Molecular Phylogenetics and Evolution 133 (2019) 256–262

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Cetartiodactyla: Updating a time-calibrated molecular phylogeny

Juan P. Zurano[a,1,*], Felipe M. Magalhães[a,1], Ana E. Asato[b], Gabriel Silva[b], Claudio J. Bidau[c], Daniel O. Mesquita[d], Gabriel C. Costa[e]

[a] Programa de Pós-Graduação em Ciências Biológicas, Universidade Federal da Paraíba-UFPB, Centro de Ciências Exatas e da Natureza, CEP 58000-000, João Pessoa, Paraíba, Brazil
[b] Programa de Pós-Graduação em Ecologia, Universidade Federal de Rio Grande do Norte-UFRN, Centro de Biociências, DECOL, CEP 59078-900, Natal, Rio Grande do Norte, Brazil
[c] Paraná y Los claveles, 3304 Garupá, Misiones, Argentina
[d] Departamento de Sistemática e Ecologia, Universidade Federal da Paraíba-UFPB, Centro de Ciências Exatas e da Natureza, CEP 58000-000, João Pessoa, Paraíba, Brazil
[e] Department of Biology and Environmental Sciences, Auburn University at Montgomery, Montgomery, AL 36124, United States

ARTICLE INFO

Keywords:
Mammals
Cetacea
Artiodactyla
Mitogenomics
Phylogenetic relationships

ABSTRACT

Cetartiodactyla comprises one of the most diverse mammal radiations. Currently, 23 families, 131 genera and more than 330 species are recognized. Several studies have been trying to resolve its phylogenetic relationships. The most comprehensive dated phylogenetic hypothesis available includes only 55% of the extant species, precluding a clear understanding of ecological and evolutionary patterns in Cetartiodactyla. Here, we gathered all mitochondrial genetic data available in GenBank to build a robust Cetartiodactyla calibrated phylogenetic tree using 21 fossil calibration points. We found mitogenomic data for 225 species and included other 93 species from which there was at least one mitochondrial gene available. Using a Bayesian approach, we generated a dated tree comprising 90% of the extant Cetartiodactyla species (n = 318). The major lineages showed robust support and families divergence times are congruent with the available fossil evidence and with previously published phylogenetic hypotheses. By making available a dated phylogeny with extensively sampled clades, we expect to foster future studies on the origin, tempo and mode of Cetartiodactyla diversification.

1. Introduction

Understanding the tempo and mode of diversification has been of utmost interest in evolutionary biology. Efforts from researchers to accurately estimate phylogenetic relationships, and the ever-growing availability of molecular sequences and paleontological data boosted the development of sophisticated tree reconstruction methods (e.g. dos Reis et al., 2015). Likewise, continuum development on molecular clock methods has increased the availability of more accurate phylogenetic dated trees (Bromham et al., 2018). Calibrated trees allows inference of diversification patterns and are essential for macroevolutionary studies and comparative phylogenetic methods (e.g. Garamszegi et al., 2014). As an example of their resourcefulness, the use of dated trees revealed that diversification rates can explain differences in species richness among clades (Scholl et al., 2016), and helped to understand the role of historical factors on evolutionary radiations events (Meredith et al., 2011).

The order Cetartiodactyla currently comprises 23 families, 131 genera and over 330 species (IUCN, 2018). This order represents an impressive example of adaptive radiation, and is a well-studied group regarding its morphology, ecology and behavior (Berta et al., 2015; Prothero and Foss, 2007). The close relationship between Artiodactyla and Cetacea, both taxa currently included within Cetartiodactyla order, has been recovered in previous studies using molecular (mitochondrial and nuclear DNA) and paleontological data (e.g. Gatesy et al., 1999; Graur and Higgins, 1994; Montgelard et al., 1997). Since then, molecular phylogenetic studies have focused on the resolution of the relationships at the suprafamilial level without considering taxa diversity within the order (Zhou et al., 2011). Another goal of more recent studies has been the reconstruction of major Cetartiodactyla lineages (e.g. Bovidae: Bibi, 2013; Ancodonta: Boisserie et al., 2011; Ruminantia: Fernández and Vrba, 2005; Suina: Frantz et al., 2016; Cervidae: Gutiérrez et al., 2017; Cetacea: Steeman et al., 2009; McGowen et al., 2009).

Molecular data for Cetartiodactyla has increased in the last ten years, and sequences of extinct species (e.g. *Bison priscus*, *Bos*

* Corresponding author.
E-mail addresses: zuranojp@gmail.com (J.P. Zurano), danmesq@dse.ufpb.br (D.O. Mesquita), gcosta@aum.edu (G.C. Costa).
[1] These authors also contributed equally to this work.

https://doi.org/10.1016/j.ympev.2018.12.015
Received 11 May 2018; Received in revised form 11 December 2018; Accepted 12 December 2018
Available online 15 December 2018

primigenius) and recently described species (e.g., *Inia araguaiaensis, Tursiops australis*) are now available. Additionally, fossil records are also abundant and representative for almost all Cetartiodactyla clades which certainly improve estimates of divergence time (Bibi, 2013). The work by Hassanin et al., (2012) is currently the most comprehensive molecular phylogenetic study in terms of species representativeness (~55%). Our goal is to use all mitochondrial information available coupled with fossil information to generate a dated tree covering more than 90% of Cetartiodactyla extant species. We hope that our results will contribute to a better understanding of Cetartiodactyla diversification patterns, and also foster studies in diverse fields such as macroevolution, biogeography and ecological studies relying on comparative phylogenetic methods.

2. Methods

2.1. Sampling of species and data collection

We collected all mitochondrial (mtDNA) available in GenBank for Cetartiodactyla (until June 2018). We aimed for complete mitochondrial genomes (mitogenomes) keeping only one per species. To favor comparisons with previously published mitogenomic phylogenetic inferences, we prioritize mitogenomes used in Hassanin et al. (2012). We gathered mitogenomes for 225 Cetartiodactyla species, and found at least one mtDNA gene for other 93 species (Table 1, Table S1 in Appendix A for GenBank accession numbers). We included the domestic species *Bubalus bubalis* (domestic form of wild *B. arnee*) and *Camelus dromedarius*, because wild forms sequences of those species were not available. Furthermore, our study includes 6 species not listed in the IUCN database and that were not included in any previous Cetartiodactyla published phylogenetic hypothesis; an extinct species (*Bison priscus*) and five recently described species (*Inia araguaiaensis*, *Tursiops australis*, *Giraffa giraffa*, *Giraffa reticulata*, and *Giraffa tippelskirchi*). As outgroup (OU), we used three mitogenomes of Perissodactyla (*Rhinoceros unicornis*, *Equus zebra* and *Tapirus indicus*), which shares a most recent common ancestor with Cetartiodactyla (Meredith et al., 2011; Wu et al., 2014).

Table 1
Species sampled to generate the time–calibrated phylogeny of Cetartiodactyla. Family: main extant families in the order. Species: Number of species currently recognized (according to IUCN, 2018) by family. Our tree: species included in our larger dataset (dataset1). Mitogenome: mitogenomes used for each family in our analyses. Family NA is a percentage of missing–data by family.

Family	Species	Our tree	Mitogenome	Family NA (%)
Tayassuidae	3	3	1	44.0
Suidae	18	14	7	36.9
Camelidae	4	4	4	0
Hippopotamidae	4	2	2	0
Balaenidae	4	4	4	0
Neobalaenidae	1	1	1	0
Balaenopteridae	8	8	8	0
Eschrichtiidae	1	1	1	0
Physeteridae	3	3	2	30.8
Platanistidae	1	1	1	0
Ziphiidae	21	21	8	58.0
Lipotidae	1	1	1	0
Pontoporidae	1	1	1	0
Iniidae	2	2	1	46.2
Monodontidae	2	2	2	0
Phocoenidae	7	7	3	49.7
Delphinidae	37	37	22	21.9
Tragulidae	10	6	4	16.5
Antilocapridae	1	1	1	0
Giraffidae	5	5	2	54.9
Cervidae	55	52	35	29.7
Moschidae	7	6	4	28.9
Bovidae	143	136	110	16.2

2.2. Phylogenetic analyses and divergence time estimation

We aligned the sequences using MAFFT algorithm default configuration (http://mafft.cbrc.jp/alignment/server/). We used Geneious v1.7.9 program (Kearse et al., 2012) to manually eliminate mitochondrial genome poorly aligned regions (non-coding genes and D-loop region). The final DNA alignment of 14,935 bp was split into five partitions according to transcript type (tRNA, 1443 bp; rRNA 2134 bp; protein-coding mRNA, 11358 bp) and all protein-coding genes subdivided into codon position (CODP1; CODP2; and CODP3, 3786 bp each). To acknowledge if the addition of missing data may have any impact on tree topology, node support and divergence time estimates, we created two datasets for all subsequent analyses: (i) dataset1 including all mitochondrial information available in Genbank (321 species with OU), being 23% of this dataset composed by missing data, and (ii) dataset2 maintaining only taxa with full mitogenome data (228 species with OU; Table 1). To account for evolutionary rate heterogeneity among partitions, we used PartitionFinder2 (PF) with the greedy algorithm, linked branch lengths along with Bayesian Information Criterion to select the best-fit partitioning scheme model of nucleotide substitution (Lanfear et al., 2016). For both datasets, PF suggested four partitions (tRNA + rRNA, CODP1, CODP2 and CODP3) and selected the GTR + Γ + I as best-fit model of nucleotide substitution for all four partitions. Finally, we estimated phylogenetic relationships along with divergence time under Bayesian approach using BEAST v.1.10.1 (Suchard et al., 2018). We analyzed both datasets under a relaxed uncorrelated lognormal clock (Drummond et al., 2006) allowing branches to have independent substitution rates. Because we implemented multiple fossil calibrations (see below), we let BEAST calculate the relative mean clock rate for each partition without providing prior information. Additionally, because extinction has a major impact on lineage diversification patterns (Quental and Marshall, 2010), we used a birth-death speciation model as tree prior.

To date the tree, we used 21 fossils (see Table S1 in Appendix B for prior distributions and fossil sources) that were either incorporated in phylogenetic analyses or recommended as calibration points by previous paleontological revisions (e.g. Bibi 2013; Benton et al., 2015). For both datasets, we ran three independent replications with 300 million generations each, sampling every 15,000 interactions. To check runs convergence and the effective sampled size (ESS > 200) we used Tracer v1.7 (http://beast.community/tracer). We combined all replications for both datasets discarding the first 20% of each run as burn-in. To generate the maximum clade credibility tree, we used TreeAnnotator v1.10.1 (http://beast.community/treeannotator). To assure that our posterior distributions were not heavily affected by prior settings, we evaluated effective priors by running the analyses without the data (sampling from prior in BEAST) to confirm that our results were mainly obtained by our data sets (Drummond and Bouckaert, 2015).

3. Results and Discussion

3.1. Phylogenetic relationships and node support

Both of our datasets yielded phylogenies with nearly identical topologies (Fig. 1, Fig. S1) that differ slightly in divergence time estimates (Table 2). Despite that, the confidence interval (sampled from the 95% HPD) of Time to Most Recent Common Ancestor (TMRCA) estimated for main Cetartiodactyla lineages are overlapping when datasets are compared (Table 2). The mean mutation rate (*ucld.mean* parameter in BEAST) estimated for both datasets was ~ 9×10^{-3} sites/Ma (95% highest posterior density [95% HPD] = 9.1×10^{-3}–1.0×10^{-2}) or roughly 2% total mitochondrial DNA divergence per Ma (see Table S2 in Appendix B, for partitions clock rates from each dataset). We used a representative number of mitogenomic sequences per lineage and multiple calibration points evenly distributed across the phylogeny, which certainly favored such concordances (Table 2). Therefore,

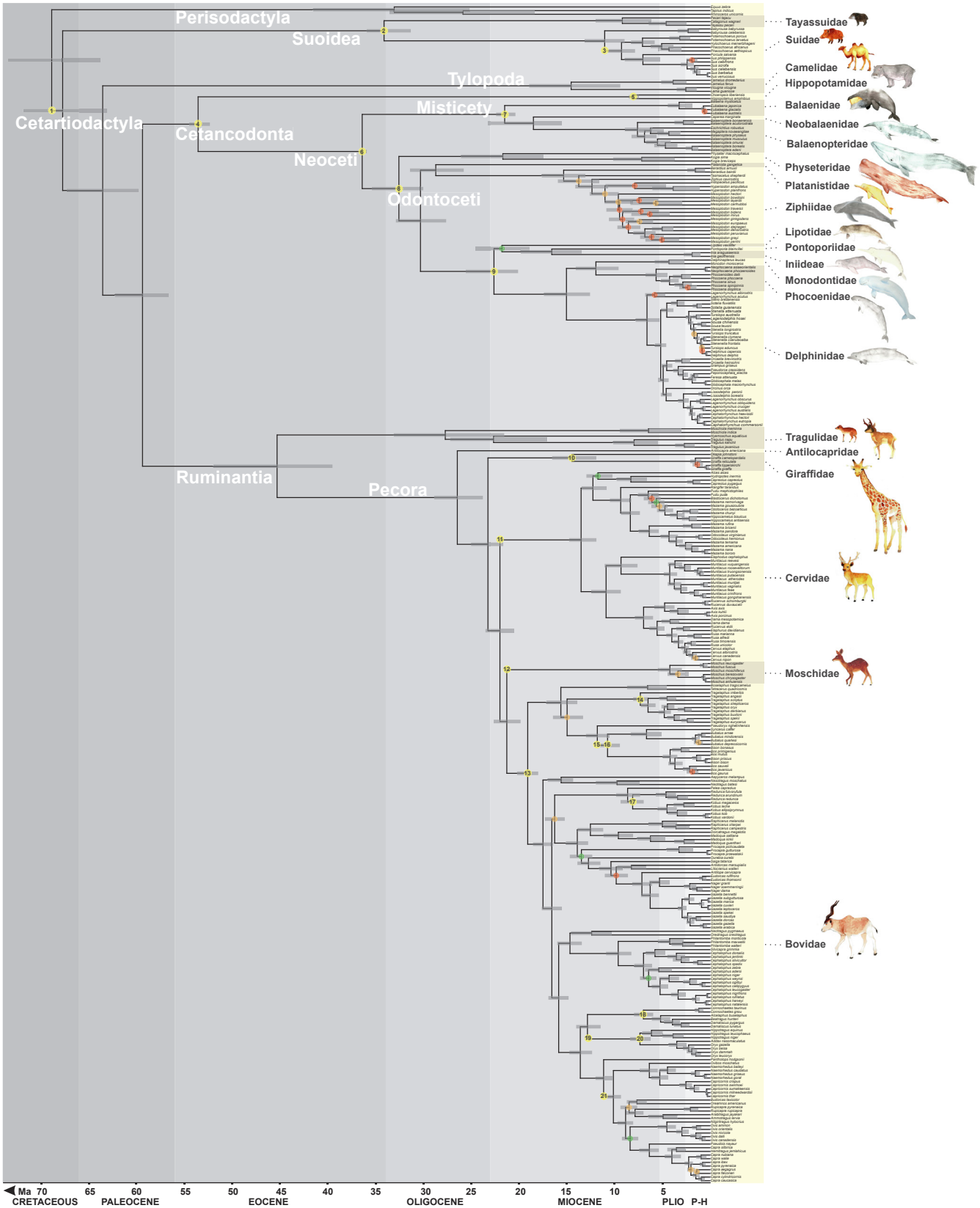

Fig. 1. Time-calibrated phylogenetic tree of Cetartiodactyla derived from BEAST. Horizontal gray bars represent the 95% highest posterior density of each node. Calibrations points (from 1 to 21) are arranged on nodes or along branches to indicate crown and stem fossils, respectively (See Table S1 in Appendix B). Filled circles in red represent nodes with posterior probabilities (PP) < 0.60; oranges circles denote nodes with PP values between 0.61 and 0.80; nodes with green circles have PP = 0.81–0.90 and all nodes without any circles have high support (PP > 0.91). In the scale of the geological era: H = Holocene, P = Pleistocene, PLIO = Pliocene. Images were drawn by Guilherme Oliveira.

Table 2
Divergence time estimates for the origin of main crown Cetartiodactyla lineages recovered in BEAST analyses. All mtDNA refers to age estimates from dataset1 and Mitogenomic from dataset2. Values are in millions of years ago and presented as mean and confidence interval (minimum and maximum values, respectively) sampled from the 95% high posterior density (95% HPD). The empty cells in dataset2 refers to nodes that lack mitogenomes representatives needed to assemble crown groups (Suidae lacks *Babyrousa*, Tayassuidae and Iniidae are represented by one species), precluding the estimates of the respective TMRCA (as in dataset1).

TAXA	ALL mtDNA		MITOGENOMIC	
	TMRCA	95% HPD	TMRCA	95% HPD
Cetartiodactyla	67.7	64.1–74.4	69.1	64.1–74.4
Suoidea	34.1	32.3–36.9	34.0	32.2–36.7
Tayassuidae	9.2	6.4–12.2	–	–
Suidae	11.0	8.9–13.1	–	–
Suinae	9.3	7.8–10.7	10.0	7.1–13.1
Camelidae	14.5	10.1–19.7	15.2	10.2–20.5
Cetancodonta	53.6	52.6–54.8	53.7	52.6–55.1
Hippopotamidae	8.1	7.6–8.7	8.1	7.6–8.8
Neoceti	36.4	36.1–36.9	36.4	36.0–37.0
Mysticeti	21.5	20.6–22.6	21.1	20.5–21.9
Balaenidae	3.2	2.0–4.6	4.2	2.7–5.7
Balaenopteridae	7.6	6.3–9.2	8.7	7.1–10.4
Odontoceti	32.6	29.7–35.1	32.9	29.8–35.5
Physeteridae	21.7	17.3–25.9	21.8	15.4–28.2
Ziphidae	15.3	13.0–17.6	16.9	13.7–20.0
Delphinida	22.6	19.7–25.2	22.5	19.6–25.6
Iniidae	1.9	1.0–2.8	–	–
Delphinoidea	15.0	12.7–17.5	15.3	12.8–18.3
Monodontidae	4.3	2.8–5.9	4.6	2.9–6.5
Phocoenidae	5.7	4.5–6.8	5.4	3.6–7.2
Delphinidae	6.6	5.4–7.9	7.0	5.8–8.2
Ruminantia	45.3	38.6–51.1	46.2	39.7–52.1
Tragulidae	27.7	22.3–33.9	27.5	21.5–33.6
Pecora	26.4	24.0–29.1	27.5	24.5–30.3
Giraffidae	14.5	12.4–17.0	14.4	12.3–17.0
Cervidae	13.5	12.2–15.1	14.2	12.4–16.1
Capreolinae	12.3	11.1–13.7	12.8	11.0–14.5
Cervinae	10.9	9.3–12.4	11.5	9.6–13.3
Moschidae	4.2	2.9–5.5	3.9	2.5–5.5
Bovidae	19.1	18.0–20.2	19.5	18.3–20.8
Bovinae	15.6	14.1–17.3	15.9	14.0–17.6
Boselaphini	6.5	4.4–8.5	7.2	4.8–9.6
Tragelaphini	7.3	6.5–8.1	7.3	6.5–8.1
Bovini	10.7	9.5–12.0	10.6	9.4–12.0
Antilopinae	17.5	16.4–18.5	18.0	16.7–19.3
Antilopini	13.9	12.7–15.0	14.5	13.0–16.0
Reduncini	8.1	6.8–9.3	7.9	6.6–9.2
Cephalophini	9.6	8.3–11.0	10.3	8.7–11.9
Alcelaphini	7.0	6.0–8.0	7.0	5.9–8.1
Hippotragini	7.3	6.3–8.5	7.2	6.1–8.3
Caprini	10.1	9.3–10.9	10.3	9.3–11.4

including taxa with missing data has little effect on divergence time estimates when multiple fossil calibration points and a comprehensive genetic database (e.g. mitogenomic in our case) are available.

Our phylogenetic tree hypothesis is congruent with most phylogenies generated using mitogenomic data and with those that have used both mitochondrial and/or nuclear markers (e.g. Hassanin et al., 2012; Zhou et al., 2011). Cetartiodactyla is a monophyletic group and the relationships among the major lineages (Suoidea, (Tylopoda, (Cetancodonta, Ruminantia)))[2] are well-supported (posterior probability [PP] = 1). Recent studies discuss a certain instability at the root of the phylogeny, varying between Suoidea or Tylopoda as the first lineage to diverge (Vislobokova, 2013 and references therein). Both of our datasets show a basal position of Suoidea (Suidae + Tayassuidae; PP = 1.0), which is concordant with previous phylogenetic inferences relying on mitochondrial DNA (Hassanin et al., 2012). Nevertheless, phylogenomic researches supports an early divergence of Tylopoda instead of Suoidea (Zhou et al., 2011; Meredith et al., 2011). Disagreements with some previously reported results could be related to differences on phylogenetic inference methods, molecular markers, taxon sampling and outgroup choice. Our results also support the monophyly of Cetancodonta (Cetacea + Ancodonta) and Cetruminantia (Cetancodonta + Ruminantia), which is in agreement with other molecular, morphological and paleontological evidence (Zhou et al., 2011; Hassanin et al., 2012; Vislobokova, 2013).

[2] Names in parentheses refer to the phylogenetic relationships of the lineages following the Newick format.

3.2. Divergence time estimates for Cetartiodactyla lineages

Because there are no significant differences between our datasets regarding TMRCA estimates, we discuss our results based on the dataset1 (with 321 species). The origin of Cetartiodactyla dates back to Cretaceous-Paleocene boundaries (mean = 67.7 Ma; 95% HPD = 64.1–74.4 Ma) (Fig. 1, Table 2). Our results are congruent with previous estimates for therian mammals (e.g. Meredith et al., 2011), but more recent compared to some Cetartiodactyla studies (e.g. ~81 Ma in Zhou et al., 2011; ~86 Ma in Hassanin et al., 2012). Differences could be due to bias imposed by sparse taxon sampling (e.g. Zhou et al., 2011; twenty one species) and fewer calibrations points (e.g. Hassanin et al., 2012; six calibration points), which has been shown to yield older divergence time estimates (Zheng and Wiens, 2015).

According to our estimates, the divergence of Suoidea occurred in late Eocene (mean = 34.1 Ma; 95% HPD = 32.3–36.9 Ma), and split estimates of both Tayassuidae (mean = 9.2 Ma; 95% HPD = 6.4–12.2 Ma) and Suidae (mean = 11.0 Ma; 95% HPD = 8.9–13.1 Ma) dated back to Middle Miocene. Phylogenetic relationships and divergence times within Suidae are congruent to those reported in previous studies that used both mitochondrial and nuclear markers (Gongora et al., 2011). *Babyrousa* genus is the first to branch off the Suidae family. The South-Asiatic lineages *(Porcula* and *Sus* genera*)* and those from sub-Saharian Africa (*Potamochoerus*, *Hylochoerus* and *Phacochoerus* genera) were recovered within the Suinae subfamily. The origin of Suinae also dated back to Middle Miocene (mean = 9.3 Ma; 95% HPD = 7.8–10.7 Ma) along with most supra-generic cladogenesis events within this subfamily (Fig. 1).

Tylopoda precedes the appearance of Cetruminantia. The fossil record suggests an ancient origin in North America (~40–45 Ma) and posterior migration to Asia and South America by the extant lineages ancestors (Cui et al., 2007). The diversification between Camelini and Lamini, the only extant lineages within Camelidae, occurred within the Early to Middle Miocene boundary (mean = 14.5 Ma; 95% HPD = 10.1–19.7 Ma). Our results show that the split between *Camelus* species occurred at the Mio-Pliocene boundary (mean = 3.9 Ma; 95% HPD = 2.6–5.2 Ma), and that of *Vicugna-Lama* in the Early Pliocene (mean = 2.5 Ma; 95% HPD = 1.6–3.8 Ma) (Fig. 1).

Estimates of TMRCA for Cetancodonta dated back to the Early Eocene (mean = 53.6 Ma; 95% HPD = 52.6–54.8 Ma). Cetaceans fossils (e.g. *Pakicetus, Himalayacetus*) and divergence time previously estimated for Cetacea - Ancodonta splitting (e.g. Orliac et al., 2010) also support our results. The TMRCA of Hippopotamidae, a family belonging to Ancodonta lineage, occurred in the Late Miocene (mean = 8.1 Ma; 95% HPD = 7.6–8.7 Ma), which is consistent with fossil records from the African Chorora formations (Boisserie et al., 2017).

The TMRCA of Neoceti (the only surviving lineage of Cetacea) was estimated in the Late Eocene (mean = 36.4 Ma; 95% HPD = 36.1–36.9 Ma). This result is broadly supported by fossil data (*Mystacodon selenensis*, Lambert et al., 2017). The TMRCA of Odontoceti was estimated in the Late Oligocene (mean = 32.6 Ma; 95% HPD = 29.7–35.1 Ma), which is also consistent with previous studies (Steeman et al., 2009). Conversely, our analyses yielded younger age estimates for the TMRCA of Mysticeti (dating back to the Early Miocene boundaries: mean = 21.5 Ma; 95% HPD = 20.6–22.6 Ma) when

compared to previous divergence times estimates (mean = 30.35 Ma; 95% HPD = 26.50–34.13, Marx and Fordyce, 2015). Differences in the crown Mysticeti fossil used as minimum constraint for calibration point (*Moronacetus parvus*, ~21–23 My) likely contributed for such discrepancies. However, divergence estimates for each family within this lineage (Table 1) are congruent to previous researches (Árnason et al., 2018).

The relationships within Mysticeti (Balaenidae, (Neobalaenidae, Balaenopteridae)) are strongly supported (PP = 1). Balaenidae is a monophyletic group that diverged very late in comparison to other families, as TMRCA dated back to the Pliocene (mean = 3.2 Ma; 95% HPD = 2.0–4.6 Ma). *Caperea marginata* is the only representative of Neobalaenidae, diverging from Balaenopteridae during the Early Miocene (mean = 18.5 Ma (95% HPD = 13.7–21.4 Ma). In our analyses, both *Megaptera* and *Eschrichtius* (the Grey whale is usually assigned to the family Eschrichtiidae) are nested within *Balaenoptera* making Balaenopteridae a paraphyletic group as currently presented. Previous studies have suggested that Balaenopteridae (TMRCA: mean = 7.6 Ma; 95% HPD = 6.3–9.2 Ma) experienced rapid radiation. In such cases, mtDNA may show phylogenetic pitfalls posed by incomplete lineage sorting and hybridization (Árnason et al., 2018). Nevertheless, this paraphyly reinforces the need for specific taxonomic revision to better assess phylogenetic uncertainty of balaenid species (Hassanin et al., 2012; Sasaki et al., 2005).

The relationships between the four major Odontoceti lineages (Physeteridae, ((Platanistidae, Ziphiidae), Delphinida)) had robust support (PP = 1). In our tree, Physeteridae (sperm whales) diverged early in comparison to remaining clades, and Platanistidae (South Asian river dolphins) is the sister group of Ziphiidae (beaked whales). Our results differ from previous researches, which instead recovered a clade formed by Ziphiidae + Delphinida using both comprehensive multilocus (Steeman et al., 2009; McGowen et al., 2009) and phylogenomic (Meredith et al., 2011; Zhou et al., 2011) datasets for cetaceans. In Ziphiidae, the addition of species with missing data yielded poorly resolved phylogenetic relationships within *Mesoplodon* genus (Fig. 1), mostly due to low genetic coverage (e.g., only one gene per species; Table S1 in Appendix A) hampering the resolution of phylogenetic inferences therein. Delphinida (including the families Lipotidae, Iniidae, Phocoenidae, Monodontidae and Delphinidae) is a well-supported clade and the relationships among families and subfamilies are congruent between our datasets. Within Delphinidae, Delphininae subfamily is the first taxa to branch off within the clade followed by the clade formed by Globicephalinae + Lissodelphininae. In addition, we recovered several polyphyletic genera (e.g., *Lagenorhynchus*, *Stenella* and *Tursiops*). Although the origin of toothed whales dated back to the Early Oligocine (~32 Ma), the most speciose and ecologically diverse groups of this major lineage only appeared in Late Miocene during the last ~10 Ma.

Ruminantia (Tragulina + Pecora) is a well-supported clade (PP = 1) that originated during the middle Eocene (mean = 45.3 Ma; 95% HPD = 38.6–51.1 Ma). The basal splitting within Ruminantia has received strong molecular and morphological support (Métais and Vislobokova, 2007). Many of the earliest ruminants are in the Tragulina infraorder, which has only one extant family (Tragulidae). Our estimates showed that Tragulidae originated during the transition from Middle to Late Oligocene (mean = 27.7 Ma; 95% HPD = 22.3–33.9 Ma), but we cannot discard a Late Eocene origin as suggested by fossil evidence (Métais and Vislobokova, 2007). Most ruminants belong to the Pecora infraorder and are grouped in five families (Antilocapridae, (Giraffidae, (Cervidae, (Moschidae,Bovidae)))). We recovered Pecora as a well-supported monophyletic group (PP = 1). The phylogenetic relationships among these families has been amply debated because lineage diversification occurred in a short time interval (during the Oligocene-Miocene transition, ~20 Ma ago). Our analyses support an early divergence of Antilocapridae within Pecora, followed by Giraffidae, Cervidae and the clade formed by Moschidae + Bovidae, which agrees with recent proposals (e.g. Bibi, 2013). The estimated TMRCA of Pecora dated back to Late Oligocene (mean = 26.4 Ma; 95% HPD = 24.0–29.1 Ma), which is supported by the fossil record (Vislobokova, 1997), and agrees with estimates obtained in phylogenetic studies using molecular and morphological characters (Fernández and Vrba, 2005).

The estimated TMRCA of Giraffidae fell within the Middle Miocene boundary (mean = 14.5 Ma; 95% HPD = 12.4 – 17.0 Ma). Additionally, we incorporated three recently described giraffe species (*Giraffa giraffa*, *G. reticulata*, and *G. tippelskirchi,* Fennessy et al., 2016*)* not included in any previous Cetartiodactyla large scale phylogenetic inference. The TMRCA of giraffes was estimated at the Pleistocene (mean = 1.5 Ma; 95% HPD = 1.0–2.2 Ma) (Fig. 1), which agrees with divergence times estimated in the species' original description (Fennessy et al., 2016).

Cervidae is one of the most diverse ruminant families and originated in the Middle Miocene (mean = 13.5 Ma; 95% HPD = 12.2–15.1 Ma) according to our estimates. Two subfamilies are recognized within Cervidae: the subfamily Cervinae (TMCRA: mean = 10.9 Ma; 95% HPD = 9.3–12.4 Ma), which includes the genera *Axis*, *Cervus*, *Dama*, *Elaphurus*, *Rucervus*, *Rusa* [tribe Cervini], and *Muntiacus*, *Elaphodus* [tribe Muntianciani]; and the subfamily Capreolinae (TMCRA: mean = 12.3 Ma; 95% HPD = 11.1–13.7 Ma), which comprises the genera *Alces* [tribe Alceini], *Capreolus* and *Hydropotes* [tribe Capreolini], *Odocoileus*, *Blastocerus*, *Hippocamelus*, *Mazama*, *Ozotoceros* and *Pudu* [tribe Odocoileini], and *Rangifer* [tribe Rangiferini]). The recognition of these subfamilies was initially proposed based on morphology and subsequently by molecular phylogenies (e.g. Heckeberg et al., 2016), and also received strong support in our analyses (PP = 1). The phylogenetic relationships of Cervidae have been thoroughly studied (e.g. Duarte et al., 2008; Gutiérrez et al., 2017), being Cervinae phylogenetically better resolved compared to Capreolinae (Heckeberg et al., 2016). In our phylogenetic hypothesis, most internal nodes within Cervidae are well-supported (PP > 0.95), except for some intrageneric relationships between Cervini and Odocoileini (Fig. 1). Furthermore, we observed that the genera *Pudu* and *Mazama* (Capreolinae subfamily), and *Rusa* and *Rucervus* (Cervinae subfamily) are found as non-monophyletic.

Divergence among Moschidae and Bovidae started in the Early Miocene (mean = 21.3 Ma; 95% HPD = 19.9–22.7 Ma). The TMCRA of extant Moschidae was estimated in the Pliocene (mean = 4.2 Ma; 95% HPD = 2.9–5.5), meanwhile diversification in Bovidae (the most speciose family in the Order) started in the Early Miocene (mean = 19.1 Ma (95% HPD = 18.0–20.2 Ma). The two most representative subfamilies within Bovidae (Antilopinae and Bovinae) started their diversification around the Late to Middle Miocene (mean = 17.5 Ma and 15.6 Ma, respectively; Fig. 1). Additionally, some Bovidae genera such as *Bos* and *Bison* (Bovinae subfamily) and *Cephalophus* and *Neotragus* (Antilopinae subfamily) are not monophyletic as reported in previous works (Bärmann and Schikora, 2014; Bibi, 2013; Hassanin et al., 2012; Johnston and Anthony, 2012).

3.3. Missing data and the importance for dense taxon sampling

Highly supported phylogenies for several invertebrates and vertebrates taxa are recurrent in the literature mostly due to the increasing availability of mitogenomes (e.g. Gibb et al., 2016). However, for most taxa, genetic information is scarce or restricted to only one or a few genes but are equally relevant for phylogenetic analyses (Jiang et al., 2014; Zheng and Wiens, 2015). Nevertheless, assembling groups of genomic data with incomplete information requires special attention, especially in a phylogenomic context (when multiple nuclear genes are used; Zheng and Wiens, 2015). Empirical and simulated data suggested that addition of taxa with incomplete data has no significant impact on divergence time and topology estimates (Zheng and Wiens, 2015), also improving phylogeny accuracy by subdividing long branches among complete sampled taxa (Wiens and Tiu, 2012). On the other hand, high levels of missing data could generate misleading phylogenies when

nucleotide matrices are small (100 ~ 1000 bp), lacks enough informative characters (Zheng and Wiens, 2015), taxa are sparsely sampled (Wiens and Tiu, 2012), or when one does not account for different evolution patterns among sites or genes (Roure et al., 2013; Wu et al., 2014). Thus, in mitogenomic phylogenies (characterized by large nucleotide datasets), the addition of taxa with incomplete information should have a limited negative effect and yield well-supported topologies (Jiang et al., 2014), especially when datasets are analyzed under proper partitioning schemes and substitution model that accounts for site heterogeneity (Roure et al., 2013). Our results support this assertion (Fig. 1, and Fig. S1, Table 2).

Most of the recent phylogenetic studies of Cetartiodactyla focused on a better phylogenetic resolution within families by increasing the number of taxa sampled. With the improvements of phylogenetic algorithms and computational power, large mitogenomic datasets can now accommodate complex evolution models among partitions and overpass some pitfalls (e.g., posed by model mis-specifications or misleading phylogenetic signals of mitochondrial DNA) yielding phylogenies that are congruent to well-established nuclear trees (Wu et al., 2014). Although we acknowledge the limitations of using one locus (mtDNA) over a multilocus approach, until now there is no dated tree that extensively and simultaneously included all Cetartiodactyla lineages in a Bayesian framework. Further, estimates produced by different analyses must be considered hypotheses open to be challenged as new fossil and genomic evidence becomes available. In our work, we make available a robust calibrated phylogeny covering more than 90% of the species of Cetartiodactyla. Our results are congruent with fossil evidence and most of previously published molecular phylogenies, representing the most comprehensive time-calibrated phylogenetic hypothesis to this date.

Acknowledgments

We thank Pablo A. Martinez, Pedro C. Estrela for comments on the manuscripts. JPZ, AAB and GS thank to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil (process number 1489599, 181075, 181077) and FMM thanks Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq), Brazil (process number 167148) for their scholarships. DOM and GCC thank CNPq for the research fellowship. We would like to thank the kind assistance of Wayne Pfeiffer and Mark Miller from *CIPRES* platform and Pedro Silva and Danilo Chaves from *High Performance Computing Center* (NPAD) at UFRN-Brazil for allowing us to access high performance clusters. We also thank to Guilherme Oliveira for drawing the species pictures in Fig. 1.

Conflict of interest

The authors declare that they have no conflict of interest.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2018.12.015.

References

Árnason, Ú., Lammers, F., Kumar, V., Nilsson, M.A., Janke, A., 2018. Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. Sci. Adv. 4, eaap9873. https://doi.org/10.1126/sciadv.aap9873.

Bärmann, E.V., Schikora, T., 2014. The polyphyly of Neotragus - results from genetic and morphometric analyses. Mamm. Biol. 79, 283–286. https://doi.org/10.1016/j.mambio.2014.01.001.

Benton, M.J., Donoghue, P.C.J., Asher, R.J., Friedman, M., Near, T.J., Vinther, J., 2015. Constraints on the timescale of animal evolutionary history. Palaeontologia Electronica. 18, 1–106. https://doi.org/10.26879/424.

Berta, A., Sumich, J.L., Kovacs, K.M., 2015. Marine Mammals: evolutionary biology, third ed. Academic Press.

Bibi, F., 2013. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. BMC Evol. Biol. 13, 166. https://doi.org/10.1186/1471-2148-13-166.

Boisserie, J.R., Fisher, R.E., Lihoreau, F., Weston, E.M., 2011. Evolving between land and water: key questions on the emergence and history of the hippopotamidae (hippopotamoidea, cetancodonta, cetartiodactyla). Biol. Rev. 86, 601–625. https://doi.org/10.1111/j.1469-185X.2010.00162.x.

Boisserie, J.R., Suwa, G., Asfaw, B., Lihoreau, F., Bernor, R.L., Katoh, S., Beyene, Y., 2017. Basal hippopotamines from the upper Miocene of Chorora. Ethiopia. J. Vertebr. Paleontol. 37. https://doi.org/10.1080/02724634.2017.1297718.

Bromham, L., Duchêne, S., Hua, X., Ritchie, A.M., Duchêne, D.A., Ho, S.Y.W., 2018. Bayesian molecular dating: opening up the black box. Biol. Rev. 93, 1165–1191. https://doi.org/10.1111/brv.12390.

Cui, P., Ji, R., Ding, F., Qi, D., Gao, H., Meng, H., Yu, J., Hu, S., Zhang, H., 2007. A complete mitochondrial genome sequence of the wild two-humped camel (Camelus bactrianus ferus): an evolutionary history of Camelidae. BMC Genomics 8, 241. https://doi.org/10.1186/1471-2164-8-241.

dos Reis, M., Donoghue, P.C.J., Yang, Z., 2015. Bayesian molecular clock dating of species divergences in the genomics era. Nat. Rev. Genet. 17, 1–10. https://doi.org/10.1038/nrg.2015.8.

Drummond, A.J., Bouckaert, R.R., 2015. Bayesian evolutionary analysis with BEAST. Cambridge University Pess.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88. https://doi.org/10.1371/journal.pbio.0040088.

Duarte, J.M.B., González, S., Maldonado, J.E., 2008. The surprising evolutionary history of South American deer. Mol. Phylogenet. Evol. 49, 17–22. https://doi.org/10.1016/j.ympev.2008.07.009.

Fennessy, J., Bidon, T., Reuss, A., Kumar, V., Elkan, P., Nilsson, M.A., Vamberger, M., Janke, A., 2016. Multi-locus analyses reveal four giraffe species instead of one. Curr. Biol. 26, 1–7. https://doi.org/10.1016/j.cub.2016.07.036.

Fernández, M.H., Vrba, E.S., 2005. A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. Biol. Rev. 80, 269–302. https://doi.org/10.1017/S1464793104006670.

Frantz, L., Meijaard, E., Gongora, J., Haile, J., Groenen, M.A.M., Larson, G., 2016. The evolution of Suidae. Annu. Rev. Anim. Biosci. 4, 61–85. https://doi.org/10.1146/annurev-animal-021815-111155.

Garamszegi, L., Paradis, E., Symonds, M.R.E., Blomberg, S.P., Mundry, R., 2014. Modern phylogenetic comparative methods and their application in evolutionary biology, first ed. Springer Heidelberg New York Dordrecht London. http://doi.org/10.1007/978-3-662-43550-2.

Gatesy, J., Milinkovitch, M., Waddell, V., Stanhope, M., 1999. Stability of cladistic relationships between Cetacea and higher-level Artiodactyl Taxa. Syst. Biol 48, 6–20. https://doi.org/10.1080/106351599260409.

Gibb, G.C., Condamine, F.L., Kuch, M., Enk, J., Moraes-Barros, N., Superina, M., Poinar, H.N., Delsuc, F., 2016. Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans. Mol. Biol. Evol. 33, 621–642. https://doi.org/10.1093/molbev/msv25.

Gongora, J., Cuddahee, R.E., Nascimento, F.F., do Palgrave, C.J., Lowden, S., Ho, S.Y.W., Simond, D., Damayanti, C.S., White, D.J., Tay, W.T., Randi, E., Klingel, H., Rodrigues-Zarate, C.J., Allen, K., Moran, C., Larson, G., 2011. Rethinking the evolution of extant sub-Saharan African suids (Suidae, Artiodactyla). Zool. Scr. 40, 327–335. https://doi.org/10.1111/j.1463-6409.2011.00480.x.

Graur, D., Higgins, D.G., 1994. Molecular evidence for the inclusion of cetaceans within the order Artiodactyla. Mol. Biol. Evol. 11, 357–364. https://doi.org/10.1093/oxfordjournals.molbev.a040118.

Gutiérrez, E.E., Helgen, K.M., McDonough, M.M., Bauer, F., Hawkins, M.T.R., Escobedo-Morales, L.A., Patterson, B.D., Maldonado, J.E., 2017. A gene-tree test of the traditional taxonomy of American deer: The importance of voucher specimens, geographic data, and dense sampling. Zookeys 697, 87–131. https://doi.org/10.3897/zookeys.697.15124.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., Jansen van Vuuren, B., Matthee, C., Ruiz-Garcia, M., Catzeflis, F., Areskoug, V., Nguyen, T.T., Couloux, A., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. C. R. Biol. 335, 32–50. https://doi.org/10.1016/j.crvi.2011.11.002.

Heckeberg, N.S., Erpenbeck, D., Wörheide, G., Rössner, G.E., 2016. Systematic relationships of five newly sequenced cervid species. PeerJ 4, e2307. https://doi.org/10.7717/peerj.2307.

IUCN, 2018. The IUCN Red List of Threatened Species. Version 2018–2. http://www.iucnredlist.org.

Jiang, W., Chen, S.Y., Wang, H., Li, D.Z., Wiens, J.J., 2014. Should genes with missing data be excluded from phylogenetic analyses? Mol. Phylogenet. Evol. 80, 308–318. https://doi.org/10.1016/j.ympev.2014.08.006.

Johnston, A.R., Anthony, N.M., 2012. A multi-locus species phylogeny of African forest duikers in the subfamily Cephalophinae: evidence for a recent radiation in the Pleistocene. BMC Evol. Biol. 12. https://doi.org/10.1186/1471-2148-12-120.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Lambert, O., Martínez-Cáceres, M., Bianucci, G., Di Celma, C., Salas-Gismondi, R., Steurbaut, E., Urbina, 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, 1535–1541. https://doi.org/10.1016/j.cub.2017.04.026.

Lanfear, R., Frandsen, 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. https://doi.org/10.1093/molbev/msw260.

Marx, F.G., Fordyce, R.E., 2015. Baleen boom and bust: a synthesis of mysticete phylogeny, diversity and disparity. R. Soc. Open Sci. 2, 140434. https://doi.org/10.1098/rsos.140434.

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. https://doi.org/10.1016/j.ympev.2009.08.018.

Meredith, R.W., Janečka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A., Eizirik, E., Simão, T.L.L., Stadler, T., Rabosky, D.L., Honeycutt, R.L., Flynn, J.J., Ingram, C.M., Steiner, C., Williams, T.L., Robinson, T.J., Burk-Herrick, A., Westerman, M., Ayoub, N.A., Springer, M.S., Murphy, W.J., 2011. Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334, 521–524. https://doi.org/10.1126/science.1211028.

Métais, G., Vislobokova, I., 2007. Basal ruminants, in: Prothero, D.R., Foss, S.E. (Eds). The evolution of artiodactyls. The Johns Hopkins University Press., Baltimore, pp. 189-212.

Montgelard, C., Catzeflis, F.M., Douzery, E., 1997. Phylogenetic relationships of artiodactyls and cetaceans as deduced from the comparison of cytochrome b and 12S rRNA mitochondrial sequences. Mol. Biol. Evol. 14, 550–559. https://doi.org/10.1093/oxfordjournals.molbev.a025792.

Orliac, M., Boisserie, J.-R., MacLatchy, L., Lihoreau, F., 2010. Early Miocene hippopotamids (Cetartiodactyla) constrain the phylogenetic and spatiotemporal settings of hippopotamid origin. Proc. Natl. Acad. Sci. 107, 11871–11876. https://doi.org/10.1073/pnas.1001373107.

Prothero, D.R., Foss, S.E., 2007. The evolution of artiodactyls. Baltimore, Maryland.

Quental, T.B., Marshall, C.R., 2010. Diversity dynamics: molecular phylogenies need the fossil record. Trends Ecol. Evol. 25, 434–441. https://doi.org/10.1016/j.tree.2010.05.002.

Roure, B., Baurain, D., Philippe, H., 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. Mol. Biol. Evol. 30, 197–214. https://doi.org/10.1093/molbev/mss208.

Sasaki, T., Nikaido, M., Hamilton, H., Goto, M., Kato, H., Kanda, N., Pastene, L.A., Cao, Y., Fordyce, R.E., Hasegawa, M., Okada, N., 2005. Mitochondrial phylogenetics and evolution of mysticete whales. Syst. Biol. 54, 77–90. https://doi.org/10.1080/10635150590905939.

Scholl, J.P., Wiens, J.J., Wiens, J.J., 2016. Diversification rates and species richness across the Tree of Life. Proc. R. Soc. B. 283. https://doi.org/10.1098/rspb.2016.1334.

Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y.W., Rabosky, D.L., Nielsen, R., Rahbek, C., Glenner, H., Sørensen, M.V., Willerslev, E., Sorensen, M.V., Willerslev, E., 2009. Radiation of extant cetaceans driven by restructuring of the oceans. Syst. Biol. 58, 573–585. https://doi.org/10.1093/sysbio/syp060.

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., Rambaut, A., 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4, 1–5. https://doi.org/10.1093/ve/vey016.

Vislobokova, I., 1997. Eocene-Early Miocene ruminants in Asia. Actes du Congr. BiochroM 97, 215–223.

Vislobokova, I.a., 2013. On the origin of Cetartiodactyla: Comparison of data on evolutionary morphology and molecular biology. Paleontol. J. 47, 321–334. https://doi.org/10.1134/S003103011303012X.

Wiens, J.J., Tiu, J., 2012. Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. PLoS One 7, e42925. https://doi.org/10.1371/journal.pone.0042925.

Wu, J., Hasegawa, M., Zhong, Y., Yonezawa, T., 2014. Importance of synonymous substitutions under dense taxon sampling and appropriate modeling in reconstructing the mitogenomic tree of Eutheria. Genes Genet. Syst. 89, 237–251. https://doi.org/10.1266/ggs.89.237.

Zheng, Y., Wiens, J.J., 2015. Do missing data influence the accuracy of divergence-time estimation with BEAST? Mol. Phylogenet. Evol. 85, 41–49. https://doi.org/10.1016/j.ympev.2015.02.002.

Zhou, X., Xu, S., Yang, Y., Zhou, K., Yang, G., 2011. Phylogenomic analyses and improved resolution of Cetartiodactyla. Mol. Phylogenet. Evol. 61, 255–264. https://doi.org/10.1016/j.ympev.2011.02.009.