

**MOLECULAR PHYLOGENETICS OF BALEEN WHALES (MYSTICETI) AND EFFECT OF DIFFERENT MARKERS ON PHYLOGENY**

**Chandrakant S Jadhav<sup>1\*</sup>, Vipin Hiremath<sup>1</sup>, Sachin Thorat<sup>1</sup>, Tejswini Sontakke<sup>1</sup>, Anand Metri<sup>2</sup>  
Sagar Biradar<sup>3</sup>**

<sup>1</sup> Paul Hebert Centre for DNA Barcoding and Biodiversity Studies,  
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad-431004 India.

<sup>2</sup> Shivchatrapati college, N-3, CIDCO, Aurangabad

<sup>3</sup> Government Institute of science. Aurangabad

<sup>1\*</sup>Corresponding Author: Chandrakant S Jadhav, Email- [chandrakant.jadhavpatil@gmail.com](mailto:chandrakant.jadhavpatil@gmail.com)

**ABSTRACT**

Baleen whales are largest cetaceans present on the earth. In this study we studied molecular phylogenetics of 15 baleen whales and studied effect of *cox1*, *cytb* and 13 concatenated protein coding genes on the phylogeny. The mitogenome size of *Caperea marginata* (16383 bp) was lowest and highest for *Balaenoptera bonaerensis* (16421 bp). All the mitogenomes of Mysticeti were AT rich. The lowest AT content was observed in the *Kogia breviceps* (56.8%) whereas the highest AT content was observed in the *Caperea marginata* (60.5%). AT skew values of all mitogenomes were positive whereas GC skew values were negative. Phylogenetic analysis suggests Balaenidae as the basal group and supports paraphyly of (Eschrichtiidae + Balaenopteridae). Moreover our analysis suggests that concatenated 13 PCG approach is always better and accurate than single gene based methods for evaluating the phylogenetic relationships of animals.

**KEYWORDS:** Cetaceans, Mitogenome, Mysticeti, Phylogenetics.

**INTRODUCTION**

Cetaceans are the marine mammals which include dolphins, whales & porpoises and are classified in the order cetartiodactyla. Though they can't live outside the water but they are originated from the terrestrial mammals and invaded into water systems where they went through rapid morphological and physiological adaptation (Thewissen *et al.*, 2002; Fordyce *et al.*, 2001; Gingerich *et al.*, 1983). In this transition from land to water odontocetes developed echolocation and toothed predation ability while mysticetes developed filter feeding ability using the baleen. Based on the morphology the order cetacea is further divided into three suborders out of which one is extinct & two are extant. The extant suborder includes Mysticeti (baleen whales) & Odontoceti (toothed whales).

Baleen whales are one of the largest animals present on the earth and comprises of 14 species (Committee on Taxonomy 2017). Mysticeti sub order comprises of 4 families i.e. Balaenopteridae (Humpback whales and rorquals) includes 8 species; Balaenidae (right whales) includes 4 species whereas Neobalaenidae (pygmy right whales) and Eschrichtiidae (grey whales) are monotypic families (Committee on Taxonomy 2017). As they are widely distributed in marine habitat and due to their large body size the studies on them are largely restricted. Previous phylogenetic analysis based on *cytb* and control region suggest Balaenidae (right whales) as the basal group among mysticetes and placed Gray whales in the Balaenopteridae clade, however the phylogenetic relationships between Eschrichtiidae (Gray whale) and Balaenopteridae (rorquals) is unresolved (Arnason *et al.*, 1993; Arnason *et al.*, 1994). According to DNA satellite based studies the Balaenidae diverged first while Neobalaenidae diverged later on and found more closely related to Eschrichtiidae + Balaenopteridae than Balaenidae (Arnason *et al.*, 1991; Arnason *et al.*, 1992). Moreover *cytb* and control region analysis suggested the paraphyly of Balaenopteridae (Arnason *et al.*, 1993a; Arnason *et al.*, 1994). Furthermore, other studies based on partial mitochondrial and nuclear DNA sequences and morphological studies also could not resolve the issue (Rychel *et al.*, 2004; Geisler *et al.*, 2003). Although different studies resulted in different inferences but there is consensus on Balaenidae as the basal group. However despite of morphological and molecular studies the phylogenetic relationships among Neobalaenidae, Eschrichtiidae, and the Balaenopteridae is unclear.

Animal mitochondrial DNA is used as a molecular marker in phylogenetic studies of mammals due to relatively rapid rate of evolution, maternal inheritance and absence of recombination (Olivio *et al.*, 1983; Arnason *et al.*, 2002; Brown

*et al.*, 1979). As individual mitochondrial DNA genes may provide different tree topologies therefore to resolve the phylogenetic relationships between baleen whales and to test the impact of different markers on their phylogeny we used both single gene (cox1 and cytb) and supergene (concatenated 13 protein coding genes) datasets for analysis (Gontcharov *et al.*, 2004; Urantowka *et al.*, 2017). In this study we used published mitogenomes of baleen whales for investigating the bayesian phylogeny.

**Table 1.** Mitochondrial genomes used in this study.

Order	Suborder	Family	Common name	Species	Accession No	References
<b>Ingroup</b>	Mysticeti	Balaenidae (Right whales)	Bowhead whale	<i>Balaena mysticetus</i>	NC_005268.1	Arnason <i>et al</i> 2004
			North Atlantic right whale	<i>Eubalaena glacialis</i>	NC_037444.1	Arnason <i>et al</i> 2018
			North Pacific right whale	<i>Eubalaena japonica</i>	NC_006931.1	Sasaki <i>et al</i> 2005
			Southern right whale	<i>Eubalaena australis</i>	NC_006930.1	Sasaki <i>et al</i> 2005
			Pygmy Right whales	<i>Caperea marginata</i>	NC_005269.1	Arnason <i>et al</i> 2004
		Eschrichtiidae	Gray whale	<i>Eschrichtius robustus</i>	NC_005270.1	Arnason <i>et al</i> 2004
			Balaenopteridae (rorquals)	North Pacific minke whale	<i>Balaenoptera acutorostrata</i> s.	NC_005271.1
		Antarctic minke whale		<i>Balaenoptera bonaerensis</i>	NC_006926.1	Sasaki <i>et al</i> 2005
		Sei whale		<i>Balaenoptera borealis</i>	NC_006929.1	Sasaki <i>et al</i> 2005
		Offshore Bryde's whale		<i>Balaenoptera brydei</i>	NC_006928.1	Sasaki <i>et al</i> 2005
		Bryde's whale		<i>Balaenoptera edeni</i>	NC_007938.1	Sasaki <i>et al</i> 2006
		Blue whale		<i>Balaenoptera musculus</i>	NC_001601.1	Arnason <i>et al</i> 1993
		Omura's whale		<i>Balaenoptera omurai</i>	NC_007937.1	Sasaki <i>et al</i> 2006
		Fin whale		<i>Balaenoptera physalus</i>	NC_001321.1	Arnason <i>et al</i> 1991
		Humpback whale	<i>Megaptera novaeangliae</i>	NC_006927.1	Sasaki <i>et al</i> 2005	
<b>Outgroup</b>	Odontoceti	Kogiidae	Pygmy sperm whale	<i>Kogia breviceps</i>	NC_005272.1	Arnason <i>et al</i> 2004
	Whippomorpha	Hippopotamidae		<i>Hippopotamus amphibius</i>	NC_000889.1	Ursing <i>et al</i> 1998

## MATERIALS AND METHODS

We selected 15 species of Mysticeti as ingroup and 2 species of cetartiodactyla as outgroup for rooting the tree. Outgroup species includes *Kogia breviceps* (toothed whale) and *H. amphibius* from whippomorpha suborder (Table 2). MAFFT version 7.273 was used for sequence alignment. For alignment The L-INS-I algorithm was used. Mitochondrial genomes were downloaded from refseq database of NCBI. Nucleotide composition was calculated using MEGA 7.0.  $AT\text{-skew} = (A - T) / (A + T)$  and  $GC\text{-skew} = (G - C) / (G + C)$  formulas were used to calculate AT and GC skew. Bayesian inference (BI) method was used to construct the Phylogeny by using MrBayes v3.2.557. Four simultaneous Markov chains were run for 1 million generations, with burn-in of 25% trees and tree sampling set after 1000 generations.

## RESULTS AND DISCUSSION

### Nucleotide composition of Mysticeti mitogenomes

The complete mitogenome size of all species under study ranges from 16383 (*Caperea marginata*) bp to 16421 (*Balaenoptera bonaerensis*) bp. All the mitogenomes exhibits the AT rich content which is typically observed in cetaceans. The lowest AT content was observed in the *Kogia breviceps* (56.8%) whereas the highest AT content was observed in the *Caperea marginata* (60.5%). Similarly AT and GC skew parameters used to study compositional nucleotide bias in the mitogenomes. AT skew values of all mitogenomes were positive which indicates percentage of A nucleotide is more than T nucleotide on the other hand GC skew values were negative for all mitogenomes which suggest percentage C nucleotide is more than G nucleotide (Table 2).

### Phylogeny

Although phylogenetic position of some taxa were different the Bayesian inference based on the cox1, cytb and concatenated 13 PCG datasets resulted in almost similar trees with different posterior probability values. The *Kogia breviceps* (toothed whale) and the closest relative of baleen whales *H. amphibius* were used as outgroup species for rooting the tree (Gatesy *et al.*, 1996). In the BI tree of cox1 gene the Balaenidae family was at the basal position which is supported by the high posterior probability values (Fig.1). Next to Balaenidae the *Caperea marginata* (Pygmy right whale) Sole member of family Neobalaenidae was diverged which is followed by Eschrichtiidae and Balaenopteridae. Similarly, BI tree of cytb gene suggests Balaenidae as the basal clade and Neobalaenidae as next branched clade. Similar to cox1 tree the Eschrichtiidae was placed in Balaenopteridae clade but the position of *Eschrichtius robustus*

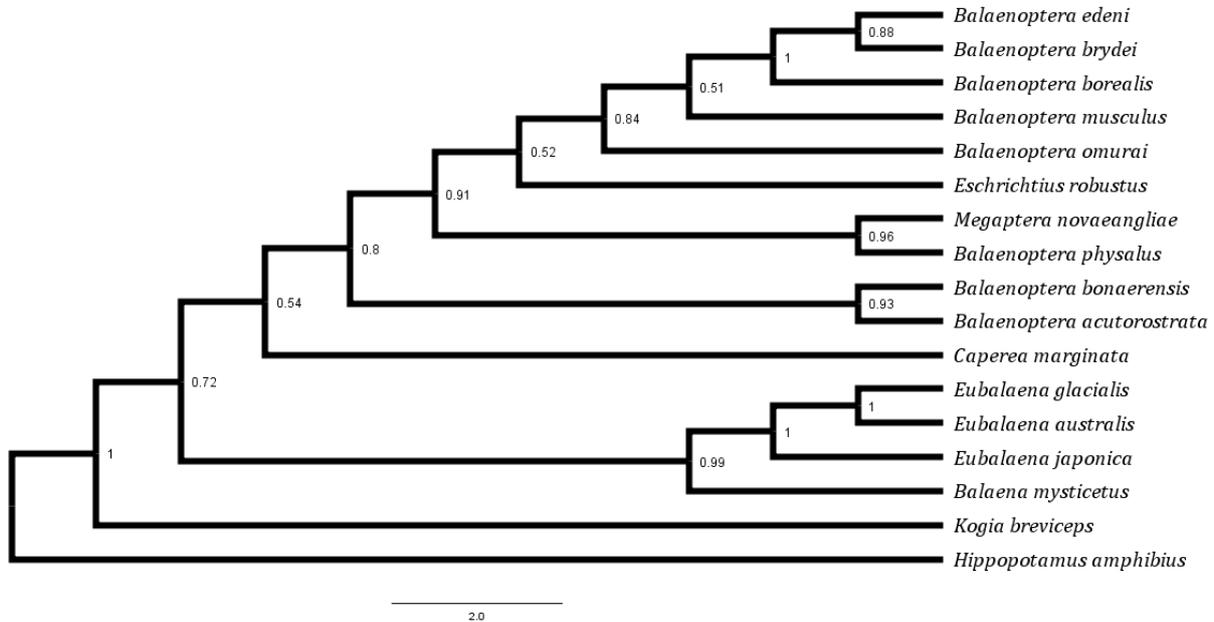
was different. In addition to that the position of *Balaenoptera omurai* and *Balaenoptera musculus* was noticed as exchanged. In The BI tree of all PCG similar relationships of four major lineages was observed. However the Gray whale is grouped with the fin and humpback whale.

**Table 2.** Nucleotide composition, GC skew and AT skew of mitochondrial genomes of all species.

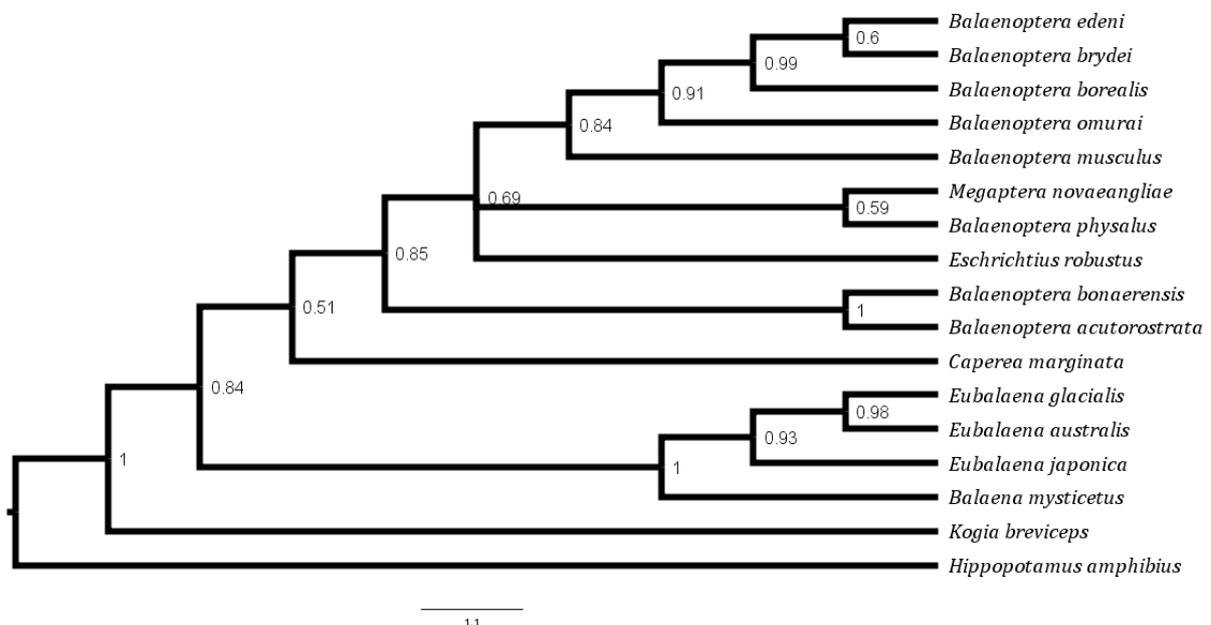
Species	T	C	A	G	Total	G+C	A+T	GC skew	AT skew
<i>Megaptera novaeangliae</i>	26.4	27.7	32.8	13.1	16398.0	41	59.2	-.4	.1
<i>Kogia breviceps</i>	25.3	29.1	31.5	14.1	16406.0	43	56.8	-.3	.1
<i>Hippopotamus amphibius</i>	24.7	28.6	32.7	14.0	16407.0	43	57.4	-.3	.1
<i>Eubalaena japonica</i>	26.2	27.9	32.7	13.2	16386.0	41	58.9	-.4	.1
<i>Eubalaena glacialis</i>	26.1	28.0	32.6	13.2	16386.0	41	58.8	-.4	.1
<i>Eubalaena australis</i>	26.1	28.0	32.6	13.2	16385.0	41	58.8	-.4	.1
<i>Eschrichtius robustus</i>	26.0	28.2	32.7	13.1	16412.0	41	58.7	-.4	.1
<i>Caperea marginata</i>	27.5	26.6	32.9	13.0	16383.0	40	60.5	-.3	.1
<i>Balaenoptera physalus</i>	26.7	27.3	32.7	13.3	16398.0	41	59.4	-.3	.1
<i>Balaenoptera omurai</i>	26.6	27.6	32.7	13.1	16404.0	41	59.3	-.4	.1
<i>Balaenoptera musculus</i>	26.6	27.6	32.8	13.0	16402.0	41	59.4	-.4	.1
<i>Balaenoptera edeni</i>	26.6	27.6	32.9	12.9	16409.0	40	59.5	-.4	.1
<i>Balaenoptera brydei</i>	26.9	27.3	32.8	13.0	16408.0	40	59.7	-.4	.1
<i>Balaenoptera borealis</i>	26.7	27.4	32.9	13.0	16410.0	40	59.6	-.4	.1
<i>Balaenoptera bonaerensis</i>	26.3	27.8	33.0	12.9	16421.0	41	59.3	-.4	.1
<i>Balaenoptera acutorostrata</i>	26.5	27.5	33.0	12.9	16417.0	40	59.6	-.4	.1
<i>Balaena mysticetus</i>	26.0	28.1	32.7	13.2	16390.0	41	58.7	-.4	.1

Overall our analysis suggest the strong monophyly of Balaenidae and it is the basal group with high posterior probability values (Fig 1, 2 and 3) this observation is consistent with previous studies (Arnason *et al.*, 1992, 1993, 1994, 1996; Mcleod *et al.*, 1993; Milinkovitch *et al.*, 1994). The monotypic Neobalaenidae family (*Caperea marginata*) diverged next to the Balaenidae and grouped as the sister taxa to the (Eschrichtiidae + Balaenopteridae) such grouping pattern is supported by previous studies (Arnason and Best, 1991; Arnason *et al.*, 1992). On the contrary our analysis supports the paraphyly of (Eschrichtiidae + Balaenopteridae). Whereas some morphological studies supports the closeness of Neobalaenidae to the Balaenidae family (Bisconti *et al.*, 2000; Geisler *et al.*, 2003). In the past studies the phylogenetic position of Eschrichtiidae was not very well resolved with high confidence. However our analysis by all methods supports (Eschrichtiidae + Balaenopteridae) paraphyly which is strongly supported by the posterior probability values (Fig 3) and previous studies (Arnason *et al.*, 1993, 1994).

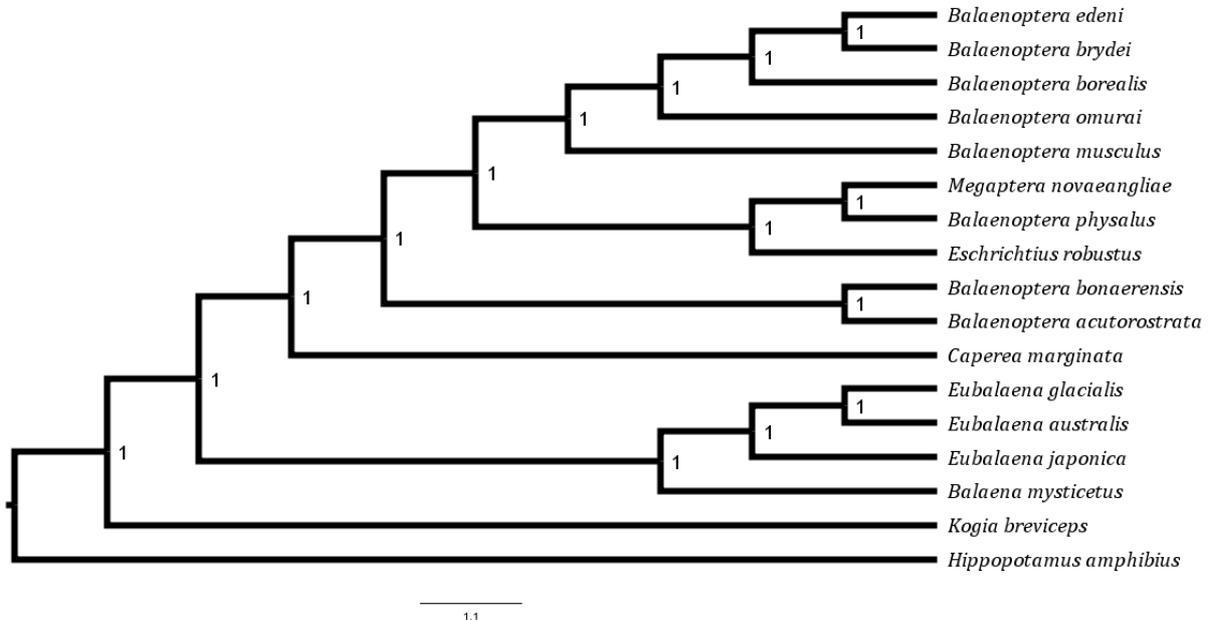
In the BI tree of all PCGS the (Eschrichtiidae + Balaenopteridae) group indicates 3 sub clades (Fig 3). The clade 1 includes the mink whales; clade 2 includes (Gray + fin + humpback whale) whereas the clade 3 includes (blue + omuras + sei + brydei whales) (Fig 3). Association of Gray whale with the fin and humpback whales strongly supports the paraphyly. On the other hand cytb and cox1 trees, indicates 4 sub clades such as clade 1 include the mink whale; clade 2 includes (fin + humpback whale); clade 3 includes the Gray whale and clade 4 includes (blue + omuras + sei + brydei whales). The Balaenopteridae family comprises the balaenopterinae and megapterinae subfamilies. Association of the humpback whale with fin whale further confirms the paraphyly of Balaenopteridae. However based on control region and cytb analysis humpback whale grouped as the sister taxa to blue whales indicating the paraphyly of Eschrichtiidae + Balaenopteridae (Arnason *et al.*, 1993, 1994). Furthermore, analysis based on morphology grouped Gray whale as sister clade to the Balaenopteridae + Balaenidae (McLeod *et al.*, 1993; Geisler *et al.*, 1996; Bisconti *et al.*, 2000).



**Figure 1.** Phylogenetic relationship of mysticetes constructed from *cox1* gene by Bayesian inference (BI) method using MrBayes v3.2.5. Value at each node indicates Bayesian posterior probability.



**Figure 2.** Phylogenetic relationship of mysticetes constructed from *cytb* gene by Bayesian inference (BI) method using MrBayes v3.2.5. Value at each node indicates Bayesian posterior probability.



**Figure 3.** Phylogenetic relationship of mysticetes constructed from concatenated sequences of 13 PCGs by Bayesian inference (BI) method using MrBayes v3.2.5. Value at each node indicates Bayesian posterior probability.

Overall the phylogenetic analysis based on single gene (cox1 and cytb) and all 13 PCG indicates that the tree topology of single gene based trees were different. Although the tree topology of all PCG and cytb tree was similar but the posterior probability values of all PCG tree is better than the single gene based trees. In addition to that it includes more characters than single gene and the phylogenetic inference will be more accurate with high confidence. Therefore concatenated 13 PCG approach is always better and accurate than single gene based methods for evaluating the phylogenetic relationships of animals.

### REFERENCES

- Arnason U. and Gullberg A. (1993).** Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature. *J. Mol. Evol.* 37:312–322.
- Arnason U., Gullberg A. and Widegren B. (1993 a).** Cetacean mitochondrial DNA control region: sequences of all extant baleen whales and two sperm whale species. *Mol. Biol. Evol.* 10:960–970
- Arnason U., Gretarsdottir S. and Widegren (1992).** Mysticete (baleen whale) relationships based upon the sequence of the common cetacean DNA satellite. *Mol. Biol. Evol.* 9:1018–1028.
- Arnason U. and Gullberg A. (1994).** Relationship of baleen whales established by cytochrome b gene sequence comparison. *Nature* 367:726–728.
- Arnason U. and Gullberg A. (1996).** Cytochrome b nucleotide sequences and the identification of five primary lineages of extant cetaceans. *Mol. Biol. Evol.* 13:407–417
- Arnason U. and Best P. B. (1991).** Phylogenetic relationships within the Mysticeti (whalebone whales) based upon studies of highly repetitive DNA in all extant species. *Hereditas* 114:263–269.
- Arnason Ulfur, Adegoko Joseph A., Bodin Kristina Born, Erik W. Esa, Yuzine B. Gullberg, Anette Nilsson, Maria Short, Roger V., Xu Xiufeng, and Janke Axel (2002).** Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc. Natl. Acad. Sci.* 99:8151–8156.
- Arnason U., Gullberg A. and Janke A. (2004).** Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene* 333:27–34.
- Arnason U., Lammers F., Kumar V., Nilsson M. A. and Janke A. (2018).** Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Sci. Adv.* 4.
- Bisconti M. (2000).** New description, character analysis and preliminary phyletic assessment of two Balaenidae skulls from the Italian Pliocene. *Palaeontogr. Ital.* 87:37–66.



- Brown W. M., George M. and Wilson A. C. (1979).** Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* 76:1967–1971.
- Committee on Taxonomy (2017).** List of marine mammal species and sub species. *Society for Marine Mammology*, <https://www.marinemammalscience.org>.
- Fordyce Ewan R. and De M C. (2001).** Evolutionary history of cetaceans: a review. In: *Mazin, J.-M., de Buffre ´nil, V. (Eds.), Secondary Adaptation of Tetrapods to Life in Water. Verlag Dr. Friedrich Pfeil, Munich.* 169-233.
- Gatesy J., Hayashi C., Cronin M., atew A. and Arctander P. (1996).** Evidence from milkcasein genes that cetaceans are close relatives of hippopotamid artiodactyls. *Mol.Biol. Evol.* 13:954–963.
- Ginerich P. D., Wells N. A., Russell D. E. and Shah S. M. I. (1983).** Origin of whales in epicontinental remnant seas: new evidence from the early Eocene of Pakistan. *Science* 220:403-406.
- Geisler J. H. and Sanders A. (2003).** Morphological evidence for the phylogeny of Cetacea. *J. Mamm. Evol.* 10:23–129.
- Geisler J. H. and Luo Z. (1996).** The petrosal and inner ear of *Herpetocetus* sp. (Mammalia: Cetacea) and their implications for the phylogeny and hearing of archaic mysticetes. *J. Paleontol.* 70:1045– 1066.
- Gontcharov A. A., Marin B. And Melkonian M. (2004).** Are Combined Analyses Better Than Single Gene Phylogenies? A Case Study Using SSU rDNA and rbcL Sequence Comparisons in the Zygnematophyceae (Streptophyta). *Mol. Biol. Evol.* 21:612–624.
- McLeod S.A., Whitmore F.C., Barnes L.G., 1993.** The bowhead whale. In: *Burns, J.J., Montague, J.J., Cowles, C.J. (Eds.), Allen Press, Lawrence, Kans,* pp. 45–70.
- Milinkovitch M. C., Meyer A. and Powell J. R. (1994).** Phylogeny of all major groups of cetaceans based on DNA sequences from three mitochondrial genes. *Mol. Biol. Evol.* 11:939–948
- Olivio P. D., Van De Walle M. J., Laipis P. J., and Hauswirth W. W. (1983).** Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. *Nature* 306:400–402.
- Rychel A. L., Reeder T. W. and Berta A. (2004).** Phylogeny of mysticete whales based on mitochondrial and nuclear data. *Mol. Phylogenet. Evol.* 32:892–901.
- Sasaki Takeshi, Nikaido Masato, Hamilton Healy, Goto Mutsuo, Kato Hidehiro, Kanda Naohisa, Pastene Luis A., Cao Ying, Fordyce R. Ewan, Hasegawa Masami, and Okada Norihiro (2005).** Mitochondrial phylogenetics and evolution of mysticete whales. *Syst. Biol.* 54: 77–90.
- Sasaki Takeshi, Nikaido Masato, Wada Shiro, Yamada Tadasu K., Cao Ying, Hasegawa Masami, and Okada Norihiro (2006).** *Balaenoptera omurai* is a newly discovered baleen whale that represents an ancient evolutionary lineage. *Mol. Phylogenet. Evol.* 41:40–52.
- Thewissen J. G. M. and Williams E. M. (2002).** The Early Radiations of Cetacea (Mammalia): Evolutionary Pattern and Developmental Correlations. *Annu. Rev. Ecol. Syst.* 33, 73–90.
- Urantowka A. D., Krocak A. and Mackiewicz P. (2017).** The influence of molecular markers and methods on inferring the phylogenetic relationships between the representatives of the Arini (parrots, Psittaciformes), determined on the basis of their complete mitochondrial genomes. *BMC Evolutionary Biology* 17.
- Ursing B. M. and Arnason U. (1998).** Analyses of mitochondrial genomes strongly support a hippopotamus-whale clade. *Proc. R. Soc. B Biol. Sci.* 265:2251–2255.