

# More DNA Support for a Cetacea/Hippopotamidae Clade: The Blood-Clotting Protein Gene $\gamma$ -Fibrinogen

John Gatesy

Laboratory of Molecular Systematics and Evolution, Department of Ecology and Evolutionary Biology, University of Arizona

Recent phylogenetic analyses of DNA sequences suggest that cetaceans (whales) and hippopotamid artiodactyls (hippos) are extant sister taxa. Consequently, the shared aquatic specializations of these taxa may be synapomorphies. This molecular view is contradicted by paleontological data that overwhelmingly support a monophyletic Artiodactyla (even-toed ungulates) and a close relationship between Cetacea and extinct mesonychian ungulates. According to the fossil evidence, molecular, behavioral, and anatomical resemblances between hippos and whales are interpreted as convergences or primitive retentions. In this report, competing interpretations of whale origins are tested through phylogenetic analyses of the blood-clotting protein gene  $\gamma$ -fibrinogen from cetaceans, artiodactyls, perissodactyls (odd-toed ungulates), and carnivores (cats, dogs, and kin). In combination with published DNA sequences, the  $\gamma$ -fibrinogen data unambiguously support a hippo/whale clade and are inconsistent with the paleontological perspective. If the phylogeny favored by fossil evidence is accepted, the convergence at the DNA level between Cetacea and Hippopotamidae is remarkable in its distribution across three genetic loci:  $\gamma$ -fibrinogen, the linked milk casein genes, and mitochondrial cytochrome *b*.

## Introduction

The evolutionary origin of Cetacea has puzzled zoologists for over a century. It generally has been assumed that there are no extant functional or anatomical intermediates to obligately aquatic cetaceans. Thus, paleontological finds have provided the critical evolutionary links between cetaceans and their terrestrial ungulate ancestors (e.g. Gingerich et al. 1983; Thewissen, Hussain, and Arif 1994).

Surprisingly, recent phylogenetic analyses of DNA sequences hint that semiaquatic hippopotamid artiodactyls are the closest extant relatives of Cetacea (fig. 1A). Both nuclear casein sequences (Gatesy et al. 1996) and mitochondrial (mt) cytochrome *b* sequences (Irwin and Arnason 1994; Arnason and Gullberg 1996; Hasegawa and Adachi 1996) favor a hippo/whale clade. This tentatively supported hypothesis begs the question of whether the superficially similar aquatic specializations of these taxa are further evidence of their close kinship.

The molecular inference is difficult to reconcile with paleontological data that favor a monophyletic Artiodactyla (Prothero, Manning, and Fischer 1988) and the derivation of Cetacea from within the mesonychian radiation of the late Paleocene/early Eocene (fig. 1B). Numerous dental and skeletal synapomorphies link Cetacea to the extinct mesonychian ungulates (Thewissen 1994; Zhou et al. 1995).

Additional data are necessary to discriminate between these contrasting scenarios of cetacean genesis. Smith et al. (1996) pointed out that "coding sequences of both mtDNA and nuclear genes have yet to provide highly convincing data [on cetacean origins], and thus ... a more fruitful area of investigation might involve

noncoding nuclear DNA." In this report, I combine new comparative sequence data for introns 2–3 and exons 2–4 of  $\gamma$ -fibrinogen with published DNA sequences for  $\kappa$ -casein,  $\beta$ -casein, and mt cytochrome *b* to assess the putative Hippopotamidae/Cetacea sister group relationship.

## Materials and Methods

### PCR, Sequencing, and Alignment

$\gamma$ -Fibrinogen is a plasma glycoprotein that interacts with the related  $\alpha$ - and  $\beta$ -fibrinogen chains in the blood coagulation process. In *Homo*, the nuclear  $\gamma$ -fibrinogen gene is divided into 10 exons and spans over 8 kb (Rixon, Chung, and Davie 1985).

A 523–581-bp fragment of  $\gamma$ -fibrinogen (exon 3, intron 2, intron 3, and sections of exons 2 and 4) was PCR-amplified, cloned, and sequenced from representatives of the six extant lineages of artiodactyls that extend to the Oligocene (Pecora, Tragulidae, Hippopotamidae, Suidae, Tayassuidae, and Camelidae), three basal groups of Cetacea (Balaenopteridae, Delphinoidea, and Physteridae), and the primary divisions of both Perissodactyla (Hippomorpha/Ceratomorpha) and Carnivora (Feloidea/Caniformia—see below). PCR, cloning, and sequencing methods were as in Gatesy et al.

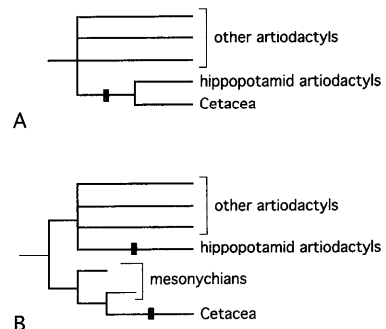


FIG. 1.—Two contrasting hypotheses of whale origins: (A) the inference from DNA sequence data and (B) the paleontological view. The bars mark the evolution of aquatic traits shared by hippos and whales.

Abbreviations: mt, mitochondrial.

Key words:  $\gamma$ -fibrinogen, Cetacea, Artiodactyla, Hippopotamidae.

Address for correspondence and reprints: John Gatesy, Laboratory of Molecular Systematics and Evolution, Department of Ecology and Evolutionary Biology, University of Arizona, Biosciences West, Tucson, Arizona 85721. E-mail: gatesy@mullis.biosci.arizona.edu.

*Mol. Biol. Evol.* 14(5):537–543. 1997

© 1997 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Downloaded from <http://mbe.oxfordjournals.org/> by guest on December 31, 2015

238-human 5'	3'	exon2	intron2
AGAAHAATGCTGCATCTTAGATG		AAAGATT	GTAAGTTACTTTTAT---GTTTCTCTC---TGTGTTTGAACGGACT---TGGGCAGAGAAATTTTATCTTATAAGAGATCTTTTAGC
<i>Ovis dalli</i>		CG	A.G
<i>Alces alces</i>		CG	G.G
<i>Giraffa camelopardalis</i>	C	CG	G
<i>Tragulus napu</i>		AG	C
<i>Delphinapterus leucas</i>		AGT	G
<i>Physeter catadon</i>		AGT	CG
<i>Balaenoptera physalus</i>	C	AGT	G
<i>Choeropsis liberiensis</i>	M	AGT	CT
<i>Sus scrofa</i>	G	AG	G
<i>Tayassu tajacu</i>	G	AG	C
<i>Camelus dromedarius</i>	M	CGT	T
<i>Tapirus indicus</i>		A	T
<i>Equus przewalskii</i>		AGT	C
<i>Crocota crocuta</i>		AGT	C
<i>Canis latrans</i>	M	AGT	C
<i>Homo sapiens</i>		AGT	C

<i>Ovis dalli</i>	AAATATCATTGAC-AGTTAGACTTTTT	-----	AAAAACATAGTCTAGGTTTTACCTGTTTTT--
<i>Alces alces</i>	T	G	AA
<i>Giraffa camelopardalis</i>	G		W
<i>Tragulus napu</i>	G	G	TTT
<i>Delphinapterus leucas</i>	G		TTTTX
<i>Physeter catadon</i>	G		TTTTT
<i>Balaenoptera physalus</i>	C	G	TTT
<i>Choeropsis liberiensis</i>	G		TTTTTX
<i>Sus scrofa</i>	C	G	G
<i>Tayassu tajacu</i>	G	C	G
<i>Camelus dromedarius</i>	G	C	G
<i>Tapirus indicus</i>	C	A	G
<i>Equus przewalskii</i>	AG	GT	A
<i>Crocota crocuta</i>	G	T	G
<i>Canis latrans</i>	G	T	G
<i>Homo sapiens</i>	C	G	C

<i>Ovis dalli</i>	---ATTAT	-----	CAGGATTTTTAA-TCACACT	-----	GTTGTATTT	
<i>Alces alces</i>	A	AGAT-CTCA	-TGTAGCTTTTT		AT	
<i>Giraffa camelopardalis</i>	A	AGAT-CTCA	-TGTAGCTTTTT		AT	
<i>Tragulus napu</i>	A	GAGAT-CTA	-TGTGGCTTTTA		ATTG	
<i>Delphinapterus leucas</i>	AG	AGAT-CTTA	-TGTAGCTTTTT		TT	
<i>Physeter catadon</i>	AGGAGAT-CTTA	-TGTAGCTTTTT		AT	A	
<i>Balaenoptera physalus</i>	AG	AGAT-CTTA	-TGTAGCTTTTT		AT	
<i>Choeropsis liberiensis</i>	AG	AGAT-CTTA	-TGTAGCTTTTT		AT	
<i>Sus scrofa</i>	A	ACAG	AGAT-TTCA	-GGTAGCTTTTT	ATG	
<i>Tayassu tajacu</i>	C	ACAG	AGAT-TTTA	-GGTAGCTTTTT	ACG	
<i>Camelus dromedarius</i>	TTXA	C	GG	GGAT-TTTA	-TGTAGCTTTTT	GT
<i>Tapirus indicus</i>	A	AGAT-TTTA	AAGTGTAGCTTTT		AT	
<i>Equus przewalskii</i>	G	AGAT-TTTA	AAGTGTAGCTTTT		AT	
<i>Crocota crocuta</i>	AG	AGATTTTA	-TGTAATTTT		AAAAATTTATCTAGTTTTAAAAAT	
<i>Canis latrans</i>	A	GGAT-TTTA	-TGCARGTTTTTTGTTGTTGTTTGTGTTTGAAT		A	
<i>Homo sapiens</i>	C	A	AGAT-TTTA	AAGAGTAGCATCT		GT

	exon3
<i>Ovis dalli</i>	TCAG GGTAGTTATTGCCCACTACCTGTGGAATGCAGACTTCTGTCTAATTTACCAACCAGTGTAGACAAGGATCTACGAAATTTGGAAGGCATCTTC
<i>Alces alces</i>	G
<i>Giraffa camelopardalis</i>	G
<i>Tragulus napu</i>	C
<i>Delphinapterus leucas</i>	T
<i>Physeter catadon</i>	T
<i>Balaenoptera physalus</i>	T
<i>Choeropsis liberiensis</i>	R
<i>Sus scrofa</i>	C
<i>Tayassu tajacu</i>	C
<i>Camelus dromedarius</i>	C
<i>Tapirus indicus</i>	C
<i>Equus przewalskii</i>	C
<i>Crocota crocuta</i>	G
<i>Canis latrans</i>	G
<i>Homo sapiens</i>	T

FIG. 2.—The final alignment of the  $\gamma$ -fibrinogen sequences. Periods represent nucleotide identity to the reference *Ovis dalli* sequence. Dashes indicate gaps introduced into the alignment. The degenerate PCR primers are shown at the 5' and 3' ends of the alignment with the positions of the primers in the *Homo*  $\gamma$ -fibrinogen gene (Rixon, Chung, and Davie 1985). @'s are above nucleotide positions that unambiguously support the Cetacea/Hippopotamidae clade for the topology in figure 4. X's in the sequences are TA cloning ambiguities or natural polymorphisms (gap or T).

(1996). Degenerate primers for  $\gamma$ -fibrinogen are shown in figure 2.

The sequences were aligned to the human  $\gamma$ -fibrinogen gene (Rixon, Chung, and Davie 1985) with MALIGN, a multiple sequence alignment program that uses parsimony as the basis for alignment choice (Wheeler and Gladstein 1994). MALIGN parameters were: leading, trailing, and internal gap cost = 3, extragaps = 2, changecost = 1, nogaps, score 3, quick, atbr, and contig. Adjustments were made to the algorithmic alignment by eye using SeqApp 1.9a (Gilbert 1992). These changes

were mainly the consolidation of adjacent gaps in intron 2 and decreased the overall cost of the alignment from 657 to 619 steps. The final alignment of 651 nucleotide positions is shown in figure 2.

In order to match the taxonomic sampling for the  $\gamma$ -fibrinogen data set, sections of  $\kappa$ -casein exon 4 and  $\beta$ -casein exon 7 were PCR-amplified and sequenced from representatives of Physeteridae and Caniformia (see below). PCR primers and methods were as in Gatesy et al. (1996). The new casein sequences were easily incorporated into published alignments for  $\kappa$ -casein and

<i>Ovis dalli</i>	TATCAGGTGAGAAACAAACATCAGAAGCCACAGAGCTGGTCAAAGCAATCAAGATCAGCTATAACCCCTGATGAAACCGTCAAAGCCAA	intron3	GTGAGACTCTAA
<i>Alces alces</i>	.....G.....C.A.....A.....A.....		.....A.....
<i>Giraffa camelopardalis</i>	.....A.....C.A.....A.....A.....		.....A.....
<i>Tragulus napu</i>	.....G.....TC.A.....G.A.....A.....		.....A.....
<i>Delphinapterus leucas</i>	CG.....G.....G.....T.....C.....CCG.T.....G.....AG.....A.....		.....A.....
<i>Physeter catadon</i>	CG.....G.....G.....T.....A.....C.....CCG.T.....G.....AG.....A.....		.....A.....
<i>Balaenoptera physalus</i>	C.G.A.....G.....G.....T.....C.....CCG.T.....G.....AG.R.....A.....		.....A.....
<i>Choeropsis liberiensis</i>	C.G.....G.....GG.....C.A.....C.....C.....AG.....A.....C.....		.....A.....
<i>Sus scrofa</i>	C.....A.....G.....G.....A.....C.....C.....A.....T.....TA.....A.....G.....A.....G.....		.....A.....G.....
<i>Tayassu tajacu</i>	C.....A.....G.R.....A.....A.....C.....C.A.....T.....C.....RG.....A.....A.....G.....A.....G.....		.....A.....G.....
<i>Camelus dromedarius</i>	C.....A.....GA.....G.....A.....A.....CG.....C.A.....CC.....G.....T.....		.....T.....
<i>Tapirus indicus</i>	A.....A.C.....A.....A.....A.....A.....C.....C.AG.....G.....A.....AA.....G.....AA.....G.....		.....AA.....G.....
<i>Equus przewalskii</i>	C.....GA.C.....A.....TC.....C.....GA.C.....A.....A.....C.....CGA.C.....T.....AC.....A.....G.....AA.....G.....		.....AA.....G.....
<i>Crocota crocuta</i>	AG.GG.A.....A.....G.....A.....AT.....A.....T.....C.AG.....C.....T.....AA.....G.....AC.....AA.....		.....AA.....
<i>Canis latrans</i>	AGA.GAA.....A.....C.....G.....A.....AG.....A.....T.....C.A.....C.....T.....C.....G.....AC.....AA.....A.....		.....AA.....A.....
<i>Homo sapiens</i>	C.....A.....A.....A.....T.....A.....C.....A.....A.....C.AC.....CT.....T.....T.A.....A.....AAA.....		.....AAA.....
<i>Ovis dalli</i>	CTGCCACTGACTGAGAGAG-----TCATCTATGAAGTGCTTTTCYGGTATTT-TAGTGGTTCTGTTTGCATTATGTAT--TTTATATCTCTGATCC		
<i>Alces alces</i>	.CC.....C.....A.....A.....T.....T.....C.....-.....T.....		
<i>Giraffa camelopardalis</i>	.....A.....A.....T.....C.....T.....-.....T.....		
<i>Tragulus napu</i>	.CA.....A.....A.....T.....T.....C.....TT.....A.....A.....C.....TT.....T.....G.....C.....		
<i>Delphinapterus leucas</i>	.CA.....ACT.....G.....T.....T.....A.....-.....C.....T.....C.....T.....G.....T.....		
<i>Physeter catadon</i>	.CA.....ACT.....A.....G.....T.....T.....T.....C.....T.....T.....AC.....T.....G.....T.....		
<i>Balaenoptera physalus</i>	.CA.....ACT.....G.....T.....T.....T.....-.....C.....T.....C.....T.....G.....C.....		
<i>Choeropsis liberiensis</i>	-.A.TG.....T.....ACT.....G.....G.....T.....T.....T.....-.....CR.....C.....T.....G.....C.....		
<i>Sus scrofa</i>	.CAG.G.....C.....A.A.....G.....GG.....T.....T.....G.....G.....T.....G.....GA.....R.....T.....T.....GY.....C.....		
<i>Tayassu tajacu</i>	.CAG.A.....C.....A.AA.....GG.....GT.....T.....T.....C.G.....G.....CC.....GA.....GA.....G.....TC.....T.....G.....		
<i>Camelus dromedarius</i>	.CA.....C.C.....G.....A.A.....GY.....T.....T.....GG.....CA.....A.....A.....T.....T.....C.....		
<i>Tapirus indicus</i>	A.....A.....G.....A.A.....A.....CT.G.....T.....T.....G.....AG.....A.C.....A.....A.....T.....G.....A.....T.....		
<i>Equus przewalskii</i>	A.A.....C.....G.....A.A.....A.....T.....G.....G.....T.....T.....G.....-.....T.....G.A.....TC.GA.C.A.-.....G.....T.....G.....		
<i>Crocota crocuta</i>	A.A.T.....C.....A.A.....-.....G.....T.....G.....TG.....R.T.....G.....-.....C.....C.A.....AC.....TG.....G.....		
<i>Canis latrans</i>	A.A.TG.C.G.....C.A.....-.....G.....TC.....C.....T.....YT.....CC.....GCC.....CA.....CC.....T.....G.....		
<i>Homo sapiens</i>	AGA.T.....CA.A.A.TAATAATA.A.....G.....T.....T.....G.....-.....T.....A.....G.....-.....T.....G.....		
<i>Ovis dalli</i>	CATATTACAG	exon4	GTAACATAGAGAGTCTACCAAGAATTACAAGAGATGATG
<i>Alces alces</i>	.....A.....C.....		.....C.....
<i>Giraffa camelopardalis</i>	.G.....C.....		.....C.....
<i>Tragulus napu</i>	.C.....AC.....G.....G.....C.....A.....		.....A.....G.....T.....
<i>Delphinapterus leucas</i>	.G.....A.....GGG.....A.....T.....C.....A.....G.....T.....		
<i>Physeter catadon</i>	.G.....G.....GGG.....C.....A.....T.....		
<i>Balaenoptera physalus</i>	.G.....A.....GGG.....C.....T.....C.....A.....T.....		
<i>Choeropsis liberiensis</i>	.....GG.....C.....A.....W.....		
<i>Sus scrofa</i>	.C.....T.....A.....GG.....C.A.....C.....G.....G.....C.....A.....T.....		
<i>Tayassu tajacu</i>	.C.G.....AC.....GG.....C.A.....C.....AG.....C.....A.....T.....		
<i>Camelus dromedarius</i>	.....GG.....G.....T.....A.....		
<i>Tapirus indicus</i>	.G.....CGGA.....T.....T.....G.....G.....C.....A.....C.....		
<i>Equus przewalskii</i>	.....C.....GA.....TGC.....T.....GAG.....G.....C.....A.....		
<i>Crocota crocuta</i>	TGCCG.....AC.....CG.....T.....G.....G.....C.....A.....		
<i>Canis latrans</i>	.....AC.....GGG.....C.T.G.....R.....G.....C.....C.....A.....		
<i>Homo sapiens</i>	T.....A.....TG.....CGC.....TTT.....G.....C.....G.....T.....T.A.....		
		804-human	
		3'	
		CTTCTTAACTTTATCTCT	5'

FIG. 2. (Continued)

$\beta$ -casein (Gatesy et al. 1996). Only one additional gap was necessary to accommodate the new sequences.

Taxa were: (g =  $\gamma$ -fibrinogen, k =  $\kappa$ -casein, b =  $\beta$ -casein, c = cytochrome *b*. New sequences for  $\gamma$ -fibrinogen,  $\kappa$ -casein and  $\beta$ -casein are marked with #'s. Other sequences are from GenBank.) Cervidae = g#b-Alces alces, k-Cervus nippon, c-Odocoileus hemionus; Bovidae = g#-Ovis dalli, k+b+c-Ovis aries; Giraffidae = g#+k+b+c-Giraffa camelopardalis; Tragulidae = g#+b+c-Tragulus napu, k-Tragulus javanicus; Delphinoidea = g#+b+c-Delphinapterus leucas, k-Delphinidae sp.; Physteridae = g#+k#b#+c-Physeter catadon; Balaenopteridae = g#+k+b+c-Balaenoptera physalus; Hippopotamidae = g#-Choeropsis liberiensis, k+b+c-Hippopotamus amphibius; Suidae = g#+k+b+c-Sus scrofa; Tayassuidae = g#+k+b+c-Tayassu tajacu; Camelidae = g#+b+c-Camelus dromedarius, k-Lama guanicoe; Hippomorpha = g#-Equus przewalskii, k+b+c-Equus grevyi; Ceratomorpha = g#+k+b-Tapirus indicus, c-Diceros bicornis; Feloidea = g#-Crocota crocuta, b+k-Panthera uncia, c-Panthera leo; Caniformia = g#-Canis latrans, k#b#+c-Ailurus fulgens; Primates = g+k+b+c-Homo sapiens. New sequences were submitted to GenBank under accession numbers U86643-U86661.

### Phylogenetic Analysis

The  $\gamma$ -fibrinogen data were analyzed cladistically in combination with DNA sequences for the linked milk casein genes (Chikuni et al. 1995; Gatesy et al. 1996) and mt cytochrome *b* (Irwin, Kocher, and Wilson 1991; Irwin and Arnason 1994; Arnason and Gullberg 1996; Ledje and Arnason 1996). The following subsets of data were analyzed:  $\gamma$ -fibrinogen exons,  $\gamma$ -fibrinogen introns,  $\gamma$ -fibrinogen introns + exons, mt cytochrome *b*, the linked  $\beta$ + $\kappa$ -caseins, the three nuclear genes, and all four genes combined. For some taxa, all four genes were not derived from a single species. Each of these "composite" terminal taxa was assumed to be monophyletic.

In higher-level comparisons of mt cytochrome *b* from ungulates, Irwin, Kocher, and Wilson (1991) and Milinkovitch, Orti, and Meyer (1995) noted a saturation of transitions at third codon positions. This class of nucleotide substitutions was ignored in phylogenetic analyses of mt cytochrome *b*.

PAUP 3.1.1 (Swofford 1993) searches were branch-and-bound or heuristic with 100 random taxon addition replicates and TBR branch swapping. Gaps were scored as missing data, and polymorphisms/PCR errors among clones were treated as ambiguities.

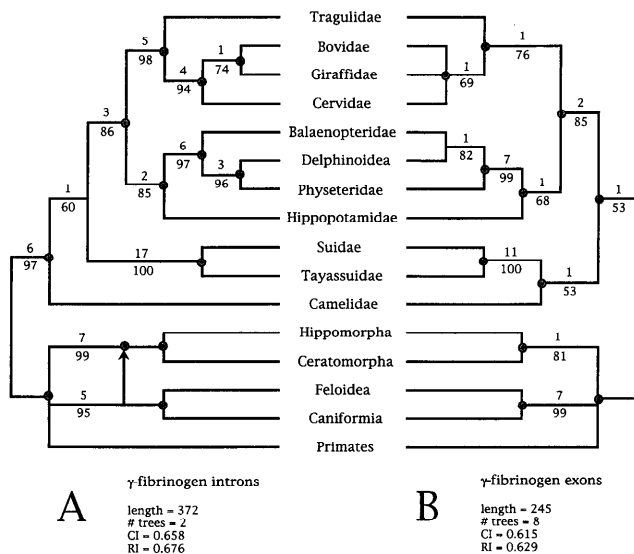


FIG. 3.—Strict consensus trees of minimum-length topologies for  $\gamma$ -fibrinogen introns (A) and exons (B). Tree lengths, the number of minimum-length trees (# trees), consistency indices (CIs—Kluge and Farris 1969) disregarding uninformative characters, and retention indices (RIs—Farris 1989) are shown. Branch support values are above internodes, and bootstrap scores greater than 50% are below internodes. Groups that are found in the simultaneous analysis of the  $\gamma$ -fibrinogen exons + introns are marked by gray dots. The arrow points to one node that is supported by the simultaneous analysis of exons + introns but is not resolved in the separate analyses of exons or introns. For the total  $\gamma$ -fibrinogen data set, the hippo/whale clade was stable to the inclusion of gaps as a fifth character state (branch support = 4, bootstrap = 93). Branch lengths are not proportional to the number of nucleotide substitutions.

Numerous independent studies support a close relationship between Cetacea and Artiodactyla with other extant mammals more distantly related (Slijper 1962; Fitch and Beintema 1990; Gingerich, Smith, and Simons 1990; Novacek 1992; Milinkovitch, Orti, and Meyer 1993; Queralt et al. 1995; Stanhope et al. 1996). Therefore, cladograms were rooted with sequences from Perissodactyla, Carnivora, and Primates.

The stability of clades in minimum-length trees was assessed through branch support estimates (Bremer 1994) and bootstrap scores (Felsenstein 1985). Branch support, the number of additional character changes necessary to collapse an internal branch, was calculated for each node using the “constraints” command in PAUP 3.1.1 with 50 random taxon addition replicates and TBR branch swapping. Each bootstrap analysis included 1,000 replications. Searches were heuristic with simple taxon addition and TBR branch swapping.

## Results

Phylogenetic results for  $\gamma$ -fibrinogen are summarized in figure 3. The  $\gamma$ -fibrinogen topologies are generally congruent with morphological estimates of mammal phylogeny in that Pecora (Giraffidae + Bovidae + Cervidae), Ruminantia (Pecora + Tragulidae), Cetacea, Suina (Suidae + Tayassuidae), Artiodactyla + Cetacea, Carnivora, and Perissodactyla are monophyletic. The odontocete whales, Physeteridae and Delphinoidea,

cluster in two of the three  $\gamma$ -fibrinogen analyses. More controversially, both the introns and exons of  $\gamma$ -fibrinogen support a hippo/whale clade and a ruminant/hippo/whale clade (fig. 3).

Only two nodes are incompatible between the strict consensus tree for the  $\gamma$ -fibrinogen exons and the strict consensus tree for the  $\gamma$ -fibrinogen introns (fig. 3). The  $\gamma$ -fibrinogen cladograms also conform well to the minimum-length topology for all four genes. In the total DNA cladogram, Pecora, Ruminantia, Cetacea, Odontoceti, Cetacea + Hippopotamidae, Cetacea + Hippopotamidae + Ruminantia, Suina, Artiodactyla + Cetacea, Carnivora, and Perissodactyla are again monophyletic (fig. 4).

None of the DNA data sets resolve a monophyletic Artiodactyla. In all analyses, Cetacea is nested two to three nodes within “Artiodactyla.” The cost of artiodactyl monophyly is 6 character steps for cytochrome *b*, 9 for  $\gamma$ -fibrinogen, 15 for the caseins, and 30 for all four genes combined. All data partitions favor a Hippopotamidae/Cetacea sister group (figs. 3 and 4). Support for this clade is extensive in the simultaneous analysis of all four genes (branch support = 15, bootstrap = 99), in the nuclear data set (branch support = 8, bootstrap = 98), and in the  $\gamma$ -fibrinogen data set (branch support = 4, bootstrap = 91). A sister group relationship between Ruminantia and Cetacea + Hippopotamidae is also strongly supported by the nuclear genes (branch support = 13, bootstrap = 99) and the  $\gamma$ -fibrinogen (branch support = 5, bootstrap = 97). According to all of the DNA data sets, ruminating artiodactyls (Pecora, Tragulidae, and Camelidae) are not monophyletic.

The mt gene, cytochrome *b*, is characterized by substantially lower consistency (Kluge and Farris 1969) and retention indices (Farris 1989) relative to the three nuclear genes (fig. 4). This pattern is likely the result of three characteristics of mt cytochrome *b* evolution in mammals: (1) a rapid overall rate of nucleotide substitution, (2) extreme rate heterogeneity at nonsynonymous sites, and (3) a high transition/transversion ratio (Irwin, Kocher, and Wilson 1991; Chikuni et al. 1995). Given the number of substitutions in mt cytochrome *b* on the total DNA evidence tree (1,256 of the 2,894 total changes), this gene contributes limited branch support in comparison to the nuclear data. For nodes found in the total DNA topology, the sum of branch support values for cytochrome *b* is 47. The sum of branch support for the three nuclear genes is 251 (fig. 4).

## Discussion

To date, portions of four genes have been sequenced for the Hippopotamidae. In sum, this DNA evidence overwhelmingly supports a close phylogenetic relationship between Hippopotamidae and Cetacea (fig. 4). The total of 2,779 nucleotide positions includes mt protein coding sequences (cytochrome *b*), exons from three nuclear genes ( $\gamma$ -fibrinogen,  $\beta$ -casein, and  $\kappa$ -casein), and nuclear introns ( $\gamma$ -fibrinogen).

The evolutionary dynamics of these DNA segments vary widely. The mt cytochrome *b* gene is characterized



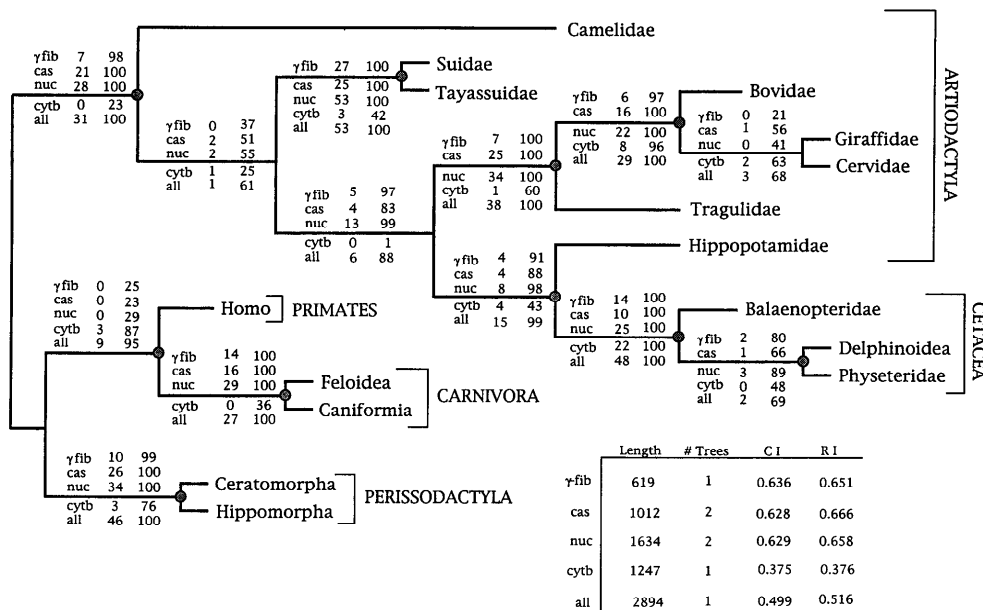


FIG. 4.—A combined DNA cladogram based on four genes: mt cytochrome *b* (1,140 nucleotide positions), the linked nuclear milk caseins ( $\beta$ -casein exon 7 [499 positions] and  $\kappa$ -casein exon 4 [489 positions]), and  $\gamma$ -fibrinogen (exons 2–4 and introns 2–3 [651 positions]). Branch support values followed by bootstrap scores are shown at internodes for  $\gamma$ -fibrinogen (g fib), the caseins (cas), the combined nuclear DNA data (nuc), mt cytochrome *b* (cytb), and the total DNA data set (all). Tree lengths, the number of minimum length trees (# trees), consistency indices disregarding uninformative characters (CI), and retention indices (RI) are shown. The total DNA topology is not altered when third-position transitions of cytochrome *b* are included. When gaps are scored as a fifth character state, the same topology applies except that Camelidae and Suidae + Tayassuidae are sister taxa. Nodes that are stable to the exclusion of all transition substitutions are marked by gray dots. Branch lengths are not proportional to the number of character changes.

by no indels, a high transition-to-transversion ratio, few amino acid replacements, and a rapid rate of synonymous nucleotide substitution (Irwin, Kocher, and Wilson 1991). The nuclear milk caseins are extreme among known protein genes in their rate of amino acid replacement (Wolfe and Sharpe 1993), are permissive to indels that are multiples of three bases, and have a low transition/transversion ratio (Cronin et al. 1996; Gatesy et al. 1996). The nuclear  $\gamma$ -fibrinogen exons are characterized by no apparent indels, are more evolutionarily conservative than the caseins at the amino acid level, and have a much slower overall rate of nucleotide substitution relative to mt cytochrome *b*. The alignment of  $\gamma$ -fibrinogen introns shows indels of various lengths and a mixture of hypervariable and conserved regions (fig. 2). If the Hippopotamidae/Cetacea grouping is spurious, the molecular convergence between these taxa is remarkable in its consistency across a diversity of genic regions. A simpler explanation for the similarities at the DNA level between hippos and whales is common ancestry.

This interpretation is complicated by fossil evidence (fig. 1B). Van Valen (1966) was the first to posit a phylogenetic link between the extinct mesonychia ungulates and Cetacea. He dismissed the possibility of a close relationship between Cetacea and Artiodactyla and argued that "it is . . . improbable that any strongly herbivorous taxon was ancestral to the highly predatory archaeocetes (early whales)." Both primitive cetaceans and mesonychia are usually considered to be carnivorous, a rarity among hoofed mammals (Van Valen 1966; Szalay 1969).

There are striking resemblances between the teeth of primitive cetaceans and those of mesonychia ungulates. The similarities are so complete that isolated teeth from early whales have been misidentified as mesonychia teeth. Thewissen (1994) showed that the following dental characters group Cetacea with mesonychia to the exclusion of artiodactyls and other hoofed mammals: upper premolar four protocone absent, upper molar trigon basin reduced, lower molar talonid basin lost, and lower third molar hypoconulid lost. These reductions in tooth complexity are thought to be functionally linked to a decrease in mediolateral grinding movements of the jaws and a transition to reliance on adduction as the principle jaw movement (Thewissen 1994).

In addition to the Cetacea/Mesonychia association, gross anatomical comparisons of fossils overwhelmingly favor a monophyletic Artiodactyla (fig. 1B). Prothero (1993) noted "a wide array of unique and bizarre morphological specializations from every part of the anatomy" as evidence that artiodactyls form a monophyletic group. A trochleated distal astragalus (Schaeffer 1948), a partial double mesoclyx in distal deciduous premolars (Gentry and Hooker 1988), an enlarged facial portion of the lacrimal, an expanded orbitosphenoid that separates the frontal from the alisphenoid, and narrow lower molar trigonids (Prothero, Manning, and Fischer 1988; Prothero 1993) have been cited as synapomorphies for Artiodactyla.

DNA evidence has no direct bearing on the phylogenetic placement of the wholly extinct mesonychia. However, the hypothesis of artiodactyl monophyly is

open to scrutiny from a molecular perspective. Numerous molecular data sets favor artiodactyl paraphyly, with Cetacea resolved as an artiodactyl subclade (Goodman, Czelusniak, and Beeber 1985; Irwin, Kocher, and Wilson 1991; Graur and Higgins 1994; Irwin and Arnason 1994; Honeycutt et al. 1995; Gatesy et al. 1996; Smith et al. 1996).

Likewise, artiodactyl monophyly is not supported by any of the DNA data sets analyzed here (figs. 3 and 4), and the cost of a monophyletic Artiodactyla is substantial in the combined analysis of all four genes. Thirty unambiguous artiodactyl skeletal "synapomorphies" would have to be added to the combined DNA data set to force the removal of Cetacea from within Artiodactyla. This inference assumes that a single nucleotide substitution carries as much weight in phylogenetic analysis as the evolution of a stable morphological feature such as the double-pulleyed astragalus of artiodactyls. I suspect this assumption is not reasonable to many paleontologists. However, at the least, the combined DNA analysis indicates the need for paleontologists to quantify all of the fossil evidence in an explicit character matrix (e.g., Theodor 1996). Until the morphological and molecular characters can be scrutinized simultaneously using widely accepted criteria for homology (Patterson 1982; De Pinna 1991), it is impossible to determine whether artiodactyl paraphyly is a "grossly unparsimonious" (Prothero 1993) hypothesis.

From the paleontological perspective, aquatic specializations of cetaceans and hippopotamids are interpreted as evolutionary convergences (fig. 1B). The DNA evidence presented here brings this view into question (fig. 1A). The following are potential synapomorphies of whales plus hippos. Most of these traits are difficult to assess in extinct taxa.

- Hippos spend a significant part of their lives in freshwater, and two of the earliest whales, *Pakicetus* and *Nalacetus*, were also apparently restricted to freshwater environments (Thewissen et al. 1996).
- Hippopotamus amphibius* and extant cetaceans both nurse their offspring underwater. This is a rare behavior among mammals (Slijper 1962, p. 381). However, to my knowledge there is no record of this behavior in *Choeropsis liberiensis* (the pygmy hippo). Field observations of *Choeropsis* are lacking, given its secretive nature.
- Hippos and whales are nearly hairless. *H. amphibius* has approximately 25 short, fine hairs per 100 cm<sup>2</sup> of skin on its back and an even sparser distribution of hair on the flanks and belly (Luck and Wright 1964). Cetaceans are almost totally hairless (Ling 1974).
- Both taxa lack sebaceous glands (Luck and Wright 1964; Ling 1974).
- The ability to communicate underwater is shared by hippos and whales (Popper 1980; Ketten 1991; Barklow 1995), but any detailed similarities between these taxa in underwater sound production or hearing are not clear as yet.
- Hippos and whales lack true scrotal testes. The testes are inguinal in hippopotamid artiodactyls (Chapman 1881; Erken, Klaver, and Frankenhuis 1994) and intraabdominal in cetaceans (Slijper 1962, p. 349; De Smet 1977). Most extant artiodactyls have true scrotal testes (Wislocki 1933). If the condition in hippopotamids is interpreted as the intermediate state, the relative position of the testes supports Cetacea + Hippopotamidae.

Given the strong evidence for a Cetacea/Hippopotamidae clade from noncoding, protein-coding, nuclear, and mt DNA, it is more difficult to argue that the common aquatic traits of these taxa are the results of convergent evolution. However, a clear conflict between DNA sequences and fossils remains. Future studies that combine all of the systematic evidence, fossils, DNA sequences, amino acid sequences, behavioral traits, and characteristics of "soft" tissues may be required to sort out this incongruence.

### Acknowledgments

G. Amato, H. Rosenbaum, M. Cronin, P. Vrana, P. Arctander, and E. Stephens donated tissue and DNA samples. A. de Queiroz, M. Hedin, M. Milinkovitch, C. Hayashi, and an anonymous reviewer commented on various stages of the manuscript. The staff of the University of Arizona automated sequencing facility significantly aided in data collection. Funding was from an NSF RTG Post Doctoral Fellowship and NSF Systematics Panel Grant #DEB-9509551.

### LITERATURE CITED

- ARNASON, U., and A. GULLBERG. 1996. Cytochrome *b* nucleotide sequences and the identification of five primary lineages of extant cetaceans. *Mol. Biol. Evol.* **13**:407–417.
- BARKLOW, W. 1995. Hippo talk. *Nat. Hist.* **104**:54.
- BREMER, K. 1994. Branch support and tree stability. *Cladistics* **10**:295–304.
- CHAPMAN, H. 1881. Observations upon the hippopotamus. *Proc. Acad. Nat. Sci. Philadelphia* **1881**:126–148.
- CHIKUNI, K., Y. MORI, T. TABATA, M. SAITO, M. MONMA, and M. KOSUGIYAMA. 1995. Molecular phylogeny based on the  $\kappa$ -casein and cytochrome *b* sequences in the mammalian suborder Ruminantia. *J. Mol. Evol.* **41**:859–866.
- CRONIN, M., R. STUART, B. PIERSON, and J. PATTON. 1996.  $\kappa$ -casein gene phylogeny of higher ruminants (Pecora, Artiodactyla). *Mol. Phylogenet. Evol.* **6**:295–311.
- DE PINNA, M. 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics* **7**:367–394.
- DE SMET, W. 1977. The position of the testes in cetaceans. Pp. 361–386 in R. HARRISON, ed. *Functional anatomy of marine mammals*. Vol. 3. Academic Press, London.
- ERKEN, A., P. KLAVER, and M. FRANKENHUIS. 1994. Castration and sterilization of an adult male *Hippopotamus*. *Verh. Ber. Erkr. Zootiere* **36**:41–43.
- FARRIS, J. 1989. The retention index and the rescaled consistency index. *Cladistics* **5**:417–419.
- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- FITCH, W., and J. BEINTEMA. 1990. Correcting parsimonious trees for unseen nucleotide substitutions: the effect of dense

- branching as exemplified by ribonuclease. *Mol. Biol. Evol.* **7**:438–443.
- GATESY, J., C. HAYASHI, M. CRONIN, and P. ARCTANDER. 1996. Evidence from milk casein genes that cetaceans are close relatives of hippopotamid artiodactyls. *Mol. Biol. Evol.* **13**:954–963.
- GENTRY, A., and J. HOOKER. 1988. The phylogeny of the Artiodactyla. Pp. 235–272 in M. BENTON, ed. *The phylogeny and classification of the tetrapods*, Vol. 2. Mammals. Clarendon Press, Oxford.
- GILBERT, D. 1992. SeqApp. Version 1.9a. Indiana University, Bloomington.
- GINGERICH, P., B. SMITH, and E. SIMONS. 1990. Hind limbs of Eocene *Basilosaurus*: evidence of feet in whales. *Science* **249**:154–157.
- GINGERICH, P., N. WELLS, D. RUSSELL, and S. SHAH. 1983. Origin of whales in epicontinental seas: new evidence from the early Eocene of Pakistan. *Science* **220**:403–406.
- GOODMAN, M., J. CZELUSNIAK, and J. BEEBER. 1985. Phylogeny of Primates and other eutherian orders: a cladistic analysis using amino acid and nucleotide sequence data. *Cladistics* **1**:171–185.
- GRAUR, D., and D. HIGGINS. 1994. Molecular evidence for the inclusion of cetaceans within the order Artiodactyla. *Mol. Biol. Evol.* **11**:357–364.
- HASEGAWA, M., and J. ADACHI. 1996. Phylogenetic position of cetaceans relative to artiodactyls: reanalysis of mitochondrial and nuclear sequences. *Mol. Biol. Evol.* **13**:710–717.
- HONEYCUTT, R., M. NEDBAL, R. ADKINS, and L. JANECEK. 1995. Mammalian mitochondrial DNA evolution: a comparison of the cytochrome *b* and cytochrome *c* oxidase II genes. *J. Mol. Evol.* **40**:260–272.
- IRWIN, D., and U. ARNASON. 1994. Cytochrome *b* gene of marine mammals: phylogeny and evolution. *J. Mamm. Evol.* **2**:37–55.
- IRWIN, D., T. KOCHER, and A. WILSON. 1991. Evolution of the cytochrome *b* gene of mammals. *J. Mol. Evol.* **32**:128–144.
- KETTEN, D. 1991. The marine mammal ear: specializations for aquatic audition and echolocation. Pp. 717–754 in D. WEBSTER, R. FAY, and A. POPPER, eds. *The biology of hearing*. Springer-Verlag, New York, N.Y.
- KLUGE, A., and J. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* **18**:1–32.
- LEDJE, C., and U. ARNASON. 1996. Phylogenetic analyses of complete cytochrome *b* genes of the order Carnivora. *J. Mol. Evol.* **42**:135–144.
- LING, J. 1974. The integument of marine mammals. Pp. 1–44 in R. HARRISON, ed. *Functional anatomy of marine mammals*. Vol. 2. Academic Press, London.
- LUCK, C., and P. WRIGHT. 1964. Aspects of the anatomy and physiology of the skin of the hippopotamus (*H. amphibius*). *Q. J. Exp. Physiol.* **49**:1–14.
- MILINKOVITCH, M., G. ORTI, and A. MEYER. 1993. Revised phylogeny of whales suggested by mitochondrial ribosomal DNA sequences. *Nature* **361**:346–348.
- . 1995. Novel phylogeny of whales revisited but not revised. *Mol. Biol. Evol.* **12**:518–520.
- NOVACEK, M. 1992. Mammalian phylogeny: shaking the tree. *Nature* **356**:121–125.
- PATTERSON, C. 1982. Morphological characters and homology. Pp. 21–74 in A. JOYSEY and A. FRIDAY, eds. *Problems of phylogenetic reconstruction*. Academic Press, London.
- POPPER, A. 1980. Sound emission and detection by delphinids. Pp. 194–223 in L. Herman, ed. *Cetacean behavior: mechanisms and functions*. John Wiley and Sons, New York, N.Y.
- PROTHERO, D. 1993. Ungulate phylogeny: molecular versus morphological evidence. Pp. 173–181 in F. SZALAY, M. NOVACEK, and M. MCKENNA, eds. *Mammal phylogeny*, Vol. 2. Placentals. Springer Verlag, New York, N.Y.
- PROTHERO, D., E. MANNING, and M. FISCHER. 1988. The phylogeny of the ungulates. Pp. 201–234 in M. BENTON, ed. *The phylogeny and classification of the tetrapods*, Vol. 2. Mammals. Clarendon Press, Oxford.
- QUERALT, R., R. ADROER, R. OLIVA, R. WINKFEIN, J. RETIEF, and G. DIXON. 1995. Evolution of protamine P1 genes in mammals. *J. Mol. Evol.* **40**:601–607.
- RIXON, M., D. CHUNG, and E. DAVIE. 1985. Nucleotide sequence of the gene for the  $\gamma$  chain of human fibrinogen. *Biochemistry* **24**:2077–2086.
- SCHAEFFER, B. 1948. The origin of a mammalian ordinal character. *Evolution* **2**:164–175.
- SLIJPER, E. 1962. Whales. Hutchinson, London.
- SMITH, M., M. SHIVJI, V. WADDELL, and M. STANHOPE. 1996. Phylogenetic evidence from the IRBP gene for the paraphyly of toothed whales, with mixed support for Cetacea as a suborder of Artiodactyla. *Mol. Biol. Evol.* **13**:918–922.
- STANHOPE, M., M. SMITH, V. WADDELL, C. PORTER, M. SHIVJI, and M. GOODMAN. 1996. Mammalian evolution and the interphotoreceptor retinoid binding protein (IRBP) gene: convincing evidence for several superordinal clades. *J. Mol. Evol.* **43**:83–92.
- SWOFFORD, D. 1993. PAUP: phylogenetic analysis using parsimony. Version 3.1.1. Illinois Natural History Survey, Champaign.
- SZALAY, F. 1969. Origin and evolution of function of the mesonychid condylarth feeding mechanism. *Evolution* **23**:703–720.
- THEODOR, J. 1996. Postcranial evolution in early non-ruminant artiodactyls. *J. Vertebr. Paleontol. Abstr.* **16**:15A.
- THEWISSEN, J. 1994. Phylogenetic aspects of cetacean origins: a morphological perspective. *J. Mamm. Evol.* **2**:157–184.
- THEWISSEN, J., S. HUSSAIN, and M. ARIF. 1994. Fossil evidence for the origin of aquatic locomotion in archaeocete whales. *Science* **263**:210–212.
- THEWISSEN, J., L. ROE, J. O'NEIL, S. HUSSAIN, A. SAHNI, and S. BAJPAI. 1996. Evolution of cetacean osmoregulation. *Nature* **381**:379–380.
- VAN VALEN, L. 1966. Deltatheridia, a new order of mammals. *Am. Mus. Nat. Hist. Bull.* **132**:1–126.
- WHEELER, W., and D. GLADSTEIN. 1994. MALIGN. Version 2.1. American Museum of Natural History, New York, N.Y.
- WISLOCKI, G. 1933. Location of the testes and body temperature in mammals. *Q. Rev. Biol.* **8**:385–396.
- WOLFE, H., and P. SHARPE. 1993. Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. *J. Mol. Evol.* **37**:441–456.
- ZHOU, X., R. ZHAI, P. GINGERICH, and L. CHEN. 1995. Skull of a new mesonychid (Mammalia, Mesonychia) from the late Paleocene of China. *J. Vert. Paleo.* **15**:387–400.

JEFFREY R. POWELL, reviewing editor

Accepted February 3, 1997