Syst. Biol. 50(3):367–390, 2001

Mining the Mammalian Genome for Artiodactyl Systematics

CONRAD A. MATTHEE,[1] JAY D. BURZLAFF, JEREMY F. TAYLOR, AND SCOTT K. DAVIS

Department of Animal Science, Texas A&M University, College Station, Texas, USA

Abstract.—A total of 7,806 nucleotide positions derived from one mitochondrial and eight nuclear DNA segments were used to provide a robust phylogeny for members of the order Artiodactyla. Twenty-four artiodactyl and two cetacean species were included, and the horse (order Perissodactyla) was used as the outgroup. Limited rate heterogeneity was observed among the nuclear genes. The partition homogeneity tests indicated no conflicting signal among the nuclear gene fragments, so the sequence data were analyzed together and as separate loci. Analyses based on the individual nuclear DNA fragments and on 34 unique indels all produced phylogenies largely congruent with the topology from the combined data set. In sharp contrast to the nuclear DNA data, the mtDNA cytochrome *b* sequence data showed high levels of homoplasy, failed to produce a robust phylogeny, and were remarkably sensitive to taxon sampling. The nuclear DNA data clearly support the paraphyletic nature of the Artiodactyla. Additionally, the family Suidae is diphyletic, and the nonruminating pigs and peccaries (Suiformes) were the most basal cetartiodactyl group. The morphologically derived Ruminantia was always monophyletic; within this group, all taxa with paired bony structures on their skulls clustered together. The nuclear DNA data suggest that the Antilocaprinae account for a unique evolutionary lineage, the Cervidae and Bovidae are sister taxa, and the Giraffidae are more primitive. [Artiodactyla; Cetacea; cytochrome *b*; indels; nuclear DNA; Ruminantia]

The order Artiodactyla has a worldwide distribution and at present comprises three morphologically diverse suborders, the Suiformes, Tylopoda, and Ruminantia. Their evolution dates to the Paleocene (~70–80 million years ago; Waddell et al., 1999), the Suiformes (currently including swine and hippopotamus) apparently being the most primitive group in the order, based on fossil (O'Leary and Geisler, 1999) and other morphological characters (nonruminating, two-/three-chambered stomach, retained upper incisors and canines, low crowned cusped molars, and absence of horns or antlers; Nowak, 1999). Although the suborder Tylopoda (camel and llama) shares many of these primitive features, the three-chambered digestive system of its members is based on rumination, and their molars are high crowned with crescents (Nowak, 1999). These latter dental features are considered to be synapomorphic, uniting the Tylopoda with the most derived suborder Ruminantia (Tragulidae–chevrotain; Antilocapridae–pronghorn; Giraffidae–giraffe and okapi; Cervidae–deer, elk, and muntjac; Bovidae–cattle, sheep, and other antelope). Except for the primitive chevrotain (which is characterized by a three-chambered stomach, presence of upper canines, and absence of horns or antlers), all species in this group have a four-chambered stomach, and in nearly all instances (see below) they have lost their upper incisors and possess paired bony structures on their skulls (horns, antlers, or ossocones), at least in the males (Eisenberg, 1981; Nowak, 1999).

Recent paleontological evidence (see Gatesy, 1998, and references therein) and genetic data from both mitochondrial (Graur and Higgins, 1994; Irwin and Arnason, 1994; Ursing and Arnason, 1998) and nuclear DNA sources (Queralt et al., 1995; Shimamura et al., 1997; Gatesy et al., 1999a; Nikaido et al., 1999) suggest that the Artiodactyla is paraphyletic. The order Cetacea (whales and dolphins) has been placed within the Artiodactyla as a sister taxon to the hippopotamus, thus creating a diphyletic suborder Suinae (for contrasting viewpoints, see Luckett and Hong, 1998; O'Leary and Geisler, 1999).

Artiodactyl relationships at lower taxonomic levels are equally problematic, a case in point involving the Miocene origin of the pronghorn (*Antilopcapra*, Antilocapridae) within the diverse suborder Ruminantia (De Blase and Martin, 1974). Several molecular attempts to place the pronghorn phylogenetically have failed, and various authors have speculated that this pecoran taxon might be a "link" between the bovids and cervids (O'Gara and Matson, 1975; Leinders

[1]Address correspondence to this author at Department of Zoology, University of Stellenbosch, Private Bag X1, Stellenbosch 7602, South Africa; E-mail: cam@maties.sun.ac.za

and Heintz, 1980; Eisenberg, 1981; Janis and Scott, 1987; Solounias, 1988). The pronghorn and members of the Cervidae share immunoglobulin antigens (Curtain and Fudenberg, 1973), karyotypic similarities (Gallagher et al., 1994), and morphological characteristics such as a lachrymal duct with two orifices, absence of process urethra, and absence of a Cowper's gland (Leinders and Heintz, 1980). On the other hand, the pronghorn and members of the Bovidae share similarities in horn structures (Leinders and Heintz, 1980) and are grouped according to the results of isoenzyme studies (Baccus et al., 1983). Moreover, systematic ambiguities are not limited to the pronghorn; similar problems surround the placement of the Giraffidae (Douzery and Catzeflis, 1995; Cronin et al., 1996; Montgelard et al., 1997; Gatesy et al., 1999a). Whereas mitochondrial DNA (mtDNA) sequence data indicate a sister taxon relationship between the Giraffidae and the Cervidae (Jermann et al., 1995; Gatesy et al., 1996), morphological synapomorphies (Janis and Scott, 1987) and immunogenetic evidence (Schreiber et al., 1990) suggest the Bovidae and Cervidae are sister taxa, with the Giraffidae being more primitive. The most recent molecular reconstruction of pecoran evolutionary relationships, using sequence data derived from several genetic markers, failed to demarcate reliably the relationships among the Cervidae, Giraffidae, and Bovidae (Gatesy et al., 1999a).

In the past, mtDNA has been extensively used for mammalian systematics at nearly all taxonomic levels. Given the high mutation rate in animal mtDNA (Brown et al., 1982), the marker has performed well in recovering phylogenies between species and genera within the same family. In marked contrast, however, mammalian mtDNA phylogenies focused at the higher taxonomic levels (between subfamilies, families, suborders, and orders) are not robust across analytical methods and most of the deeper nodes in the resulting trees are either unresolved or characterized by low (<70%) bootstrap support (Irwin et al., 1991; Allard et al., 1992; Gatesy et al., 1992; Nedbal et al., 1994; Douzery and Catzeflis, 1995; Honeycutt et al., 1995; Milinkovitch et al., 1996; Matthee and Robinson, 1997; Montgelard et al., 1997; Flynn and Nedbal, 1998).

The inability of mtDNA to recover mammalian phylogenies is commonly attributed to a rapid evolutionary radiation within the respective groups (Kraus and Miyamoto, 1991; Allard et al., 1992; Gatesy et al., 1992; Douzery and Catzeflis, 1995; Lara et al., 1996; Matthee and Robinson, 1997; Halanych et al., 1999; Halanych and Robinson, 1999). In an attempt to overcome this limitation, various sensitivity analyses were performed to identify regions or characters that potentially contained sufficient phylogenetic signal. Typically, areas characterized by extensive homoplasy were excluded in an effort to reveal any signal present in the data (Allard et al., 1992; Swofford et al., 1996; Montgelard et al., 1997; Matthee and Robinson, 1997; Flynn and Nedbal, 1998; Luckett and Hong, 1998). It has been suggested that the cladogenesis of mammalian orders most likely will be resolved only by using a combination of data sets from multiple sources (Allard et al., 1992) and possibly by including more mtDNA sequence data (Douzery and Catzeflis, 1995).

Nuclear DNA sequence data offer a potentially powerful alternative to mtDNA. The nuclear genome of mammals is ~166,000 times larger than the mitochondrial genome and also provides sets of markers that segregate independently. Studies of nuclear DNA genes remain limited and the majority of the previous investigations have been based on small fragments with only a limited number of taxa. In part, the systematic emphasis on mtDNA studies is driven by the availability of so-called "universal mtDNA primers." What is not widely realized among the systematic community is that similar nuclear genome resources are becoming available for most plant and animal groups, a result of the ongoing efforts in genomics. In an attempt to utilize this resource and to provide a well-supported phylogeny for members of the order Artiodactyla, we sequenced eight nuclear DNA fragments from 1 perrisodactyl, 24 artiodactyl, and 2 cetacean specimens. The primers were selected from a suite of gene mapping markers currently used to detect target inserts in bovine and ovine bacterial artificial chromosome (BAC) libraries. Because primers of this type are located within conserved coding regions and are developed by using comparative sequence data (typically from the human and mouse genomes), their utility is in most instances extendible to other, distantly related mammalian taxa (Venta et al., 1996; Lyons et al., 1997).

TABLE 1. A conventional classification scheme (order, suborder, and family) of the mammalian species used in the present study (following Nowak, 1999). Common names of taxa are followed by the specific names in parentheses.

Perissodactyla	
Hippomorpha	
Equidae	Horse (*Equus caballus*)
Cetacea	
Odontoceti	
Physeteridae	Pygmy sperm whale (*Kogia breviceps*)
Mysticeti	
Balaenidae	Bowhead whale (*Balaena mysticetus*)
Artiodactyla	
Suiformes	
Suidae	Domestic pig (*Sus scrofa*)
Tayassuidae	Collared peccary (*Tayassu tajacu*)
Hippopotamidae	Pygmy hippopotamus (*Hexaprotodon liberiensis*)
Tylopoda	
Camelidae	Llama (*Llama glama*)
	Camel (*Camelus bactrianus*)
Ruminantia	
Tragulidae	Asiatic chevrotain (*Tragulus meminna*)
Antilocapridae	Pronghorn (*Antilocapra americana*)
Cervidae	Reeve's muntjac (*Muntiacus reevesi*)
	Mule deer (*Odocoileus hemionus*)
	Reindeer/caribou (*Rangifer tarandus*)
Giraffidae	Giraffe (*Giraffa camelopardalis*)
	Okapi (*Okapia johnstoni*)
Bovidae	Cow (*Bos indicus*)
	Nilgai (*Boselaphus tragocamelus*)
	Lesser kudu (*Tragelaphus imberbis*)
	Sable antelope (*Hippotragus niger*)
	Thomson's gazelle (*Gazella thomsoni*)
	Kirk's dik-dik (*Madoqua kirkii*)
	Impala (*Aepyceros melampus*)
	Waterbuck (*Kobus ellipsiprymnus*)
	Tsessebe (*Damaliscus lunatus*)
	Sheep (*Ovis aries*)
	Goat (*Capra hircus*)
	Muskox (*Ovibos moschatus*)

MATERIALS AND METHODS

Taxon Sampling

Twenty-four artiodactyl taxa were selected representing all three extant suborders (Nowak, 1999; Table 1). Because previous molecular evidence had suggested a paraphyletic origin for the order Artiodactyla, two whale specimens representing both cetacean suborders were included (Table 1). The horse (order Perissodactyla) was used as the outgroup for all the DNA analyses. In several instances, artiodactyl specimens were chosen to reduce potential problems with long branch attraction (Felsenstein, 1978; Swofford et al., 1996; Huelsenbeck, 1997) and to decrease the possible affects of taxon sampling on the phylogeny (Hillis, 1996, 1998; Kim, 1996; Graybeal, 1998). The species-rich suborder Ruminantia, which comprises five extant families, was extensively sampled.

Data Collection

Total genomic DNA was extracted from fibroblast cells or frozen tissue by using standard phenol/chloroform/isoamyl alcohol procedures or a QIAamp DNA purification kit (Qiagen Ltd.). Initially, a DNA panel of 12 representative taxa was tested for amplification with ~60 primer pairs designed to screen bovine and ovine BAC libraries. Several of the conserved gene-specific universal mammalian primers from Venta et al. (1996) were also used. For the present investigation, polymerase chain reaction (PCR) primers were selected on the basis of a combination of criteria. A relatively small amplicon product (<800 base pairs [bp]) facilitates accurate

automated sequencing results without requiring the use of multiple internal primers for proofreading. We also preferred primers that amplify specifically across a wide range of taxa within the group of interest (in our case, the Perrisodactyla, Artiodactyla, and Cetacea), yielding fragments that represent evolutionary markers and segregate independently. Knowledge of the evolutionary rate of the included genes would improve the phylogenetic results: Clearly, slowly evolving segments should be used for older evolutionary events, whereas rapidly evolving segments are more appropriate for recent events. To obtain information regarding the evolutionary rates of the genes used in this study, we selected, sequenced, and analyzed a subset of taxa to test for phylogenetic signal. Finally eight, presumably independent, nuclear DNA segments, all located on different chromosomes in cattle, were included in our investigation (Table 2). Although independence among markers is difficult to assess with certainty (linkage groups can change radically over evolutionary time; see Robinson et al., 1998, and references therein), a previous chromosomal study including representatives of all the advanced pecorans showed numerous monobranchial autosomal homologies among taxa (Gallagher et al., 1994). In cases where available primers failed to amplify the DNA of a taxon of interest, new primers were designed from the artiodactyl sequences at hand to attain these amplicons.

PCR reaction mixes (50 μl total volume) contained ~100–200 ng of total genomic DNA. Thirty-five amplification cycles were performed with denaturation at 94°C for 30 sec, annealing between 54°C and 65°C for 30 sec (Table 2), and extension at 72°C for 60 sec. Concentrations of $MgCl_2$ were varied (1.0–2.5 mM; Table 2) to achieve optimal amplification. A 5-μl sample of each PCR reaction mix was screened on 1% agarose gels along with a size standard (λ-DNA cut with Hind III) before sequencing. The gene-specific PCR products were purified by using a QIAquick PCR purification kit (Qiagen Ltd.) and cycle-sequenced by using BigDye terminator chemistry according to the recommendations of the manufacturer (Perkin-Elmer Applied Biosystems). Sequencing products were cleaned by using Centrisep spin columns (Princeton Separations) and analyzed with an ABI 377XL automated sequencer. Sequences obtained from the nuclear DNA segments were compared with sequences in GenBank by using BLASTN, and the exon/intron boundaries were defined from the published data for each gene. Both strands were sequenced to improve the accuracy of the base identification. Heterozygous changes in the nuclear DNA data occurred in low frequency (<0.5% of the cases), mostly involving transitional changes. These nucleotides were arbitrarily designated as one of the two possible states, based on the frequency in the other individuals. Irrespective of the designation, none of the changes involved parsimony-informative sites and therefore did not influence our conclusions. In some instances (fewer than five taxa per gene), nonspecific PCR amplification produced multiple fragments. All amplified fragments that corresponded roughly to the expected product size were extracted from an agarose gel by using a commercial DNA gel extraction kit (Qiaex II gel extraction kit; Qiagen Inc.), reamplified, and sequenced. In each case, only one product showed substantial (>90%) sequence identity with the target fragment.

Sequence Data and Alignment

All nuclear sequences were aligned manually by using the conserved exon sequences to "anchor" the intron alignment. No deletions or insertions were present in the coding regions, and all alignment gaps were introduced in the introns. These gaps varied in size and ranged from 1 to >1,500 bp long. Areas of ambiguous alignment were excluded before the phylogenetic analyses were performed. The boundaries of the uncertain regions were defined by pruning the sequences to the last and first shared homologous character present among all taxa. Although special care was taken to ensure accurate and optimal sequence alignment, it is possible, considering the size of the data set, that some investigator error may have occurred. We strongly believe, however, that minor alterations to our alignment will not change the outcome of this study. The sequences of the 215 nuclear fragments generated by this study were deposited in GenBank (accession numbers AF165596 to AF165811), and the aligned data file can be accessed from EMBL (accession number DS39739). For the cytochrome *b* data, 26 of the 27 genera included in the nuclear study

TABLE 2. Abbreviations and full names of the gene fragments used in the present investigation. The mapped chromosomal location of each segment in *Bos taurus* is indicated in parentheses. PCR annealing temperature and $MgCl_2$ concentrations are given for each primer pair.

Gene	Locus name	PCR conditions		Forward primer	Reverse primer
		Temp	$MgCl_2$, mM		
MGF (*Bta*5)	Stem cell factor	54	1.5	A[a]-ATCCATTGATGCCTTCAAGG C - GTGAATGTGTGATTTCTTCAA E - GATTTCTTCAACATTAAGTC	B[a]-CTGTCATTCCTAAGGGAGCTG D - CTGGAATCTAAATAGAAAGC
TG (*Bta*14)	Thyroglobulin	55	2.5	A[a] - AGGGTCAGCCTATTCCTGG C - GAGCCCAAGCAATGTAAGTCT E - GCCYATTCCTGGAACTCGAAGT	B[a] - AGCTGCTTCTCTGTAGCTCATG D - ATACCACTCGAAGGCCTGCTC F - CCTGGCCAGCACTGTTCTGAG
PRKCI (*Bta*X)	Protein kinase C, iota	60	2.5	A[b] - GGGTAATAGGAAGAGGAAGTT C - AGTTATGCTAAAGTACTGTTG	B[b] - CCAACAAGGAAAGGATGAT D - GGACGCCTGTTCAAAGACATG
SPTBN1 (*Bta*11)	β-Spectrin, nonerythrocytic 1	57	1.5	A[c] - TCTCAAGACTATGGCAAACA C - GAAGACCTGTTACAGAAGCA	B[c] - CTGCCATCTCCCAGAAGAA D - TCTGCTGCCAACTGGCAAAGC
STAT5 (*Bta*19)	Signal transducer & activator of transcription 5A	64	1.0	A[d]- GAAGAAACATCACAAGCCCC C - GTGCTCCGTGCACCCAGGGC	B[d] - AGACCTCATCCTTGGGCC D - ATGGAGAAATCCCGCGTGGT
TNF-A (*Bta*23)	Tumor necrosis factor - α	62	1.5	A[e] - GAAGAGGTGAGTTTCTGGCCGGC C - CACGCTCTTCTGCCTGCTGCA	B[e] - CCACCTGGGGACTGCTGGGGAGA D - ACTGCTGGCTGCTTCTCTATC
K-CAS (*Bta*6)	κ-Casein	65	1.0	A[f] - CAAGCTCTGGGCAAAGTGGTT C - TTTGAAACTAATGTTATTTTTA E - GTGGAAGGAAGATGTACAAATC	B[f] - GATAAATGTGGGTGTGGGTGACG D - CTAACTGCAACTGGCTTTGCATA
TH (*Bta*3)	Thyrotropin	56	2.0	A - GCATGTGGGCARRCAATGTC	B - ATGGCYTCATGTATRCAGTCA

[a]Lyons et al. (1997).
[b]Marcos De Donato, Department of Animal Science, Texas A&M University, Texas, USA.
[c]Venta et al. (1996).
[d]Eric Antoniou, School of Animal and Microbial Sciences, Reading, UK.
[e]Alan Teale, International Livestock Research Institute, Nairobi, Kenya.
[f]Tracie Webster, Victorian Institute of Animal Science, Victoria, Australia.

were available (*Equus caballus*: D32190, *Tragulus javanicus*: D32189 [Chikuni et al., 1995]; *Sus scrofa*: AB015079 [Watanobe et al., unpubl.]; *Tayassu tajacu*: U66289: [Theimer and Keim, 1998]; *Llama glama*: U06429, *Camelus bactrianus*: U06427 [Stanley et al., 1994]; *Hexaprotodon liberiensis*: Y08814 [Montgelard et al., 1997]; *Kogia breviceps*: U72040 [Milinkovitch et al., 1996]; *Balaena mysticetus*: X75588 [Arnason and Gullberg, 1994]; *Odocoileus hemionus*: X56291, *Antilocapra americana*: X56286, *Giraffa camelopardalis*: X56287: [Irwin et al., 1991]; *Muntiacus* sp: AJ000023, *Rangifer tarandus*: AJ000029 [Randi et al., 1998]; *Ovibos moschatus*: U90303 [Groves and Shields, 1997]; *Boselaphus tragocamelus*: AJ222679 [Hassanin and Douzery, 1999]; *Ovis aries*: AB006800 [Takada et al., unpubl.]; *Capra Hircus*: D84201 [Arai et al., unpubl.]; *Bos taurus*: J01394 [Anderson et al., 1982]; *Kobus ellipsiprymnus*: AF022059, *Damaliscus lunatus*: AF016635, *Tragelaphus imberbus*: AF022064, *Hippotragus niger*: AF022061, *Gazella dama*: AF025954, *Madoqua kirkii*: AF022070 [Matthee and Robinson, 1999a]). To obtain complete taxon representation for the mtDNA analyses, the cytochrome *b* gene of the okapi was sequenced and included in the final data analyses (GenBank accession number AF181470). To do so, we used published primers L14841 and H15915 (for details, see Matthee and Robinson, 1999a).

Phylogenetic Approach

Alignment gaps not excluded from our data were treated as missing characters. Unique indels that involved two or more consecutive base pairs, that had clearly defined alignment borders, and that were present in at least two or more taxa were scored as being present or absent and coded into a separate data set (Table 3). The unique indel data set was treated independently and was not included in the nucleotide analyses. The phylogenetic signals contributed by each of the eight nuclear DNA sequence data sets were compared by using the partition-homogeneity tests in PAUP* 4.0b2a written by David L. Swofford. Subsequently, all nuclear DNA fragments were analyzed as a single data set. These analyses revealed several consistently retrieved nodes that were well-supported by bootstrap analyses (>80%). To evaluate the ability of nuclear DNA subsets to recover the same phylogeny, independent gene analyses were performed. Congruence among trees was scored on the basis of 13 selected nodes recovered by the combined analyses. The selection of these nodes was by subordinal and familial classification, and nodes so defined were labeled A through M. To further explore inconsistency among the independent data sets included in this study, we calculated incongruence length differences (Farris et al., 1995; Baker and De Salle, 1997) and determined the significance of these values by using PAUP* 4.0b2a.

Phylogenetic Methods

Because phylogenetic accuracy can be improved substantially by selecting methods on the basis of appropriate models of evolutionary change (Swofford et al., 1996), several analyses were performed to define the characteristics of the sequence data included in this study. Saturation analyses of both nuclear and mtDNA data were performed by plotting the number of transitions (TI) and transversions (TV) against the percentage sequence divergence. To test for among-site rate heterogeneity, we estimated a gamma shape parameter and calculated the empirical TI:TV ratio using maximum likelihood (because of computational constraints, the parsimony topology was used each time as a reference). Base frequencies were incorporated and the nuclear DNA data were tested for rate heterogeneity among lineages by using a relative rate test (Tajima, 1993). The horse was used as reference taxon and the evolutionary rate of each taxon was compared with that of the 25 remaining lineages. Maximum likelihood branch length estimates were used to indicate whether the evolutionary rate of a taxon of interest was faster or slower than that of the other taxa. Significance values ($\alpha = 0.1$) for the chi-square distribution were adjusted from $\chi^2 < 3.841$ to $\chi^2 < 8.2838$ by using the Bonferroni correction for multiple comparisons. To condense the results, taxa were considered to be significantly different in their rate of evolution if they differed significantly in more than three of the pairwise comparisons.

TABLE 3. The presence (1) or absence (0) of a deletion in the nuclear fragments used in this study (see text for details). The gene fragment names correspond to the abbreviations given in Table 2 and the letters a to H represent the 34 parsimony-informative indels. N = missing data for the chevrotain. The length of each indel, expressed as the number of base pairs, is a follows: a = 15; b = 21; c = 2; d = 2; e = 3; f = 4; g = 14; h = 11; i = 3; j = 4; k = 14; l = 20; m = 254; n = 100; o = 298; p = 18; q = 2; r = 268; s = 4; t = 2; u = 269; v = 4; w = 3; x = 3; y = >1,200; z = 2; A = 4; B = 4; C = 4; D = 2; E = 2; F = 8; G = 5; H = 4.

	TH						*SPTBN1*						*TG*				*MGF*					*PRKC1*				*TNF-A*					*STAT5*		*K-CAS*	
	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	A	B	C	D	E	F	G	H
Horse	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pig	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
Peccary	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
Llama	1	0	0	1	0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	0	0	1	0	0	1	0	0	0	1	1	0	1	0	0
Camel	1	0	0	1	0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	0	0	1	0	0	1	0	0	0	1	1	0	1	0	0
Bowhead whale	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
Pigmy whale	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
Hippopotamus	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Chevrotain	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0	N	N	N	N	N	0	0	0	0
Pronghorn	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1	0	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
Muntjac	0	0	0	1	0	1	0	1	0	0	0	1	1	0	0	1	0	1	0	1	0	0	1	0	0	0	1	0	1	0	1	0	0	1
Muledeer	0	0	0	1	0	1	0	1	0	0	0	1	1	0	0	1	0	1	0	1	0	0	1	1	0	0	1	0	1	0	1	0	0	1
Reindeer	0	0	0	1	0	1	0	1	0	0	0	1	1	0	0	1	0	1	0	1	0	0	1	1	0	0	1	0	1	0	1	0	0	1
Giraffe	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1	0	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
Okapi	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1	0	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
Cow	0	0	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
Nilgai	0	0	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
Lesser kudu	0	0	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
Sable	0	1	0	1	0	0	0	0	0	0	1	1	1	0	0	0	1	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
Gazelle	0	0	0	1	0	0	0	0	1	1	1	1	1	0	0	0	1	1	0	1	0	0	1	0	0	0	1	1	1	0	0	0	0	0
Dik-dik	0	0	0	1	0	0	0	0	1	1	1	1	1	0	0	0	1	1	0	1	0	0	1	0	0	0	1	1	1	0	0	0	0	0
Waterbuck	0	0	0	1	0	0	0	0	1	0	1	1	1	0	0	0	1	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
Tsessebe	0	1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	1	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
Impala	0	0	0	1	0	0	0	0	1	0	1	1	1	0	0	0	1	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
Sheep	0	1	0	1	0	0	0	0	0	0	1	1	1	0	0	0	1	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
Goat	0	1	0	1	0	0	0	0	0	0	1	1	1	0	0	0	1	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0
Muskox	0	1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	1	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0

Maximum parsimony, neighbor joining, and maximum likelihood methods were used with PAUP* 4.0b2a. The parsimony analyses utilized the heuristic search option with 10 random-addition sequences and TBR branch swapping. On the basis of the characteristics of the nuclear sequence data included in this study, the data were treated as unordered and no TI:TV weightings were applied. On the other hand, clear saturation of TIs in the mtDNA cytochrome *b* data required additional weighting schemes to reduce the noise in these data. In the latter instance transitional changes were either downweighted (the empirical weighting of 1:2.6 for TI:TV estimated by maximum likelihood), completely excluded, or masked by translating the gene into amino acids. Distance and maximum likelihood analyses for all data were based on the HKY85 correction (Hasegawa et al., 1985), which adjusts for the differences in TI:TV ratio and for unequal base frequencies. Among-site rate heterogeneity and the estimated gamma shape value were incorporated in the distance and maximum likelihood calculations. The empirical TI:TV ratios were used as input values in all maximum likelihood analyses. Finally, nodal support for the parsimony and neighbor joining analyses was assessed from 1,000 bootstrap replicates, and 100 iterations were performed for the combined maximum likelihood analyses by using the neighbor joining tree as reference. Competing hypotheses of branching patterns were tested with the Kishino and Hasegawa (1989) test and the maximum likelihood values of alternating hypotheses.

Attributes of the DNA Data

Apart from the systematic focus of the present study, we were interested in the ability of nuclear DNA sequence data to resolve the phylogeny of the Artiodactyla, a problematic evolutionary question that had proven nearly intractable using mtDNA analyses. The combined nuclear DNA sequence data were partitioned into regions of noncoding introns and coding exons for comparison. Because these two data sets are under different evolutionary constraints, we anticipated they might not be equivalent in recovering the phylogeny of the Artiodactyla. Additionally, we wanted to estimate the number of nuclear DNA sequence characters necessary to recover the phylogeny obtained from all 6,666 positions. Starting with the complete data set, characters were randomly excluded in increments of 5% by using the jackknifing approach (100%, 95%, 90%, etc.), and the reduced data sets were analyzed by both parsimony and neighbor joining methods (see also Springer et al., 1999). This process was repeated 1,000 times for each exclusion level and the resulting trees were evaluated to determine how often each node in the total data solution was recovered by the partial analyses. Nodes were collapsed when they were retrieved $<50\%$ of the time.

To test the sensitivity of the analyses to taxon sampling, taxa were randomly excluded from the data sets (Hillis, 1996, 1998; Kim, 1996; Graybeal, 1998), and neighbor joining and parsimony solutions were recomputed. The horse, hippopotamus, chevrotain, and pronghorn were represented by single exemplars from within their taxonomic groups and were included in all analyses. A single individual from the remaining monophyletic evolutionary lineages was randomly included each time (pig or peccary; llama or camel; bowhead or pygmy sperm whale; muntjac, muledeer, or reindeer; giraffe or okapi; and one representative of the family Bovidae). The procedure was repeated 100 times and the resulting topologies were evaluated to determine how often all nodes in the total data solution were recovered by the partial analyses.

Finally, we examined the influence of missing data on the artiodactyl nuclear DNA phylogeny. Because the effects of missing data are also influenced by taxon sampling (Wiens, 1998), both the number of taxa and the amount of missing data were varied. In total, 100 random replicates of 36 different data sets were generated from the total nuclear DNA data. The incremental data sets included 3, 6, 9, 12, 15, and 18 randomly selected taxa in which 600, 1,200, 1,800, 2,400, 3,000, or 3,600 randomly selected characters were replaced by missing data. Both parsimony and neighbor joining analyses were performed on all 3,600 data sets and all of the generated trees were saved to outfiles. A consensus tree was generated in PAUP* 4.0b2a for all of the outfiles. The strict consensus trees were compared with the total data solution, and the number of nodes affected by the missing data were recorded (nodes not present

TABLE 4. Characteristics of the nuclear gene segments used in this investigation (gene name abbreviations correspond to those in Table 2). Values for each segment include the number of characters excluded because of ambiguous alignment (EXCL), the number of coding (EXON) and noncoding (INTRON) characters, the number of unique indels (INDEL), and the number of parsimony-informative (PARS INF) and variable characters (VARIABLE). The α-values estimated by maximum likelihood from the parsimony tree are given for each data set.

Gene	Total	EXCL	EXON	INTRON	INDEL	PARS INF	VARIABLE	α-Value
MGF	1073	74	54	945	5	187	404	1.798
TG	1487	37	249	1201	5	268	529	1.173
PRKCI[a]	1048	68	156	824	4	122	259	1.412
SPTBN1	997	194	219	584	7	280	416	0.760
STAT5	733	145	204	384	2	163	275	0.775
TNF-A	545	0	0	545	4	166	321	2.965
K-CAS	597	0	270	327	1	142	251	2.376
TH	741	37	270	434	7	163	288	0.903
All data	7221	555	1422	5244	35	1491	2743	1.094
Intron	5244	555	0	0	35	1249	2341	2.140
Exon	1422	0	0	0	0	242	398	0.147
Cytb	1140	0	1140	0	0	478	586	0.266

[a] An insertion of ~1500 bp in the Cetacea intron is not included in the presented statistics.

100% of the time were scored as affected). To estimate the influence of the number of taxa showing missing data versus the number of missing characters, pairwise comparisons were performed between the replicates for which the exact number of excluded characters was the same (e.g., a situation of three taxa having 1,200 characters missing was compared with six taxa having 600 characters missing).

RESULTS

Characteristics of the DNA Data

The total aligned nuclear DNA data set, including all indels, comprised 7,221 bp, of which 555 were excluded because of uncertain alignment (Table 4). These data constitute 1,422 bp in exon regions and 5,244 bp in introns or other noncoding regions. As typically observed in coding regions, most substitutions (66%) were present at third-codon positions, with first-codon (19%), and second-codon (15%) changes being more conservative. In sharp contrast to the mtDNA, which has strong selection against guanine within the family Bovidae (Matthee and Robinson, 1999a; Rebholz and Harley, 1999), the exon regions of the nuclear DNA showed little bias toward any base (A = 25.5%, C = 25.5%, G = 26.2%, T = 22.8); a slight bias towards adenine and thymine was apparent for the intron sequences (A = 28.4%, C = 18.8%, G = 23.2%, T = 29.6%). In contrast, although the exon regions were characterized by among-site rate heterogeneity ($\alpha = 0.147$), the intron regions showed nearly complete absence of among-site rate variation ($\alpha = 2.14$; Table 4).

The independent nuclear DNA fragments ranged in length between 545 and 1487 bp and contained between 251 and 529 variable characters (Table 4). Approximately half of the variable characters were parsimony-informative and 34 unique indels were scored across the nuclear fragments (Table 4). In total, 23.8% of the characters in the intron regions were parsimony-informative versus only 17.0% of the characters in the exon regions. Total nuclear DNA sequence divergence values among ingroup taxa ranged from 1.81% (between the goat and the muskox) to 19.58% (between the peccary and the chevrotain). Average sequence divergence values between the outgroup and the ingroup taxa were estimated as 18.50% (±0.57%) for the nuclear DNA comparisons and 20.31% (±1.23%) for the parallel mtDNA data set. For the ingroup comparisons the nuclear DNA data set was characterized by an average sequence divergence of 10.69% (±5.37%), whereas the mtDNA sequences evolved at a distinctly faster rate, having an average sequence divergence of 17.53% (±2.92%).

Pairwise nuclear DNA comparisons between the 27 taxa showed a nearly perfect linear increase in TI and TV over time (Fig. 1a). Comparisons involving the outgroup also reflected a distinct clustering separate from most ingroup comparisons, a scenario clearly indicating that these data

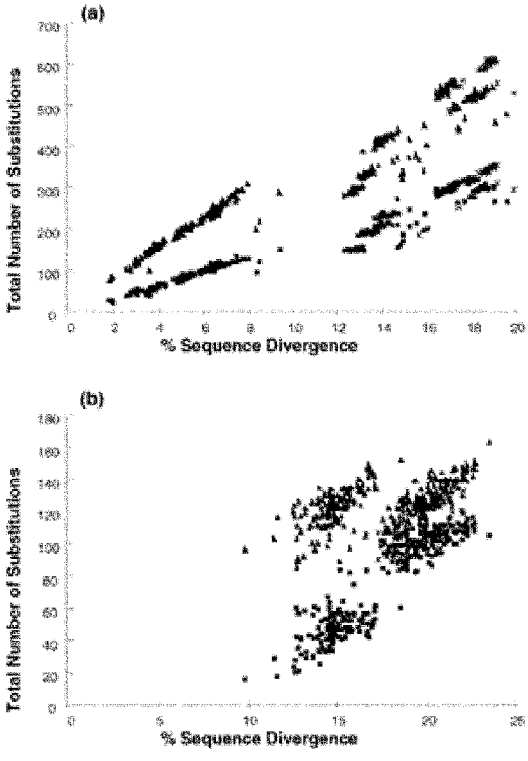

FIGURE 1. Saturation plots for the total number of transitions and transversions plotted against the uncorrected sequence divergence values. (a) Nuclear DNA data. (b) Mitochondrial cytochrome *b* DNA data.

have not reached saturation (Griffiths, 1997; Matthee and Robinson, 1999a). In sharp contrast to the nuclear DNA comparisons, where transitional changes are not saturated even at 20% sequence divergence, the cytochrome *b* sequence data showed saturation for transitions at roughly 10% sequence divergence (Fig. 1b; see also Brown et al., 1982; Irwin et al., 1991; Matthee and Robinson, 1999a). In fact, no marked difference was apparent in the total number of transition substitutions between ingroup/ingroup and ingroup/outgroup comparisons for the cytochrome *b* data.

Pairwise comparisons between the taxa studied revealed a generally constant rate of sequence substitution among lineages; differences in the rate of evolution were mainly restricted to the Cetacea, Tylopoda, and Suidae (Table 5). Both whales showed a significantly slower rate of evolution at three of the genes (*MGF*, *SPTBN1*, and *TH*), and the bowhead whale was also slow for *PRKCI*. Where suids showed significant rate differences, they both showed an increased rate, whereas the Tylopoda were characterized by one fast and one slowly evolving gene (Table 5).

Combined Analyses and the Artiodactyl Phylogeny

The partition-homogeneity test indicated no significant conflicting phylogenetic signal among the eight independent nuclear DNA segments used in this study ($P = 0.66$). Combined analyses using both neighbor joining and parsimony resulted in identical topologies at the subordinal and familial level, and a single most-parsimonious tree was found (Fig. 2). Parsimony bootstrap support for the nodes defining suborders and families ranged from 72% to 100% (Fig. 2). The maximum likelihood tree (Fig. 3) was identical to the topology in Figure 2 except for the placement of the pronghorn, which grouped as a

TABLE 5. Taxa showing significant rate heterogeneity ($\alpha = 0.1$) for each gene when their rate of sequence evolution is compared with that of other exemplars used in this study (see text for details).[a]

Gene name	No. of taxa	Taxon names
MGF	3	Pig$^+$, bowhead whale$^-$, pygmy sperm whale$^-$
TG	3	Llama$^+$, camel$^+$, chevrotain$^+$
PRKCI	1	Bowhead whale$^-$
SPTBN1	5	Llama$^-$, camel$^-$, bowhead whale$^-$, pygmy sperm whale$^-$, hippopotamus$^-$
STAT5	0	No taxa
TNFA	1	Giraffe$^-$
K-CAS	0	No taxa
TH	3	Peccary$^+$, bowhead whale$^-$, pygmy sperm whale$^-$

[a] $^+$a significantly faster mutation rate, $^-$a significantly slower rate.

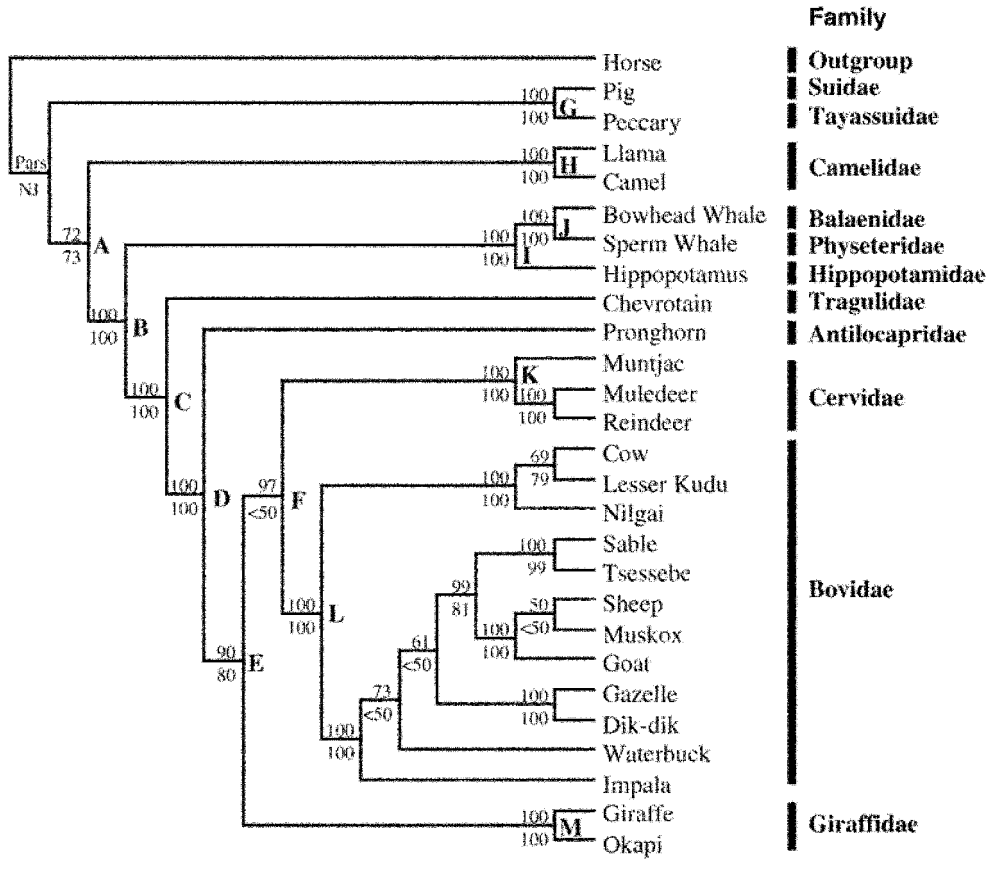

FIGURE 2. Phylogenetic tree (parsimony and neighbor joining) showing the evolutionary relationships among the 26 ingroup taxa sampled in the present study. The tree is 4703 steps long and has a consistency index value of 0.73 and a retention index value of 0.74. The numbers above and below the branches represent bootstrap values obtained using parsimony (Pars) and neighbor joining (NJ), respectively. The nodes (A–M) indicate the 13 familial and subordinal associations and correspond to the designations in Tables 6 and 8.

sister taxon to the Giraffidae. Apart from this node, the bootstrap support for the nodes defining suborders and families was once again high (ranging between 87% and 100%; Fig. 3). There was no difference ($t = 0.5005$; $P = 0.6167$) between the log-likelihood for the optimum tree ($-\ln L = 33784.90144$) and for a constrained topology ($-\ln L = 33788.64009$) that was identical to the tree presented in Figure 2. Therefore, support for the placement of the pronghorn is limited, a result that probably reflects the short internal branch separating the pronghorn from the remainder of the pecorans (Fig. 3).

All combined analyses placed the morphologically primitive suborder Suiformes (excluding the hippopotamus) basal in the artiodactyl radiation. The ruminants were always the most-derived group in the assemblage, and the suborder Tylopoda was intermediate between the Ruminantia and the Suiformes (excluding the hippopotamus; Figs. 2 and 3). In accord with previous molecular suggestions (reviewed in Gatesy et al., 1999a), the order Cetacea clustered as a sister lineage to the hippopotamus; our analyses consistently suggest that this clade is basal to the Ruminantia and is more derived than the Tylopoda.

Two of the nodes (A and F; Fig. 2) that define the associations between suborders and families in our topology had relatively weak bootstrap support (<80%) in at least one of the phylogenetic methods of analysis. A topology that places the camelids basal to the suids requires eight additional steps under parsimony, and the log-likelihood values of the two alternative topologies (one placing

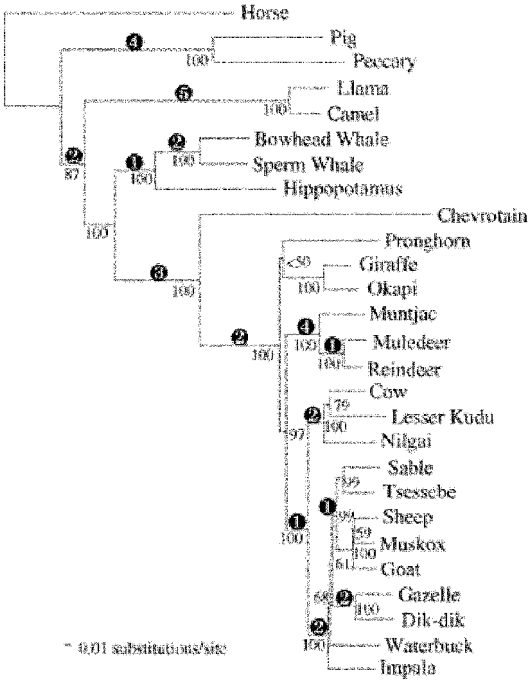

FIGURE 3. Maximum likelihood gene tree based on all 6,666 characters included in this study. Branch lengths are drawn proportionally to the evolutionary distances (see scale given), and the numbers in black circles indicate synapomorphic indels. The bootstrap support for each node is indicated either below the branch or to the right of each node.

the pig/peccary basal and the other placing the camel/llama basal) differ by 12.17 (not significant $t = 1.166$; $P = 0.2436$). Likewise, a topology including a Cervidae and Giraffidae sister relationship also requires an additional eight steps. In this instance, the difference in log-likelihood values between the two topologies approaches significance ($t = 1.78$; $P = 0.075$).

Differences in evolutionary rates among lineages have the potential to create long branches that might bias gene trees (Felsenstein, 1978; Swofford et al., 1996; Huelsenbeck, 1997). Both the Tylopoda and the Suiformes (excluding the hippopotamus) lineages show a greater or a slower rate of evolution for some of the genes (*MGF*, *TG*, *TH*, *SPTBN1*). If these four genes are removed from the analyses to exclude the effects of rate heterogeneity, then all of the remaining genes except *TNF-A* support the suids as being basal and the camelids as being more derived.

Analyses of Individual Genes

Independent gene analyses not only account for the possible influences of different evolutionary processes shaping the evolution of genes and taxa but also allow congruence among individual gene trees to be used as support for the "correct" phylogeny (Hillis et al., 1996; Gatesy et al., 1999a). Maximum parsimony (Fig. 4a–h), neighbor joining, and maximum likelihood analyses of each of the eight independent nuclear DNA fragments, together with the parsimony analyses of the unique indels, resulted in compatible evolutionary associations when compared with the combined data analyses (Fig. 2; Table 6). Combining each of the eight nuclear DNA fragments in a pairwise fashion resulted in one to seven additional steps (Table 7). None of these values were statistically significant, indicating the absence of conflicting signal among the nuclear DNA segments.

Unfortunately, several PCR attempts (using different primer pairs) and cloning methods failed to produce sequence data for the chevrotain *TNF-A* region; this taxon was therefore excluded from the *TNF-A* independent gene analyses. Although most parsimony searches using data from individual genes resulted in more than one tree (Table 6), the topologies differed primarily because of branch swapping within the derived cluster representing the family Bovidae. Six of the 13 selected nodes (C, D, G, H, K, L) were consistently recovered by all of the analyses and three additional nodes (I, J, M) were recovered in >80% of the analyses (Table 6). A single node (E) was recovered by less than half of the analyses (44%). Interestingly, this node, defining the pronghorn basal to the Giraffidae, was the only node that varied between phylogenetic methods in analyses of the complete data set (compare Figs. 2 and 3).

Indel Analysis

Analyses based on the 34 unique indels (Table 3; Fig. 3) recovered 9 of the 13 selected nodes that identify relationships among suborders and families, and 8 of these had bootstrap support >70% (Table 6). Some of these indels represent the insertion of specific SINE elements of >150 bp. One example was found in the thyroglobulin region, for which a 254-bp insertion is present in both

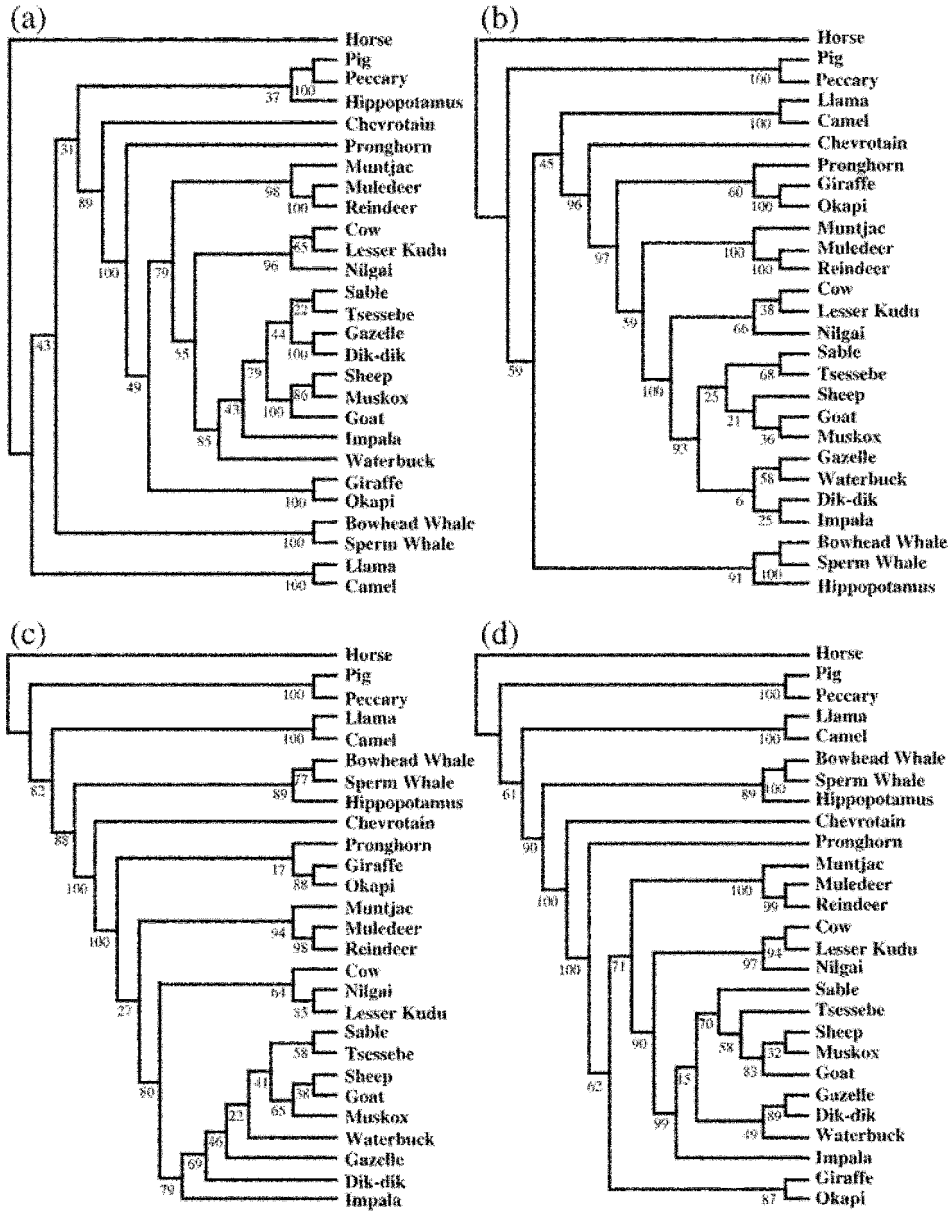

FIGURE 4. Bootstrap parsimony trees for the eight nuclear DNA fragments sequenced in this study. (a) *MGF*; (b) *TG*; (c) *PRKCI*; (d) *SPTBN1*; (*Continued*)

members of the Cetacea, as well as in the hippopotamus. In the same region, a different 290-bp sequence was inserted in the peccary; this insert has 85% sequence identity with the porcine SSPRE sine sequence. Unfortunately, the indel analyses failed to resolve associations in which the internal branches were short (Fig. 3; Table 6), which is probably a reflection on the rare nature of indel events. For example, although a single indel placed the suids basal in the phylogeny, an unresolved trichotomy was found among the Cetacea/hippopotamus clade, the Tylopoda, and the Ruminantia. Nevertheless, the consistency index value for these data was high (0.919; Table 6), indicating that unique indels are clearly useful characters for evolutionary studies (see Fig. 3; Edwards and Wilson, 1990; Giribet and Wheeler, 1999; Matthee and Robinson, 1999b; Nikaido et al., 1999).

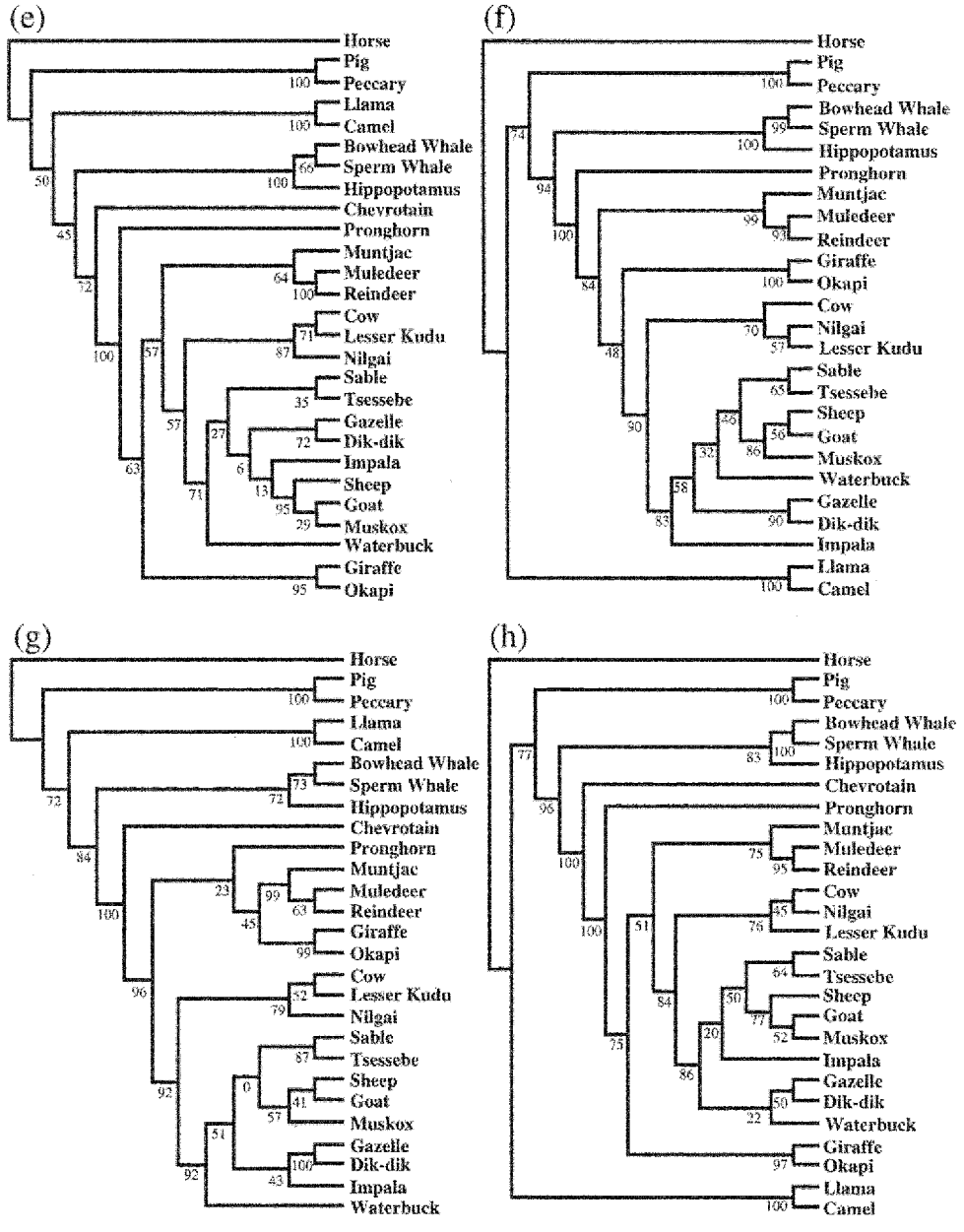

FIGURE 4. (Continued) (e) *STAT5*; (f) *TNF-A*; (g) *K-CAS*; (h) *TH*. Values below the branches indicate bootstrap support for the nodes and abbreviations for gene fragments correspond to those given in Table 2.

Importantly, however, the indel data set was not homoplasy-free (also see Kishino et al., 1990; Hillis, 1999), indicating that even conservative characters can sometimes be misleading from an evolutionary perspective.

Robustness of the Nuclear DNA Data

A striking result to emerge from this study was that parsimony analyses of the complete data set remained consistent when as many as 90% of characters were randomly excluded. This implies that ~700 nuclear DNA characters, drawn at random from the genome, are capable of recovering the phylogeny in Figure 2. In our case, even the exclusion of 95% of the characters (such that 333 characters remained for parsimony analyses) resulted in the collapse of just three nodes (A, E, and F; Fig. 2). The neighbor joining

TABLE 6. Summary of the topologies obtained by the independent gene analyses in the present study. The node letters (A–M) correspond to those in Figure 2 and the numerical values indicate the percentage of times each node was recovered with >70% bootstrap support by parsimony (Pars.) or neighbor joining (NJ) analyses. NF indicates nodes that were not found. The number of trees, tree length (Steps), consistency index value (CI), and retention index (RI) value are given for each parsimony tree. Bootstrap values were not computed under maximum likelihood (ML), so only the presence (√) or absence (×) of nodes for the maximum likelihood solution is indicated.

Data set	Method	No. of trees	Steps	CI	RI	Nodes													% with >70% bootstrap
						A	B	C	D	E	F	G	H	I	J	K	L	M	
MGF	Pars.	24	621	0.783	0.757	NF	NF	89	100	<50	79	100	100	NF	100	98	55	100	8
	NJ					NF	NF	85	100	NF	<50	100	100	NF	65	96	<50	100	6
	ML					×	×	√	√	√	√	√	√	×	√	√	√	√	10
TG	Pars.	3	918	0.709	0.700	59	NF	96	97	NF	59	100	100	91	100	100	100	100	9
	NJ					NF	56	86	100	NF	NF	100	100	78	100	100	98	100	9
	ML					√	×	√	√	×	√	√	√	×	√	√	√	√	11
PRKCI	Pars.	66	403	0.769	0.786	82	88	100	100	NF	<50	100	100	89	77	94	80	88	11
	NJ					NF	<50	100	100	NF	NF	100	100	90	98	98	59	85	8
	ML					√	√	√	√	×	√	√	√	√	√	√	√	√	12
SPTBN1	Pars.	9	833	0.682	0.754	61	90	100	100	62	71	100	100	89	100	100	90	87	11
	NJ					NF	90	99	100	66	57	100	100	98	100	100	96	100	10
	ML					√	√	√	√	×	√	√	√	√	√	√	√	√	12
STAT5	Pars.	69	505	0.710	0.720	50	<50	72	100	63	57	100	100	100	66	64	57	95	6
	NJ					NF	NF	65	100	72	<50	100	100	100	69	83	60	97	7
	ML					√	√	√	√	√	√	√	√	√	√	√	√	√	13
TNF-A	Pars.	12	523	0.790	0.780	NF	94	—	100	84	NF	100	100	100	99	99	90	100	10
	NJ					NF	92	—	99	90	NF	100	100	99	84	94	84	100	10
	ML					×	√	—	√	√	√	√	√	√	√	√	√	√	12
K-CAS	Pars.	22	392	0.791	0.791	72	84	100	96	NF	NF	100	100	72	73	99	92	99	11
	NJ					<50	59	94	94	NF	NF	100	100	89	NF	100	94	99	8
	ML					√	√	√	√	×	×	√	√	√	√	√	√	√	11
TH	Pars.	10	466	0.751	0.790	NF	96	100	100	75	51	100	100	83	100	75	84	97	11
	NJ					54	89	100	100	NF	NF	99	100	NF	96	84	88	100	9
	ML					×	√	√	√	√	√	√	√	√	√	√	√	√	12
Indels	Pars.	16	37	0.919	0.972	85	NF	71	93	NF	NF	97	100	67	96	99	75	NF	8
% of times nodes were recovered						52	72	100	100	44	56	100	100	80	96	100	100	96	

TABLE 7. Incongruent length differences (ILDs) for the pairwise comparisons among the fragments included in this study. The values above the diagonal represent the number of extra steps introduced by combining the partitions (Farris et al., 1995; Baker and DeSalle, 1997); the significance values (P) are given below the diagonal.

Genes	*MGF*	*TG*	*PRKCI*	*SPTBN1*	*STAT5*	*TNF-A*	*K-CAS*	*TH*	*Cyt-b*
MGF	—	7	5	4	3	5	5	5	31
TG	0.32	—	4	6	5	7	4	6	29
PRKCI	0.17	0.80	—	3	5	2	2	3	23
SPTBN1	0.40	0.63	0.41	—	2	3	3	4	34
STAT5	0.80	0.91	0.46	0.83	—	4	1	3	22
TNF-A	0.12	0.38	0.73	0.71	0.75	—	4	1	25
K-CAS	0.58	0.80	0.87	0.56	0.94	0.41	—	4	23
TH	0.25	0.80	0.73	0.49	0.72	0.96	0.73	—	22
Cyt-b	0.02*	0.02*	0.02*	0.01*	0.16	0.03*	0.35	0.09	—

*,ILD is statistically significant ($P < 0.05$).

analyses were less successful in recovering the phylogeny in Figure 2 from the reduced data sets and required at least four times as many characters (in this case 2,666). However, the same three nodes that collapsed during the parsimony analyses (A, E, and F) were always involved when the trees were not congruent.

The nuclear DNA data were extremely insensitive to taxon sampling (Hillis, 1996, 1998; Kim, 1996; Graybeal, 1998;). A impressive 99% of the random 10-taxa parsimony trees reflected the same evolutionary relationships that were found by the complete data set. The only tree that differed from this finding was obtained from a data set that included the horse, peccary, camel, bowhead whale, hippopotamus, chevrotain, pronghorn, muntjac, giraffe, and lesser kudu. In this instance, the giraffe (family Giraffidae) and muntjac (family Cervidae) clustered as sister taxa to the exclusion of the lesser kudu (family Bovidae). The neighbor joining analyses were more sensitive to taxon sampling and recovered the suggested topology in only 35% of the random-taxa data sets. All of the ambiguities were similarly restricted to the short internal nodes within the Ruminantia (D, E, and F; Fig. 2).

The missing data simulations performed in the present study clearly indicated that phylogenetic resolution decreases as the number of missing characters (or the number of taxa with missing data) increases (Fig. 5a,b). Once again there was a marked distinction between parsimony and neighbor joining methods, with the parsimony analyses being less sensitive than the distance analyses to missing data (Fig. 5a,b). For example, when 12 of 26 taxa had >36% (2,400 bp) of the data missing, 30% of the nodes were affected in parsimony analysis (Fig. 5a). In sharp contrast, when only three or more taxa had 27% (1,800 bp) missing data, >30% of the nodes recovered by neighbor joining were affected (Fig. 5b). Moreover (data not shown), relationships defined by short branches were the ones most influenced by missing data. For example, the placement of the waterbuck, the relationship between sheep, goat, and muskox and the placement of the gazelle and dik-dik clade were all particularly sensitive to missing data (Fig. 3). For neighbor joining, a few taxa with a large number of missing characters show a more noticeable effect than a large number

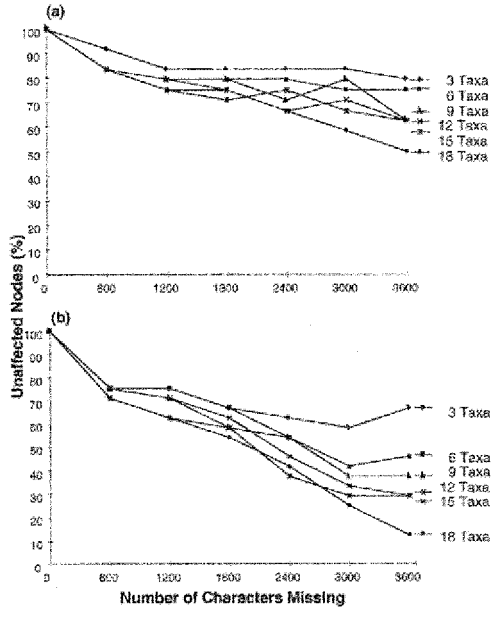

FIGURE 5. Percentage of nodes unaffected by missing data for parsimony (a) and neighbor joining (b) analyses. The number of randomly excluded taxa for each simulation is indicated at right (see text for details).

of taxa in which each has a few characters missing. For example, in three taxa missing 1,800 characters, 33.3% of the nodes are affected, whereas in nine taxa with 600 characters missing, only 25% of the nodes are affected (Fig. 5b). The parsimony analyses do not show this relationship, and the number of nodes affected was remarkably similar between comparisons; for example, three taxa having 1,800 characters missing and nine taxa with 600 characters missing both have 16.7% of their nodes affected (Fig. 5a).

Nuclear Versus Mitochondrial DNA Data

The nuclear exonic regions were, as expected, characterized by fewer substitutions than introns, exhibiting 27.94% and 44.65% variable positions, respectively. Despite this difference, both the nuclear DNA intron and exon regions were consistent in recovering the same topology (Table 8). Both parsimony and neighbor joining analyses recovered at least 10 of the 13 nodes in each instance, and bootstrap support ranged between 52% and 100% (Table 8). Both introns and exons were characterized by low homoplasy indices (<35%) and both performed equally well-in recovering a well-supported phylogeny for the order Artiodactyla (Table 8).

The mtDNA cytochrome *b* bootstrap phylogeny was poorly resolved, and several associations were found that conflicted with both the nuclear DNA data and the morphological evidence (Fig. 6). Five of the eight nuclear DNA genes were significantly incongruent with the mtDNA data (Table 7), reflecting differences in the evolution of cytoplasmic and nuclear DNA. Unfortunately, attempts to reduce the homoplasy in the cytochrome *b* data through differential weighting of TIs and TVs, or translating the gene into amino acids, did not improve the resolution or consistency among methods (Table 8). We therefore conclude that the cytochrome *b* gene failed to recover a robust phylogeny for the members of the order Artiodactyla. The retention index value for the parsimony topology was only 0.316—in sharp contrast to the high retention index values from the different nuclear DNA topologies, which ranged from 0.700 to 0.791 for sequence data and 0.972 for indels (Table 6). Only the maximum likelihood analysis of the cytochrome *b* data recovered most of the well-established groups (9 of 13 nodes). However, support for the nodes was generally low; moreover, the topology differed from those retrieved from the nuclear DNA and previous morphological analyses by placing the Giraffidae as a sister taxon to the Bovidae, with the Cervidae being basal among pecorans.

Given the inability of the complete cytochrome *b* data set to recover the artiodactyl phylogeny accurately and with high statistical support, we are not surprised that the mtDNA data were also severely affected by taxon sampling (also see Philippe and Douzery, 1994; Milinkovitch et al., 1996). Only 11% of the neighbor joining trees produced by 10 exemplar taxa were consistent with the mtDNA neighbor joining tree obtained when all 27 taxa were included. The parsimony analyses were even more sensitive to the random sampling of taxa, with only 6% of the subsample trees agreeing with the parsimony tree obtained for the complete data set.

DISCUSSION

Mitochondrial Versus Nuclear DNA Data

The nuclear DNA sequence data displayed several advantageous features for phylogeny reconstruction at the familial/subordinal level in Artiodactyla. In addition to the low amount of among-site rate heterogeneity for the intron regions (Table 4), and the absence of meaningful bias in base composition, plotting the total number of substitutions against the sequence divergence values gave no indication of saturation (Fig. 1a). These factors suggest a low level of homoplasy in the data, which in turn is reflected by the high retention index values for the nuclear parsimony analyses and high bootstrap support for most nodes in all of the analyses (Fig. 4; Table 6).

An interesting finding that emerges from this study is that despite using a greater number of characters in the intron (5,244 bp) than in the exon analyses (1,422 bp), both intron and exon sequences performed equally well in recovering the artiodactyl phylogeny. We deem it reasonable, therefore, to argue that both coding and noncoding nuclear DNA segments have an equal ability to retain phylogenetic signal over the evolutionary timeframe of this study. Furthermore, the strength of the phylogenetic signal in the nuclear data was such that the random exclusion of taxa

TABLE 8. Summary of the topologies obtained by the intron, exon, and cytochrome *b* analyses in the present study. The node letters (A–M) correspond to those in Figure 2 and the numerical values indicate the percentage of times each node was recovered with >70% bootstrap support by parsimony or neighbor joining analyses. Abbreviations as in Table 6. The differential weighting method for the cytochrome *b* (Cyt-b) data set is also shown (see text for details). Only the presence (√) or absence (×) of nodes for the maximum likelihood solution is indicated.

Data set	Method	No. of trees	Steps	CI	RI	Nodes													% with >70% bootstrap
						A	B	C	D	E	F	G	H	I	J	K	L	M	
Intron	Pars.	1	3981	0.747	0.753	63	100	100	100	88	98	100	100	100	100	100	100	100	12
	NJ					NF	100	100	100	52	74	100	100	100	100	100	100	100	11
	ML					√	√	√	√	×	√	√	√	√	√	√	√	√	12
Exon	Pars.	7	716	0.671	0.720	65	92	100	100	57	NF	100	100	88	94	100	93	98	10
	NJ					55	76	100	100	NF	NF	100	100	83	91	100	90	100	10
	ML					√	√	√	√	√	×	√	√	√	√	√	√	√	12
Cyt-B	Pars. (1:1)	1	2820	0.338	0.316	NF	NF	57	68	62	NF	<50	100	56	100	96	NF	79	4
	Pars. (1:2.6)					NF	NF	74	58	NF	NF	<50	99	53	100	96	NF	81	4
	Pars. (1:0)					NF	NF	90	80	NF	NF	<50	100	67	100	100	NF	94	5
	Pars. (protein)					NF	NF	NF	NF	NF	NF	<50	98	NF	100	69	<50	NF	2
	NJ					NF	NF	72	91	<50	NF	61	100	<50	100	94	NF	83	6
	ML					×	×	√	√	×	×	√	√	√	√	√	√	√	9

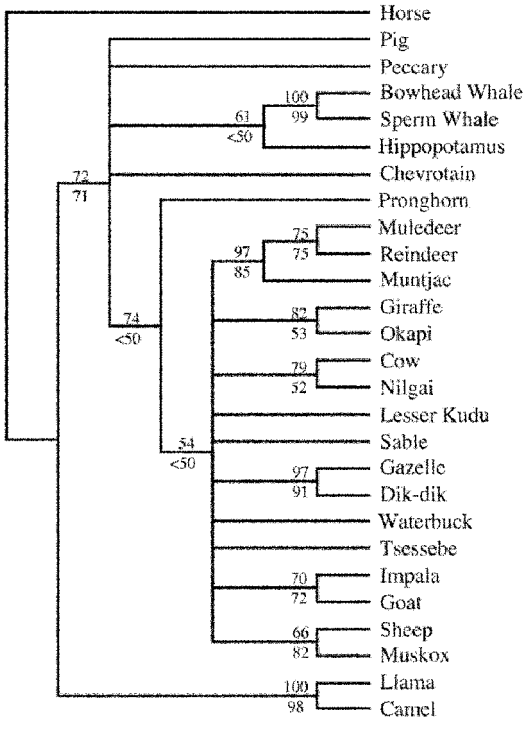

FIGURE 6. Parsimony bootstrap consensus tree based on the mtDNA cytochrome *b* data in this study. The numbers above and below the branches represent bootstrap values obtained by using parsimony and neighbor joining, respectively. Nodes with <50% support in both analyses are collapsed.

did not influence the nuclear DNA results obtained with parsimony anlysis (Hillis, 1996, 1998; Kim, 1996; Graybeal, 1998). Finally, jacknifing results suggesting that <1,000 bp of data are needed for phylogenetic inference by parsimony analysis is promising for future evolutionary studies. This number is in good agreement with previous studies such as that of Gatesy et al. (1996), in which ~900 characters from a single nuclear DNA gene were used to produce an artiodactyl phylogeny essentially identical to the one we obtained here.

In sharp contrast to the nuclear DNA data, the mtDNA cytochrome *b* gene had almost twice the proportion of variable characters seen in the nuclear DNA exons (1,422 bp exon = 28% variable; 1,140 bp cytochrome *b* = 51% variable; Table 4). However, this difference did not reflect phylogenetic utility. The mtDNA data were plagued by homoplasy (Milinkovitch et al., 1996) with clear evidence of saturation (Fig. 1b), a low retention index for the most-parsimonious tree (0.316), and extreme sensitivity to taxon sampling. The saturation present in the mtDNA data adversely affected the retention of phylogenetic signal over time and thus the mtDNA topologies were either unresolved or resolved in an arbitrary manner. Moreover, most nodes were not supported by bootstrap analyses (for more evidence, see Kraus and Miyamoto, 1991; Douzery and Catzeflis, 1995; Gatesy et al., 1996; Matthee and Robinson, 1996; Milinkovitch et al., 1996; Gatesy, 1998).

The retrieval of phylogenetic information from mtDNA sequence data has been widely considered as being confounded by rapid radiations within mammalian groups. Examples abound for the Artiodactyla (Allard et al., 1992; Matthee and Robinson, 1999a; Rebholz and Harley, 1999), the Rodentia (Lara et al., 1996; Matthee and Robinson, 1997), the Carnivora (Flynn and Nedbal, 1998), and the Lagomorpha (Halanych and Robinson, 1999), and uncertainties persisted even when complex weighting schemes were used in an effort to reduce homoplasy/noise (Table 8). Two groups that formed part of the focus of this study, the pecorans and the species-rich family Bovidae, are known to have experienced rapid speciation (Vrba, 1985; Allard et al., 1992; Cronin et al., 1996), and relationships within these groups remain problematic. However, most portions of the nuclear gene phylogeny were well-supported (Fig. 2; Table 6), suggesting that the failure to produce a well-supported phylogeny is actually the result of the inability of mtDNA sequence data to resolve taxonomic questions at higher levels rather than the result of rapid radiations.

Missing Data and the Artiodactyl Phylogeny

The missing data simulations performed in this study provide insight into the effect of database mining on phylogeny reconstruction. Combined nuclear DNA analyses often rely on data derived from the Entrez database. As a result, usually several taxa have missing data because not all taxa have been sequenced for completely overlapping sequence fragments. For example, in the recent artiodactyl analysis by Gatesy et al. (1999a) one of the matrixes (whippo-1) had

~25% missing data (roughly 28,000 of the 117,000 characters) and the other matrix (whippo-2) had at least 55% missing data (roughly 355,000 of the 642,000 characters; both data sets are available at http://www.utexas.edu/ftp/depts/systbiol/48_1/vol48_1.html). Our analyses of missing data made it evident there was a decline in phylogenetic accuracy as the amount of missing data increased. The effect was more apparent for the neighbor joining analyses, where inconsistent results were obtained for nearly all nodes (87.5%) when 37% (64,800 of the 174,000 characters) of the data were missing. Under the same scenario, parsimony analyses were inconsistent at only 50% of the nodes. Importantly, the nodes were more likely to be unstable when internal branches were short, a scenario probably explaining the weak molecular resolution previously obtained for the pecoran lineages (Gatesy et al., 1999a). In contrast, the nodes defined by the long branches are remarkably insensitive to missing data and replacements of taxa and characters, and it is these relationships that are congruent between the present and the previous artiodactyl study (Gatesy et al., 1999a). Although the parsimony analyses did not reflect any clear trend, the neighbor joining results obtained herein support Wiens' (1998) suggestion that phylogenetic accuracy is increased by adding fewer characters scored for all taxa than by adding a larger number of characters for a subset of taxa.

More Characters or More Taxa?

In a recent chloroplast DNA investigation of the Rubiaceae, Bremer et al. (1999) concluded it is better to add more characters than more taxa if the purpose is to obtain a strongly supported phylogeny for a problematic data set. The artiodactyl DNA study presented here does not fully support this contention and points to a more complicated situation (also see Graybeal, 1998). Although the well-supported nuclear DNA artiodactyl phylogeny was not sensitive to taxon sampling, and thus in broad terms supports the conclusions reached by Bremer et al. (1999), the weakly supported cytochrome *b* phylogeny was severely affected by the number and identity of taxa included. Increasing the number of taxa improved resolution in the latter instance, but increasing the number of noninformative characters derived from the mtDNA molecule did not provide significantly better resolution for many mammalian uncertainties (Gatesy, 1998; Waddell et al., 1999). We think it reasonable that more characters will benefit only analyses in which the phylogenetic signal is present but weak.

Artiodactyla Systematics

The combined nuclear DNA data presented here excelled in recovering the relationships within the problematic suborder Ruminantia. Apart from the basal placement of the chevrotain, which is well-established on morphological (Eisenberg, 1981; Nowak, 1999) and molecular grounds (Douzery and Catzeflis, 1995; Cronin et al., 1996; Gatesy et al., 1999a), our data suggest a sister taxon relationship between the families Bovidae and Cervidae, with the Giraffidae being more basal. These findings are consistent with some morphological (Janis and Scott, 1987) and previous genetic data (Schreiber et al., 1990; Gatesy et al., 1999a). Moreover, the nuclear DNA and some of the mtDNA data suggested that the pronghorn is a primitive member of the group and is not closely related to any of the other Pecoran families. The uniqueness of the pronghorn is in accord with previous molecular suggestions (Cronin et al., 1996; Gatesy et al., 1996) and supports the recognition of Antilocapridae as a valid family within the suborder Ruminantia (also see Nowak, 1999). If this finding holds, it will imply that the morphological characters that have been used in the past to include the pronghorn within the Cervidae or Bovidae (Leinders and Heintz, 1980) are problematic for phylogenetic inference in this group.

Analyses of morphological data derived from extinct and extant artiodactyl and cetacean taxa (O'Leary and Geisler, 1999), and a combined assessment of morphological and molecular evidence (Luckett and Hong, 1998), suggest that the order Artiodactyla is monophyletic and that members of the order Cetacea do not convincingly cluster as a sister taxon to the hippopotamids. In sharp contrast to these studies, our results, based on both mitochondrial and nuclear DNA data, add credence to the strong previous molecular evidence that the order Artiodactyla is paraphyletic and the suborder Suiformes is not monophyletic (Graur

and Higgins, 1994; Irwin and Arnason, 1994; Queralt et al., 1995; Gatesy et al., 1996, 1999a, 1999b; Shimamura et al., 1997; Gatesy, 1998; Stanhope et al., 1998; Ursing and Arnason, 1998). The lineages with the most primitive two-chambered stomach, the Suidae (pigs) and Tayassuidae (peccary), were placed basal in the nuclear phylogeny. The Hippopotamidae, the members of which possess a three-chambered stomach (Nowak, 1999), has historically grouped with these taxa in the suborder Suiformes but is more derived in the molecular phylogenies. The hippopotamus invariably formed part of an aquatic assemblage with the cetaceans and, although several anatomical features could be used to unite these taxa (a distinctive mechanism of penile erection, three primary lung bronchi, lack of scrotal testes, and lost sebaceous glands; see Gatesy, 1998, and references therein), not all of these features are uniquely derived (Luckett and Hong, 1998). Analyses of 17 data sets indicated that the inclusion of the Cetacea within the Artiodactyla was supported by >100 synapomorphic characters (Gatesy et al., 1999a), an association corroborated by uniquely inserted short (SINE) and long (LINE) interspersed elements (Nikaido et al., 1999). The present study strengthens these findings and adds an additional 118 unique derived characters to support the inclusion of the Cetacea within the Artiodactyla. In total, 235 molecular synapomorphic characters drawn from >20 independent loci throughout the genome now support the inclusion of the Cetacea within the Artiodactyla.

The node defining the placement of the suids and camelids within the Artiodactyla is more problematic. An immunogenic study based on comparative determinant analysis (Schreiber et al., 1990), combined with a three-chambered digestive system, high crowned molars, and a highly specialized rumination process (Nowak, 1999; also see Gentry and Hooker, 1988, for additional morphological evidence), suggests that the Tylopoda is a derived lineage within the Artiodactyla. This placement is strongly supported by recent morphological analyses of extinct and extant artiodactyls (O'Leary and Geisler, 1999). In the present study all three methods of phylogenetic inference as well as the unique indel analysis suggest that the Suiformes (excluding the hippopotamus) are the most primitive artiodactyl group, and the Tylopoda is more derived (bootstrap support >72% in the combined analyses). The Kishino–Hasegawa test found no difference between these alternative hypotheses. Closer investigation of the nuclear DNA data revealed that six of the eight gene fragments place the Suiformes basal in the phylogeny, whereas the remaining two gene fragments put the Tylopoda in this position (Fig. 4). In concert, these findings contradict the suggestion of Gatesy et al. (1999a) that molecular data predicate a basal placement of the Tylopoda. Although the more-derived position of the Suiformes (excluding the hippopotamus) is also corroborated by a single SINE element insertion (Nikaido et al., 1999), it is important to realize that both the data sets supporting the basal position of the Tylapoda were characterized by extensive missing data, which can potentially lead to false conclusions (Maddison, 1991; Hillis, 1999). Incomplete lineage sorting will also obscure SINE/LINE data, particularly at this level of resolution (Hillis, 1999; Nikaido et al., 1999). On the other hand, the outcome of the present study (placing the Suiformes basal) is certainly weakened by the use of a single outgroup. At this stage, these factors preclude any definitive statement on the phylogenetic position of the Tylapoda and Suidae.

Apart from this single node, the combined topology presented here and that of Gatesy and colleagues' Figure 7 (Gatesy et al., 1999a) were in excellent agreement. This result is important from a systematic viewpoint. The data in our study were drawn from several different loci and differ substantially from those of Gatesy et al. (1999a) in the form of taxon sampling and the amount of missing data. The combined analyses of Gatesy et al. (1999a) included 79 taxa, multiple outgroups, and unfortunately a large amount of missing data. On the other side, the present study included 27 taxa and limited missing data but was weakened by the use of only one outgroup. Despite these differences, the strong congruence among the studies is heartening. Given the outcome of this comparison, we believe that increased taxon sampling will not affect the phylogenetic placement of the Cetacea within the Artiodactyla (for a contrasting view, see Luckett and Hong, 1998) and conclude that the paraphyly of the Artiodactyla is a real phenomenon (Gatesy et al., 1999a).

Molecules Versus Morphology and the Artiodactyl Phylogeny

In broad terms the molecular data are in remarkable agreement with previous artiodactyl morphological classification schemes based on complex structures such as dentition, multichambered stomachs, and horns/antlers/ossicones (Nowak, 1999). They do, however, disagree with results from two uniquely derived morphological character complexes of the astragalus and the deciduous dentition, which are found only in Artiodactyla (Luckett and Hong, 1998). Complicated processes such as the presence or absence of rumination have long been used for the classification of mammals. Unfortunately, the molecular data confirm that these complex morphological and physiological adaptations are not without homoplasy and can certainly be lost (and to a lesser extent gained) in parallel through evolutionary time. Regardless of the exact evolutionary position of the suids, the group as a whole is characterized by a two-chambered stomach, whereas the genus *Babyrousa* is characterized by a three-chambered stomach (Davis, 1940). Along the same lines, the hippopotamus and Cetacea are more derived than the Tylopoda but are not characterized by rumination. Even complex horn/antler/ossicone structures, which are in most instances evolutionarily informative (monophyly of the Pecora, Bovidae, and Giraffidae), are not fixed within entire evolutionary lineages. Both musk deer (*Moschus*) and Chinese water deer (*Hydropotes*) lack antlers and have enlarged canines similar to the more primitive chevrotain (Nowak, 1999), making this character homoplastic for the Cervidae.

With all these data at hand, clearly even the most complex structures and unlikely events (for example, sequence indels) cannot be used unquestioningly to explain evolutionary relationships among taxa. Although the wealth of nuclear DNA sequences and primers emerging from genomic efforts offer an unprecedented opportunity to collect parallel informative molecular data sets from independent nuclear loci in different taxa, no single data set can provide definitive answers to all the evolutionary and systematic uncertainties.

ACKNOWLEDGMENTS

We thank the following people and agencies who provided tissue/DNA: Alexander Rooney, Dan Gallagher Jr., Eddie Louis, and the staff of CRES at the San Diego Zoo (Oliver Ryder and Arlene Kumamoto). Dan Gaile assisted with simulations of missing data. Clare Gill, Terry Robinson, and Sara Davis provided helpful editorial comments, and Patrick Luckett and Masami Hasegawa are acknowledged for constructive criticisms of earlier drafts of this manuscript. We thank John Gatesy, Scott Edwards, and one anonymous reviewer for their comments. The National Research Foundation (South Africa) and the Skye Foundation and Charitable Trust provided financial support to C.A.M.

REFERENCES

ALLARD, M. W., M. M. MIYAMOTO, L. JARECKI, F. KRAUS, AND M. R. TENNANT. 1992. DNA systematics and the evolution of the artiodactyl family Bovidae. Proc. Natl. Acad. Sci. USA 89:3972–3976.

ANDERSON, S., M. H. DE BRUIJN, A. R. COULSON, I. C. EPERON, F. SANGER, AND I. G. YOUNG. 1982. Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J. Mol. Biol. 156:683–717.

ARNASON, U., AND A. GULLBERG. 1994. Relationship of baleen whales established by cytochrome *b* gene sequence comparison. Nature 367:726–728.

BACCUS, R., N. RYMAN, M. H. SMITH, C. REUTERALL, AND D. CAMERON. 1983. Genetic variability and differentiation of large grazing mammals. J. Mammal 64:109–121.

BAKER, R. H., AND R. DESALLE. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46:654–673.

BREMER, B., R. K. JANSEN, B. OXELMAN, M. BACKLUND, H. LANTZ, AND K. KIM. 1999. More characters or more taxa for a robust phylogeny—case study from the coffee family (Rubiaceae). Syst. Biol. 48:413–435.

BROWN, W. M., E. M. PRAGER, A. WANG, AND A. C. WILSON. 1982. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. J. Mol. Evol. 18:225–239.

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. A., R. STUART, B. J. PIERSON, AND J. C. PATTON. 1996. κ-Casein gene phylogeny of higher ruminants (Pecora, Artiodactyla). Mol. Phylogenet. Evol. 6:295–311.

CURTAIN, C. C., AND H. H. FUDENBERG. 1973. Evolution of the immunoglobulin antigens in the Ruminantia. J. Biochem Genet. 8:301–308.

DAVIS, D. D. 1940. Notes on the anatomy of the babirusa. Field Mus. Nat. Hist. Zool. 22:363-411.

DE BLASE, A. F., AND R. E. MARTIN. 1974. Artiodactyla. Pages 235–243 *in* A manual of mammalogy. WCB Publishers, Dubuque, Iowa.

DOUZERY, E., AND F. M. CATZEFLIS. 1995. Molecular evolution of the mitochondrial 12S rRNA in Ungulata (Mammalia). J. Mol. Evol. 41:622–636.

EDWARDS, S. V., AND A. C. WILSON. 1990. Phylogenetically informative length polymorphism and sequence variability in mitochondrial DNA of Australian songbirds (*Pomatostomus*). Genetics 126:695–711.

EISENBERG, J. F. 1981. The grandorder Ungulata. Pages 171–209 *in* The mammalian radiations: An analysis of trends in evolution, adaptation, and behavior. Univ. of Chicago Press, Chicago.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1995. Constructing a significance test for incongruence. Syst. zool. 44:570–572.

FELSENSTEIN, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Biol. 27:401–410.

FLYNN, J. J., AND M. A. NEDBAL. 1998. Phylogeny of the Carnivora (Mammalia): Congruence vs incompatibility among multiple data sets. Mol. Phylogenet. Evol. 9:414–426.

GALLAGHER, D. S. JR., J. N. DERR, AND J. E. WOMACK. 1994. Chromosome conservation among the advanced pecorans and determination of primitive bovid karyotype. J. Hered. 85:204–210.

GATESY, J., C. HAYASHI, M. A. CRONIN, AND P. ARCTANDER. 1996. Evidence from milk casein genes that cetaceans are close relatives of hippopotamid artiodactyls. Mol. Biol. Evol. 13:954–963.

GATESY, J., M. MILINKOVITCH, V. WADDELL, AND M. STANHOPE. 1999a. Stability of cladistic relationships between Cetacea and the higher-level artiodactyl taxa. Syst. Biol. 48:6–20.

GATESY, J., P. O'GRADY, AND R. H. BAKER. 1999b. Corroboration among data sets in simultaneous analysis: Hidden support for phylogenetic relationships among higher level artiodactyl taxa. Cladistics 15:271–313.

GATESY, J., D. YELON, R. DE SALLE, AND E. S. VRBA. 1992. Phylogeny of the Bovidae (Artiodactyla, Mammalia), based on mitochondrial ribosomal DNA sequences. Mol. Biol Evol. 9:433–446.

GATESY, J. 1998. Molecular evidence for the phylogenetic affinities of Cetacea. Pages 63–112 *in* The emergence of whales (Thewissen, ed.). Plenum Press, New York.

GENTRY, A. W., AND J. J. HOOKER. 1988. The phylogeny of the Artiodactyla. Pages 235–272 in The phylogeny and classification of the Tetrapods, volume 2, (M. Benton, ed.). Clarendon Press, Oxford.

GIRIBET, G., AND W. C. WHEELER. 1999. On gaps. Mol. Phylogenet. Evol. 13:132–143.

GRAUR, D., AND D. G. HIGGINS. 1994. Molecular evidence for the inclusion of cetaceans within the order Artiodactyla. Mol. Biol. Evol. 11:357–364.

GRAYBEAL, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47:9–17.

GRIFFITHS, C. S. 1997. Correlation of functional domains and rates of nucleotide substitution in cytochrome *b*. Mol. Phylogenet. Evol. 7:352–365.

GROVES, P., AND G. F. SHIELDS. 1997. Cytochrome *b* sequences suggest convergent evolution of the Asian takin and Arctic muskox. Mol. Phylogenet. Evol. 8:363–374.

HALANYCH, K. M., J. R. DEMBOSKI, B. JANSEN VAN VUUREN, D. R. KLEIN, AND J. A. COOK. 1999. Cytochrome *b* phylogeny of North American hares and jackrabbits (*Lepus*, Lagomorpha) and the effects of saturation in ingroup taxa. Mol. Phylogenet. Evol. 11:213–221.

HALANYCH, K. M., AND T. J. ROBINSON. 1999. Multiple substitutions affect the phylogenetic utility of cytochrome *b* and 12S rDNA data: Examining a rapid radiation in leporid (Lagomorpha) evolution. J. Mol. Evol 48:369–379.

HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 21:160–174.

HASSANIN, A., AND E. J. P. DOUZERY. 1999. The tribal radiation of the family Bovidae (Artiodactyla) and the evolution of the mitochondrial cytochrome *b* gene. Mol. Phylogenet. Evol. 13:227–243.

HILLIS, D. M. 1996. Inferring complex phylogenies. Nature 383:130–131.

HILLIS, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47:3–8.

HILLIS, D. M. 1999. SINEs of the perfect character. Proc. Natl. Acad. Sci. USA 96:9979–9981.

HILLIS, D. M., B. K. MABLE, AND C. MORITZ. 1996. Applications of molecular systematics. Pages 515–544 *in* Molecular systematics (D. M. Hillis, B. K. Mable, and C. Moritz, eds.). Sinauer Associates, Sunderland, Massachusetts.

HONEYCUTT, R. L., M. A. NEDBAL, R. M. ADKINS, AND L. 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.

HUELSENBECK, J. C. 1997. Is the Felsenstein zone a fly trap? Syst. Biol. 46:69–74.

IRWIN, D. M., AND U. ARNASON. 1994. Cytochrome *b* gene of marine mammals: Phylogeny and evolution. J. Mammal. Evol. 2:37–55.

IRWIN, D. M., T. D. KOCHER, AND A. C. WILSON. 1991. Evolution of the cytochrome *b* gene of mammals. J. Mol. Evol. 32:128–144.

JANIS, C. M., AND K. M. SCOTT. 1987. The interrelationships of higher ruminant families with special emphasis on the members of the Cervoidea. Am. Mus. Novit. 2893:1–85.

JERMANN, T. M., J. G. OPITZ, J. STACKHOUSE, AND S. A. BENNER. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. Nature 374:57–59.

KIM, J. 1996. General inconsistency conditions for maximum parsimony: Effects of branch lengths and increasing numbers of taxa. Syst. Biol. 45:363–374.

KISHINO, H., T. MIYATA, AND M. HASEGAWA. 1990. Maximum-likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. 31:151–160.

KISHINO, H. T., AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29:170–179.

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

LARA, M. C., J. L. PATTON, AND M. N. F. DA SILVA. 1996. The simultaneous diversification of South American echimyid rodents (Hystricognathi) based on complete cytochrome *b* sequences. Mol. Phylogenet. Evol. 5:403–413.

LEINDERS, J. J. M., AND E. HEINTZ. 1980. The configuration of the lacrimal orifices in pecorans and tragulids (Artiodactyla, Mammalia) and its significance for the distinction between Bovidae and Cervidae. Beaufortia 30:155–161.

LUCKETT, W. P., AND N. HONG. 1998. Phylogenetic relationships between the orders Artiodactyla and Cetacea: A combined assessment of morphological and molecular evidence. J. Mammal. Evol. 5:127–182.

LYONS, L. A., T. F. LAUGHLIN, N. G. COPELAND, N. A. JENKINS, J. E. WOMACK, AND S. J. O'BRIEN. 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nat. Genet. 15:47–56.

MADDISON, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. Syst. Zool. 40:315–328.

MATTHEE, C. A., AND T. J. ROBINSON. 1997. Molecular phylogeny of the springhare, *Pedetes capensis*, based on mitochondrial DNA sequences. Mol. Biol. Evol. 14:20–29.

MATTHEE, C. A., AND T. J. ROBINSON. 1999a. Cytochrome *b* phylogeny of the family Bovidae: resolution within the Alcelaphini, Antilopini, Neotragini, and Tragelaphini. Mol. Phylogenet. Evol. 12:31–46.

MATTHEE, C. A., AND T. J. ROBINSON. 1999b. Mitochondrial DNA population structure of roan and sable antelope: Implications for the translocation and conservation of the species. Mol. Ecol. 8:227–238.

MILINKOVITCH, M. C., R. G. LEDUC, J. ADACHI, F. FARNIR, M. GEORGES, AND M. HASEGAWA. 1996. Effects of character weighting and species sampling on phylogeny reconstruction: A case study based on DNA sequence data in cetaceans. Genetics 144:1817–1833.

MONTGELARD, C., F. M. CATZEFLIS, AND E. DOUZERY. 1997. Phylogenetic relationships of artiodactyls and cetaceans as deduced from the comparison of cytochrome *b* and 12S rRNA mitochondrial sequences. Mol. Biol. Evol. 14:550–559.

NEDBAL, M. A., M. W. ALLARD, AND R. L. HONEYCUTT. 1994. Molecular systematics of the Hystricognath rodents: Evidence from the mitochondrial 12S rRNA gene. Mol. Phylogenet. Evol. 3:206–220.

NIKAIDO, M., A. P. ROONEY, AND N. OKADA. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: Hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. USA 96:10261–10266.

NOWAK, R. M. 1999. Order Artiodactyla. Pages 1051–1238 *in* Walker's mammals of the world, volume 2. John Hopkins Univ. Press, London.

O'GARA, B. W., AND G. MATSON. 1975. Growth and casting of horns by pronghorns and exfoliation of horns by bovids. J. Mammal. 56:829–846.

O'LEARY, M. A., AND J. H. GEISLER. 1999. The position of Cetecae within Mammalia: Phylogenetic analysis of morphological data from extinct and extant taxa. Syst. Biol. 48:455–490.

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

QUERALT, R., R. ADROER, R. OLIVA, R. WINKFEIN, J. RETIEF, AND G. DIXON. 1995. Evolution of the protamine P1 genes in mammals. J. Mol. Evol. 40:601–607.

RANDI, E., N. MUCCI, M. PIERPAOLI, AND E. DOUZERY. 1998. New phylogenetic perspectives on the Cervidae (Artiodactyla) are provided by the mitochondrial cytochrome *b* gene. Proc. R. Soc. Lond. B 265:793–801.

REBHOLZ, W., AND E. HARLEY. 1999. Phylogenetic relationships in the bovid subfamily Antilopinae based on mitochondrial DNA sequences. Mol. Phylogenet. Evol. 12:87–94.

ROBINSON, T. J., W. R. HARRISON, F. A. PONCE DE LEÓN, S. K. DAVIS, AND F. F. B. ELDER. 1998. A molecular cytogenetic analysis of *X* chromosome repatterning in the Bovidae: Transpositions, inversions, and phylogenetic inference. Cytogenet. Cell Genet. 80:179–184.

SCHREIBER, A., D. ERKER, AND K. BAUER. 1990. Artiodactylan phylogeny: An immunogenetic study based on comparative determinant analysis. Exp. Clin. Immunogenet. 7:234–243.

SHIMAMURA, M., H. YASUE, K. OHSHIMA, H. ABE, H. KATO, T. KISHIRO, M. GOTO I, MUNECHIKA, AND N. OKADA. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388:666–669.

SOLOUNIAS, N. 1988. Evidence from horn morphology on the phylogenetic relationships of the pronghorn (*Antilocapra americana*). J. Mammal. 69:140–143.

SPRINGER, M. S., H. M. AMRINE, A. BURK, AND M. J. STANHOPE. 1999. Additional support for Afrotheria and Paenungulata, the performance of mitochondrial versus nuclear genes, and the impact of data partitions with hetergeneous base composition. Syst. Biol. 48:65–75.

STANHOPE, M. J., O. MADSEN, V. G. WADDELL, G. C. CLEVEN, W. W. DE JONG, AND M. S. SPRINGER. 1998. Highly congruent molecular support for a diverse superordinal clade of endemic African mammals. Mol. Phylogenet. Evol. 9:501–508.

STANLEY, H. F., M. KADWELL, AND J. C. WHEELER. 1994. Molecular evolution of the family Camelidae: A mitochondrial DNA study. Proc. R. Soc. Lond. B. 256:1–6.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, AND D. M. HILLIS. 1996. Phylogenetic inference. Pages 407–514 *in* Molecular systematics, 2nd edition (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer, Sunderland, Massachusetts.

TAJIMA, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135:599–607.

THEIMER, T. C., AND P. KEIM. 1998. Phylogenetic relationships of peccaries based on mitochondrial cytochrome *b* DNA sequences. J. Mammal. 79:566–572.

URSING, B. M., AND U. ARNASON. 1998. Analyses of mitochondrial genomes strongly support a hippopotamus–whale clade. Proc. R. Soc. Lond. B 265:2251–2255.

VENTA, P. J., J. A. BROUILLETTE, V. YUZBASIYAN-GURKAN, AND G. J. BREWER. 1996. Gene-specific universal mammalian sequence-tagged sites: Application to the canine genome. Biochem. Genet. 34:321–341.

VRBA, E. 1985. African Bovidae: Evolutionary events since the Miocene. S. A. J. Sci. 81:263–266.

WADDELL, P. J., N. OKADA, AND M. HASEGAWA. 1999. Towards resolving the interordinal relationships of the placental mammals. Syst. Biol. 48:1–5.

WIENS, J. J. 1998. Does adding characters with missing data increase or decrease phylogenetic accuracy? Syst. Biol. 47:625–640.

Received 14 January 2000; accepted 14 May 2000
Associate Editor: S. Edwards