Revised phylogeny of whales suggested by mitochondrial ribosomal DNA sequences

Michel C. Milinkovitch, Guillermo Ortí
& Axel Meyer

Revised phylogeny of whales suggested by mitochondrial ribosomal DNA sequences

Michel C. Milinkovitch[*][†], **Guillermo Ortí**[‡]
& Axel Meyer[‡][§]

* Institut de Recherche Interdisciplinaire en Biologie Humaine et Nucléaire, Campus Erasme, Université Libre de Bruxelles, Brussels 1070, Belgium
‡ Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794, USA

LIVING cetaceans are subdivided into two highly distinct suborders, Odontoceti (the echolocating toothed whales) and Mysticeti (the filter-feeding baleen whales), which are believed to have had a long independent history. Here we report the determination of DNA sequences from two mitochondrial ribosomal gene segments (930 base pairs per species) for 16 species of cetaceans, a perissodactyl and a sloth, and construct the first phylogeny for whales and dolphins based on explicit cladistic methods. Our data (and earlier published myoglobin sequences) confirmed that cetaceans are closely related to artiodactyls and that all families and superfamilies of cetaceans are monophyletic. A surprising finding was that one group of toothed whales, the sperm whales, is more closely related to the baleen whales than to other odontocetes. The common ancestor of baleen whales and sperm whales might have lived only 10–15 million years ago. The suggested paraphyly of toothed whales has many implications for classification, phylogeny and our understanding of the evolutionary history of cetaceans.

The transition from a terrestrial to the fully aquatic lifestyle of cetaceans required a complete remodelling of many biological systems. Understanding of the origin and evolutionary relationships of cetaceans was hindered by their highly specialized morphology. A link between cetaceans and ungulates was first suggested more than a hundred years ago[1] and recently confirmed by palaeontological[2-5] and molecular data[6-8] hinting at a closer relationship of cetaceans with artiodactyl than with perissodactyl ungulates. Living cetaceans have been assigned to the suborder Mysticeti (11 species) and the suborder Odontoceti (about 67 species), both thought to have separated from the extinct suborder Archeoceti more than 35–40 million years ago[9]. Fossils of the extant whales are much younger[9,10]. The fossil record of cetaceans is incomplete and has not provided clear evidence on the origin of and the relationships among the three suborders of cetaceans (reviewed in refs 9, 10). The taxonomic distinction between odontocets and mysticets seems well supported by morphological differences, which include the presence of teeth in the former and baleen, or whalebone in the latter. The ancient origin of baleen whales from a lineage of toothed whales was postulated because teeth are found in the condylarth mesonychids, an extinct group of ungulates linked to the origin of cetaceans[2-5] and because of the embryological potential of mysticetes to express teeth[2,10].

A strongly supported phylogeny, based on cladistic methods, of the major lineages of living cetaceans is currently not available. We present a phylogenetic hypothesis for all major groups of cetaceans (except river dolphins) based on mitochondrial ribosomal DNA sequences. These genes have been used successfully in a range of molecular phylogenetic studies (see for example refs 11–14). Using the polymerase chain reaction (PCR) we amplified and directly sequenced portions of the two (12S and 16S) mitochondrial ribosomal genes for 16 species of cetaceans, the cow (an artiodactyl)[15], the wild ass (a perissodactyl), a human[16] and a sloth. Data were analysed by maximum

parsimony[17], neighbour-joining[18] and maximum likelihood[19] methods, and the robustness of the phylogenetic hypotheses tested by bootstrapping[20]. Edentata (sloths, anteaters and armadillos) are the most primitive order of eutherian mammals (reviewed in ref. 21) and our sloth sequences were therefore used as root in the phylogenetic analyses.

These analyses resolved some open questions in cetacean systematics with confidence (Fig. 1). All families and superfamilies of cetaceans were found to be monophyletic: the beaked whales (Ziphioidea), the baleen whales (family Balaenopteridae), the sperm whales (Physeteroidea), the Delphinoidea, the dolphins (Delphinidae) and the porpoises (Phocoenidae) (Fig. 1). In agreement with previous morphological and molecular hypotheses[4-7], we find that cetaceans appear to be more closely related to artiodactyls than to perissodactyls (node a, Fig. 1). The monophyly of cetaceans is also strongly supported (node b, Fig. 1). Within cetaceans, the phylogenetic position of beaked whales must remain tentative because their placement as the sister group to all living whales is somewhat uncertain because of a low bootstrap value (node c, Fig. 1). In some of the analyses, beaked whales are placed with the sperm plus baleen whale group. We find strong support for the monophyly of the Delphinoidea (node e, Fig. 1) and the Delphinidae (node not labelled in Fig. 1). The relationships among the three families of the Delphinoidea are not well resolved, yet porpoises appear to be closer to dolphins than to belugas (node f, Fig. 1). The phylogeny within the Delphinidae remains largely unresolved (at the 50 per cent majority-rule bootstrap level). The confirmation of groupings based on previous palaeontological and molecular data raises confidence in the reliability of the phylogenetic information contained in these molecules.

A surprising outcome is the suggested sister group relationship between sperm whales and baleen whales. Sperm whales appear to be more closely related to baleen whales than they are to any other group of toothed whales (node d, Fig. 1). This phylogenetic hypothesis is at odds with the traditional odotocete-mysticete division[9]. The suggested paraphyly of odontocetes, making them an unnatural (not monophyletic) group, would seem to call for a reclassification of cetaceans. The taxonomic rank of baleen whales would need to be lowered if one subscribes to the cladistic view that groups of organisms that are given equal taxonomic ranks (that is, the suborders Mysticeti and Odontoceti) must each include an ancestor and all of its descendants. Evolutionary taxonomists, however, attempt to portray in the classification not only the branching of lineages but also their subsequent morphological divergence. No consensus has been reached among the schools of cladistic and evolutionary classification on this question.

More important than the potential impact on the classification of whales are the many biological and evolutionary implications of this finding. The mode and rate of morphological evolution might have proceeded differently than was previously assumed. For example, echolocation is believed to occur in all toothed whales but not in baleen whales. If our phylogenetic hypothesis is correct, it is more parsimonious to postulate that baleen whales lost the capacity for echolocation (and other behavioural, physiological and morphological adaptations related to acquisition of food), rather than to assume that they never evolved it.

Interestingly, Schlötterer et al.[22] found surprisingly low amounts of sequence divergence (3.2%) in DNA sequences of a nuclear simple-sequence locus between the odontocetes and mysticete balaenopterids; a very low divergence for an estimated split of 35–40 Myr[9,10]. They presented two alternative explanations[22] for their finding: (1) whale DNA evolves at the slowest reported substitution rate for neutral nucleotide positions or (2) a re-evaluation of the age of modern whales is in order. On the basis of these nuclear DNA data Schlötterer et al.[22] suggest that the origin of all modern whales might date back to only 20–25 Myr and that whales might still have a much slower mutation rate than primates. Obviously, if whales had typical substitution

FIG. 1 Majority rule bootstrap consensus tree based on 394 base pairs of the small (12S) and 536 bp of the large (16S) mitochondrial ribosomal genes of 16 cetacean species. We sequenced PPHOC; *Phocoena phocoena* (harbour porpoise); PSPIN, *Phocoena spinipinnis* (Burmeister's porpoise); DDELP, *Delphinus delphis* (common dolphin); TTRUN, *Tursiops truncatus* (bottlenose dolphin); CEUTR, *Cephalorhynchus eutropia* (black dolphin); LOBSC, *Lagenorhynchus obscurus* (dusky dolphin); LALBI, *Lagenorhynchus albirostris* (white-beaked dolphin); LPERO, *Lissodelphis peronii* (southern right whale dolphin); GMELA, *Globicephala melas* (long-finned pilot whale), DLEUC, *Delphinapterus leucas* (beluga); KBREV; *Kogia breviceps* (pygmy sperm whale); PCATO, *Physeter catodon* (sperm whale); MEURO, *Mesoplodon europaeus* (Gervais' beaked whale); MPERU, *Mesoplodon peruvianus* (peruvian beaked whale); BPHYS, *Balaenoptera physalus* (fin whale); MNOVA, *Megaptera novaeangliae* (humpback whale); and two outgroup species, wild Ass (the kulan; *Equus hemionus*) and a sloth (*Choloepus didactylus*). Tissue samples were obtained from stranded or by-catch animals, or from delphinaria. Sequences were aligned with CLUSTAL[26] and secondary structure models of the cow ribosomal genes[27,28]. For the fin whale, the 16S sequence is from ref. 24; for the 12S ribosomal gene, one position (201 in our sequence) differs from ref. 24. DNA was extracted from blood, skin, spleen or liver tissue

from frozen or specimens preserved in dimethyl sulphoxide. The primers used for the PCR amplification and direct sequencing[29,30] of part of the 12S gene were modified[31] L1091 (ref. 29) and H1478 (ref. 29) primers, and those for the 16S gene were 16sar-L and 16sbr-H[31]. Single-stranded amplifications were done directly from whole genomic DNA, followed by three phenol–chloroform extractions and two ethanol precipitations; further details of the protocol have been reported[29-31]. Most of the sequences were determined (at IRBHN) with an automatic sequencer (Applied Biosystems 373A) following the manufacturer's protocols. In all cases both strands were sequenced and some species were sequenced manually as well as automatically. Sequences have been deposited in EMBL under accession numbers Z18632–Z18666 and can also be obtained from the authors. Numbers indicate bootstrap values; above the branches from a neighbour-joining[18] (NJ) analysis (1,000 replications; program from T. S. Whittam), below the branches numbers from a parsimony bootstrap analysis (PAUP[17], 100 replications). The sequence divergences were corrected for multiple substitutions according to the observed threefold higher frequency of transitions over transversions (TS/TV), based on regression analyses done among delphinids, delphinoids and all cetaceans under 8% uncorrected sequence divergence. All regression analyses yielded slopes between 2.9 and 3.5, hence the use of a 1:3 TS/TV in NJ and parsimony analyses. A similar TS/TV for the same genes has been found among some groups of ungulates[13]. Three NJ analyses were performed with combined data sets and distance matrices calculated based on Kimura, maximum likelihood (PHYLIP, Version 3.4.1, DNADIST[19]) or Jukes–Cantor distances. All three distance matrices yielded NJ trees in agreement with the topology shown here. In all analyses (NJ and parsimony) nodes a, b, d and e are stable and occur in all most parsimonious trees and in the bootstrap trees. Parsimony analyses with a TS/TV 1:1, 1:10 and transversion parsimony also yielded congruent topologies to the bootstrap tree shown underscoring the robustness of the
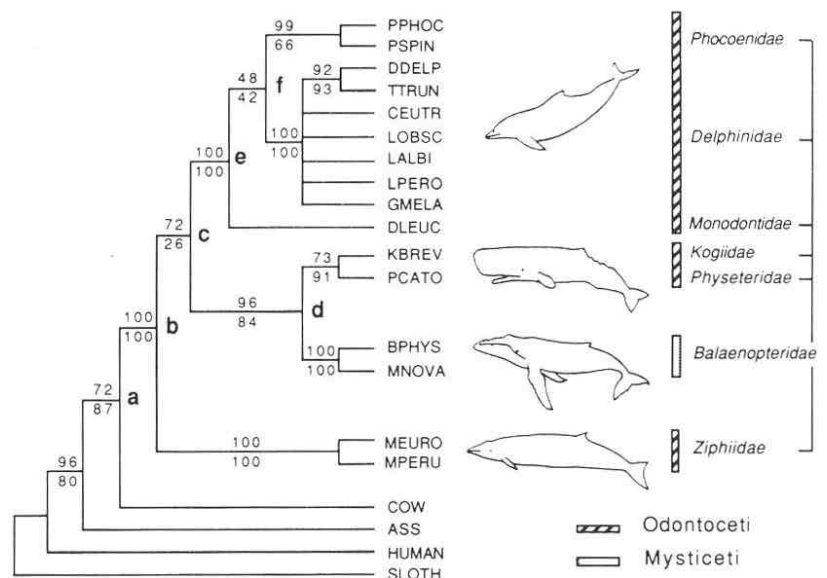
result. For example, in an unweighted parsimony analysis, six equally shortest trees were found (congruent with the topology shown here), tree length = 822, CI = 0.66, CI excluding uninformative characters = 0.54. In all parsimony analyses the following options were used: heuristic search (for bootstrapping; branch-and-bound for most other searches), MULPARS option in effect, MAXTREES = 100, and TBR branch swapping. Indels were coded as single characters, irrespective of their length. When indels of different length overlapped each size class of indel was assigned a particular character state. Further variations of the analyses were to use the cow as closer outgroup instead of the sloth, to include and exclude gaps, to include and exclude compensatory changes in the putative stems, and to analyse 12S and 16S separately. All variations always yield a congruent topology with the one shown here. Nodes c and f are unstable, having low bootstrap values, and hence are regarded as tentative. In some parsimony analyses beaked whales are grouped with the sperm-baleen whale group. Node f was the only unstable node when data were submitted to the successive approximation approach to character-weighting of Farris. None of the most parsimonious solutions ever contained a monophyletic Odontoceti group. Constraining Odontoceti monophyly produced a solution eight steps longer than the shortest trees (tree length = 461 steps, bovine as outgroup, unweighted). Bootstrap analysis (NJ: 1,000 replications and parsimony: 100 replications) yielded 0% support for the monophyly of odontocetes. With parsimony analysis node d is more strongly supported by the 16S data (bootstrap value of 95%) than by the 12S data (60%), and than by both genes combined (84%). Assigned branch lengths (ACCTRAN option in PAUP, unweighted) between the major nodes are: a–b: 38, b–c: 16, c–d: 17, c–e: 24, e–f: 11. A maximum likelihood analysis (DNAML[19]) produced a topology in agreement with the one shown, with all branch lengths significantly positive ($P < 0.01$) except those among delphinids.

---

rates, the estimated age of divergence would have been even younger. According to our data, baleen whales do not appear to have experienced a long independent history from toothed whales and might indeed be even younger than had been suggested by Schlötterer *et al.*[22]. The transversion divergence rate for mitochondrial ribosomal genes of ungulates has been calibrated to be about 0.14% per million years[12,23,24]. We find a 1.5 per cent transversion divergence between sperm and baleen whales. If, *prima facie*, cetaceans have a similar molecular divergence rate as ungulates then the split between sperm whales and Balaenopteridae might be only around 10–13 Myr. This is not in disagreement with the palaeontological data because the oldest fossils of balaenopterids are from Late Miocene deposits (5–10 Myr)[9]. The distance matrix of Schlötterer *et al.*[22] does not seem to rule out a close relationship of sperm whales with baleen whales. Of all groups of odontocetes, the sperm whales have the shortest mean distance to baleen whales (2.5%) (beaked (3.0%), dolphins (3.4%), porpoise (3.5%)). Furthermore, diver-

gences among families of odontocetes are higher than 2.5%. But these values are quite similar and based on a relatively small number of substitutions. Our mitochondrial data agree with these nuclear data[22] in suggesting the possible need for a re-evaluation of the age of modern whales, and go further to suggest that odontocetes might be paraphyletic and that baleen whales and sperm whales share a recent common ancestor.

Further support for the sperm plus baleen whale relationship derives from other nuclear data. Published myoglobin amino-acid sequences of cetaceans (10 species), cow, horse, hedgehog and human were phylogenetically analysed (PAUP[17]). The monophyly of the superfamilies sperm whales, beaked whales, dolphins plus porpoises (no white whale sequence is available) is supported (Fig. 2). The phylogeny is weaker[20], but largely the same as the one based on our mitochondrial ribosomal sequences: sperm and baleen whales are grouped together (Fig. 2). Previous analyses[25] of cetacean relationships based on myoglobin and haemoglobin sequences gave equivocal results
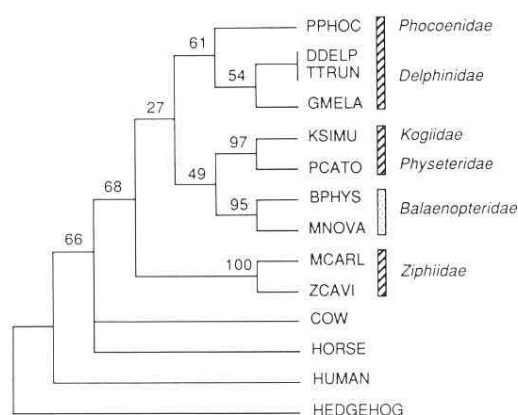
FIG. 2 Bootstrap consensus tree based on published myoglobin sequences. The majority rule bootstrap consensus tree favours the grouping of the Ziphiidae with the sperm whale–baleen whale clade at a bootstrap value of 36%. Numbers represent bootstrap values (100 replications). Analysis was done with PAUP[17] using a stepmatrix (PROTPARS) which gives the minimum number of amino-acid replacement substitutions needed to convert one amino acid to another. Abbreviations of the species names are as in Fig. 1, except MCARL, *Mesoplodon carlhubbsi* (Hubbs' beaked whale); ZCAVI, *Ziphius cavirostris* (Goose-beaked whale). Species were chosen to be as close as possible to the ones used for our mitochondrial sequence analysis. The inclusion of other cetacean species for which myoglobin sequences are available (*Balaenoptera acutorostrata, B. borealis, Eschrichtius gibbosus, Inia geoffrensis, Phocoena dalli, Stenella attenuata, Orcinus orca*) and further outgroups (sheep and red deer) did not change the topology but tended to lower the bootstrap values.

but hinted that toothed whales might not be monophyletic. We anticipate that more molecular data will be collected that will test the phylogeny of cetaceans. □

1. Flower, W. H. *Proc. zool. Soc. Lond.* **1883,** 466–513 (1883).
2. Van Valen, L. *Bull. Am. Mus. nat. Hist.* **132,** 1–126 (1966).
3. Szalay, F. S. *Am. Mus. nat. Hist. Nov.* **2361,** 1–26 (1969).
4. Gingerich, P. D., Wells, N. A., Russell, D. E. & Shah, S. M. I. *Science* **220,** 403–406 (1983).
5. Gingerich, P. D. Smith, B. H. & Simons, E. L. *Science* **249,** 154–157 (1990).
6. Goodman, M., Czelusniak, J. & Beeber, J. E. *Cladistics* **1,** 171–185 (1985).
7. Irwin, D. M., Kocher, T. D. & Wilson, A. C. *J. molec. Evol.* **32,** 128–144 (1991).
8. Milinkovitch, M. C. *J. evol. Biol.* **5,** 149–160 (1992).
9. Barnes, L. G., Doming, D. P. & Ray, C. E. *Mar. Mammal Sci.* **1,** 15–53 (1985).
10. Barnes, L. G. & Mitchell, E. in *Evolution of African Mammals.* (ed. Maglio, V. J. & Cooke, H. B. S.) 582–602 (Harvard University Press, London, 1978).
11. Mindell, D. P. & Honeycutt, R. L. *A. Rev. Ecol. Syst.* **21,** 541–566 (1990).
12. Allard, M. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **89,** 3972–3976 (1992).
13. Miyamoto, M. M., Kraus, F. & Ryder, O. A. *Proc. natn. Acad. Sci. U.S.A.* **87,** 6127–6131 (1990).
14. Gatsey, J., Yelon, D., DeSalle, R. & Vrba, E. S. *Molec. Biol. Evol.* **9,** 433–446 (1992).
15. Anderson, S. *et al. J. molec. Biol.* **156,** 683–717 (1982).
16. Eperon, I. C. *et al. Nature* **286,** 460–467 (1980).
17. Swofford, D. L. *Phylogenetic Analysis Using Parsimony* Version 3.0s (Illinois Natural History Survey, Champaign, 1991).
18. Saitou, N. & Nei, M. *Molec. Biol. Evol.* **4,** 406–425 (1987).
19. Felsenstein, J. *PHYLIP* Version 3.4.1. (University of Washington, Seattle, 1991).
20. Felsenstein, J. *Evolution* **39,** 783–791 (1985).
21. Novacek, M. J. *Nature* **356,** 121–125 (1992).
22. Schlötterer, C., Amos, B. & Tautz, D. *Nature* **354,** 63–65 (1991).
23. Kraus, F. & Miyamoto, M. M. *Syst. Zool.* **40,** 117–130 (1991).
24. Arnason, U., Gullberg, A. & Widegren, B. *J. molec. Evol.* **33,** 556–568 (1991).
25. Czelusniak, J. *et al.* in *Current Mammalogy* Vol. 2 (ed. Genoways, H. H.) 545–572 (Plenum, New York, 1990).
26. Higgins, D. G. & Sharp, P. M. *Gene* **73,** 237–244 (1988).
27. Gutell, R. R. & Fox, G. E. *Prog. Nucleic Acid Res.* **16** (suppl.), R175–R313 (1988).
28. Gutell, R. R., Weiser, B., Woese, C. & Noller, H. F. *Nucleic Acids Res.* **32,** 155–215 (1985).
29. Kocher, T. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **86,** 6196–6200 (1989).
30. Meyer, A., Kocher, T. D., Basasibwaki, P. & Wilson, A. C. *Nature* **347,** 550–553 (1990).
31. Palumbi, S. R. *The Simple Fool's Guide to PCR* (Department of Zoology, University of Hawaii, Honolulu, 1991).