Phylogenetic Evidence from the IRBP Gene for the Paraphyly of Toothed Whales, with Mixed Support for Cetacea as a Suborder of Artiodactyla

Marta R. Smith,* Mahmood S. Shivji,* Victor G. Waddell,† and Michael J. Stanhope†

*Oceanographic Center, Nova Southeastern University; and †Biology and Biochemistry, Queen's University, Belfast

Introduction

Recent molecular phylogenetic analyses of Cetacea and of their affinities with Artiodactyla have suggested two surprising possibilities: (1) sperm whales are more closely related with baleen whales (suborder Mysticeti) than with the other toothed whales (suborder Odontoceti; Milinkovitch, Orti, and Meyer 1993; Milinkovitch, Meyer, and Powell 1994) and (2) Cetacea may represent a suborder of Artiodactyla (Graur and Higgins 1994). The first of these phylogenetic hypotheses was based initially on an analysis of portions of mitochondrial 12S and 16S genes (Milinkovitch, Orti, and Meyer 1993) and subsequently on a tandem alignment of portions of the 12S,16S, and cytochrome *b* loci from 21 species of Cetacea, and several artiodactyl outgroups (Milinkovitch, Meyer, and Powell 1994). In a separate analysis involving the complete cytochrome *b* locus, Arnason and Gullberg (1994) presented data suggesting a further possibility, which was that toothed whales might be closer relatives of baleen whales than of sperm whales. Adachi and Hasegawa (1995) have since reanalyzed the cytochrome *b* data using maximum likelihood and show that the conclusion is highly sensitive to choice of outgroup taxa, with the majority of situations supporting the baleen/sperm whale clade. Most recently, Arnason and Gullberg (1996), in a further analysis of the complete cytochrome *b* gene, including representatives of all currently recognized cetacean families, present evidence for a monophyletic Odontoceti; however, this conclusion was not strongly supported. There is at present no published account of this issue from the perspective of a nucleotide sequence analysis of a single-copy nuclear gene.

The Graur and Higgins hypothesis suggests that cows are more closely related to cetaceans than to pigs, which disrupts the monophyly of Artiodactyla, resulting in either the interpretation of a paraphyletic Artiodactyla or an Artiodactyla that includes Cetacea as a suborder. This hypothesis was based on the analysis of 11 nuclear encoded protein sequences and five mitochondrial genes from two artiodactyl suborders, one species of cetacean, and either mouse, seal, or mouse and seal as outgroup. We are aware of no published account that addresses this issue from the perspective of a single-copy nuclear gene at the DNA sequence level.

Earlier papers of ours have demonstrated the utility of exon 1 sequences from the gene encoding interphotoreceptor retinoid binding protein (IRBP) for addressing higher level systematics in mammals (Stanhope et al. 1992, 1993, 1996). The purpose of this article is to present information that addresses both these phylogenetic hypotheses regarding Cetacea and their affinities with Artiodactyla, with sequences derived from this same 1.2-kb 5' region of exon 1 of the IRBP gene (between 261 and 15 15 of the published human sequence, Fong et al. 1990).

Materials and Methods

The common and scientific names of the species included in the analyses, the order within which they are classified, and number of base pairs of nucleotide sequence represented for each are as follows: rough-toothed dolphin, Steno *bredanensis* (Cetacea; 1,158); pilot whale, *Globicephala macrorhynchus* (Cetacea; 630); gray whale, *Eschrichtius robustus* (Cetacea; 1,078); minke whale, *Balaenopteru acutorostrata* (Cetacea; 1,073); pygmy sperm whale, *Kogia breviceps* (Cetacea; 1,241); giant sperm whale, *Physeter catodon* (Cetacea; 1,241); cow, *Bos taurus* (Artiodactyla; 1,241); pig, *Sus scrofa* (Artiodactyla; 1,24 1); horse, *Equus caballus* (Perissodactyla; 1,177); cat, *Felis catus* (Carnivora; 1,150); mouse, *Mus domesticus* (Rodentia; 1,24 1); dugong, *Dugong dugon* (Sirenia; 1,046); hyrax, *Procavia capensis* (Hyracoidea; 1 ,014); African elephant, *Loxodonta africana* (Proboscidea; 1,078); and three-toed sloth, *Bradypus tridactylus* (Edentata; 1,011). Horse, cat, mouse, the paenungulates (dugong, hyrax, elephant), and sloth served as outgroups, utilized in various combinations (described below). With this choice of cetacean taxa we have represented two species of each suborder of Cetacea (pilot whale and rough-toothed dolphin, family Delphinidae, suborder Odontoceti; gray whale, family Eschrichtiidae, suborder Mysticeti; minke whale, family Balaenopteridae, suborder Mysticeti), as well as a representative of each of the two families of sperm whales (Kogiidae and Physeteridae). DNA sequences were determined for giant sperm whale, pygmy sperm whale, minke whale, gray whale, and pilot whale by direct sequencing of PCR amplified fragments. Primers for amplification of this region were the same as those previously described (Stanhope et al. 1992). Dye terminator cycle sequencing reactions were loaded on an Applied Biosystems 373A automatic sequencer, following the manufacturer's protocols. Internal sequencing primers were designed as necessary. All templates were sequenced at least twice on both strands. Sequences for rough-toothed dolphin, pig, horse, the paenungulates, and sloth come from our most recent IRBP analysis (Stanhope et al. 1996); the cow sequence comes from a published account of the entire gene (Borst et al. 1989);

mouse and cat come from our earlier analysis of this gene (Stanhope et al. 1992).

Relationships within Cetacea were first examined with the tree rooted at pig, cow, and pig/cow together. The artiodactyl–cetacean relationship was then examined using the six species of cetaceans and two artiodactyls, concomitant with one of the following taxa or combinations of taxa as outgroups: (1) horse; (2) cat; (3) mouse; (4) horse, cat, tree rooted at sloth; (5) paenungulates (African elephant, hyrax, dugong), tree rooted at sloth. These various choices of outgroups for the artiodactyl/cetacean relationship also provide a further perspective on the relationships within Cetacea since they provide additional character polarization possibilities. Our recent analysis of IRBP sequences from 25 taxa encompassing all eutherian orders (Stanhope et al. 1996) excepting Pholidota (pangolin) provide convincing evidence for a monophyletic grouping of Cetacea and Artiodactyla, in agreement with a wide range of other sources of data (see, e.g., Gingerich, Smith, and Simons 1990; Novacek 1992; Milinkovitch, Orti, and Meyer 1993). They also strongly support the monophyly of the Paenungulata, again in agreement with a wide range of other sources of data (see, e.g., Novacek 1992; Honeycutt and Adkins 1993; Springer and Kirsch 1993; Porter, Goodman, and Stanhope 1996), suggesting the possibility that choice of this superorder as a sister group to the Artiodactyla/Cetacea clade may facilitate the breaking up of this long branch, allowing better character polarization. Rooting of the tree at sloth in two of the outgroup comparisons is based on the widely held belief that edentates are the most primitive eutherian (Novacek 1992).

The maximum-parsimony (MP), neighbor-joining (NJ; Saitou and Nei 1987), and maximum-likelihood (DNAML; Felsenstein 1981) programs within PHYLIP (Felsenstein 1993) were used for analysis. Kimura two-parameter distances (Kimura 1980) with a transition/transversion ratio of 2.0 were used as input for the NJ analyses. Jukes-Cantor distances yielded very similar results. The DNAML analyses used a transition/transversion ratio of 2.0; varying this parameter to settings of 1.5/1, 3/1, and 4/1, on several preliminary runs of two data sets, had little effect on the outcome. For all choices of outgroup, 1,000 bootstrap replicates (Felsenstein 1985) for the MP and NJ analyses and 100 replicates for the DNAML analysis were used to assess the relative strengths of each of a number of competing hypotheses regarding the cetaceans and their artiodactyl relatives. The various hypotheses were: (1) baleen whale/sperm whale, (2) baleen whale/toothed whale, (3) toothed whale/sperm whale, (4) cow/Cetacea, (5) pig/Cetacea, (6) cow/pig (fig. 1).

Results and Discussion

The results, in agreement with Adachi and Hasegawa's analysis, indicate that choice of outgroup had a significant influence on the outcome, but that this was most evident with regard to the artiodactyl/cetacean affinity (table 1). All of the MP and NJ analyses prefer-
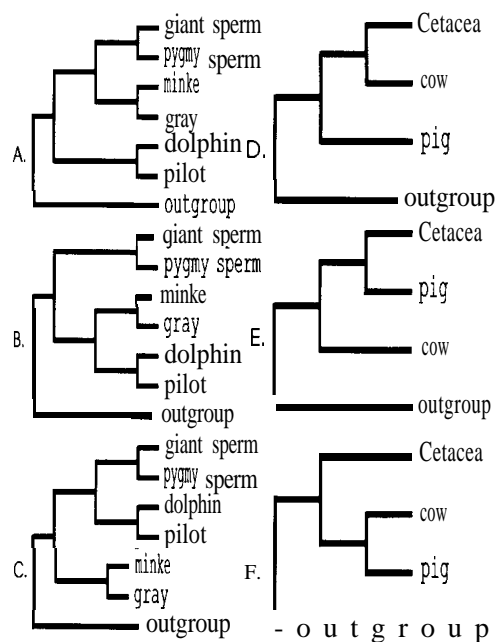

FIG. 1 .-The topologies of the six phylogenetic hypotheses under investigation. Each of the above branching arrangements corresponds with the following column headings in table 1: *A:* baleen/sperm; *B:* baleen/toothed; *C:* toothed/sperm; *D:*cow/Cetacea;*E:*pig/Cetacea;*F:* cow/pig.

entially supported a baleen whale/sperm whale clade, with bootstrap support ranging from 5 1%–95%. Ten of the 16 analyses supported this clade in excess of 80%, with multiple outgroup taxa generally resulting in higher bootstrap support. The DNAML analyses, however, did not agree with this assessment, with the majority of tests split approximately evenly between baleen/sperm whale and baleen/toothed whale. In contrast to the Adachi and Hasegawa analysis of the cytochrome *b* data, the traditional hypothesis of toothed/sperm whale received very little support in any of our IRBP tests. All tests supported a monophyletic Cetacea, and Artiodactyla/Cetacea at 100%; the monophyly of baleen whales, sperm whales, and toothed whales (Delphinidae), respectively, received bootstrap support of 97%-100%.

Less of a consensus was evident regarding the precise nature of an artiodactyl/cetacean affinity. A pig/cetacean clade was the favored grouping in 7 of the 15 analyses, the only convincing bootstrap support arising when rodents were used as outgroup (bootstrap support of 75%–91%). A cow/cetacean clade was the favored option in the remaining eight analyses, generally with relatively high bootstrap support. The monophyly of Artiodactyla was never the favored grouping.

These results provide further support for the hypothesis of a paraphyletic Odontoceti, with the highest bootstrap support for monophyly falling at only 20%, and most analyses generally below 10%. We also feel, however, that because of the limited representation of taxa this should remain a tentative hypothesis. These IRBP sequences suggest that the paraphyly of Odontoceti is due to a common ancestry of baleen whales and sperm whales; however, since the maximum-likelihood

Table 1
IRBP Bootstrap Support for Various Cetacean and Artiodactyl/Cetacean Clades

Outgroup(s)	Test		Baleen/ Sperm	Baleen/ Toothed	Toothed/ Sperm	Cow/ Cetacea	Pig/ Cetacea	Cow/ Pig
Cow	M	P	72	22	6	**NA**	**NA**	NA
		DNAML	42	44	14	**NA**	**NA**	NA
		NJ	85	15	0	**NA**	**NA**	NA
Pig.	M	P	74	21	5	**NA**	**NA**	NA
		DNAML	46	43	11	**NA**	**NA**	NA
		NJ	**51**	43	6	**NA**	**NA**	NA
Cow/pig	M	P	**77**	18	5	**NA**	**NA**	NA
		DNAML	46	52	2	**NA**	**NA**	NA
		NJ	77	20	3	NA	**NA**	NA
Horse	M	P	91	8	1	78	5	17
		DNAML	64	22	14	86	0	14
		NJ	86	12	2	70	12	18
Cat	M	P	**81**	II	8	86	6	8
		DNAML	50	34	16	78	10	12
		NJ	95	4	1	63	7	30
M o u s e................		MP	**81**	14	5	20	75	5
		DNAML	41	46	13	9	88	3
		NJ	**74**	18	8	2	91	1
Horse, cat, root at sloth		MP	86	4	10	73	8	19
		DNAML	63	33	4	60	2	38
		NJ	89	7	4	26	44	30
Paenungulates, root at sloth		MP	83	10	7	8	53	39
		DNAML	42	38	20	11	46	43
		NJ	86	II	3	6	54	40

Nom-Bootstrap support is expressed as a percentage. MP: maximum parsimony; DNAML: maximum likelihood; NJ: neighbor joining; NA: not applicable. The favored option in each case is indicated with bold type.

results do not resoundingly corroborate this view we would also regard this more specific conclusion as one in need of further examination. Evidence for convincing phylogenetic associations must come from a consensus of different data sets and methods of analysis. At present there is no such convincing consensus. Two of the disconcerting factors in regarding any of these three possible clades with confidence are that the proposed groupings can be easily disrupted in any of these data sets or analyses (including this one) with only a few substitutions to the most parsimonious tree, and that the maximum-likelihood bootstrap figures reported here and in the Adachi and Hasegawa study for the favored associations are often near the 50% range.

An additional area of recent discussion regarding cetacean evolution concerns attempts at dating the cetacean radiation. Milinkovitch, Orti, and Meyer (1993) originally estimated that the common ancestor of sperm and baleen whales may have existed 10–15 MYA. They subsequently modified this estimate to approximately 25 MYA (Milinkovitch, Meyer, and Powell 1994). The available fossil evidence suggests a rapid diversification into the principal lineages of cetaceans some 30-34 MYA (Fordyce and Barnes 1994; Arnason and Gullberg 1996). The discovery of Eocene whales with vestigial limbs resembling those of the even-toed artiodactyls, places a common ancestry of Cetacea and Artiodactyla at about 60 MYA (Gingerich, Smith, and Simons 1990). The IRBP gene in all of the cetaceans represented in this analysis appears to be evolving at approximately the same rate, with mean cetacean/artiodactyl percentage sequence divergences of 12.0 (gray whale), 12.9 (minke), 11.0 (pilot whale), 12.2 (dolphin), 11.4 (giant sperm

whale), and 11.8 (pygmy sperm whale). The mean divergence figure between sloth and all of the various taxa included in this analysis is 24.3%. An estimated time for the split between edentates and the rest of the placentals is about 115 MYA (Novacek 1992), resulting in rate of sequence change of about 0.21%/Myr. If we take a different calibration point we get a roughly similar figure. For example, the estimated divergence between Sirenia and Proboscidea is about 65 MYA (Novacek 1992); IRBP sequence divergence between dugong and elephant is 12.5%, yielding a rate of about 0.19%/Myr. If we use the 60 Myr figure for common ancestry between cetaceans and artiodactyls, the rate of sequence change is 0.20%/Myr. Conversely, if we use as our calibration point a figure of 34 Myr as representative of the earliest point in the diversification of Cetacea, the rate of IRBP sequence change drops to 0.12%/Myr, suggesting a dramatic evolutionary rate slowdown during the adaptive radiation of Cetacea. This figure would result in a date of 32 MYA for the diversification of baleen and sperm whales. This rate estimate, however, of 0.12%/Myr, based on the 34 MYA calibration point, is at odds with other means of estimating evolutionary rates in Cetacea. For example, it has been suggested that about 11MYA, the Delphinoidea diversified into the various lineages that we see today as porpoises, dolphins, and beluga (Barnes, Doming, and Ray 1985; Milinkovitch, Meyer, and Powell 1994). We only have members of one of these families included in this analysis (pilot whale and rough-toothed dolphin), but they show about 1.6% sequence divergence, yielding a highly conservative rate of at least 0.15%/Myr. This figure is more in line with rates estimated on the basis of a com-

parison involving the outgroup taxa included in this analysis, which suggests that the cetaceans as a group are evolving anywhere from 7%-27% slower than the other orders of mammals. This in turn would yield rate figures of 0.15-O. 19%/Myr. This suggests that the more reasonable figure for Cetacea is about 0.17%/Myr, dating the split between baleen and sperm whales at about 22 MYA, and for the cetacean radiation at about 25 MYA. This latter estimate is in approximate agreement with that suggested by Milinkovitch, Meyer, and Powell (1994) and Schlötterer, Amos, and Tautz (1991), and remains curiously at odds with the paleontological view of 30-34 MYA. Elsewhere Milinkovitch (1995) has argued that some of the 34-Myr-old fragmentary fossils are questionable representatives of early mysticetes. Whatever the precise dates, all of the fossil data, as well as the short internodes typical of the molecular analyses, do seem to point to a diversification into the various principal lineages over a surprisingly short period of time.

In regard to the artiodactyl-cetacean affinities, the main consensus emerging from this analysis is that the present IRBP data do not support a monophyletic Artiodactyla. It is not clear, however, whether this is due to a paraphyletic Artiodactyla that has cow more closely related to cetaceans or pig more closely related to cetaceans. The principal set of analyses supporting pig/Cetacea with any degree of bootstrap support comes with rodents as outgroup, which are not as appropriate a choice of sister group to an Artiodactyla/Cetacea clade as are horse or cat (see, e.g., Li et al. 1990; Arnason and Johnson 1992; Honeycutt and Adkins 1993; Stanhope et al. 1993, 1996; Honeycutt et al. 1995), suggesting that the more reliable conclusion may be cow/Cetacea.

In our opinion, both these phylogenetic issues remain far from settled. From the perspective of IRBP, a more rigorous test of a paraphyletic Odontoceti, of the strength of the sperm whale/baleen whale clade, and of the possible subordinal status of cetaceans within an expanded concept of Artiodactyla awaits the acquisition of additional samples that will increase the phylogenetic density of each of the main groupings. Coding sequences of both mtDNA and nuclear genes have yet to provide highly convincing data and thus we suggest a more fruitful area of investigation might involve noncoding nuclear DNA. The sequence divergence of exon 1 of IRBP between the artiodactyls and cetaceans suggests that intron sequences between these taxa might be unambiguously aligned. Similar levels of divergence, for example, are present between several of the main groupings of primates, and intron 1 sequences can be easily aligned between such taxa (Harada et al. 1995; unpublished data). Such regions of more rapidly evolving nuclear data could well represent an important additional source of phylogenetic information for both these systematic problems.

Sequence Availability

The nucleotide sequence data reported in this paper appear in GenBank under the following accession numbers: gray whale, *Eschrichtius robustus,* U50649; giant sperm whale, *Physeter catodon,* U508 18; pygmy sperm whale, *Kogia breviceps,* U50819; minke whale, *Balaenoptera acutorostrata,* U50820; pilot whale, *Globicephala macrorhynchus,* U50821.

LITERATURE CITED

ADACHI, J., and M. HASEGAWA. 1995. Phylogeny of whales: dependence of the inference on species sampling. Mol. Biol. Evol. **12**:177–179.

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

———. 1996. Cytochrome *b* nucleotide sequences and the identification of five primary lineages of extant cetaceans. Mol. Biol. Evol. **13**:407–417.

ARNASON, U., and E. JOHNSON. 1992. The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina.* J. Mol. Evol. **33**:556–568.

BARNES, L. G., D. I? DOMING, and C. E. RAY. 1985. Status of studies on fossil marine mammals. Mar. Mamm. Sci. 1:15-53.

BORST, D. E., T. M. REDMOND, J. E. ELSER, M. A. GONDA, B. WIGGERT, G. J. CHADER, and J. M. NICKERSON. 1989. Interphotoreceptor retinoid binding protein: gene characterization, protein repeat structure, and its evolution. J. Biol. Chem. 264:1115-1 123.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368-376.

———. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–79 1.

———. 1993. PHYLIP (phylogeny inference package). Version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

FONG, S.-L., W.-B. FONG, T. A. MORRIS, K. M. KEDZIE, and C. D. B. BRIDGES. 1990. Characterization and comparative structural features of the gene for human interstitial retinol-binding protein. J. Biol. Chem. **265**:3648–3653.

FORDYCE, R. E., and L. G. BARNES. 1994. The evolutionary history of whales and dolphins. Annu. Rev. Earth Planet. Sci. **22**:419–455.

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

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

HARADA, M. L., H. SHCHNEIDER, M. P. C. SCHNEIDER, I. SAMPAIO, J. CZELUSNIAK, and M. GOODMAN. 1995. DNA evidence on the phylogenetic systematics of New World monkeys: support for the sister grouping of *Cebus* and *Saimiri*

from two unlinked nuclear genes. Mol. Phylogenet. Evol. **4**:33 l-349.

HONEYCUTT, R. L., and R. M. ADKINS. 1993. Higher level systematics of eutherian mammals: an assessment of molecular characters and phylogenetic hypotheses. Annu. Rev. Ecol. Syst. **24**:279–305.

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.

KIMURA, M. 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16: l 1 l-120.

LI, W.-H. M. GOUY, P. M. SHARP, C. O'HUIGIN, and Y.-W. YANG. 1990. Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla and Camivora and molecular clocks. Proc. Natl. Acad. Sci. USA **87**:6703–6707.

MILINKOVITCH, M. C. 1995. Molecular phylogeny of cetaceans prompts revision of morphological transformations. Trends Ecol. Evol. **10**:305–345.

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

NOVACEK, M. J. 1992. Mammalian phylogeny: shaking the tree. Nature **356**:121–125.

PORTER, C. A., M. GOODMAN, and M. J. STANHOPE. 1996. Evidence on mammalian phylogeny from sequences of exon 28 of the von Willebrand factor gene. Mol. Phylogenet. Evol. **5**:89–101.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SCHLÖTTERER, C., B. AMOS, and D. TAUTZ. 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature **354**:63–65.

SPRINGER, M. S., and J. A. W. KIRSCH. 1993. A molecular perspective on the phylogeny of placental mammals based on mitochondrial 12S rDNA sequences, with special reference to the problem of the Paenungulata. J. Mamm. Evol. **1**:149–166.

STANHOPE, M. J., W. J. BAILEY, J. CZELUSNIAK, M. GOODMAN, J.-S. SI, J. NICKERSON, J. G. SGOUROS, G. A. M. SINGER, and T. K. KLEINSCHMIDT. 1993. A molecular view of primate supraordinal relationships from the analysis of both nucleotide and amino acid sequences. Pp. 25 l-292 in R. D. E. MACPHEE, ed. Primates and their relatives in phylogenetic perspective. Plenum, New York.

STANHOPE, M. J., J. CZELUSNIAK, J.-S. SI, J. NICKERSON, and M. GOODMAN. 1992. A molecular perspective on mammalian evolution from the gene encoding interphotoreceptor retinoid binding protein, with convincing evidence for bat monophyly. Mol. Phylogenet. Evol. **1**:148–160.

STANHOPE, M. J., M. R. SMITH, V. G. WADDELL, C. A. PORTER, M. S. SHIVJI, and M. GOODMAN. 1996. Mammalian evolution and the IRBP gene: convincing evidence for several superordinal clades. J. Mol. Evol. (in press).