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, Husseing 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; Ilaegawa 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 Physeteridae), 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.

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.

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

![Diagram](https://example.com/diagram.png)

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.
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 γ-fibrinogen data set, sections of κ-casein exon 4 and β-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 κ-casein and
β-casein (Gatesy et al. 1996). Only one additional gap was necessary to accommodate the new sequences.

**Taxa were:** (g = γ-fibrinogen, k = κ-casein, b = β-casein, c = cytochrome b) New sequences for γ-fibrinogen, κ-casein and β-casein are marked with #. Other sequences are from GenBank. (Cervidae = g# + k + b + c-Cervus nippon, c-Docodoea hemionus, Bovidae = g#-Ovis dalli, k+b+c-Ovis arctoy; Giraffidae = g# + k + b + c-Giraffa camelopardalis; Tragulidae = g# + b + c-Tragulus napu, k-Tragulus javanicus; Delphinidae; Phystedidae = g# + k + b + c-Physted catadon; Balaenopteridae = g# + k + b + c-Balaenoptera physalus; Hippopotamidae = g#-Choeropsis liberiensis, k+b+c-Hippopotamus amphibius; Suidae = g# + k + b + c-Sus scrofa; Gattayusidae = g# + k + b + c-Tayassu tajacu; Canididae = g# + b + c-Canis dromedarius, k-Lama guanicoe; Hipposmopha = g#-Equus przewalskii, k+b+c-Equus grevyi; Ceratomorpha = g# + k + b + c-Tapirus indicus, c-Dicerorhinus bicornis; Felidae = g#-Crocuta crocuta, b + c-Panthera uncia, c-Panthera leo; Caniformia = g#-Canis latrans, k+b+c-Canis lupus fulgenus; Primates = g# + k + b + c-Homo sapiens under accession numbers U86643-U86661.

**Phylogenetic Analysis**

The γ-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 Armano 1994; Arnason and Gullberg 1996; Ledje and Arnason 1996). The following subsets of data were analyzed: γ-fibrinogen exons, γ-fibrinogen introns, γ-fibrinogen introns + exons, mt cytochrome b, the linked β + κ-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.
Results

Phylogenetic results for γ-fibrinogen are summarized in figure 3. The γ-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 Delphinoida, cluster in two of the three γ-fibrinogen analyses. More controversially, both the introns and exons of γ-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 γ-fibrinogen exons and the strict consensus tree for the γ-fibrinogen introns (fig. 3). The γ-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 γ-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 γ-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 γ-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 nonhomonymous 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 (γ-fibrinogen, β-casein, and κ-casein), and nuclear introns (γ-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 (nut), 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.

There are striking resemblances between the teeth of primitive cetaceans and those of mesonychian ungulates. The similarities are so complete that isolated teeth from early whales have been misidentified as mesonychian teeth. Thewissen (1994) showed that the following dental characters group Cetacea with mesonychians 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 mesocylix 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 mesonychians. 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.

1. Hippos spend a significant part of their lives in freshwater, and two of the earliest whales, *Pakicetus* and *Naralacetus*, were also apparently restricted to freshwater environments (Thewissen et al. 1996).
2. *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.
3. Hippos and whales are nearly hairless. *H. amphibius* has approximately 25 short, fine hairs per 100 cm² of skin on its back and an even sparser distribution of hair on the flank and belly (Luck and Wright 1964). Cetaceans are almost totally hairless (Ling 1974).
4. Both taxa lack sebaceous glands (Luck and Wright 1964; Ling 1974).
5. 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.
6. Hippos and whales lack true scrotal testes. The testes are inguinal in hippopotamid artiodactyls (Chapman 1881; Erken, Klaver, and Frankenhuys 1994) and infraabdominal 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. Arcander, and E. Stephens donated tissue and DNA samples. A. de Queiroz, M. Hedin, M. Milinkovitch, Ch. 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

FITCH, W., and J. BEINTEMA. 1990. Correcting parsimonious trees for unseen nucleotide substitutions: the effect of dense


