Evidence from Milk Casein Genes that Cetaceans are Close Relatives of Hippopotamid Artiodactyls

John Gatesy, Cheryl Hayashi,† Mathew A. Cronin,‡ and Peter Arctander§*

*Laboratory of Molecular Systematics and Evolution, Department of Ecology and Evolutionary Biology, University of Arizona; †Department of Biology, Yale University, and Department of Invertebrates, American Museum of Natural History; ‡LGL Ecological Genetics, Bryan, Texas; and §Department of Population Biology, University of Copenhagen

The inferred transition from terrestrial hoofed mammal to fully aquatic cetacean has been intensively studied with fossil evidence. However, large sections of this remarkable evolutionary sequence are missing. Phylogenetic analysis of extant taxa may help to fill in some of these gaps. In this report, κ-casein (exon 4) and β-casein (exon 7) milk protein genes from cetaceans and other placental mammals were PCR-amplified, sequenced, and aligned to previously published sequences. Phylogenetic analyses of the casein data suggest that hippopotamid artiodactyls are more closely related to cetaceans than to other artiodactyls (even-toed hoofed mammals). An analysis of the nuclear casein sequences combined with published mitochondrial cytochrome *b* DNA sequences also supports the Cetacea/Hippopotamidae sister group. This affinity implies that some of the aquatic traits of cetaceans were derived in the common ancestor of Cetacea and Hippopotamidae. An extant "missing link" to Cetacea may have been overlooked by science since the description of the semiaquatic *Hippopotamus* in 1758. Paleontological information is grossly inconsistent with this hypothesis. If the casein phylogeny is accurate, large gaps in the fossil record as well as extensive morphological reversals and convergences must be acknowledged.

Introduction

Extant cetaceans (toothed whales and baleen whales) are highly transformed relative to all other extant mammals. Cetaceans are characterized by a nearly hairless body, reduced hind limbs, retracted nasal bones, caudal flukes, and a wholly aquatic lifestyle (Slijper 1962, pp. 58-92). Predictably, the wide gap in morphology between cetaceans and other mammals is narrowed when extinct taxa are considered. Recently described Eocene whale material reveals functional and anatomical intermediates between obligately aquatic cetaceans and their terrestrial ungulate relatives (Gingerich et al. 1983, 1994; Gingerich, Smith, and Simons 1990; Berta 1994; Thewissen, Hussain, and Arif 1994). These extinct taxa have obvious implications for cetacean origins. Less obvious is the notion that anatomical comparisons between whales and their closest extant relatives should also provide insights into the ancestral whale morphotype. Such clues are impossible to derive from fossils that do not preserve skin, reproductive organs, viscera, and behavior.

Unfortunately, the phylogenetic history of Cetacea is unresolved, making inferences of evolutionary transformation problematic. Several molecular analyses favor a recent common ancestor for Artiodactyla (even-toed ungulates) plus Cetacea (Boyden and Gemeroy 1950; Fitch and Beintema 1990; Milinkovitch, Orti, and Meyer 1993; Queralt et al. 1995). Selected anatomical characters, such as a distinctive mechanism of penile erection, three primary bronchi of the lung, and a paraxonic arrangement of the metatarsals (Slijper 1962, pp. 136, 35 1, 352; Gingerich, Smith, and Simons 1990; Thewissen 1994), are consistent with this hypothesis. Other molecular data sets support the derivation of Cetacea from within Artiodactyla, thus rendering the latter group paraphyletic. However, in these studies there is no consensus on the artiodactyl lineage that is closest to the ancestry of Cetacea (Goodman, Czelusniak, and Beeber 1985; Irwin, Kocher, and Wilson 1991; Graur and Higgins 1994; Irwin and Arnason 1994; Honeycutt et al. 1995). From a morphological perspective, the interpretation of cetaceans as derived artiodactyls is considered outlandish and has been strongly criticized. Some anatomical data sets suggest that artiodactyls are not among the closest extant relatives of Cetacea (Prothero, Manning, and Fischer 1988; Novacek 1989; Thewissen 1994).

Because of the apparent rapidity of the ungulate radiation in the early Cenozoic and the mode of evolution in the mitochondrial (mt) genes that have been sampled thus far, Philippe and Douzery (1994) argued that "close to 600,000 nucleotides, i.e, 20,000 nucleotides for thirty species, must be sequenced before molecular phylogenies provide a clear resolution about Artiodactyla monophyly." This estimate is greater than the entire mt genome in these organisms. As an alternative to mtDNA, we sampled nuclear casein genes whose pattern of nucleotide substitution may be more appropriate for this difficult systematic problem.

Caseins are nutritional milk proteins that are unique to mammals. In the milk of the domestic cow, κ-casein and three calcium-sensitive caseins (α_{s1}, α_{s2}, and β) form stable aggregates called milk micelles. These micelles increase the solubility of calcium phosphate in milk and permit the efficient transfer of nutrients from mother to offspring (Mercier, Vilotte, and Provot 1990). The casein genes form a tight linkage group in the domestic cow (Threadgill and Womack 1990). However, while the calcium-sensitive caseins compose a small gene family (Hobbs and Rosen 1982), κ-casein is thought to be more closely related to the blood protein,

Abbreviations: mt, mitochondrial; cytb, cytochrome *b*.

Key words: Cetacea, Artiodactyla, ***Hippopotamus,*** milk protein, casein.

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. 13(7):954–963. 1996

y-fibrinogen (Jolles, Loucheux-Lefebvre, and Henschen 1978).

Caseins evolve exceptionally rapidly at the amino acid level (e.g., Wolfe and Sharpe 1993). This pattern is assumed to result from the relaxed evolutionary constraints on nutritional milk proteins that lack enzymatic function. Our alignments of casein DNA sequences from closely related species show that the rates of synonymous (silent) and nonsynonymous (replacement) nucleotide substitution are not significantly different, and that changes are evenly distributed over the three codon positions (unpublished data). Numerous sites are free to evolve, yet the average rate of substitution is much slower than that in commonly sequenced mt genes (Chikuni et al. 1995; unpublished data).

In this report, κ-casein and p-casein DNA sequences are analyzed cladistically to assess the phylogenetic placement of Cetacea relative to the major extant artiodactyl lineages and to five additional orders of placental mammals.

Materials and Methods

Data Collection

For the caseins, PCR amplification is not trivial. In alignments of published casein cDNA sequences, there are no long, contiguous stretches of conserved nucleotides where nondegenerate PCR primers can be implemented. The exon/intron structure of the caseins further hinders the design of effective PCR primers. The genes are over 90% intron, and the small exons are interspersed between long, presumably hypervariable, introns (Alexander et al. 1988; Bonsing et al. 1988; Mercier, Vilotte, and Provot 1990). Fortunately, κ-casein exon 4 and p-casein exon 7 contain the majority of the protein coding sequences and were targeted for PCR. The following primers are far from universal in mammals. For κ-casein, primers are (5' to 3', numbers represent approximate positions in the *Bos taurus* κ-casein gene sequence-Alexander et al. 1988): κL10595-GTG CTGAGYAGGTATCCTAG and κR11450-GTAGAG TGCAACAACACTGG from Pinder et al. (1991), KL10528-GCYRTGAGARWGATRAAAGA, KL10563-AAANHDYCAARTATATCCCA, κL10596-TGCTGA RTAKSTHTCCTAGT, κR11018-TGYRTTGTSYTCT-TYGAT, κR11022-TTGTSYTCTTYGATRTYTCCTT, and κR11024-GTCTTCTTTGATGTCTCCT. For β-casein, primers are (5' to 3', numbers represent approximate positions in the *Bos taurus* p-casein gene sequence-Bonsing et al. 1988): βL6307-AGRAKRA-ACDCCAGRATAAA, βL64 11 -TVTYCCACARAA-CATCC, βL6436-RCCTNTYCYTCAKCCTGAA, βR6677-AGGCAKRAMTTTGGVCTGAG, βR6750-AGGMTCCTGGTASAGCAGAA.

PCR was by standard procedures using annealing temperatures ranging from 47 to 56°C. In most cases, various primer combinations and annealing temperatures had to be tested before there was positive amplification. PCR products were then cloned and manually sequenced according to previously reported protocols (Wray, Lee, and DeSalle 1993). Some sequences were determined using an ABI automated sequencer. Two to five clones were sequenced for each casein fragment. An exception is the peccary p-casein sequence, for which we could sequence only one homologous clone after multiple PCRs, primer combinations, ligations, and sequencing reactions. In this case, primers βL6436 and βR6677 were utilized. Other p-casein sequences range in length from 354 to 436 nucleotides. The κ-casein sequences range from 378 to 462 nucleotides.

Taxa that were sequenced for this study and those from GenBank are (κ = κ-casein, β = p-casein, * = Genbank sequences, # = sequenced for this study): goat κ+β= *Capra hircus**, sheep κ+β = *Ovis aries**, cow κ+β = *Bos taurus**, deer κ = *Cervus nippon**, deer β = *Alces alces#*, giraffe κ+β= *Giraffa camelopardalis#*, pronghorn κ+β= *Antilocapra americana#*, chevrotain κ = *Tragulus javanicus**, chevrotain β = *Tragulus napu#*, hippo β+κ= *Hippopotamus amphibius#,* toothed whale κ = Delphinidae sp.#, toothed whale β = *Delphinapterus leucas#*, baleen whale κ+β = *Balaenoptera physalus#*, camel κ = *Lama guanicoe#*, camel β= *Camelus dromedarius#*, pig κ+β= Sus *scrofa**, peccary κ + β= *Tayassu tajacu#*, tapir κ + β = *Tapirus indicus#*, zebra κ+β= *Equus grevyi#*, leopard κ+β = *Panthera uncia#*, human κ+β= *Homo sapiens**, rabbit κ+β = *Oryctolagus cuniculus**, guinea pig κ = *Cavia cutleri**, mouse κ+β = *Mus musculus**, rat κ+β = *Rattus norvegicus**. All new sequences were deposited into GenBank under accession numbers U53885–U53906.

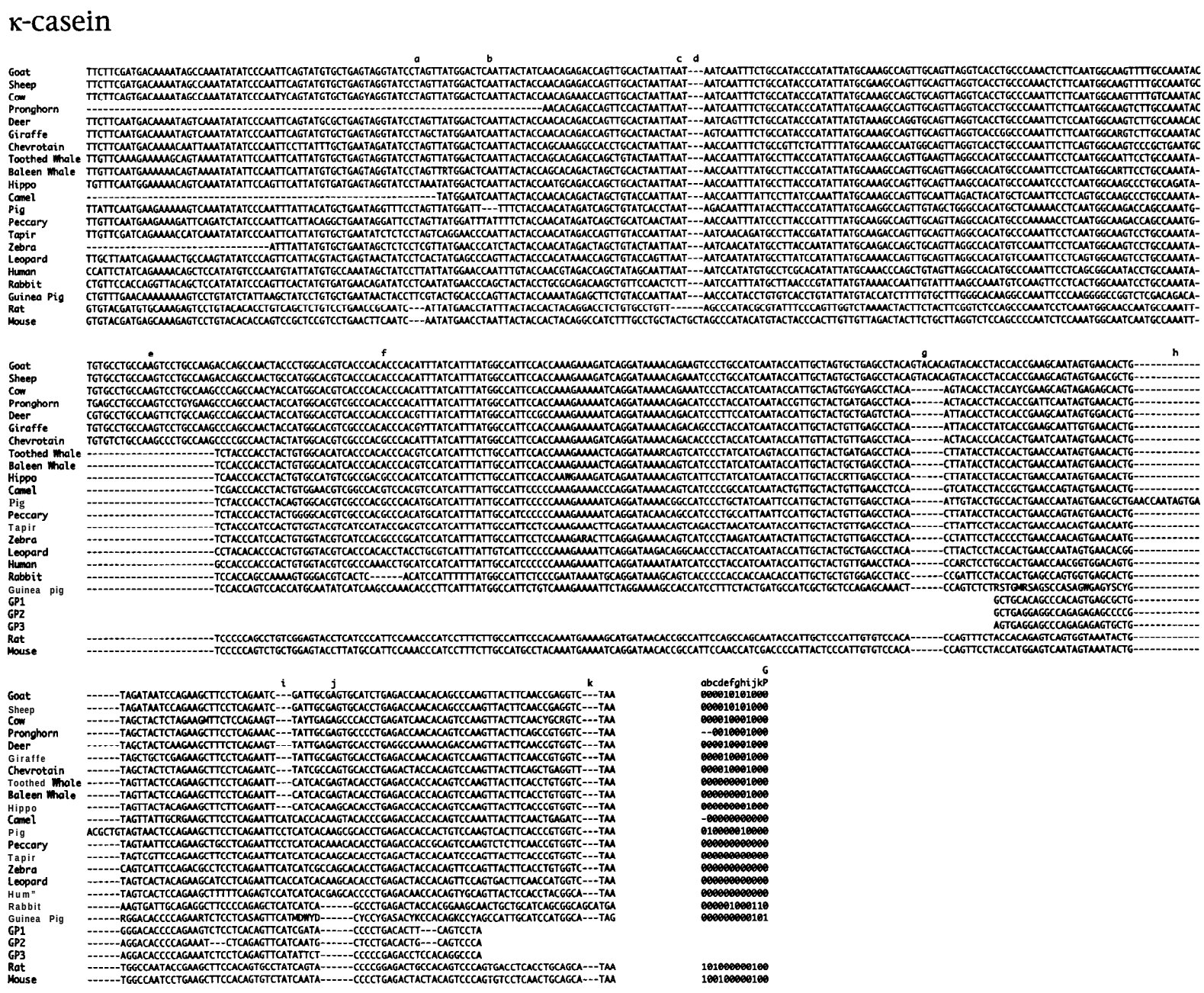
Sequence Alignment/Character Coding

Differences between clone sequences (polymorphism plus PCR and cloning error) as well as polymorphism in published sequences were coded as IUPAC ambiguities. Similarly, the three long direct repeat copies at the 3' end of the *Cavia* κ-casein cDNA (Hall 1990) were aligned to each other, and sites with differences were assigned appropriate ambiguity codes (fig. 1).

The casein sequences were algorithmically aligned with the parsimony-based alignment program MALIGN (Wheeler and Gladstein 1994) using a range of gap cost parameters (gap cost from two to five, nucleotide substitution cost equal to one, and the extragaps option set to one less than the gap cost). These computer-generated alignments were then adjusted so that all gaps in the final alignment were multiples of three and therefore did not disrupt the reading frame. The final alignments are shown in figure 1.

Gaps of all lengths were coded as single characters, except when gaps of different lengths overlapped in the alignment. In these cases, contiguous gaps were divided into separate characters. For example, given a taxon A with a six-base gap, a taxon B with a three-base gap that overlaps the last three positions of the gap in taxon A, and a taxon C with no gaps, the first three gap positions would be coded as one character that is present in A and absent in B and C. The last three gap positions correspond to a second character that is present in A and B and absent in C. Codes for gap characters are shown following the final alignments in figure 1.

κ-casein


a b c d
Goat
Sheep
Cow
Pronghorn
Deer
Giraffe
Chevrotain
Toothed Whale
Baleen Whale
Hippo
Camel
Pig
Peccary
Tapir
Zebra
Leopard
Human
Rabbit
Guinea Pig
Rat
Mouse
e f g h
Goat
Sheep
Cow
Pronghorn
Deer
Giraffe
Chevrotain
Toothed Whale
Baleen Whale
Hippo
Camel
Pig
Peccary
Tapir
Zebra
Leopard
Human
Rabbit
Guinea pig
GP1
GP2
GP3
Rat
Mouse
i j k G
abcdefghijkP
Goat 000010101000
Sheep 000010101000
Cow 000010001000
Pronghorn --0010001000
Deer 000010001000
Giraffe 000010001000
Chevrotain 000010001000
Toothed Whale 000000001000
Baleen Whale 000000001000
Hippo 000000001000
Camel -000000000000
Pig 010000010000
Peccary 000000000000
Tapir 000000000000
Zebra 000000000000
Leopard 000000000000
Hum" 000000000000
Rabbit 000001000110
Guinea Pig 000000000101
GP1
GP2
GP3
Rat 101000000100
Mouse 100100000100

β-casein

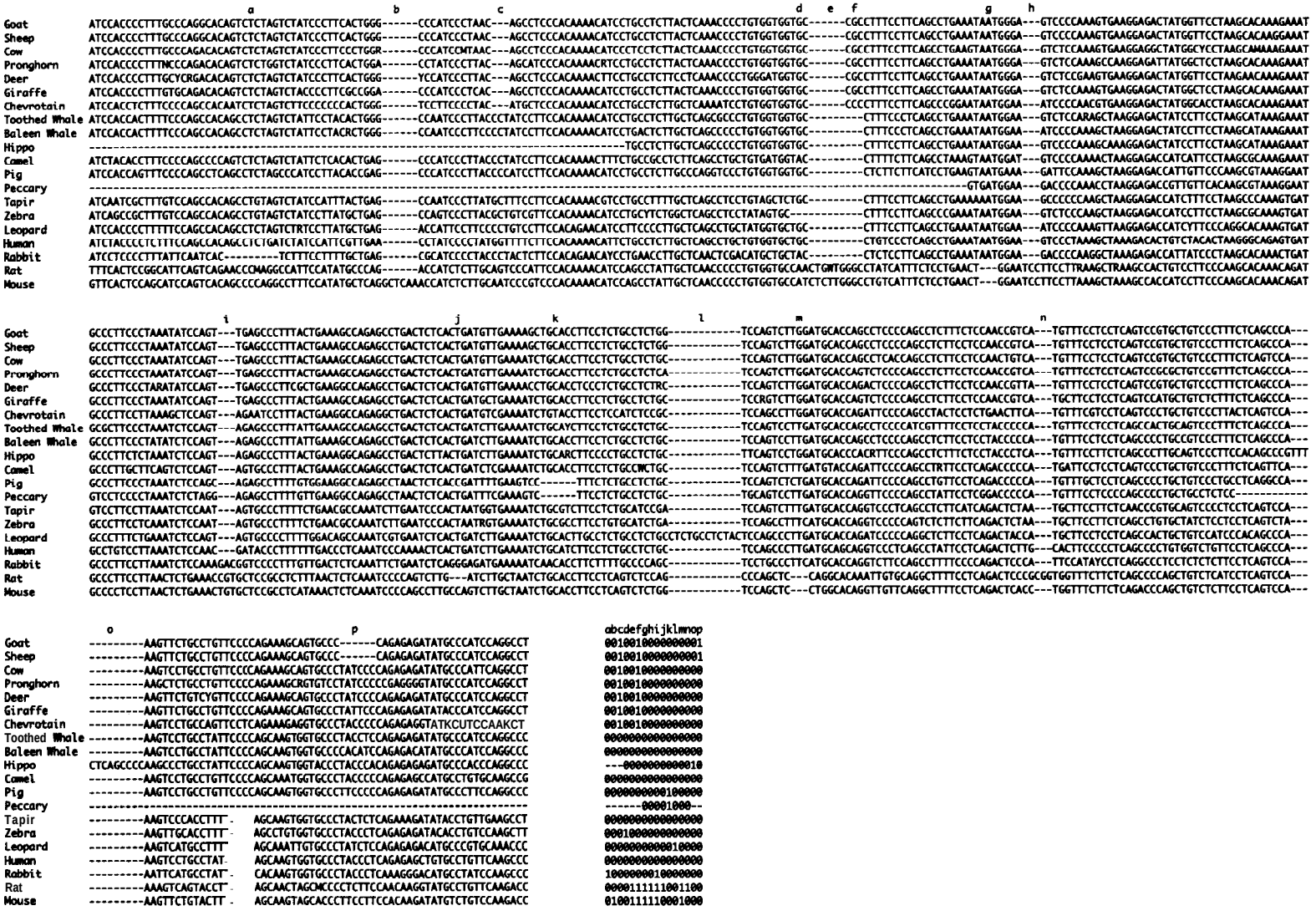


FIG. *I.-The* final alignments of κ-casein (exon 4) and β-casein (exon 7) sequences. The three direct repeat copies at the 3' **end of** the ***Cavia*** κ-casein cDNA (Hall 1990) are aligned below the ambiguity coding for these repeats. Gap characters are marked by lowercase letters above the alignments, and codes for gap characters (0, I) follow each alignment. Dashes indicate where gaps were introduced into the alignments. Dashes at the ends of sequences and among the gap characters are missing data. The gap character labeled GP accounts for the direct repeats in *Cavia* κ-casein that are minimally the result of one evolutionary event. Taxa are as described in Materials and Methods. Nucleotide positions correspond to 10550-I 1012 in the *Bos taurus* κ-casein sequence of Alexander et al. (1988) and 63266750 in the *Bos taurus* β-casein sequence of Bonsing et al. (1988).

Phylogenetic Analysis

The individual casein genes were analyzed separately and combined using PAUP 3.1.1 (Swofford 1993) and NONA 1.0 (Goloboff 1993*a*). All characters were equally weighted. Within a parsimony framework, an equal weighting of characters necessarily implies a minimization of the number of character transformations. Such statements show the best fit between initial hypotheses of homology and final estimates of homology (Farris and Kluge 1985).

PAUP heuristic parsimony searches were 1,000 random addition replicates with TBR branch swapping. Alternative branch-swapping routines did not find shorter cladograms. PAUP results were checked using NONA with 1,000 random addition replicates and TBR branch swapping. All cladograms were rooted according to the morphological hypothesis of McKenna (1975). That is, rodents and lagomorphs were used to root the remainder of the taxa. This rooting does not contradict the morphological consensus of Novacek (1989).

Branch support (Bremer 1994), the number of extra steps required to collapse a particular node, was estimated for each resolved clade using the "constraints" command in PAUP and 50 random addition replicates with TBR branch swapping. Bootstrapping is encouraged by *Molecular Biology and Evolution.* Bootstrap scores (Felsenstein 1985) were derived from 1,000 replicates using PAUP with simple taxon addition and TBR branch swapping.

Mt cytochrome *b* (cytb) is the only DNA sequence that has been sampled for all of the taxa in the casein data set (Irwin, Kocher, and Wilson 1991; Ma et al. 1993; Irwin and Arnason 1994). Mt cytb data (1,140 nucleotide positions) were analyzed with the caseins in a combined phylogenetic analysis. All characters were given equal weight. Mt cytb sequences used in this study are: cow (=*Bos taurus*), sheep (=*Ovis aries*), goat (=*Capra hircus*), pronghorn (=*Antilocapra americana),* giraffe (= *Giraffa camelopardalis),* deer (= *Odocoileus hemionus),* chevrotain (= *Tragulus napu),* camel (=*Camelus dromedarius),* pig (=*Sus scrofa),* peccary (= *Tayassu tajacu),* hippo (=*Hippopotamus amphibius),* toothed whale (=*Stenella attenuata),* baleen whale (=*Balaenoptera physalus),* zebra (= *Equus grevyi),* rhino (=*Diceros bicornis),* seal (=*Phoca vitulina),* human (=*Homo sapiens),* rabbit (= *Oryctolagus cuniculus),* mouse (=*Mus domesticus),* and rat (=*Rattus norvegicus)* from GenBank and guinea pig (= *Cavia porcellus)* from Ma et al. (1993).

In analyses where we found more than one minimum-length cladogram, we reduced the number of equal-length topologies by Goloboff weighting with character fit (*k* from 1 to 6) using PEEWEE (Goloboff 19936). This procedure weights individual characters according to their homoplasy. The different *k* values specify the concavity of the weighting curve (Goloboff 1993c).

Results

Independent cladistic analyses of the two casein genes produce broadly congruent results (figs. 2A and B). Cetacea is solidly entrenched within Artiodactyla. No rootings of the minimum-length cladograms are consistent with artiodactyl monophyly. For both genes, the sister group of Cetacea is Hippopotamidae, although support is not overwhelming. Much of the topological disagreement between the two gene trees is within the Pecora (bovids, cervids, giraffids, and antilocaprids), a group that historically has been resistant to consistent hierarchical subdivision (Gentry and Hooker 1988; Kraus and Miyamoto 1991).

When the β- and κ-casein data are analyzed simultaneously, Artiodactyla remains paraphyletic, and the sister group of Cetacea again is Hippopotamidae (fig. 2C). As in the analyses of the individual genes, the *Hippopotamus*/Cetacea group clusters with Ruminantia (Pecora + Tragulidae-chevrotains). Eleven extra evolutionary steps are required to make Artiodactyla monophyletic. When this constraint is applied, Cetacea is the sister group of Artiodactyla. Forty-four additional steps are necessary for the data to conform to some morphological estimates of ungulate phylogeny in which Cetacea is more closely related to Perissodactyla (odd-toed ungulates) than to Artiodactyla (Prothero, Manning, and Fischer 1988; Novacek 1989; Thewissen 1994).

The relative quality of the casein data set is indicated in several measures:

1. There are many previously hypothesized clades in all topologies (figs. 2*A–C*). As these groups were erected on the basis of morphological evidence, the presence of these clades is an indication of the consistency of the casein data with gross anatomical data.
2. In the combined β- plus κ-casein cladogram, branch support values (Bremer 1994) and bootstrap scores (Felsenstein 1985) are high for most components (fig. 2*C*).
3. There is much character congruence between the caseins. One percent of the total character incongruence in the combined casein tree is accounted for by conflict between the κ- and β-casein data sets (Mickevich and Farris 1981). The combined casein cladogram is only five steps longer than the sum of the individual gene cladogram lengths (figs. 2*A–C*).
4. There is also much taxonomic congruence between the two casein gene topologies. Twelve of 17 comparable nodes are consistent between the κ- and β-casein topologies (Fig. 2B).
5. For all codon positions, ensemble consistency indices (Kluge and Farris 1969) and retention indices (Farris 1989) are higher in the caseins than in mt cytb (fig. 3). This is the case whether these statistics are calculated from the combined casein plus mt cytb tree (fig. 2*D*—see below) or from independent phylogenetic analyses of each codon position in isolation from other characters (analyses not shown). Nucleotide substitutions in the caseins are spread relatively evenly among the three codon positions (fig. 3). An even distribution of substitutions across the three codon positions may limit the number of multiple hits at individual sites.

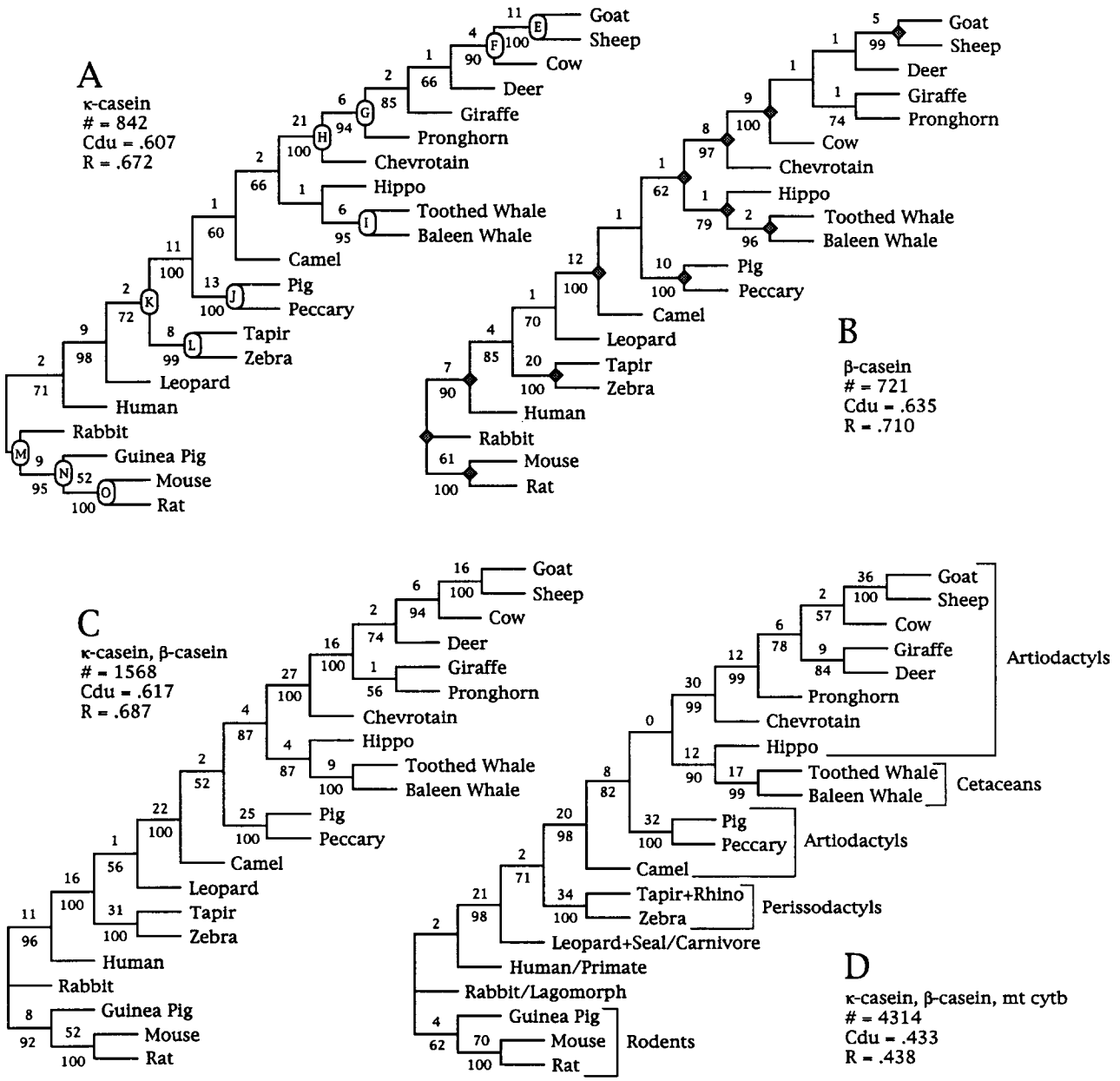


FIG. 2.—Minimum-length cladograms for the (*A*) κ-casein, (*B*) β-casein, and (*C*) combined casein data sets and (*D*) the fittest (Goloboff 1993*c*) of the two minimum-length cladograms for the casein plus mt cytb data. Tree length (#), consistency index disregarding uninformative characters (Cdu), and retention index (R) are indicated for each topology. Branch support values (Bremer 1994) are shown above internal branches, and bootstrap scores (Felsenstein 1985) greater than 50% are below internodes. Previously hypothesized clades that are consistent with some or all of the trees are marked by capital letters at nodes in *A* (E = Caprini, F = Bovidae, G = Pecora, H = Ruminantia, I = Cetacea, J = Suina, K = Ungulata, L = Perissodactyla, M = Glires, N = Rodentia, O = Murinae). Groups that are consistent between the κ- and β-casein cladograms are marked by gray diamonds at nodes in *B*. Ordinal assignments of taxa are shown in *D*. Branch lengths are not proportional to the number of character changes.

6. The bias toward transitions is not extreme in the caseins examined here. The transition/transversion ratio is approximately 1.5:1 in κ-casein and 1.4:1 in β-casein as estimated by parsimony on the combined casein topology (fig. 2*C*) using MacClade 3.03 (Maddison and Maddison 1992). Comparisons of closely related taxa show that the low transition/transversion ratio is not due to the saturation of transitions.

When the casein data were analyzed with the mt cytb data, two equally parsimonious trees were discovered. These trees were reduced to the one in figure 2*D* by Goloboff weighting. This topology is stable for *k* values from 1 to 6. Character conflict between the caseins and mt cytb explains only 1.5% of the total character incongruence in the combined analysis. Thirty-one extra steps are necessary for the mt cytb to conform to the combined topology of figure 2*D*, while only two extra steps are required to fit the casein data to this same tree.

The addition of mt cytb data to the caseins does not change the broad phylogenetic conclusions of the

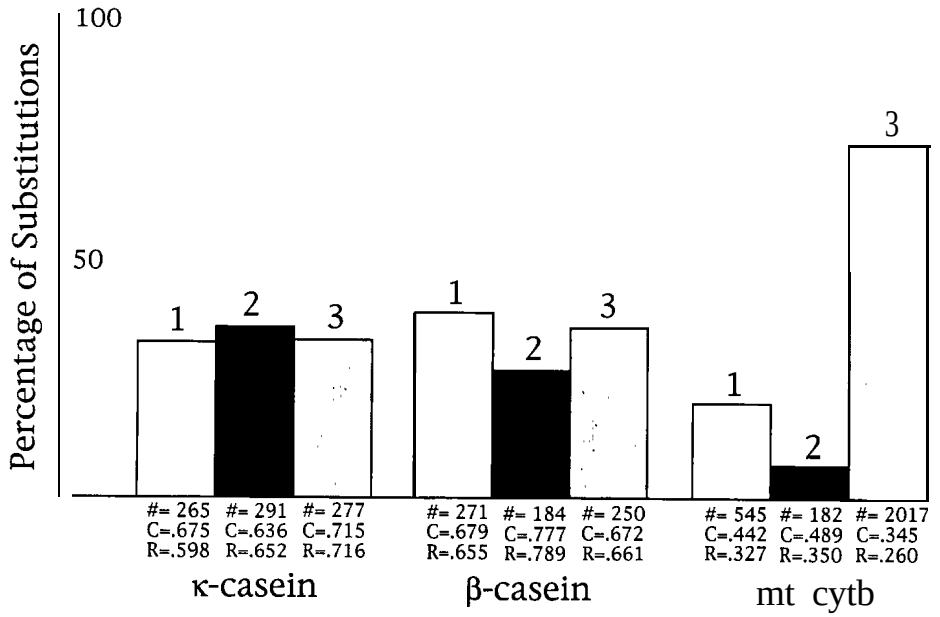


FIG. 3.—The percentage of nucleotide substitutions (vertical bars) at each of the three codon positions in κ-casein, β-casein, and mt cytb for the taxa listed in the methods section. Using PAUP and MacClade, the minimum number of nucleotide substitutions at each position (#) was estimated by parsimony on the fittest topology for the combined casein plus mt cytb data (fig. 20). The average rate of nucleotide substitution is lowest for position 2 of mt cytb. The consistency index (C) and retention index (R) for each codon position is shown. In mt cytb, the excess of substitutions at third codon positions and the scarcity of substitutions at second positions are indicative of strong selective constraints to maintain amino acid sequence (Irwin, Kocher, and Wilson 1991). In contrast, the pattern of substitution in the caseins is more balanced among the three codon positions and suggests that the synonymous and nonsynonymous rates of nucleotide substitution are not as disparate as in mt cytb. For the above analyses, the number of aligned nucleotide positions for each gene was: κ-casein = 495, β-casein = 471, and mt cytb = 1,140.

other analyses (figs. 2*A–C*), but total support, the sum of all branch support values over the tree (Bremer 1994), increases (figs. 2C and D). The following taxonomic groups are consistent with the combined casein/mt cytb topology: Caprini, Bovidae, Pecora, Ruminantia, Cetacea, Suina, Ungulata, Perissodactyla, Glires, Rodentia, and Murinae (fig. 20). The cost of a monophyletic Artiodactyla, i.e., the number of steps required to remove Cetacea from within Artiodactyla, is 19 character changes. In addition to the odd placement of Cetacea, the extreme basal position of Camelidae (camels and llamas) within Artiodactyla is contrary to morphological evidence (Gentry and Hooker 1988).

When the mt cytb data are combined with the κ- and β-casein data, branch support for the hippo/whale clade increases from 4 to 12 (figs. 2C and *D*). This group is not resolved in an equally weighted analysis of the mt cytb data alone (i.e., branch support = 0). On the fittest topology for the combined DNA data (fig. 2*D*), 2 unambiguous character transformations in κ-casein, 5 in β-casein, and 15 in mt cytb support the hippo/whale clade.

Discussion

In all of the casein trees (fig. 2), the tight relationship of Ruminantia, artiodactyls that chew the cud, with Cetacea is supported. The implication is that a domestic goat is more closely related to a giant baleen whale than to a dromedary or a domestic pig. A recent reanalysis of published mtDNA sequences and nuclear amino acid sequences for quartets of species found "statistical support" for this same grouping (Graur and Higgins 1994).

The Cetacea/Hippopotamidae clade (fig. 2), which was first detected in weighted analyses of mt cytb (Irwin and Arnason 1994), is perhaps more intriguing. Because of the numerous extinctions along the stem lineage of cetaceans, most major insights into the transformation from a terrestrial ungulate to a fully aquatic cetacean will come from fossil taxa. Regardless, the cladograms in figure 2 suggest that neontological data may provide some unique information. Extant cetaceans are characterized by a nearly hairless body, the absence of sebaceous glands, a thick layer of insulating fat, offspring that nurse while submerged, a lack of scrotal testes, and a general aquatic habit (Slijper 1962, pp. 58-69, 296-299, 349, 380–382; Nowak and Paradiso 1983, pp. 969-973). Although differing substantially in form, all of the above features are seen in one or both extant hippopotamid species (Crisp 1867; Chapman 1881; Slijper 1962, p. 381; Luck and Wright 1964; Nowak and Paradiso 1983, pp. 1347-135 1; Erken, Klaver, and Frankenhuis 1994). Given a close Hippopotamidae/Cetacea alliance, the principle of parsimony implies that at least some of these characteristics were acquired in the common ancestor of this clade. Thus, many of the aquatic traits that are seen in extant cetaceans may have evolved long before the emergence of the most primitive whale, *Pakicetus*, in the Eocene 49.0-52.5 million years ago (Gingerich et al. 1994).

Both cetaceans and hippopotamids vocalize and apparently communicate underwater (Barklow 1995).

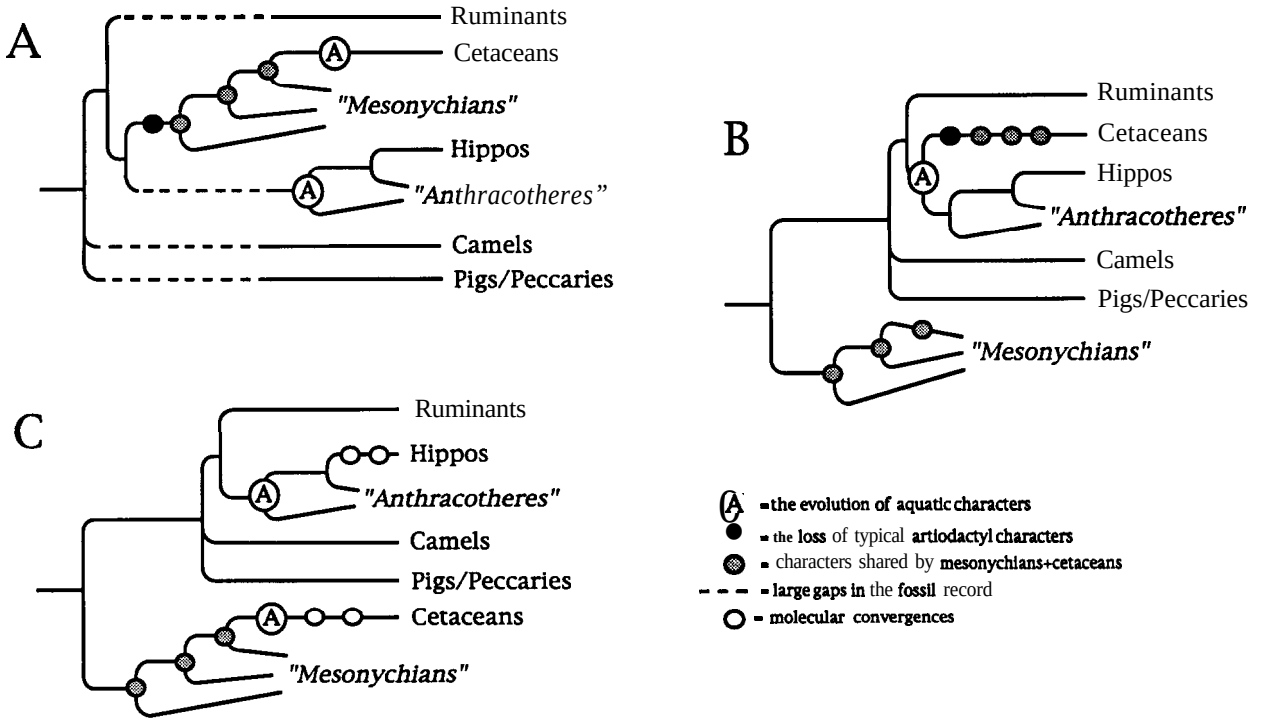


FIG. 4.-Three scenarios of cetacean origins that have different implications for the evolution of aquatic features. *(A)* If cetaceans + mesonychians are derived artiodactyls, aquatic specializations of cetaceans and hippopotamids can be interpreted as convergences. (B) If cetaceans are derived artiodactyls but mesonychians are not, aquatic specializations shared by hippos and cetaceans are potential homologues. (*C*) Alternatively, both the molecular and aquatic similarities of whales and hippos may be convergences. The lack of broken lineages in scenarios *B* and C do not imply an absence of gaps in the fossil record. However, the gaps are much reduced relative to hypothesis *A*. Branch lengths are not proportional to the radiometric time scale.

Barklow (1995) has noted that "the most intriguing of the hippos' underwater sounds are the various types of clicks they make, usually in a series. These 'click trains' bring to mind similar sounds of cetaceans ...." Traditionally, echolocation has been interpreted as a specialization of odontocetes, toothed whales. Some molecular evidence suggests that Odontoceti is paraphyletic and that the ability to echolocate was present in the common ancestor of all extant cetaceans (Milinkovitch, Orti, and Meyer 1993; Milinkovitch, Meyer, and Powell 1994; Milinkovitch 1995). Given our phylogenetic results, it might be profitable to test whether echolocation has an even more general distribution within ungulates. Although there are speculative accounts of echo-ranging in *Hippopotamus* (Longhurst 1966), at present, clear evidence is lacking.

The close relationship of amphibious hippos to whales makes the evolutionary transition from land to sea seem less daunting. However, whether the Hippopotamidae should be considered a functional intermediate to Cetacea critically hinges on the phylogenetic placement of extinct relatives of cetaceans and hippopotamids, the detection of aquatic habits in these extinct taxa, and a reconciliation of the complicated paleontological and molecular evidence.

If Hippopotamidae is the extant sister group of Cetacea, and their common aquatic traits are synapomorphies, the extinct relatives of hippos and whales should also exhibit aquatic specializations. Although alternative hypotheses are actively entertained (Pickford 1983), the two extant hippopotamid species generally are presumed to be relicts of the anthracothere lineage of artiodactyls (Colbert 1935; Gentry and Hooker 1988). Anthracotheres extend back to the Eocene (Ducrocq 1994) and at least some are thought to have been semiaquatic (Pickford 1983). Cetaceans are considered to be most closely related to Mesonychia (as delimited by Thewissen 1994), a Paleocene to late Eocene grade of carnivorous/omnivorous hoofed mammals (Osborn 1924; Szalay and Gould 1966). Most mesonychians are not characterized by obvious aquatic features and some are interpreted as being highly cursorial (Zhou, Sanders, and Gingerich 1992, but see O'Leary and Rose 1995).

If the casein trees are accurate, even the most straightforward scenarios of cetacean origins would require either imposing gaps in the fossil record (fig. 4A) or an excess of morphological convergence (fig. 4*B*).

For example, if cetaceans + mesonychians (the Cete of Thewissen 1994) are derived artiodactyls, aquatic specializations of cetaceans and hippopotamids can be interpreted as convergences (fig. 4A). This scenario entails large gaps in the fossil record. The oldest mesonychians are found in the Paleocene about 60 million years ago while the first artiodactyls appear in the earliest Eocene approximately 54 million years ago (Thewissen 1994).

In contrast, if cetaceans are derived artiodactyls but mesonychians are not (fig. 4*B*), aquatic specializations

shared by hippos and cetaceans are at least potentially homologous. In this interpretation, there has been considerable convergence in the anatomy of early cetaceans and mesonychians (fig. 4*B*). With reference to the morphological character matrix of Thewissen (1994), the cost of removing Cetacea from within Mesonychia and linking Cetacea with Artiodactyla is not trivial (six steps). This cost does not include the number of extra evolutionary transformations that are necessary to imbed Cetacea within Artiodactyla. Both of the above schemes (figs. 4A and *B*) require the loss of numerous typical artiodactyl characters in basal cetaceans or their close relatives (Prothero, Manning, and Fischer 1988).

Alternatively, the casein data may be the problem. Perhaps there has been extensive convergence in the milk proteins of mammals that sometimes (hippos) or always (whales) nurse in the water. One could speculate that the close genetic linkage of κ-casein and β-casein as well as the functional interaction of these two proteins in the milk micelle may account for some of the character correlation in our combined casein trees (figs. 2C and *D*). Furthermore, it could be imagined that nucleotide substitution rate differences between lineages may mislead (Felsenstein 1978).

Regardless, the few DNA data sets that have sampled both Hippopotamidae and Cetacea-to our knowledge, the 2,107 nucleotide positions analyzed here—support a close relationship of these taxa. If the casein trees are accurate, detailed comparisons of extant ruminants and hippopotamids with whales should reveal additional anatomical characters that justify the nesting of cetaceans within the Artiodactyla. Furthermore, new fossil discoveries should include basal cetaceans that have typical artiodactyl skeletal characters. Fossil and other systematic evidence will test whether milk, a defining feature of mammals, offers powerful information for mammalian phylogenetics.

Acknowledgements

We thank R. DeSalle, W. Wheeler, M. Hammer, and the staff of the LMSE at the University of Arizona for laboratory space and patience. I? Vrana, E. Avery Stephens, H. Rosenbaum, and the Wildlife Conservation Society-Bronx, N.Y., provided tissue and DNA samples. G. Gould, M. Milinkovitch, A. Brower, L. Waters. H. Rosenbaum, G. Amato, I? Walsh, R. DeSalle, P. Vrana, K. Pedersen, I? Meinke, S. Gatesy, B. Normark, J. Alroy, A. de Queiroz, and M. Norell commented on various stages of the manuscript. D. Irwin provided alignments of mt cytb sequences. Funding for this research was from an National Science Foundation Research Training Grant Postdoctoral Fellowship (University of Arizona), a University of Arizona Small Grant, and a National Science Foundation Systematics Panel Grant awarded to J. Gatesy. C. Hayashi was funded by an American Museum of Natural History graduate fellowship. The project was further funded by the Danish Research Council's Centre for Tropical Biodiversity.

LITERATURE CITED

ALEXANDER, L., A. STEWART, A. MACKINLAY, T. KAPELINSKAYA, T. TKACH, and S. GORODETSKY. 1988. Isolation and characterization of the bovine κ-casein gene. Eur. J. Biochem. **178**:395–401.

BARKLOW, W. 1995. Hippo talk. Nat. Hist. **104**:54.

BERTA, A. 1994. What is a whale? Science 263: 180-181.

BONSING, J., J. RING, A. STEWART, and A. MACKINLAY. 1988. Complete nucleotide sequence of the bovine β-casein gene. Aust. J. Biol. Sci. **41**:527–537.

BOYDEN, A., and D. GEMEROY. 1950. The relative position of the Cetacea among the orders of Mammalia as indicated by precipitin tests. Zoologica 35: 145-15 1.

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 κ-casein and cytochrome b sequences in the mammalian suborder Ruminantia. J. Mol. Evol. **41**:859–866.

COLBERT, E. 1935. The phylogeny of the Indian Suidae and the origin of the Hippopotamidae. Am. Mus. Novit. 799: 1–24.

CRISP, E. 1867. On some points connected with the anatomy of the hippopotamus *(Hippopotamus amphibius)*. Proc. Zool. Soc. Lond. **39**:601–612.

DUCROCQ, S. 1994. The Paleogene anthracotheres from Thailand: paleogeography and phylogeny. C. R. Acad. Sci. Paris **318**:549–554.

ERKEN, A., P. KLAVER, and M. FRANKENHUIS. 1994. Castration and sterilization of an adult male *Hippopotamus*. Verh. Ber. Erkrg. Zootiere **36**:41–43.

FARRIS, J. 1989. The retention index and the rescaled consistency index. Cladistics **5**:417–419.

FARRIS, J., and A. KLUGE. 1985. Parsimony, synapomorphy, and explanatory power: a reply to Duncan. Taxon 34: 130-135.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27: 401-410.

———. 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.

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, volume 2: mammals. Clarendon Press, Oxford.

GINGERICH, P., S. RAZA, M. ARIF, M. ANWAR, and X. ZHOU. 1994. New whale from the Eocene of Pakistan and the origin of cetacean swimming. Nature **368**:844–847.

GINGERICH, I?, B. SMITH, and E. SIMONS. 1990. Hind limbs of Eocene *Basilosaurus*: evidence of feet in whales. Science 249: 154-157.

GINGERICH, I?, 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.

GOLOBOFF, I? 1993*a*. NONA. Version 1.0. American Museum of Natural History, New York.

———. 1993*b*. PEEWEE. Version 2.0, American Museum of Natural History, New York.

———. 1993*c*. Estimating character weights during tree search. Cladistics **9**:83–91.

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.

HALL, L. 1990. Nucleotide sequence of guinea pig κ-casein cDNA. Nucleic Acids Res. 18:6129.

HOBBS, A., and J. ROSEN. 1982. Sequence of rat α- and γ-casein mRNAs: evolutionary comparison of the calcium-dependent rat casein multigene family. Nucleic Acids Res. 10: 8079-8098.

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. Mammal. 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.

JOLLES, I?, M. LOUCHEUX-LEFEBVRE, and A. HENSCHEN. 1978. Structural relatedness of κ-casein and fibrinogen y-chain. J. Mol. Evol. 11:271–277.

KLUGE, A., and J. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. 18:1–32.

KRAUS, F., and M. MIYAMOTO. 1991. Rapid cladogenesis among the pecoran ruminants: evidence from mitochondrial DNA sequences. Syst. Zool. 40:117–130.

LONGHURST, W. 1966. Observations on apparent echo-ranging by *Hippopotamus* in Uganda. Pp. 210-214 in Proceedings of the third annual conference on biological sonar and diving mammals. Stanford Research Institute, Fremont, Calif.

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.

MA, D., A. ZHARKIKH, D. GRAUR, J. VANDEBERG, and W. LI. 1993. Structure and evolution of opossum, guinea pig, and porcupine cytochrome b genes. J. Mol. Evol. 36:327–334.

MADDISON, W., and D. MADDISON. 1992. MacClade 3.03. Sinauer, Sunderland, Mass.

MCKENNA, M. 1975. Toward a phylogenetic classification of the Mammalia. Pp. 21-46 *in* W. LUCKETT and F. SZALAY, eds. Phylogeny of the Primates. Plenum, New York.

MERCIER, J., J. VILOTTE, and C. PROVOT. 1990. Structure and function of milk protein genes. Pp. 233-258 *in* H. GELDERMAN and F. ELLENDORF, eds. Genome analysis in domestic animals. Weinheim, New York.

MICKEVICH, M., and J. FARRIS. 198 1. The implications of congruence in *Menidia.* Syst. Zool. 30:351–370.

MILINKOVITCH, M. 1995. Molecular phylogeny of cetaceans prompts revision of morphological transformations. Trends Ecol. Evol. 10:328–334.

MILINKOVITCH, M., A. MEYER, and J. POWELL. 1994. Phylogeny of all major groups of cetaceans based on DNA sequences from three mitochondrial genes. Mol. Biol. Evol. 11:939–948.

MILINKOVITCH, M., G. ORTI, and A. MEYER. 1993. Revised phylogeny of whales suggested by mitochondrial ribosomal DNA sequences. Nature 361:346–348.

NOVACEK, M. 1989. Higher mammal phylogeny: the morphological-molecular synthesis. Pp. 42 l-435 *in* B. FERNHOLM, K. BREMER, and H. JORNVALL, eds. The hierarchy of life. Elsevier Science Publishers.

NOWAK, R. and J. PARADISO. 1983. Walker's mammals of the world. The Johns Hopkins University Press, Baltimore.

O'LEARY, M., and K. ROSE. 1995. Postcranial skeleton of the early Eocene mesonychid *Pachyaena* (Mammalia: Mesonychia). J. Vertebr. Paleontol. 15:401–430.

OSBORN, H. 1924. *Andrewsarchus,* giant mesonychid of Mongolia. Am. Mus. Novit. 146:1–5.

PHILIPPE, H., and E. DOUZERY. 1994. The pitfalls of molecular phylogeny based on four species, as illustrated by the Cetacea/Artiodactyla relationships. J. Mamm. Evol. 2: 133–152.

PICKFORD, M. 1983. On the origins of Hippopotamidae together with descriptions of two new species, a new genus, and a new subfamily from the Miocene of Kenya. Geobios 16: 193-217.

PINDER, S., B. PERRY, C. SKIDMORE, and D. SAVVA. 1991. Analysis of polymorphism in the bovine casein genes by use of the polymerase chain reaction. Anim. Genet. 22: 1 1–20.

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.

SLIJPER, E. 1962. Whales. Hutchinson, London.

SWOFFORD, D. 1993. PAUP: phylogenetic analysis using parsimony. Version 3.1.1. Illinois Natural History Survey, Champaign.

SZALAY, F., and S. GOULD. 1966. Asiatic Mesonychidae (Mammalia, Condylarthra). Bull. Am. Mus. Nat. Hist. 132:127-174.

THEWISSEN, J. 1994. Phylogenetic aspects of cetacean origins: a morphological perspective. J. Mammal. 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.

THREADGILL, D., and J. WOMACK. 1990. Genomic analysis of the major bovine milk protein genes. Nucleic Acids Res. 18:6935–6942.

WHEELER, W., and D. GLADSTEIN. 1994. MALIGN. Version 2.1. American Museum of Natural History, New York.

WOLFE, H., and P. SHARPE. 1993. Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. J. Mol. Evol. 37:441–456.

WRAY, C., J. LEE, and R. DESALLE. 1993. Extraction and enzymatic characterization of foraminiferal DNA. Micropaleontology 39:69–73.

ZHOU, X., W. SANDERS, and P. GINGERICH. 1992. Functional and behavioral implications of vertebral structure in *Pachyaena ossifraga* (Mammalia, Mesonychia). Contrib. Mus. Paleontol. Univ. Michigan 28:289–3 19.

JEFFREY R. POWELL, reviewing editor

Accepted May 1, 1996