated with all concatenated analyses of Dataset A using RAxML or ExaBayes, regardless of model or partitioning scheme. In addition, topologies resulting from all analyses of Dataset B agreed with those of Dataset A when Phylogenetic Analysis
and B. omurai were pruned. The phylogenomic tree resulting from the RAxML analysis with 3,191 separate partitions by gene is shown in Figure 2. Support scores only differed among the separate RAxML analyses at eight nodes (indicated by red dots; Fig. 2), otherwise they showed 100% bootstrap (BS) support (Fig. 2). Of the eight nodes that differed, only four of these had support scores less than 90% BS, two within balaenopterids and two within delphinine dolphins (Fig. 2). Both independent runs using ExaBayes showed evidence of convergence (all ESS values > 204; Supplementary Fig. S2 available on Dryad) and resulted in a topology with all nodes supported by Bayesian PP of 1.0. All species in which there were 2+ representatives were supported as monophyletic with high support (all BS 100; PP 1.0).

The species tree generated by the coalescence analysis using ASTRAL (Supplementary Fig. S3 available on Dryad) differs from the concatenated analyses at only three nodes, all within Delphinidae (Fig. 1). In addition, all nodes save three have support scores > 0.99 (Supplementary Fig. S3 available on Dryad). The ASTRAL tree places Lissodelphis and Sagmatias obladiadens + S. obscurus in a clade to the exclusion of Cephalorhynchus + Sagmatias australis with posterior probability of 0.8, as well as displacing T. truncatus from the clade including T. aduncus, Stenella attenuata + frontalis with high support (Fig. 1 and Supplementary Fig. S3 available on Dryad). Two additional nodes within Delphininae (Nodes 1 and 2; Fig. 2) are supported by the ASTRAL tree but show PP of 0.42 and 0.89, respectively (Supplementary Fig. S3 available on Dryad), agreeing with the varying support among the same nodes in the concatenated analyses. The ASTRAL species tree has a final normalized quartet score of 0.869, representing the proportion of quartets for individual gene trees that are satisfied by the species tree.
Figure 2. The best concatenated RAxML maximum likelihood tree derived from Dataset A using 3191 partitions of each protein-coding gene and 6,527,596 bp (lnL = -24,850,884.943687). The phylogenetic relationships of Delphininae are shown more clearly in the box on the left. All concatenated RAxML and Bayesian analyses using Dataset A retrieved the same topology. All nodes have 100% BS values (RAxML) or 1.0 PP (Exabayes) with the exception of the numbered nodes represented by red dots. These nodes have BS values and PP shown in the table in the upper left. The ASTRAL species tree topology only differed from the concatenated topology at three nodes shown in blue. Taxa in bold are those with data derived from previously existing genomes, transcriptomes, or GenBank sequences. Illustrations are by Carl Buell and represent (top to bottom) T. truncatus (common bottlenose dolphin), Feresa attenuata (pygmy killer whale), L. albirostris (white-beaked dolphin), Inia geoffrensis (Amazon river dolphin), Mesoplodon layardii (strap-toothed whale), Kogia sima (dwarf sperm whale), B. bonaerensis (Antarctic minke whale), and B. taurus (domestic cow).
All analyses supported the monophyly of Cetroruminantia (Cetacea + Hippopotamidae + Ruminantia), Ruminantia, Whippomorphpha (Cetacea + Hippopotamidae), Cetacea, Odontoceti, Mysticeti, Synhina, Delphinida, Physeteroidea, Inioidea, Delphinoidae, Lipotes + Inioidea, Phocoenidae + Monodontidae (Monodontoidae sensu Geisler et al. 2011), and all recognized cetacean families with the exception of Balaenopteridae. Several genera were well-supported as polyphyletic, including Balaenoptera, Sagittatus (sensu Leduc et al. 1999; Vollmer et al. 2019), Cephalorhynchus, and Stenella. Dataset A clearly supports P. gangetica, the South Asian river dolphin, as a separate lineage from the other “river” dolphins (Lipotes, Iniata, Pontoporia) and also supports its exclusion from the clade Ziphidae + Delphinida.

Divergence Dating

Table 1 shows the results of the Bayesian model selection analysis, which was used to determine the best-fitting model in subsequent MCMCTree analyses. For all sampled alignments, the AR model had the highest posterior probability (~1.0 in all cases) and was interpreted as the best-fitting model to our data. All comparable MCMCTree runs for the total data set using each partition scheme and model showed evidence of convergence with an ESS for each parameter >200 (Supplementary Figs. S4–S9 available on Dryad). The timetree of Cetacea obtained using the 6-partition AR model is shown in Figure 3 with the posterior probability distributions of both AR and IR models shown above each node. For comparison, we ran MCMCTree using a 3-partition and 10-partition model, and the timetrees for both are shown in Supplementary Figures S10 and S11 available on Dryad with posterior probability distributions of both AR and IR models above each node. Precise dates (mean and 95% CI) for both AR and IR models for the 3-, 6-, and 10-partitions of the complete data set are listed in Supplementary Table S3 available on Dryad. Divergence dates for distinct nodes are similar among the three-partition schemes for each model (AR and IR), although within Delphinida divergence times generally decrease slightly with the increase in partitions for the AR model (Supplementary Table S3 available on Dryad). Differences in posterior mean times do not differ drastically between using 10 genes (SortaDate) and using the entire data set (Fig. 4, Supplementary Figs. S13 and S14, Table S3 available on Dryad); however, variances are generally larger when using the reduced data set (Fig. 5, Supplementary Fig. S15, Table S3 available on Dryad). For example, when analyzing the 10 SortaDate genes as three partitions under the AR model (Fig. 5a'), the slope of the regression line in the infinite-sites plot is 0.192, implying that for every 1 Ma of divergence, 0.192 Ma of uncertainty are added to the 95% CI. When we include all the data in three partitions (Fig. 5a), the regression slope falls to 0.173, and it falls further to 0.115 and 0.074 when analyzed as 6 and 10 partitions, respectively (Fig. 5b,c). Thus, the analysis using the whole data set provides time estimates with the narrowest credibility intervals. The same trend is seen with the exclusion of the root (Fig. 5), as well as using the IR model (Supplementary Fig. S15 available on Dryad). We note that in none of the plots do points form a straight line. This indicates that uncertainty in time estimates are due both to limited data as well as uncertainties in the fossil calibrations (Rannala and Yang 2007).

Outside of Cetacea (and exclusive of the root), the mean age of nodes using the 6-partition model decreased using the AR model with respect to IR by an average of 8.29 Ma, with the mean age of Bovidae shifting by 14.47 Ma (Fig. 5; Supplementary Table S3 available on Dryad). Within Cetacea, nodes increased using the AR model with respect to the IR model by an average of 1.08 Ma, although most nodes within Ziphiidae, as well as Kogia decreased by >0.61 (Fig. 5; Supplementary Table S3 available on Dryad). At least eight nodes within Cetacea increased by over 3 Ma when using the AR as compared with the IR model, including Crown Balaenidae (10.61 Ma vs. 4.79 Ma), Delphinidae (19.78 Ma vs. 16.44 Ma), and Balaenopteridae (15.74 Ma vs. 10.99 Ma). Nodes within mysticetes differed widely between the analyses by an average of 3.07 Ma. Results are comparable when using the other partitioning schemes (Supplementary Table S3 available on Dryad).

Using the 6-partition AR model, we obtained a mean age for Whippomorpha (t = 53.92 Ma), less than a million years before the earliest stem cetacean, Himalalacetus subthanuensis (Bajpai and Gingerich 1998; Supplementary Table S3 available on Dryad). The age of Crown Cetacea is much more recent (t = 36.72 Ma), which is less than half a million years older than the oldest-known crown cetacean fossil, the stem mysticete Mystacodon selensis from the Late Eocene of Peru (Lambert et al. 2017). The diversification of Crown Odontoceti began before the end of the Eocene (t = 34.13 Ma), whereas the emergence of Crown Mysticeti (t = 25.73 Ma) is more than 8 myr more recent, firmly within the Oligocene. All lineages leading to modern cetacean families were present by the Middle Miocene, Balaenopteroidae, Ziphidae, Monodontidae + Phocoenidae, and Delphinidae began to diversify in the Early-to-Middle Miocene, with diversification of the speciose genus Mesoplodon and the delphinid subfamilies beginning in the Late Miocene.

DISCUSSION

The evolution of the cetaceans, from their early transition to an aquatic niche to their rapid diversification, has been the subject of numerous studies, yet several aspects of their systematics have remained unresolved. Using a targeted sequence
FIGURE 3. Timetree of Cetacea analyzed in the MCMCTree package of PAML 4.9h using six partitions and approximate likelihood (Yang 2007). A time scale in Ma (millions of years) is shown above the tree, with geologic periods labeled below the tree for reference (Q = Quaternary). Above each node the posterior distributions of the AR model (purple) and IR model (white) are shown. Raw numbers for the mean and 95% confidence intervals for each node and each model are shown in Supplementary Table S3 available on Dryad. Red circles at each node represent calibration points listed in Table 2. Illustrations are by Carl Buell and represent (top to bottom) T. truncatus (common bottlenose dolphin), S. chinensis (Indo-Pacific humpback dolphin), O. orca (killer whale), Delphinapterus leucas (beluga), L. vexillifer (Yangtze river dolphin), Mesoplodon layardii (strap-toothed whale), Ziphius cavirostris (Cuvier’s beaked whale), P. macrocephalus (sperm whale), B. physalus (fin whale), and B. taurus (domestic cow). This tree with posterior node estimates was generated with the mcmc.tree.plot function from the R package MCMCtreeR (Puttick 2019).

approach, in which we constructed RNA baits for exons based on the O. orca and T. truncatus genomes, we were able to obtain sequences for an average of 74% of target exons from most of the cetacean species as well as their closest relatives. Our final data set of 38,167 exons contains the first ever large-scale genomic data for at least 58 cetacean species, as well as for the pygmy hippopotamus, and allowed us to produce a fully resolved, time-calibrated tree that was able to elucidate with confidence several problematic relationships.

Divergence Dating

Several studies have indicated that molecular rates within cetaceans, especially within mysticetes, are much slower than those in other mammals (Bininda-Emonds 2007; Jackson et al. 2009; Dornburg et al. 2012). However, estimates of rates and times may be substantially affected by the relaxed-clock model used (e.g. dos Reis et al. 2018), and thus it is important to select the most appropriate clock model. For example, Dornburg et al. (2012) found that uncorrelated rate models perform poorly compared
FIGURE 4. Scatterplot of the estimated posterior mean times (and 95% confidence intervals) for the six-partition scheme of both AR (a) and IR (b) models for the SortaDate data set (x-axis) against all data (y-axis).

FIGURE 5. Infinite-sites plots showing the estimated posterior mean times in Ma (x-axis) plotted against the estimated posterior confidence interval (CI) widths in Ma (y-axis) for the AR model for both data sets (all data, 10 genes SortaDate) using the three different partition schemes, three partitions (a, a’), six partitions (b, b’), and 10 partitions (c, c’). The solid line represents the regression line including the root and the dotted line represents the regression line excluding the root. $R^2$ is the coefficient of determination for each comparison, whereas below each are the equations of the regression lines with and without the root.

with local clocks when there is significant rate variation between lineages. An appealing property of the AR model implemented in MCMCTree is that it allows for local clocks in more closely related species whereas allowing for rates to vary more widely in deeper nodes, a property that appears biologically realistic (dos Reis et al. 2018). In this study, we find that like dos Reis et al. (2018), the AR model was preferred based on analysis of a subset of the data (Table 1), and estimates of times are substantially affected depending on the rate.
model used (Fig. 3; Supplementary Table S3 available on Dryad). However, unlike primate estimates in dos Reis et al. (2018), the AR divergence estimates produced more recent estimates for deeper nodes than the IR model, whereas tending to skew older within cetaceans (Fig. 3). For some dates along the trunk of the cetacean tree (i.e., Delphinidae), the AR analysis obtained slightly older dates than previous analyses which have used divergence dating analyses with uncorrelated rates (McGowen et al. 2009; Slater et al. 2010; Zhou et al. 2011) or penalized likelihood (Steeman et al. 2009) although all used substantially less data. The approximate likelihood method (dos Reis and Yang 2011) has a disadvantage in that the marginal likelihood cannot be directly computed. Only by carrying out the MCMC analysis using the exact likelihood can we then estimate the marginal likelihood, but this is not computationally feasible with very large data sets. We have attempted to overcome this computational limitation by using Bayesian model selection on various subsets of our data (Table 1). Our results show that for all subsets analyzed, the preferred model is AR, although more powerful computational approaches are needed so we can validate these results with the whole data set. Nevertheless, our analyses based on the stepping stone method seem to indicate that the AR model better explains the rate variation among cetaceans and outgroups.

Computational difficulties in analyzing large-scale genomic data sets have led some to suggest using a reduced set of loci (Smith et al. 2018). However, Rannala and Yang (2007) show that uncertainties due to finite-data sampling and the relaxed-clock model approach zero asymptotically as both the number of loci (i.e., the number of partitions) and the number of sites in each locus approach infinity. Rannala and Yang (2007) also show that uncertainties due to fossil calibrations cannot be eliminated. Thus, to reduce uncertainty in time estimates, they recommend analysis of large data sets. Rannala and Yang (2007) and Inoue et al. (2010) suggest using the infinite-sites plot to assess whether uncertainty in a Bayesian clock-dating analysis is dominated by the fossil calibrations or by errors due to the finite-data samples. In the infinite-sites plot, points asymptotically approach a straight line as the number of partitions and number of sites in the analysis is increased. For some large data sets, data points should approximate a straight line, indicating that any uncertainties are due to fossil calibrations (e.g. felid data in Inoue et al. 2010). In such cases, including further molecular data in the analysis will not improve the time estimates, as the sampling errors due to finite data are zero.

We note that in the approximate likelihood method, which we used here to estimate all divergence times, computational time depends on the size of the Hessian matrix used in the approximation, which in turn depends on the number of species analyzed (dos Reis and Yang 2011), but not on the number of sites in an alignment. Thus, MCMC sampling of 100 species takes approximately the same time whether we analyze 10^3, 10^6, or 10^9 sites. This would not have been the case under exact likelihood computation where computational time is proportional to the number of site-patterns in the alignment, and under which genome-scale data cannot be analyzed. Given that this analysis produces estimates with considerably less uncertainty than those obtained using the 10 SortaDate genes (Fig. 5), we suggest that the whole data estimates should be preferred in our cetacean analysis.

**Relationships among Major Cetartiodactylan Lineages**

We obtained 100% resolution for the relationships of the five major lineages of Cetartiodactyla, with ruminants (Vivipara, Camelus) as the most basal lineage, followed by Suiformes (Sus), Ruminantia (Bos, Toxelaphus, Gazella, Oryx, Oris, Pantholops), Anocodonta (Hippopotamus, Cheoropsis), and Cetacea. Phylogenomic analysis unequivocally supported a monophyletic clade which included both hippopotamuses and cetaceans (Whippomorpha) to the exclusion of other cetartiodactyls, with a mean age of 53.92 Ma (Fig. 3). The mean age of Crown Cetacea (ξ = 36.72 Ma; Fig. 3) agrees with some previous analyses (McGowen et al. 2009; Steeman et al. 2009; Slater et al. 2010), but is slightly older than some other divergence dating analyses of the group (Nikaido et al. 2001; Sasaki et al. 2006; Xiong et al. 2009; Meredith et al. 2011; Hassanin et al. 2012; although see Marx and Fordyce 2015). This is despite using the recently described stem mysticete M. sekenensis as a calibration for Crown Cetacea (Lambert et al. 2017), the age of which (36.4 Ma) is estimated to be more than 2 Ma older than Llanoecetus dentraticatus (34.2 Ma), a stem mysticete previously identified as the oldest crown cetacean and used as a calibration point in numerous divergence dating analyses of cetaceans.

**Phylogeny and Evolution of Mysticeti**

The age of Crown Mysticeti (ξ = 25.75 Ma; 95% CI: 25.22–26.72 Ma) is younger than most estimates including those using mitochondrial genomes and low-coverage genomes of mysticetes (Sasaki et al. 2006; McGowen et al. 2009; Steeman et al. 2009; Aronson et al. 2018). However, many of these analyses used OU 2244 (previously identified as an archaic right whale ∼28 Ma) as a constraint for Crown Mysticeti. Marx and Fordyce (2015) identified OU 2244 as falling outside of Crown Mysticeti, and Mauicetus parki (the calibration used here; Table 1) as the earliest crown mysticete.

Within baleen whales, our results demonstrate unequivocal support for the monophyly of Balaenidae (ξ = 10.61 Ma), as well as for the clade of Neobalaenidae plus Balanopteroidea (Pliocugula sensu Geisler et al. 2011; ξ = 22.11 Ma), despite some morphological analyses which support a relationship between...
Neobalaenidae and Balaenidae (Bouletel and de Muizon 2006; Bisconti 2007, 2014; Steeman 2007; Churchill et al. 2011; El Adli et al. 2014). In agreement with results presented here, most molecular and some morphological analyses support Plicogulae (Árnason and Gullberg 1994; Rychel et al. 2004; Sasaki et al. 2005, 2006; Agnarsson and May-Collado 2008; Demére et al. 2008; Steeman et al. 2009; McGowen 2011; Hassanin et al. 2012; Fordyce and Marx 2013; Marx and Fordyce 2015). The North Pacific and North Atlantic right whales (Eubalaena japonica and Eubalaena glacialis, respectively) form an unequivocally well-supported clade (Fig. 2), which conflicts with some weakly supported mtDNA and ncDNA analyses that placed E. japonica in a clade with Eubalaena australis for the exclusion of E. glacialis (Rosenbaum et al. 2000; Gaines et al. 2005). Eubalaena glacialis was split into two species (E. glacialis and E. japonica) partly based on the assumption that E. japonica was more closely related to E. australis (Rosenbaum et al. 2000). The status of species within Eubalaena may have to be reevaluated in light of these results, but the mean age of species events within the genus (ε = 4.35 and θ = 2.62 Ma; Supplementary Table S3 available on Dryad) intimate that the status of E. japonica and E. glacialis as separate species is warranted.

Our analyses show that the gray whale (Eschrichtius robustus; Eschrichtiidae) is firmly nested within Balaenopteridae, in agreement with previous studies based on morphological (Árnason and Gullberg 1994; Rychel et al. 2004; Sasaki et al. 2005, 2006; Nishida et al. 2007; Demére et al. 2008; Agnarsson and May-Collado 2008; Steeman et al. 2009; McGowen 2011; Hassanin et al. 2012). In addition, we find that the genus Balaenoptera is polyphyletic with both Eschrichtius and Megaptera nested within the genus. We show high support for minke whales (B. acutorostrata + Balaenoptera bonaerensis) diverging earliest within the Balaenopteridae, followed by E. robustus (although BS scores of ML analyses range from 64 to 78). According to our analyses, the radiation of Balaenopteridae began ~15.74 Ma, and divergence dates within this group are slightly older than some recent analyses (McGowen et al. 2009; Árnason et al. 2018). Indeed divergence dates within Balaenopteridae vary greatly between our AR and IR analyses (Fig. 3), with the IR analyses showing more recent dates of up to 4.75 Ma.

Árnason et al. (2018) sequenced new low-coverage genomes from six mysticete species and discovered a similar arrangement of balaenopteroid species based on coalescence analyses of trees derived from >30,000 20-kb genomic segments, with the exception that the gray whale is sister to M. novaeangliae + B. physalus. The branch supporting this relationship is incredibly short, and support for conflicting trees is high; their analysis of quartet scores showed that no arrangement between the Balaenoptera musculus group, M. novaeangliae + B. physalus, and E. robustus could be significantly rejected. These results were interpreted as implying that large-scale hybridization played a part early in balaenopteroid evolution. Analysis of retrotransposon insertion events using the same genomes reveals a similar pattern (Lammers et al. 2019). Whatever the cause, Árnason et al. (2018) and the results presented here both agree that a formal redescription of Balaenopteridae needs to be conducted with the clear inclusion of E. robustus within the family Balaenopteridae. Conflicting relationships represented by the variable support values of Nodes 6 and 7 can be explained by the missing data in B. omurai, as the clade B. musculus + Balaenoptera was supported by 100% BS values in RAxML analyses using Dataset B (Supplementary Fig. S1 available on Dryad).

Phylogeny and Evolution of Odontoceti

We find continued robust support for a monophyletic Odontoceti (Fig. 2), which is consistent with unique synapomorphies such as the lateral expansion of the maxilla coinciding with the development of echolocation (Geisler et al. 2011). Sperm whales (Physeteridae + Kogiidae) split from other extant odontocetes in the Latest Eocene (ε = 34.13 Ma; Fig. 3); however, there is no evidence of either stem or crown odontocetes present in the Eocene (Marx et al. 2016). Physeteridae and Kogiidae diverged from one another in the Late Oligocene or Early Miocene (ε = 22.11 Ma; 95% CI: 20.58–24.08; Supplementary Table S3 available on Dryad). This agrees with other earlier divergence analyses (McGowen et al. 2009; Steeman et al. 2009; Meredith et al. 2011; Hassanin et al. 2012). Kogia and Physeter have sometimes been included in the same family (Physeteridae), but the divergence between these genera is earlier than the diversification of Superfamily Delphinoidea, and these genera likely warrant placement in separate families. This deep divergence also coincides with evidence of the existence of fossil kogiids in the Early Miocene (Velez-Juarbe et al. 2013).

Platynisia gangetica is a freshwater odontocete found in the river systems of South Asia (Indus, Ganges, and Brahmaputra) and a relict species that is part of a lineage that was much more diverse in the past, with multiple fossils known from marine deposits (Geisler et al. 2011; Marx et al. 2016). Although we included Platynista sequences from only 72 genes, we recovered strong support for P. gangetica as the sister taxon of all other odontocetes excluding sperm whales (Fig. 2), placing them in a distinct clade from the other “river dolphins” (Inia, Liptode, and Pontoporia). Some analyses of mainly mitochondrial data have united Platynista and Ziphidae (Cassens et al. 2000; Hassanin et al. 2012), but analyses integrating significant nuclear-derived data have agreed with results presented here (McGowen et al. 2009; Steeman et al. 2009; Chen et al. 2011; Geisler et al. 2011; Meredith et al. 2011; Nikaido et al. 2001; Zhou et al. 2011). Phylogenetic relationships among ziphids are fully resolved and well-supported with all but three species...
not included in our analysis (Mesoplodon loutaula, Mesoplodon traversii, Indopacificus; Fig. 2). The strong resolution obtained here is in contrast to the most comprehensive analyses of ziphiids hitherto undertaken, with results in many weakly supported nodes (Agnarsson and May-Collado 2008; Dalebout et al. 2008, 2014; McGowen et al. 2009; Steeman et al. 2009). Ziphiids started to diversify in the Early or Middle Miocene (t = 15.61 Ma; 95% CI: 13.65–17.79; Fig. 3), with Berardius as the most basal genus in the family, followed by Tasmacetus, Ziphius, Hyperoodon, and Mesoplodon (Fig. 2). Our results nest Tasmacetus shepherdi, a beaked whale with multiple functional teeth in both jaws of both sexes within a clade that has a reduced dentition of two to four mandibular teeth (Ellis and Mead 2017). We find no support for a traditional division of ziphiids into two subfamilies: Ziphiinae (Berardius, Ziphius, and Tasmacetus) and Hyperoodontinae (Indopacificus, Hyperoodon, and Mesoplodon), although the monophyly of Hyperoodon + Mesoplodon is well-supported. Mesoplodon (the most speciose genus of all cetaceans with 15 recognized species) experienced a rapid radiation beginning in the Late Miocene, with at least 13 species arising in the span of less than 5 Ma (Fig. 3). Within Mesoplodon, we find support for three major clades: the “bidens” lineage which contains bidens, gingkodens, europaeus, and mirus; the “layardii” lineage including layardii, narwhali, and hookeri; and the “hectori” lineage which includes hectori, grayi, stejnegeri, densirostris, perrini, and peruvianus. Although weakly supported in some analyses, the layardii clade and some species in the hectori clade have been recovered by either mtDNA, nuclear introns, or both (Dalebout et al. 2003, 2007, 2008, 2014; McGowen et al. 2009; Steeman et al. 2009); here, we find robust unequivocal evidence for their support. The gingkodens + mirus + europaeus clade is well-supported in analyses of both mt and ncDNA, and our results also place bidens in a clade with these species (Fig. 2). This finding differs from many previous studies in which M. bidens was placed in a basal position with respect to all other Mesoplodon species, although this arrangement received mostly weak support (Dalebout et al. 2008, 2014; McGowen et al. 2009; Steeman et al. 2009). Previous molecular analyses differed as to the phylogenetic relationships among the remaining “river dolphin” species (Inia, Pontoporia, and Lipotes). In some molecular analyses, Lipotes was placed as the most basal taxon of the Delphinida (Delphinoidea + Iniidae + Pontoporidae + Lipotidae; Cassens et al. 2009; Hamilton et al. 2001) or weakly supported as sister to Inia + Pontoporia (Agnarsson and May-Collado 2008; McGowen et al. 2009; Steeman et al. 2009; Geisler et al. 2009; Hassanan et al. 2012). In this study, we find strong evidence that all three species form a distinct clade (Fig. 2), which originated in the Late Oligocene (t = 23.97 Ma; 95% CI: 23.03–24.92). Evidence from previous analyses integrating molecular, morphological, and fossil data reveal that the two solely freshwater species, Lipotes and Inia, invaded freshwater separately, as they are more closely related to fossil taxa from marine sediments (Geisler et al. 2011). Delphinoidea (Monodontidae + Phocoenidae + Delphinidae) is well-supported with Monodontidae more closely related to Phocoenidae, as noted in previous analyses (Waddell et al. 2000; Cassens et al. 2000; Agnarsson and May-Collado 2008; McGowen et al. 2009; Steeman et al. 2009; McGowen 2011; Zhou et al. 2011; Hassanan et al. 2012). Crown delphinoids originated in the Early Miocene (t = 19.78 Ma; 95% CI: 18.81–20.76). Fossil lineages grouped in the “Kentriodontidae” have been tied to the early diversification of Delphinina and Delphinoidea, but revision of this group is in process (Murakami et al. 2014; Tebedo et al. 2018). Both Crown Phocoenidae and Crown Monodontidae originated in the Late Miocene (Fig. 3). Within phocoenids, we strongly recovered Neophocaena as the most basal genus as well as a monophyletic Phocoena. The monophyly of Phocoena conflicts with multiple analyses that have placed Phocoena phocoena and P. dalli as sister species to the exclusion of other members of Phocoena; however, relationships between these species were usually weakly supported and/or dominated by mitochondrial data (Pichler et al. 2001; McGowen et al. 2009; Steeman et al. 2009).

Our phylogenomic reconstruction provided for a clear picture of the evolutionary relationships within Delphinidae, with high statistical support and agreement between analyses for most clades (Fig. 2). The mean age of Crown Delphinidae using the AR model (t = 12.72 Ma) is older by almost 3 Ma than the IR model (t = 9.86 Ma), which is similar in age to some previous clock analyses (McGowen et al. 2009; Steeman et al. 2009; Slater et al. 2010; Hassanan et al. 2012). Most other divergences within the clade occur either shortly, thereafter, or in a somewhat simultaneous burst in the Late Miocene/Pliocene that corresponds to the major subfamilies (Fig. 3). Relationships within Delphinidae, the most speciose cetacean family, have been notoriously difficult to resolve, although several large-scale analyses in recent years have improved resolution markedly (McGowen et al. 2009; Steeman et al. 2009; McGowen 2011). We find continued support for three species at the base of Delphinidae (Leucopakrella acutas, O. orca, and Lagenorhynchus albinos; Fig. 2), although their configuration differs from an earlier large-scale analysis of the group, which places O. orca as sister to the remaining delphinids to the exclusion of both L. acutas and L. albinos (McGowen 2011). Neither L. acutas nor L. albinos is closely related to the other former members of Lagenorhynchus, now included within the genus Sagmatias (sensu Leduc et al. 1999; Vollmer et al. 2019; Fig. 2). Exclusive of these three species, we find overwhelming support for three major clades that roughly correspond to previously identified subfamilies: Lissodelphininae, Globicephalinae (with the inclusion of Grampus, Orcaca, and Steno), and Delphininae (with the inclusion of Sousa and Sotalia).
Relationships within Lissodelphininae (Lissodelphis, Sagmatias, Cephalorhynchus) resemble those of previous studies using both mitochondrial and nuclear data (McGowen et al. 2009; McGowen 2011). We were unable to include three lissodelphinine species here (Sagmatias cruciger, Cephalorhynchus eutropia, C. hectori); however, C. eutropia and C. hectori are consistently allied with Cephalorhynchus commersonii, and S. cruciger is well established as the sister species to S. australis (Pichler et al. 2001; Harlin-Cogno and Honeycutt 2006; May-Collado and Agranasson 2006; McGowen et al. 2009; Steeman et al. 2009; McGowen 2011; Banguera-Hinestroza et al. 2014; Vollmer et al. 2019). Both our concatenated and coalescence results imply that the current genera Cephalorhynchus and Sagmatias are paraphyletic and need further taxonomic revision. Cephalorhynchus heavisidii is more closely related to S. australis than to other members of Cephalorhynchus (May-Collado and Agranasson 2006; McGowen 2011), and it is likely that S. australis (and S. cruciger) will need to be transferred to Cephalorhynchus pending more complete sampling.

We find overwhelming support for the inclusion of the genera Orcella, Steno, and Grampus within the subfamily Globicephalinae, with Orcella and Steno diverging from other globicephalines in the Late Miocene (Fig. 3). Previous analyses of nuclear data supported the alliance of these genera with what were traditionally called the “Striped dolphins” (Globicephala, Sousa, Pseudorca, Peponocephala, Caballero et al. 2008; McGowen et al. 2008, 2009; Steeman et al. 2009; Banguera-Hinestroza et al. 2014). In addition, our results provide clear evidence that neither Orcella nor the “blackfish” is closely allied to Orcinus; globicephalines had been linked to Orcinus in the past (Leduc et al. 1999), and some mtDNA analyses linked Orcinus and Orcella (Leduc et al. 1999; Agranasson and May-Collado 2008). Although Steno is overwhelmingly grouped with globicephalines based on nuclear data, complete mt genomes have strongly supported its sister relationship with Sotalia and alliance with Delphininae (Cunha et al. 2011; Vilstrup et al. 2011), demonstrating extreme mitonuclear discordance that may have resulted from ancient introgression of the mitochondrial lineage. There has been difficulty in resolving the ∼ 14 currently recognized species within the subfamily Delphininae (Tursiops, Stenella, Sousa, Sotalia, Lagenodelphis, and Delphinus), likely due to rapid speciation and the documented presence of viable intergeneric hybrids in this group (Fig. 1; Perrin et al. 2013; Berubé and Palsbøll 2018).

In this study, we find high support for the polyphly of Stenella, with Tursiops, Delphinus, and Lagenodelphis nested within the genus (Fig. 2), as suggested by multiple previous molecular studies using mtDNA, nuDNA, or both (Fig. 1; Leduc et al. 1999; Caballero et al. 2008; Kingston et al. 2009; McGowen et al. 2009; Steeman et al. 2009; Xiong et al. 2009; McGowen 2011; Amaral et al. 2012). However, our analysis shows high support for Sotalia and Sousa as the most basal delphinine genera, as well as strong support for at least three other lineages that differ from previous analyses. One lineage of dolphins contains species with a distinctive contrasting pattern of patches and stripes, a “striped dolphin” lineage: Delphinus, Lagenodelphis, S. coeruleoalba, S. longirostris, and Stenella clymene. The other two lineages include the spotted dolphins (S. attenuata + S. frontalis) and the bottlenose dolphins (T. truncatus and T. aduncus), with both forming a monophyletic group with respect to the “striped” dolphin lineage. These three clades were recovered by Amaral et al. (2012) using coalescence analyses of 13 nuclear loci and mtDNA (Fig. 1), but relationships among these clades differed based on method (Fig. 2), and inclusion of more data could change the patterns presented here. The monophyly of the “striped” dolphin lineage has morphological support from at least seven cranial characters including a rostrum which is dorsoventrally compressed distally, small temporal fossae compared with other delphinines, and a grooved or slightly grooved palate (Perrin et al. 1981). Perrin et al. (1987) noted similarities in characters of the spotted and bottlenose dolphins including external coloration and cranial characters such as a smooth palate and large temporal fossae. Many previous studies have suggested synonymizing Lagenodelphis, Stenella, Tursiops, and sometimes Sousa with Delphinus (Leduc et al. 1999; Caballero et al. 2008; McGowen et al. 2009; McGowen 2011; Perrin et al. 2013), however, a formal redescription has sometimes been attempted due to the instability of relationships among phylogenetic studies. We suggest a less disruptive option by referring all species in the “striped” lineage (L. hosei, S. coeruleoalba, S. clymene, S. longirostris) to Delphinus, and retaining Stenella for the spotted dolphins (S. attenuata is the type species of Stenella [Perrin et al. 1987]) and Tursiops for the bottlenose dolphins.

Recently the validity of the species Delphinus capensis, the “long-beaked common dolphin”, has been called into question (Natoli et al. 2006; Kingston et al. 2009; Cunha et al. 2015; Farias-Curtidor et al. 2017), and the Society of Marine Mammalogy Committee on Taxonomy has recommended the use of D. delphis for all members of the genus (as used here) until further detailed analyses can be completed (Society for Marine Mammalogy Committee on Taxonomy 2017). The subspecies bairdii and tropicalis, both included in this study, had been referred to capensis, as they represented morphologically long-beaked forms (Heyning and Perrin 1994; Jepson and Van Waerebeek 2002). In agreement with Cunha et al. (2015), we find that the putative species D. capensis is paraphyletic with respect to D. delphis, although only four individuals of the genus are represented here. We find that the long-beaked dolphin of the Indian Ocean (subspecies tropicalis previously included within D. capensis [Jepson and Van Waerebeek 2002]) is more closely related to the representative short-beaked D. delphis delphis from the UK included here than either are to the long-beaked-type from California (D. delphis bairdii; 79929, 108471; Fig. 2).
One other difficult issue within Delphininae is the putative hybrid origin of the species *Stenella clymene*. Here, *S. clymene* is unequivocally well-supported as the sister species of *S. longirostris* based on substantial genomic data (Fig. 2). *Stenella clymene* was redescribed by Perrin et al. (1981), where it was noted that its external characteristics resembled *S. longirostris* but its skull resembled *S. coerulea*.*walba*. Molecular analyses of cytochrome *b* showed that *S. clymene* and *S. coerulea*.*walba* grouped together, adding support to a potential hybrid origin (Leduc et al. 1999), and subsequent analyses with mtDNA have allied *S. clymene* and *S. coerulea*.*walba* (Leduc et al. 1999, May-Collado and Agnarsson 2006; McGowen et al. 2009). Contrary to these findings, nuclear DNA in the form of AFLPs grouped *S. clymene* strongly with *S. longirostris* (Kingston et al. 2009; Fig. 1). Amaral et al. (2014) sequenced multiple individuals of all three species and showed that the cytochrome *b* sequence of most individuals of *S. clymene* were more closely related to *S. coerulea*.*walba*, but others were closer to *S. longirostris*; however, nuclear DNA from five loci showed little differentiation between the three taxa and could easily be explained by ancestral polymorphism. Our study sequenced only two individuals of *S. clymene*, but whole genome-scale sequencing of multiple representatives of all three species will likely be needed to properly address the question of its potential hybrid origin.

**CONCLUSION**

We targeted and assembled 3191 protein-coding genes from 68 species of cetaceans, two hippopotamids and three ruminants new to this study, and combined them with 18 existing genomes to produce the most comprehensive phylogenetic tree of cetaceans to date, in terms of the intersection of sequence data (38,167 exons; >6.5 million bp) and species (77 out of 89 total). Every node was well-resolved, including those within the problematic Delphinidae (true dolphins), although three nodes within the family differed between concatenated and coalescence analyses. Our results give clarity to a long debate over the contentious relationships among species currently contained within *Stenella, Lagenodelphis, Delphinus*, and *Tursiops*. Further analyses will seek to include the remaining 12 species; however, some of these were only recently split off from taxa represented here (e.g., *Sotalia fluvistilis, Sousa plumbea, S. sahulensis*), or are incredibly rare and represented by few specimens (*Mesoplodon traversii, M. hotulua*). Cetaceans are well-represented in the fossil record, and further studies will combine these new data with morphological and fossil data to produce a holistic view of cetacean evolution.

**SUPPLEMENTARY MATERIAL**

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.jq40bof

**ACKNOWLEDGMENTS**

We would like to thank W. Perrin for his input on an earlier version of this manuscript. We thank A. Devault and J. Enk (MYcroarray, now Arbor Biosciences) for help with constructing baits, as well as A. Lakeryd and J. Potter (QMUL) and M. Miller (CIPRES) for help with bioinformatics issues. We thank M. Bertelsen for the tissue of *Hippopotamus amphibius* from the Copenhagen Zoo, the Taipei Zoo for Arabian oryx and Arabian gazelle tissues, Zoological Society of London for access to bongo tissue, and K. Robertson (Southwest Fisheries Science Center [SWFSC]) for extractions and her help in obtaining approval from some of the original sample donors. We would like to acknowledge L. Balance, J. Barlow, K. Forney (SWFSC), T. Jefferson (Clymene Enterprises), the late P. Best (South African Museum), C. S. Baker (Oregon State University), R. Small and L. Quackenbush (Alaska Department of Fish and Game), E. Oleson (NMFS, Pacific Islands Fishery Science Center), K. West (University of Hawai‘i), M. P. Heide Jørgensen (Greenland Institute of Natural Resources), and D. Kreb (Yayasan Konservasi RASi) for the use of samples obtained through SWFSC. We thank the Taranaka Iwi and the New Zealand Cetacean Tissue Archive (NZCeTA), initiated by C.S. Baker and M.L. Dalebout, and maintained by the University of Auckland (R. Constantine, curator) for directly providing the sample of *Mesoplodon gingkodens* and the help of D. Steel (Oregon State University) for obtaining this sample. For all samples derived from Tasmania, we acknowledge Kris Carlson and Rachael Alderman (Princess Melikoff Trust, Marine Conservation Program, Tasmanian Department of Primary Industries, Parks, Water and Environment [DPIPWE]). We thank C. Buell for the illustrations and J. Gatesy for their use.

**FUNDING**

This research was funded by a Royal Society Newton International Fellowship to M.R.M. and S.J.R. and a European Research Council Starting grant (310482, EVOGENO) awarded to S.J.R.

**AUTHOR’S CONTRIBUTIONS**

M.R.M. and S.J.R. conceived of the study; R.D., P.D.J., S.J., A.P., and P.A.M. supplied tissues and logistical support; M.R.M. compiled genes and designed baits with the assistance of MYcroarray (now Arbor Biosciences); M.R.M. and M.S. performed laboratory work; M.R.M. and G.T. performed all bioinformatics analyses; M.R.M. performed all phylogenetic analyses; S.A.-C., M.d.R., and M.R.M. performed divergence dating analyses. M.R.M. and S.J.R. wrote the paper with contributions from the other authors.
### APPENDIX

#### TABLE A1. List of tissue sampled and taxa new to this study

| Species                              | Common name                  | ID              | Location                  | Lending institution | Institution of origin | Origin ID |
|--------------------------------------|------------------------------|-----------------|---------------------------|---------------------|------------------------|-----------|
| Cetacea                              |                              |                 |                           |                     |                        |           |
| *Balaenoptera bonaerensis*            | Antarctic minke whale        | 199219648       | Tasmania, Australia       | AAD                 |                        |           |
| *Balaenoptera borealis*              | Sei whale                    | SW2012/413      | Northumberland, England  | IOZ                 |                        |           |
| *Balaenoptera edeni*                 | Bryde’s whale                | 1380856971      | Tasmania, Australia       | AAD                 |                        |           |
| *Balaenoptera edeni*                 | Blue whale                   | Z66737          | Off Sinaloa, Mexico      | SWFSC               |                        |           |
| *Balaenoptera physalus*              | Fin whale                    | Z49099          | California, USA          | SWFSC               |                        |           |
| *Balaenoptera musculus*              |                               | SW1995-105      | Kent, England             | IOZ                 |                        |           |
| *Berardius arnuxii*                  | Arnoux’s beaked whale        | Z9128           | New Zealand               | SWFSC               | NZCeTA                 | BAR02     |
| *Berardius bairdii*                  | Baird’s beaked whale         | Z76728          | West Coast, USA           | SWFSC               |                        |           |
| *Cephalorhynchus commersonii*        | Commerson’s dolphin          | Z40             | Captive                   | SWFSC               |                        |           |
| *Cephalorhynchus heavisidii*         | Heaviside’s dolphin          | Z7320           | Yzerfontein, South Africa | SWFSC               | PBB 9622                |           |
| *Delphinapterus leucas*              | Beluga                       | Z55860          | Alaska, USA               | SWFSC               | ADFG BB0064-44         |           |
| *Delphinus delphis bairdii*          | N. Pac. long-beaked common dolphin | Z79929       | California, USA           | SWFSC               |                        |           |
| *Delphinus delphis leucas*           | Short-beaked common dolphin  | SW1999-92       | Devon, England             | IOZ                 |                        |           |
| *Delphinus delphis tropicalis*       | Indo-Pacific common dolphin  | Z4525           | Indian Ocean, off Oman    | SWFSC               |                        |           |
| *Eubalaena australis*                | Southern right whale         | Z13066          | Massachusetts, USA        | SWFSC               |                        |           |
| *Eubalaena glacialis*                | North Pacific right whale    | Z43864          | Alaska, USA               | SWFSC               |                        |           |
| *Feresa attenuata*                   | Pygmy killer whale           | Z145042         | Eastern North Pacific     | SWFSC               |                        |           |
| *Globicephala macrocephalus*         | Short-finned pilot whale     | Z39091          | California, USA           | SWFSC               |                        |           |
| *Globicephala melas*                 | Long-finned pilot whale      | SW1997-162      | Northumberland, England   | IOZ                 |                        |           |
| *Inia geoffrensis*                   | Amazon river dolphin         | Z305            | Acre, Brazil               | SWFSC               | USNM 57166              |           |
| *Kayia sinu*                         | Dwarf sperm whale            | Z12696          | Florida, USA               | SWFSC               |                        |           |
| *Kayia breviceps*                    | Pygmy sperm whale            | SW1997-159      | PembrokeShire, Wales      | IOZ                 |                        |           |
| Species                        | Common name          | ID      | Location            | Lending institution | Institution of origin | Origin ID |
|-------------------------------|----------------------|---------|---------------------|---------------------|------------------------|-----------|
| Lagenodelphis hosei           | Fraser’s dolphin     | Z452    | North Pacific       | SWFSC               | USNM                   | 500354    |
| Lagenodelphis hosei           | Fraser’s dolphin     | Z30470  | Hawaii, USA         | SWFSC               |                        |           |
| Lagenorhynchus australis      | Peale’s dolphin      | Z4926   | Cabo Espiritu Santo, Chile | SWFSC               |                        |           |
| Lagenorhynchus obliquidens     | Pacific white-sided dolphin | Z31902 | California, USA    | SWFSC               |                        |           |
| Lagenorhynchus obtusus        | Dusky dolphin        | Z2318   | Peru                | SWFSC               |                        |           |
| Lagenorhynchus (Leucocephalus) acutus | Atlantic white-sided dolphin | SW1998-90 | North Yorkshire, England | IOZ                 |                        |           |
| Lagenorhynchus albirostris    | White-beaked dolphin | SW1999-201A | Humberstone, England | IOZ                 |                        |           |
| Lisodelphis bairdii           | Southern right-whale dolphin | Z113034 | California, USA    | SWFSC               |                        |           |
| Mesoplodon bidens             | Sowerby’s beaked whale | Z9109   | New Zealand         | SWFSC               | NZCeTA                 | ZCA01     |
| Mesoplodon brevirostris       | Hubbs’s beaked whale | Z1363   | California, USA     | SWFSC               |                        |           |
| Meophodon danaeirostris        | Blainville’s beaked whale | SW1993-78 | Dyfed, Wales        | IOZ                 |                        |           |
| Mesoplodon eurypena           | Gervais’s beaked whale | Z7444   | Florida, USA        | SWFSC               |                        |           |
| Mesoplodon ginkgodens         | Ginkgo-toothed beaked whale | MgnNZ03 | Taranaki, New Zealand | NZCeTA               |                        |           |
| Mesoplodon grayi              | Gray’s beaked whale  | Z10210  | Tasmania, Australia  | AAD                 |                        |           |
| Mesoplodon hectori            | Hector’s beaked whale | Z9115   | New Zealand         | SWFSC               |                        |           |
| Mesoplodon hyenoides          | Strap-toothed whale  | 1763273011 | Tasmania, Australia | AAD                 |                        |           |
| Mesoplodon mirus              | True’s beaked whale  | Z4972   | New Jersey, USA     | SWFSC               | USNM                   | 504612    |
| Mesoplodon perini             | Perrin’s beaked whale | Z4976   | California, USA     | SWFSC               | USNM                   | 504259    |
| Mesoplodon peruvianus         | Pygmy beaked whale   | Z23629  | California, USA     | SWFSC               |                        |           |
| Mesoplodon stejnegeri         | Stejneger’s beaked whale | Z107244 | Alaska, USA         | SWFSC               |                        |           |
| Monoaon monodon               | Narwhal              | Z8293   | Uummannaq, Greenland | SWFSC               | GINR                   | GF16213   |
| Nipponochara phacocephalus    | Indo-Pacific finless porpoise | Z61334 | Hong Kong           | SWFSC               |                        |           |
| Orcella brevirostris          | Irrawaddy dolphin    | Z7205   | Mekong River, Laos  | SWFSC               |                        |           |
| Orcella heinsohni             | Australian snubfin dolphin | Z2907  | Queensland, Australia | SWFSC               |                        |           |
| Pinnipedia hectori            | Melon-headed whale   | Z4110   | Hawaii, USA         | SWFSC               |                        |           |
| Pinnipedia hookeri            | Spectacled porpoise  | Z981    | Est. Las Viñetas, Argentina | SWFSC               |                        |           |
| Phocoena phocoena              | Harbor porpoise      | SW2000-104 | Ceredigion, Wales  | IOZ                 |                        |           |
| Phocoena spinipinnis           | Burmeister’s porpoise | Z1092   | Peru                | SWFSC               |                        |           |
| Species                  | Common name               | ID     | Location            | Lending institution       | Institution of origin | Origin ID |
|-------------------------|---------------------------|--------|---------------------|---------------------------|------------------------|-----------|
| Phocoenoides dalli      | Dall’s porpoise           | Z4824  | California, USA     | SWFSC                     | SWFSC                  |           |
| Pontoporia blainvillei  | Franciscana               | Z7349  | Pygmy               | SWFSC                     |                        |           |
| Sotalia guianensis      | False killer whale        | Z123188| Molokai, Hawaii, USA| SWFSC                     | KW                     | KW2010019 |
| Sousa chinensis         | Guiana dolphin            | Z9837  | Natal, Brazil       | SWFSC                     | HKB42                  |           |
| Phocoenoides dalli      | Dall’s porpoise           | Z4824  | California, USA     | SWFSC                     | SWFSC                  |           |
| Stenella attenuata      | Pantropical spotted dolphin| Z18473 | Tropical Eastern Pacific | SWFSC                   |                        |           |
| Stenella attenuata      | Pantropical spotted dolphin| Z38219 | Tropical Eastern Pacific | SWFSC                   |                        |           |
| Stenella clymene        | Clymene dolphin           | Z1724  | Gulf of Mexico      | SWFSC                     |                        |           |
| Stenella clymene        | Clymene dolphin           | Z1726  | Gulf of Mexico      | SWFSC                     |                        |           |
| Stenella soreola bella  | Striped dolphin           | Z18431 | Tropical Eastern Pacific | SWFSC                   |                        |           |
| Stenella frontalis      | Atlantic spotted dolphin  | Z1724  | Gulf of Mexico      | SWFSC                     |                        |           |
| Stenella longirostris    | Spinner dolphin           | Z16012 | Tropical Eastern Pacific | SWFSC                   |                        |           |
| Stenella longirostris    | Spinner dolphin           | Z24923 | Tropical Eastern Pacific | SWFSC                   |                        |           |
| Steno bredanensis       | Rough-toothed dolphin     | Z18431 | Tropical Eastern Pacific | SWFSC                   |                        |           |
| Steno bredanensis       | Rough-toothed dolphin     | Z116871| Saipan, N. Marianas Islands | PIFSC                   | PIC130720.01B          |           |
| Tasmacetus shepherd     | Shepherd’s beaked whale  | Z4971  | Chubut, Argentina   | SWFSC                     | USNM                   | 484878    |
| Tursiops aduncus         | Indo-Pacific bottlenose dolphin | Z79924 | Berau Archipelago, Indonesia | SWFSC                   | YK-RASI                | TAD080423 |
| Ziphius cavirostris      | Cuvier’s beaked whale     | Z18431 | Tropical Eastern Pacific | SWFSC                   |                        |           |
| Hippopotamidae           | Common                    | Captive| CZ                  |                           |                        |           |
| Hippopotamus amphibius   | hippopotamus              | WHMO71/0546/50 | Captive| ZSL |           |
| Choeropsis libyensis    | Pygmy                     | Captive| Taipei Zoo          |                           |                        |           |
| Oryx leucoryx           | Arabian oryx              | zm693/05/4369 | Captive| ZSL |           |
| Trigalepus surculus     | Bongo                     | 20080367M10 | Captive| Taipei Zoo |           |
| Gazella subaica         | Arabian gazelle           | zm634/0821/7/08 | Captive| Taipei Zoo |           |

AAD, Australian Antarctic Division; ADFG, Alaska Department of Fish and Game; CZ, Copenhagen Zoo; CINR, Greenland Institute of Natural Resources; IOZ, Institute of Zoology; Zoological Society of London; KW, Kristi West, University of Hawai‘i; NZCET, New Zealand Cetacean Tissue Archive; PBB, Peter Best, South African Museum; PIFSC, NOAA, Pacific Islands Fisheries Science Center; SWFSC, NOAA, Southwest Fisheries Science Center; TJ, Thomas Jefferson, Clymene Enterprises; USNM, Smithsonian National Museum of Natural History; YK-RASI, Yayasan Konservasi RASI; ZSL, Zoological Society of London.
whale from the Upper Pliocene (Piacenzian) of California, USA, with observations on the evolution and relationships of the Cetotheriidae. Zool. J. Linn. Soc. 170:400–466.

Ellis R, Mead JG. 2017. Beaked whales: a complete guide to their biology and conservation. Baltimore (MD): Johns Hopkins University Press. 194 pp.

Farias-Olivares, ạ, Barragán-Díaz, C. Chávez-Carreño PA, Jiménez-Pinedo C, Palacios D.M, Caicedo D, Trujillo F, Caballero S. 2017. Range extension for the common dolphin (Delphinus sp.) to the Colombian Caribbean, with taxonomic implications from genetic barcoding and phylogenetic analyses. PLoS One 12:e0178100.

Foote AD, Liu Y, Thomas G.W, Vinal T, Alfoldi J, Dong J, Dugan D, van Els C.E, Hunter M.E, Joshi V, Khan Z, Kvarc S, Lee SL, Lindblad-Toh K, Manica A, Nielsen R, Qin X, Qu J, Raey B, Vijay R, Wei W, Wong D.M, Worthy K.C, Gilbert M.T.P, Gibbs R.A. 2015. Convergent evolution of the genomes of marine mammals. Nat. Genet. 47:272–275.

Fordyce RE, Mars P.G. 2013. The pygmy right whale Cepaea mirabilis: the last of the cetotheres. Proc. R. Soc. B.280:201222645.

Gaines C, Hare M, Beck S, Rosenbaum H.B. 2005. Nuclear markers confirm taxonomic status and relationships among highly endangered and closely related right whale species. Proc. R. Soc. B 272:533–542.

Gasc C, Peyretaille E, Peyret P. 2016. Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. Nucleic Acids Res. 44 4504–4518.

Gatesy J, Geisler JH, Chang J, Buell C, Barta E, Meredith RW, Springer MS, McGowen MR. 2013. A phylogenetic blueprint for a modern whale. Mol. Phylogenet. Evol. 66:479–506.

Gatesy J, O’Leary MA. 2001. Deciphering whale origins with molecules and fossils. Trends Ecol. Evol. 16:562–570.

Ge R-L, Cai Q, Shen Y-Y, San A, Ma L, Zhang Y, Yi X, Chen Y, Yang L, Huang Y, He R, Hui M, Wang Y, Wu X, Xu J, Zhang Y, Wu K, Gong C, Zhou W, Zhou T, Irwin DM, Yang Y, Ying L, Bao H, Kim J, Larkin DM, Ma J, Lewin HA, Xing J, Platt RN, Ray DA, Ausili L, Capitanu B, Zhang X, Zhang G, Murphy RW, Wang J, Zhang YP, Wang J. 2013. Draft genome sequence of the Tibetan antelope. Nat. Commun. 4:1858.

Geisler JH, Colbert MW, Carew JL. 2014. A new fossil species of dolphins in the subfamily Lissodelphininae: character synergy improves phylogenetic resolution. BMC Evol. Biol. 6:87.

Hasegawa M, Kishino H, Yano T. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

Hassanin A, Deluc S, Roopieti A, Hammer C, Jansen van Vuuren B, Matthews C, Ruiz-Caracha M, Castréfis F, Arekoukou V, Nguyen TT, Coulois A. 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial and nuclear DNA. C. R. Biol. 335:52–50.

Heyning JE, Perrin WF. 1994. Evidence for two species of common dolphins (genus Delphinus) from the eastern North Pacific. Contrib. Sci. 442:1–35.

Iooze J, Donoghue PJC, Yang Z. 2010. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. Syst. Biol. 59:74–89.

Jackson JA, Baker CS, Vant M, Steel D J, Medrano-Gonzalez L, Palumbi SR. 2009. Big and slow: phylogenetic estimates of molecular evolution in baleen whales (Suborder Mysticeti). Mol. Biol. Evol. 26:2427–2440.

Jefferson TA, Van Waerebeek K. 2002. The taxonomic status of the nominal dolphin species Delphinus tropicalis Van Bree, 1971. Mar. Mamm. Sci. 18:507–418.

Jefferson TA, Webber MA, Pitman RL. 2015. Marine mammals of the world: a comprehensive guide to their identification. 2nd ed. London: Academic Press. 666 pp.

Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzyr DM, Li Y, Zhang W, Stanton J-A, Brauning R, Barris WC, Hourtien T, Aken B-L, Searle S.MJ, Adelson DL, Bian C, Cam GR, Chen Y, Cheng S, Desilva U, Dosen K, Dong Y, Fan G, Franklin JR, Fu S, Fuentes-Utrilla P, Guan R, Highland MA, Holder M.E, Huang G, Ingham AB, Jiangani SN, Krafla D, Kvarc C-L, Lee SL, Liu W, Liu X, Lu C, Lv T, Mathew T, McWilliam S, Menzies M, Pas S, Robert D, Servin D, Bowler T, Wang D, Wang W, Wei B, White SN, Yang X, Ye C, Yue Y, Zeng F, Zhou Q, Hansen JB, Kristensen K, Gibbs RA, Fluck F, Waskup CC, Jones HE, Oddy VH, Nicholas FW, McEwan JW, Kajas W, Wang J, Worthy KC, Archibald AL, Cockett N, Xu X, Wang W, Dalrymple BF. 2014. The sheep genome illuminates biology of the ruminant and lipid metabolism. Science 344:1168–1173.

Kabat K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–790.

Keane M, Semkejs J, Webb AE, Li YJ, Quesada V, Craig T, Madsen L.B, van Dam S, Bravard C, Miao-See R, Michalak P, Kang L, Bhat J, Yim HS, Grishin NV, Nielsen NH, Heide-Jorgensen MP, Ozolins EM, Matson CW, Church GM, Stewart GW, Patton JC, George JC, Suydam JR, Larsen K, Lopez-Otin C, O’Connell MJ, Bichakjian W, Thomson B, de Magallanes P. 2015. Insights into the evolution of longevity from the bovid sheep genome. Cell Rep. 10:112–122.

Kellogg R. 1927. Kentriodon pernix, a Miocene pinniped from Maryland. Proc. U.S. Nat. Mus. 69:1–14.

Kingston SE, Adams LD, Rosel PE. 2009. Testing mitochondrial sequences and anonymous nuclear markers for phylogeny reconstruction in a rapidly radiating group: molecular systematics of the Delphinidae (Cetacea: Odontoceti: Delphinini). BMC Evol. Biol. 9:245.

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.
Lambert O., Martínez-Cáceres M., Bianucci G., Di Celma C., Salas-Gismondi R., Steurbeaut M., Urbino M., de Muizon C. 2017. Earliest mysticete from the Late Eocene of Peru sheds new light on the origin of baleen whales. Curr. Biol. 27:1533–1541.e2

Lammers F., Blumer N., Rücklé C., Nilsson M. 2019. Retrophylogenetics inCORDs indicate large ancestral population sizes and a rapid radiation. Mol. DNA 10.5

Lanjier F., Fransen P.B., Wright A.M., Senfeld T., Calcott B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34:772–773.

Ledoux R.C., Perrin W.F., Dehn A.E. 1999. Phylogenetic relationships among the delphinid cetaceans based on full cytochrome b sequence. Mar. Mam. Sci. 15:619–648.

Lindblad-Toh K., Garber M., Zuk O., Lin M.F., Parker B.J., Washietl S., Kheradpour P., Ernst J., Kent W.J., K saber A., Li Y., Hinds J., Marth G., Myers R., Hurles M., Durbin R. 2005. Evolution and genomic diversity in a model for mammals. Nature 437:1193–1197.

Lindblad-Toh K., Garber M., Zuk O., Lin M.F., Parker B.J., Washietl S., Kheradpour P., Ernst J., Kent W.J., K saber A., Li Y., Hinds J., Marth G., Myers R., Hurles M., Durbin R. 2005. Evolution and genomic diversity in a model for mammals. Nature 437:1193–1197.

McGowen M.R., Spaulding M., Gatesy J. 2009. Divergence date estimation with many hundreds of taxa and thousands of genes. Syst. Biol. 58:24–37.

McGowen M.R., Gatesy J., Wildman D.E. 2014. Molecular evolution strategies for next-generation sequencing. Nat. Methods 7:111–118.

McGowen M.R., Spaulding M., Gatesy J. 2016. A phylogeny of birds based on over 15,000 loci collected by target enrichment and high-throughput sequencing. PLoS One 11:e0164889.

McGowen M.R. 2011. Toward the resolution of an explosive radiation—a molecular phylogeny of oceanic dolphins (Delphinidae). Mol. Phylogenet. Evol. 60:345–357.

McGowen M.R., Clark C., Gatesy J. 2008. The vestigial olfactory receptor subgenome of odontocete whales: phylogenetic congruence between gene-tree reconciliation and supertree methods. Syst. Biol. 57:524–540.

McGowen M.R., Gatesy J., Wilmot D.E. 2014. Molecular divergence tracks macroevolutionary transitions in Cetacea. Trends. Biol. 29:126–136.

McGowen M.R., Spaulding M., Gatesy J. 2009. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. Mol. Phylogenet. Evol. 53:891–906.

Meredith R.W., Janecke J.E., Geyte S., Ryder O.A., Fisher C.A., Teeling E.C., Goodall A., Ezitzk E., Simao T.L.L., Ladler T., Reader D., Hart R., Purcell L., Jones C., Grigore C.M., Steinke C., Williams T.L., Robinson T.J., Burke Herrick A., Westerman M., Ayoub N.A., Murphy S.M., Murphy W.J. 2011. Impacts of the Cretaceous—Tertiary extinction on mammal diversification. Science 334:521–524.

Miller, M.A., Pfeiffer W., Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of whale phylogeny, diversity and disparity. R. Soc. Open Sci. 32:140434.

Monte S., Hasegawa M., Okada N. 2005. Baleen whale phylogeny and the paraphyly of river dolphins. Proc. Nat. Acad. Sci. USA 92:7384–7389.

Moran N.A., Springer M.S., Murphy W.J. 2011. Impacts of the Cretaceous—Tertiary extinction on mammal diversification. R. Soc. Open Sci. 32:140434.

Nikaido M., Pakhomov E., Brownell R.L., Cauoy C., Ding W., Zuyan Z., Schiold A.M., Brandweck R.E., Hasegawa M., Okada N. 2001. Repropon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. Mol. Phylogenet. Evol. 20:943–954.

Nishida S., Goto, M., Kanda N., Sasaki T., Takeda T. 2014. A new kentriodontid (Cetacea: Odontoceti) from the early Miocene Mouaca Astorita Formation and a revision of the stem delphinid family Kentriodontidae. J. Vert. Paleontol. 34:711375.

Perrin W.F., Mitchell E.D., Moad J.G., Caldwell D.K., Caldwell M.C., Bree FF.H., Dawhin W.H. 1987. Resuration of the spotted dolphins. Sterna spp. Mar. Mam. Sci. 3:199–207.

Perrin W.F., Mitchell E.D., Moad J.G., Caldwell D.K., van Bree FF.H. 1981. Sterna clavი, a rediscovered tropical dolphin of the Atlantic. J. Mammal. 56:583–596.

Perrin W.F., Rose PE, Cipriano F. 2013. How to contend with paraphyly in the taxonomy of the delphinine cetaceans? Mar. Mam. Sci. 29:656–718.

Pichler FB, Robineau D, Godall N.P., Meyer M.A., Oltavari C., Baker C.S. 2001. New ctenoid (Cetacea: Odontoceti) from the early Miocene Mouaca Astorita Formation and a revision of the stem delphinid family Kentriodontidae. J. Vert. Paleontol. 34:1311375.

Portik D.M., Smith L.L., Bi K. 2016. An evaluation of transcriptome-based exon capture for phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). Mol. Ecol. Resour. 16:1069–1083.

Puttick M.N. 2019. MCMCtree: functions to prepare MCMCtree analyses and visualize posterior ages on trees. Bioinformatics, https://doi.org/10.1093/bioinformatics/btz554.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M. A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904.

Reid R.C., Bree FF.H., Cipriano F. 2013. How to contend with paraphyly in the taxonomy of the delphinine cetaceans? Mar. Mam. Sci. 29:656–718.

Rohde D., Smith L.L., Bi K. 2016. An evaluation of transcriptome-based exon capture for phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). Mol. Ecol. Resour. 16:1069–1083.

Puttick M.N. 2019. MCMCtree: functions to prepare MCMCtree analyses and visualize posterior ages on trees. Bioinformatics, https://doi.org/10.1093/bioinformatics/btz554.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901–904.
Smith S.A., Brown J.W., Walker J.F. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. PLoS One 13:e0197433.

Society for Marine Mammalogy Committee on Taxonomy. 2017. List of marine mammal species and subspecies. Society for Marine Mammalogy. Available from: www.marinemammalscience.org.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Stemman M.J. 2007. Cladistic analysis and a revised classification of fossil and recent mysticetes. Zool. J. Linn. Soc. 150:875-894.

Steeman M.E. 2007. Cladistic analysis and a revised classification of cetartiodactyl mammals. R. Soc. Open Sci. 2:150156.

Surve T.P., Pitman R.L., Taylor B.L., Barlow J., Akamatsu T., Barrett T., Tsagkogeorga G., McGowen M.R., Davies K.T.J., Jarman S., Polanowski A., Thomas L., Jaramillo-Legorreta A., Cardenas-Hinojosa G., Nieto-Thewissen J.G.M., Cooper L.N., George J.C., Bajpai S. 2009. From land to water: the origin of whales, dolphins, and porpoises. Evol. Educ. Outreach 2:272–289.

Thomas L., Jamrillo-Legorreta A., Cardenas-Hinojosa G., Nieto-Garcia E., Rojas-Bracho L., Ver Hoef J.M., Moore J., Taylor B., Barlow J., Tregenza T. 2017. Last call: passive acoustic monitoring shows continued rapid decline of critically endangered vaquita. J. Acoust. Soc. Am. 142:EL512–EL517.

Tsagkogeorga G., McGowen M.R., Davies K.T.J., Jarman S., Polanowski A., Bertelsen M.F., Rossiter S.J. 2015. A phylogenomic analysis of the dolphin genus Lagenorhynchus. Mar. Mammal Sci. 35:957–1057.

Vilstrup J.T., Ho S.Y., Foote A.D., Morin P.A., Kreb D., Kruzen M., Parra V., Anderson S., Robertson K.M., de Stephanis R., Verboogh P., Willerslev E., Orlando L., Gilbert M.T.P. 2011. Mitogenomic phylogenetic analyses of the Delphinidae with an emphasis on the Globicephalinae. BMC Evol. Biol. 11.

Vollmer N.L., Ashe E., Brownell R.L. Jr, Cipriano F., Mead J.G., Reeves RR., Soldevilla M.S., Williams R. 2019. Taxonomic revision of the dolphin genus Lagenorhynchus. Mar. Mammal Sci. 35:957–1057.

Waddell V.G., Milinkovitch M.C., Bérubé M., Stanhope M.J. 2000. Molecular phylogenetic examination of the Delphinidae: 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.

Warren W.C., Kuderna L., Alexander A., Catchen J., Perez-Silva J.G., Lopez-Otin C., Quesada V., Miró P., Tomlinson C., Montague M.J., Farias F.H.G., Walter R.B., Marques-Bonet T., Glenn T., Kieran T.J., Wise S.S., Wise J.P. Jr, Waterhouse R.M., Wise J.P. Sr. 2017. The novel evolution of the sperm whale genome. Genome Biol. Evol. 9:3260–3264.

Wilson, L.E. 1973. A delphinid (Mammalia, Cetacea) from the Miocene of Palos Verdes Hills, California. Univ. Calif. Publ. Geol. Sci. 103:1–33.

Wex M., Lewis P.O., Fan Y., Kuo L., Chen M.-H. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60:150–160.

Xiong Y., Brandley M.C., Xu S., Zhou K., Yang G. 2009. Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales. BMC Evol. Biol. 9:20.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1596–1599.

Yang Z., Ranalla B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol. 23:212–226.

Yim H.-S., Cho Y.S., Guan X., Kang S.G., Jeong J.-Y., Cha S.-S., Oh H.-M., Lee J.-H., Yang E.C., Kwon K.K., Kim Y.J., Kim T.W., Kim W., Jeon J.H., Kim S.-J., Choi D.H., Joo S., Kim H.-M., Ko J., Kim H., Shin Y.-A., Jung H.-J., Zheng Y., Wang Z., Chen Y., Chen M., Jiang A., Li E., Zhang S., Hou H., Kim T.H., Yu L., Liu S., Ahn K., Cooper J., Park S.-G., Hong C.P., Iin W., Kim H.-S., Park C., Lee K., Chun S., Morin P.A., O’Brien S.J., Lee H., Kimura J., Moon D.Y., Moná N., Edwards J., Kim B.C., Kim S., Wang J., Bhab J., Lee H.S., Lee J.-H. 2013. Minke whale genome and aquatic adaptation in cetaceans. Nat. Genet. 46:88–92.

Zhang C., Rabbee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Biol. 19:153.

Zhou X., Guang X., Guan X., Sun D., Xu S., Li M., Seim I., Jie W., Yang L., Zhu Q., Xu J., Gao Q., Kaya A., Dou Q., Chen B., Ren W., Li S., Zhou K., Gladsgyve VN., Nielsen R., Fang X., Yang G. 2018. Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. Nat. Commun. 9:1276.

Zhou X., Guang X., Fu D., Kuo L., Chen M.-H., Feng Z., Wang J., Pan S., Fang X., Li M., Wei F., Xu X., Zhou K., Wang J., Yang G. 2013. Baiji genomes reveal low genetic variation and new insights into secondary aquatic adaptations. Nat. Commun. 4:2708.

Zhou X., Xu S., Yang Y., Zhou K., Yang G. 2011 Phylogenomic analyses and improved resolution of Cetartiodactyla. Mol. Phylogenet. Evol. 61:255–264.