Journal of Heredity 2004:95(6):459–473
doi:10.1093/jhered/esh054

A Comprehensive and Validated Molecular Taxonomy of Beaked Whales, Family Ziphiidae

M. L. Dalebout, C. S. Baker, J. G. Mead, V. G. Cockcroft, and T. K. Yamada

From the School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand (Dalebout and Baker); National Museum of Natural History, Mail Stop NHB 108, Smithsonian Institution, Washington, DC 20560 (Mead); Center for Dolphin Studies, Box 1856, Plettenberg Bay 6600, South Africa (Cockcroft); and National Science Museum, Tokyo, 3-23-1 Hyakunin-cho, Shinjuku-ku, Tokyo 169-0073, Japan (Yamada).
M. L. Dalebout is currently at the Biology Department, Dalhousie University, Halifax, Nova Scotia B3H 4JI, Canada.

Address correspondence to C. Scott Baker at the address above, or e-mail: cs.baker@auckland.ac.nz.

Abstract

DNA sequences from orthologous loci can provide universal characters for taxonomic identification. Molecular taxonomy is of particular value for groups in which distinctive morphological features are difficult to observe or compare. To assist in species identification for the little known family Ziphiidae (beaked whales), we compiled a reference database of mitochondrial DNA (mtDNA) control region (437 bp) and cytochrome *b* (384 bp) sequences for all 21 described species in this group. This mtDNA database is complemented by a nuclear database of actin intron sequences (925 bp) for 17 of the 21 species. All reference sequences were derived from specimens validated by diagnostic skeletal material or other documentation, and included four holotypes. Phylogenetic analyses of mtDNA sequences confirmed the genetic distinctiveness of all beaked whale species currently recognized. Both mitochondrial loci were well suited for species identification, with reference sequences for all known ziphiids forming robust species-specific clades in phylogenetic reconstructions. The majority of species were also distinguished by nuclear alleles. Phylogenetic comparison of sequence data from "test" specimens to these reference databases resulted in three major taxonomic discoveries involving animals previously misclassified from morphology. Based on our experience with this family and the order Cetacea as a whole, we suggest that a molecular taxonomy should consider the following components: comprehensiveness, validation, locus sensitivity, genetic distinctiveness and exclusivity, concordance, and universal accessibility and curation.

Recognition and delineation of species is of fundamental importance in taxonomic classification and conservation. Although the question "What is a species?" is far from resolved (e.g., Wheeler and Meier 2000), the assignment of an individual to a particular species based on morphology is generally straightforward for well-known taxa. For lesser-known or rare taxa, however, determining the species identity of an individual can be problematic. Due to the declining abundance of many species, access to complete anatomical specimens is becoming a vanishing luxury. With rare or partial voucher specimens, it can be difficult to judge whether characters are diagnostic (i.e., synapomorphies) or represent variable traits shared with other species. For many animal taxa, the lethal collection of such voucher specimens would now also be considered unethical. If species are cryptic morphologically with more common taxa and threatened directly by exploitation or indirectly by habitat loss, extinction could occur before their distinctiveness is recognized and sufficient material can be gathered for formal classification (e.g., Wakeham-Dawson et al. 2002).

Genetic information in the form of DNA sequences can serve as a universal character set for the taxonomic identification of organisms. Such genetic characters are particularly useful for species in which distinctive morphological features are difficult to observe or compare. Genetic databases have become increasingly common in the monitoring of trade and investigation of poaching (e.g., Baker et al. 1996; Baker and Palumbi 1994; DeSalle and Birstein 1996; Malik et al. 1997; Roman and Bowen 2000), but their use in addressing questions of basic organismal taxonomy or the discovery of new vertebrate species remains rare (Baker et al. 2002; Dalebout et al. 2002; Giao et al. 1998; Smith et al. 1991). For microorganisms, however, these techniques are widely used to investigate species diversity

and identity (e.g., Fuhrman et al. 1992; Moon-van der Staay et al. 2001; Pace 1997). It is becoming increasingly obvious that a molecular taxonomy would be valuable for all organisms (Dalebout and Baker 2002; Hebert et al. 2003; Tautz et al. 2003).

Genetic identification of specimens of known species differs from the usual goal of organismal phylogenetics, that of defining sister taxa (Hennig 1966). Given a database of "reference" sequences, unknown "test" specimens can be identified to species based on their phylogenetic grouping with sequences from recognized species to the exclusion of sequences from other species (Baker et al. 1996; Baker and Palumbi 1994). Reference sequences should only be obtained from validated specimens—animals examined and identified by experts for which diagnostic skeletal material or photographs of such features are collected (as discussed by Dizon et al. 2000). Problems can occur, however, if taxon sampling is incomplete (missing species) or within-species sampling is not sufficiently representative of diversity. In cases of deep intraspecific diversity (e.g., Lento et al. 1994; Wayne et al. 1990) or shallow interspecific divergence (e.g., Dizon et al. 2000; Milinkovitch et al. 2002), an unknown test sequence could group with the next most closely related species. Baker et al. (1996) suggested that identification of test sequences should be considered conclusive only if they nest within the diversity of reference sequences for a species. It is important therefore that levels of genetic diversity within and divergence between species in a group of interest are first assessed. It should also be noted that recently diverged species might not be reciprocally monophyletic at all loci (e.g., Baker et al. 1996; Hare et al. 2002).

Beaked whales (family Ziphiidae) are among the least known of all vertebrates (Wilson 1992). Found in deep-water regions worldwide, many of the 21 species currently recognized have been described from only a small number of stranded specimens. For several, there has yet to be a confirmed sighting of a living animal. Even in the last decade or so, three new ziphiids have been described. Species identification in this group is based primarily on features of cranial and tooth morphology, which for some species are diagnostic only in mature animals (Moore 1968). Many species are also very similar in external appearance (Mead 1989). As a result, stranded animals are frequently misidentified (Dalebout et al. 1998) and those seen at sea are often identifiable only to the level of family or genus.

Here we present a comprehensive and validated mitochondrial DNA (mtDNA) reference database of control region and cytochrome *b* sequences for beaked whales (Dalebout et al. 1998; Henshaw et al. 1997; this article). These mtDNA datasets, including earlier partial versions, have been used previously to correctly identify specimens misidentified from morphology and discover a new species of beaked whale (Dalebout et al. 1998, 2002, 2003; van Helden et al. 2002). These datasets are presented here in full for the first time. We also present a complementary database of nuclear DNA (nDNA) actin intron sequences for 17 of the 21 beaked whale species and consider the concordance of evolutionary patterns among these alleles with the mtDNA phylogeny.

Materials and Methods

Sample Collection and DNA Extraction

Samples were obtained from stranded or beach-cast animals, victims of incidental fisheries takes (by-catch), and museum collections (Dalebout 2002; Dalebout et al. 1998, 2003; Henshaw et al. 1997; van Helden et al. 2002; this article). Reference sequences were generated only from validated specimens (control region, $n = 42$; cytochrome *b*, $n = 41$; actin intron, $n = 31$) (Table 1). Specimens not validated by expert morphological examination or diagnostic material/documentation were considered "test" specimens and subjected to phylogenetic comparisons to the reference sequences to confirm their species identity ($n = 9$; see Discussion). Total genomic DNA was isolated from fresh tissue samples using proteinase K digestion, following standard methods (Davis et al. 1986), as modified by Baker et al. (1994). For historic museum specimens represented by teeth or bone, DNA was extracted using the silica-based method (Boom et al. 1990; Höss and Pääbo 1993) as modified by Matisoo-Smith et al. (1997), and techniques described in Pichler et al. (2001).

Polymerase Chain Reaction (PCR) Amplification and Sequencing

Using the PCR, the following loci were amplified: a 500 bp fragment of the 5′ end of the mtDNA control region (D-loop) using the primers M13-Dlp1.5-L [t-Pro whale] and Dlp5-H (Dalebout et al. 1998), a 424 bp fragment of the 5′ end of the mtDNA cytochrome *b* gene using the primers GLUDG-L and CB2-H (Palumbi 1996), and a 950 bp fragment of the first intron of the nuclear muscle actin gene using the primers ACT3-F and ACT1385-R (Baker et al. 1998) (Table 2). For the mtDNA, the internal primers discussed below were used for sequence verification for some samples. For the actin intron, the internal primers ACT5-F and M13-ACT5R-R (Baker CS and Palumbi SR, unpublished data) were used for sequence verification. Amplification of these segments followed standard protocols (Palumbi 1996).

For the majority of specimens represented only by tooth or bone, full-length mtDNA sequences were not amplified successfully, as expected from uncontaminated DNA extractions from such material (Höss and Pääbo 1993). For the mtDNA control region, these specimens are instead represented either by a 301 bp fragment amplified using the primers M13-Dlp1.5-L and Dlp4-H (Baker CS, unpublished data) or a 249 bp fragment amplified using the primers Dlp10-L (Baker et al. 1993) and Dlp4-H. For cytochrome *b*, these specimens are instead represented either by a 289 bp fragment amplified using the primers CB1-L and CB2-H (Palumbi 1996) or a 196 bp fragment amplified using the primers CYBMF-L and CYBMR-H (Dalebout 2002). The latter primer pair was designed specifically for beaked whales to avoid amplification of modern human mtDNA, a frequent contaminant of historical osteological material, or that of other cetacean species. See Table 2 for all primer sequences

Table 1. Validated specimens used to generate the mtDNA and nDNA sequences that comprise the beaked whale reference databases

Species	Sample code	Geographic origin	Alleles	GenBank accession no.	Source
Panel A: mtDNA control region					
1 *Mesoplodon bidens*, Sowerby's beaked whale	MbiSAC1309-3	NA (UK)		AY579507	a
	MbiSAC1880	NA (UK)		AY579508	a
2 *Mesoplodon bowdoini*, Andrews' beaked whale	MbowTMAG1593	SP (Australia)		AY579509	b
	MbowNMNZ619	SP (NZ)		AY579510	c
3 *Mesoplodon carlhubbsi*, Hubbs' beaked whale	McaSW1127	NP (USA)		AY579511	d
	McaSW1154	NP (USA)		AY579512	d
4 *Mesoplodon densirostris*, dense-beaked whale	MdeNZ01	SP (NZ)		AY579513	e
	MdeNHM-UK	NA (UK)		AY579514	f
5 *Mesoplodon europaeus*, Gervais' beaked whale	MeuSW4120	NA (USA)		AY579515	d
	MeuSW3853	NA (USA)		AY579516	d
6 *Mesoplodon ginkgodens*, ginkgo-toothed beaked whale	Mgin01	NP (Taiwan)		AY579517	g
	TSM8744 (H)	NP (Japan)		AY579518	h
7 *Mesoplodon grayi*, Gray's beaked whale	Mgr05	SP (NZ)		AY579519	e
	Mgr11	SP (NZ)		AY579520	e
8 *Mesoplodon hectori*, Hector's beaked whale	MheNZ02	SP (NZ)		AY579521	e
	MheNMNZ2173	SP (NZ)		AY579522	c
9 *Mesoplodon layardii*, straptooth whale	MlaySAM18078	SP (Australia)		AY579523	i
	MlaySAM9788	SP (Australia)		AY579524	i
10 *Mesoplodon mirus*, True's beaked whale	MmiSW4972	NA (USA)		AY579525	d
	MmiSW4968[a]	NA (USA)		U70465	d
11 *Mesoplodon peruvianus*, lesser beaked whale	MpeJCR1926	SP (Peru)		AF492413	j
	MpeLAM95654	NP (USA)		AY579526	k
12 *Mesoplodon stejnegeri*, Stejneger's beaked whale	MstSW10402	NP (USA)		AY579527	d
	MstSW9491	NP (USA)		AY579528	d
13 *Mesoplodon traversii*, spade-toothed whale[*]	NMNZ546 (H)	SP (NZ)		AF439992	c,1
	MNHNC1156	SP (Chile)		AF439994	l,1
14 *Mesoplodon perrini*, Perrin's beaked whale	USNM504853 (H)	NP (USA)		AF441256	m,2
	TMMC-C75	NP (USA)		AF441258	d/n,2
15 *Indopacetus pacificus*, Longman's beaked whale	QM-J2106 (H)	TP (Australia)		AY162435	o,3
	MZUF 1956	IO (Somalia)		AY162436	p,3
16 *Hyperoodon ampullatus*, northern bottlenose whale	HamSH9711	NA (Canada)		AF350440	q,4
	HamSH9601	NA (Canada)		AF350437	q,4
17 *Hyperoodon planifrons*, southern bottlenose whale	Hpl01[b]	SP (NZ)		AF036224	e
	Hpl04	SP (NZ)		AY579529	e
18 *Ziphius cavirostris*, Cuvier's beaked whale	Zca11	SP (NZ)		AY579530	e
	ZcaSAC0356	NA (UK)		AY579531	a
19 *Tasmacetus shepherdi*, Shepherd's beaked whale	Tsh01[b]	SP (NZ)		AF036226	e
	Tsh02[b]	SP (NZ)		AF036227	e
20 *Berardius arnuxii*, Arnoux's beaked whale	Bar02[b]	SP (NZ)		AF036229	e
	BarPEM28	IO (South Africa)		AY579532	r
21 *Berardius bairdii*, Baird's beaked whale	BbaSW4965[a]	NP (Japan)		U70467	d
	BbaBC9220	NP (Canada)		AY579533	s
Panel B: MtDNA cytochrome *b*					
1 *Mesoplodon bidens*, Sowerby's beaked whale	MbiSW3854	NA (USA)		AY579534	d
	MbiSW3858	NA (USA)		AY579535	d
2 *Mesoplodon bowdoini*, Andrews' beaked whale	MbowSAM18047	SP (Australia)		AY579536	i
	Mbow04	SP (NZ)		AY579537	e
3 *Mesoplodon carlhubbsi*, Hubbs' beaked whale	McaSW3804	NP (USA)		AY579538	d
	McaSW1154	NP (USA)		AY579539	d
4 *Mesoplodon densirostris*, dense-beaked whale	MdeSW4010	NP (USA)		AY579540	d
	MdeNZ02	SP (NZ)		AY579541	e
5 *Mesoplodon europaeus*, Gervais' beaked whale	MeuSW4120	NA (USA)		AY579542	d
	MeuSW7443	NA (USA)		AY579543	d
6 *Mesoplodon ginkgodens*, ginkgo-toothed beaked whale	Mgin01	NP (Taiwan)		AY579544	g
7 *Mesoplodon grayi*, Gray's beaked whale	Mgr29	SP (NZ)		AY579545	e
	MgrH04	SP (NZ)		AY579546	e
8 *Mesoplodon hectori*, Hector's beaked whale	MheTAS	SP (Australia)		AY579547	t
	MheSAM16387	SP (Australia)		AY579548	i
9 *Mesoplodon layardii*, straptooth whale	Mlay13	SP (NZ)		AY579549	e
	Mlay04	SP (NZ)		AY579550	e

Table 1. Continued

Species	Sample code	Geographic origin	Alleles	GenBank accession no.	Source
10 *Mesoplodon mirus*, True's beaked whale	MmiSW4972	NA (USA)		AY579551	d
	MmiSW4968	NA (USA)		AY579552	d
11 *Mesoplodon peruvianus*, lesser beaked whale	MpeJCR1926	SP (Peru)		AF492414	j
	Mpe-U13141	SP (Peru)		U13141	5
12 *Mesoplodon stejnegeri*, Stejneger's beaked whale	MstSW4962	NP (USA)		AY579553	d
	MstSW6481	NP (USA)		AY579554	d
13 *Mesoplodon traversii*, spade-toothed whale*	NMNZ546 (H)	SP (NZ)		AY579555	c
	MNHNC1156	SP (Chile)		AY579556	l
14 *Mesoplodon perrini*, Perrin's beaked whale	USNM504853 (H)	NP (USA)		AF441261	m,2
	TMMC-C75	NP (USA)		AF441263	d/n,2
15 *Indopacetus pacificus*, Longman's beaked whale	QM-J2106 (H)	TP (Australia)		AY162440	o,3
	MZUF 1956	IO (Somalia)		AY162441	p,3
16 *Hyperoodon ampullatus*, northern bottlenose whale	HamSH9717	NA (Canada)		AY579557	q
	HamIC454	NA (Canada)		AY579558	u
17 *Hyperoodon planifrons*, southern bottlenose whale	Hpl01	SP (NZ)		AY579559	e
	Hpl02	SP (NZ)		AY579560	e
18 *Ziphius cavirostris*, Cuvier's beaked whale	ZcaNC0296	NA (USA)		AY579561	m
	ZcaSW8398	NP (USA)		AY579562	d
19 *Tasmacetus shepherdi*, Shepherd's beaked whale	Tsh02	SP (NZ)		AY579563	e
	Tsh04	SP (NZ)		AY579564	e
20 *Berardius arnuxii*, Arnoux's beaked whale	Bar02	SP (NZ)		AY579565	e
	BarNMNZ580	SP (NZ)		AY579566	c
21 *Berardius bairdii*, Baird's beaked whale	Bba-X92541	NP (Japan)		X92541	6
	BbaBC9220	NP (Canada)		AY579567	s
Panel C: Nuclear actin intron					
1 *Mesoplodon bidens*, Sowerby's beaked whale	MbiSW3858	NA (USA)	Mbi-a1/Mbi-a1	AY579473 (1)	d
	MbiSAC1309-1	NA (UK)	Mbi-a1/Mbi-a2	AY579474 (2)	a
2 *Mesoplodon bowdoini*, Andrews' beaked whale	MbowSAM18047	SP (Australia)	Mbow-a2/Mbow-a2	AY579475 (2)	i
3 *Mesoplodon carlhubbsi*, Hubbs' beaked whale	McaSW1563	NP (USA)	Mca-a1/Mca-a1	AY579476 (1)	d
	McaSW73	NP (USA)	Mca-a2/Mca-a2	AY579477 (2)	d
4 *Mesoplodon densirostris*, dense-beaked whale	MdeNHM-UK	NA (UK)	Mde-a1/Mde-a2	AY579478 (1)	f
	MdeNZ01	SP (NZ)	Mde-a1/Mde-a2	AY579479 (2)	e
5 *Mesoplodon europaeus*, Gervais' beaked whale	MeuSW7443	NA (USA)	Meu-a1/Meu-a1	AY579480 (1)	d
	MeuSW4120	NA (USA)	Meu-a1/Meu-a2	AY579481 (2)	d
6 *Mesoplodon ginkgodens*, ginkgo-toothed beaked whale	Mgin01	NP (Taiwan)	Mgin-a1/Mgin-a2	AY579482 (1)	g
	—	—	—	AY579483 (2)	g
7 *Mesoplodon grayi*, Gray's beaked whale	Mgr01	SP (NZ)	Mgr-a1/Mgr-a1	AY579484 (1)	e
	MgrH04	SP (NZ)	Mgr-a1/Mgr-a2	AY579485 (2)	e
8 *Mesoplodon hectori*, Hector's beaked whale	MheNZ02	SP (NZ)	Mhe-a1/Mhe-a2	AY579486 (1)	e
	—	—	—	AY579487 (2)	e
9 *Mesoplodon layardii*, straptooth whale	Mlay10	SP (NZ)	Mlay-a1/Mlay-a2	AY579488 (1), AY579489 (2)	e/v
	Mlay09	SP (NZ)	Mlay-a2Mlay-/a3	AY579490 (3)	e/v
10 *Mesoplodon mirus*, True's beaked whale	MmiSW4968	NA (USA)	Mmi-a1/Mmi-a1	AY579491 (1)	d
	MmiSW4972	NA (USA)	Mmi-a1/Mmi-a2	AY579492 (2)	d
11 *Mesoplodon peruvianus*, lesser beaked whale	MpeLAM95654	NP (USA)	Mpe-a1Mpe-a2	AY579493 (1)	k
	—	—	—	AY579494 (2)	k
12 *Mesoplodon stejnegeri*, Stejneger's beaked whale	MstTSM30135	NP (Japan)	Mst-a1/Mst-a2	AY579496 (1), AY579497 (2)	h
	MstSW9491	NP (USA)	Mst-a2/Mst-a3	AY579498 (3)	d
13 *Mesoplodon traversii*, spade-toothed whale*	—	—	—	—	—
14 *Mesoplodon perrini*, Perrin's beaked whale	USNM504259	NP (USA)	Mpi-a1/Mpi-a1	AY579495 (1)	d/m,2
	TMMC-C75	NP (USA)	Mpi-a1/Mpi-a1	—	d/n,2
15 *Indopacetus pacificus*, Longman's beaked whale	—	—	—	—	—
16 *Hyperoodon ampullatus*, northern bottlenose whale	HamMD	NA (Canada)	Ham-a1/Ham-a1	AY579499 (1)	w
	HamSH9717	NA (Canada)	Ham-a1/Ham-a1	—	q
17 *Hyperoodon planifrons*, southern bottlenose whale	Hpl01	SP (NZ)	Hpl-a1/Hpl-a2	AY579501 (2)	e
	Hpl02	SP (NZ)	Hpl-a1/Hpl-a1	AY579500 (1)	e
18 *Ziphius cavirostris*, Cuvier's beaked whale	Zca06	SP (NZ)	Zca-a1/Zca-a1	AY579504 (1)	e
	ZcaSW1120	NP (USA)	Zca-a1/Zca-a2	AY579505 (2)	d

Table 1. Continued

Species	Sample code	Geographic origin	Alleles	GenBank accession no.	Source
19 *Tasmacetus shepherdi*, Shepherd's beaked whale	Tsh01	SP (NZ)	Tsh-a1Tsh-a2	AY579503 (2)	e
	Tsh02	SP (NZ)	Tsh-a1/Tsh-a1	AY579502 (1)	e
20 *Berardius arnuxii*, Arnoux's beaked whale	Bar02	SP (NZ)	Bar-a1Bar-/a1	AY579506 (1)	e
	—	—	—	—	—
21 *Berardius bairdii*, Baird's beaked whale	—	—	—	—	—
	—	—	—	—	—

Aligned sequence files for the mtDNA loci and further information on the specimens used to generate these sequences are available from http://www.dna-surveillance.auckland.ac.nz. H, holotype specimen; IO, Indian Ocean; NA, North Atlantic; NP, North Pacific; NZ; New Zealand; SP, South Pacific; TP, Tropical Pacific.

* *Mesoplodon traversii* was resurrected and recognized as synonymous with *M. bahamondi* based on results of phylogenetic comparisons using the mtDNA reference database and methods discussed in this article (van Helden et al. 2002).

Source of specimens and samples:

a B. Reid, Scottish Agricultural College, Wildlife Unit, Inverness, Scotland, UK
b D. Pemberton, Tasmanian Museum and Art Gallery, Hobart, Tasmania, Australia
c A. van Helden, Museum of New Zealand Te Papa Tongarewa, Wellington, New Zealand
d A. E. Dizon, K Robertson, NMFS South West Fisheries Science Centre, California, USA
e University of Auckland (courtesy of field staff, NZ Department of Conservation and A. van Helden, Museum of New Zealand)
f P. Jepson, Natural History Museum, London, England, UK
g J. Y. Wang, Department of Zoology, McMaster University, Hamilton, Ontario, Canada
h T. K. Yamada, National Science Museum, Tokyo, Japan
i C. Kemper, South Australian Museum, Adelaide, South Australia
j J. C. Reyes, ACOREMA/K. Van Waerebeek, Centre for Cetacean Research (CEPEC), Peru
k J. E. Heyning, Los Angeles County Museum of Natural History, California, USA
l G. P. Sanino-Vattier and J. Yañez, Centre for Marine Mammals Research-LEVIATHAN/Museo Nacional de Historia Natural, Santiago, Chile
m C. W. Potter and J. G. Mead, Smithsonian Institution National Museum of Natural History, Washington, DC, USA
n M. Haulena, The Marine Mammal Centre (TMMC), California, USA
o S. van Dyck, Queensland Museum, Brisbane, Queensland, Australia
p P. Agnelli, Natural History Museum, Zoological Section "La Specola" University of Florence, Italy
q S. K. Hooker and R. W. Baird, Biology Department, Dalhousie University, Halifax, Nova Scotia, Canada
r V. G. Cockcroft and G. Watson, Port Elizabeth Museum, Cape Province, South Africa
s R. W. Baird, Marine Mammal Research Group, Box 6244, Victoria, British Columbia, Canada
t K. Evans, School of Zoology, University of Tasmania, Hobart, Tasmania, Australia
u I. Christensen, Marine Research Institute, Nordnes, Norway
v J. Lilley, Marine Watch, Christchurch, New Zealand
w M. Dillon, Biology Department, Dalhousie University, Halifax, Nova Scotia, Canada

GenBank sequences

1 van Helden et al. (2002)
2 Dalebout et al. (2002)
3 Dalebout et al. (2003)
4 Dalebout et al. (2001)
5 Milinkovitch et al. (1994)
6 Arnason and Gullberg (1996)

[a] Shorter mtDNA control region sequences from these specimens were published by Henshaw et al. (1997).

[b] Shorter mtDNA control region sequences from these specimens were published by Dalebout et al. (1998).

and Figure 1 for a primer map. The shorter fragments nest within the larger fragments characterized for the mtDNA loci and cover the most variable segments (see below). PCR amplification and sequencing of these smaller DNA fragments followed standard protocols, except for the addition of bovine serum albumin (BSA) solution (0.5–1 μg/μl) to overcome the effect of inhibitors (Pääbo 1990). Amplification of nuclear fragments was not attempted from tooth and bone specimens. PCR products were sequenced in both directions for at least one specimen per species on an ABI 377 or modified ABI 373 automated sequencer (Applied Biosystems Inc.) using BigDye dye terminator chemistry. All reference sequences have been deposited in GenBank (see Table 1 for accession numbers).

Sequences were aligned using the program PILEUP, available in the GCG package (Deveraux et al. 1984), with a gap penalty of 2 and extension penalty of 0.3 and further checked by eye. For the nDNA dataset, variable sites were assessed by eye for each beaked whale species. Heterozygous individuals were inferred from "double peaks" (i.e., the occurrence of two different nucleotides in the same position in the sequence, indicating the presence of two different alleles) using the program Sequencher 3.1.1 (Gene Codes Corp.), and confirmed through comparison of forward and

Table 2. Primers used for amplification of sequence fragments from modern and historical specimens

	Primer sequence	Reference
mtDNA control region		
M13-Dlp1.5-L	5′-TGTAAAACGACGGCCAGTTCACCCAAAGCTGRARTTCTA-3′	Dalebout et al. (1998)
Dlp5-H	5′-CCATCGWGATGTCTTATTTAAGRGGAA-3′	Dalebout et al. (1998)
Dlp4-H	5′-GCGGGWTRYTGRTTTCACG-3′	Baker CS, unpublished
Dlp10-L	5′-CCACAGTACTATGTCCGTATT-3	Baker et al. (1993)
mtDNA cytochrome *b*		
GLUDG-L	5′-TGACTTGAARAACCAYCGTTG-3′	Palumbi (1996)
CB2-H	5′-CCCTCAGAATGATATTTGTCCTCA-3′	Palumbi (1996)
CB1-L	5′-CCATCCAACATCTCAGCATGATGAAA-3′	Palumbi (1996)
CYBMF-L	5′-GAACTATAAGAACACTAATGACCAA-3′	Dalebout (2002)
CYBMR-H	5′-TGATTCAGCCATAGTTAACGTCTCGAC-3′	Dalebout (2002)
Nuclear actin intron		
ACT3-F	5′-GGTTATCTGATGTATTCC-3′	Baker CS and Palumbi SR, unpublished
ACT1385-R	5′-CTTGTGAACTGATTACAGTCC-3′	c/o C. Conway; Baker et al. (1998)
ACT5-F	5′-CCACTACTTTAGGCAG-3′	Baker CS and Palumbi SR, unpublished
M13-ACT5R-R	5′-TGTAAAACGACGGCCAGTCTGCCTAAAGTAGTGG-3′	Baker CS and Palumbi SR, unpublished

reverse sequences. Alleles were identified either directly from homozygous individuals or by "subtracting" from heterozygotes using the principle of parsimony (i.e., for a given species most alleles from slowly evolving nuclear loci are likely to differ by only 1 or 2 bp). While only two individuals per species are included in the nDNA database presented here, actin intron sequences were amplified from multiple specimens for the majority of beaked whale species. These additional sequences, generated for other analyses, were used to assist in the identification of alleles (Dalebout 2002).

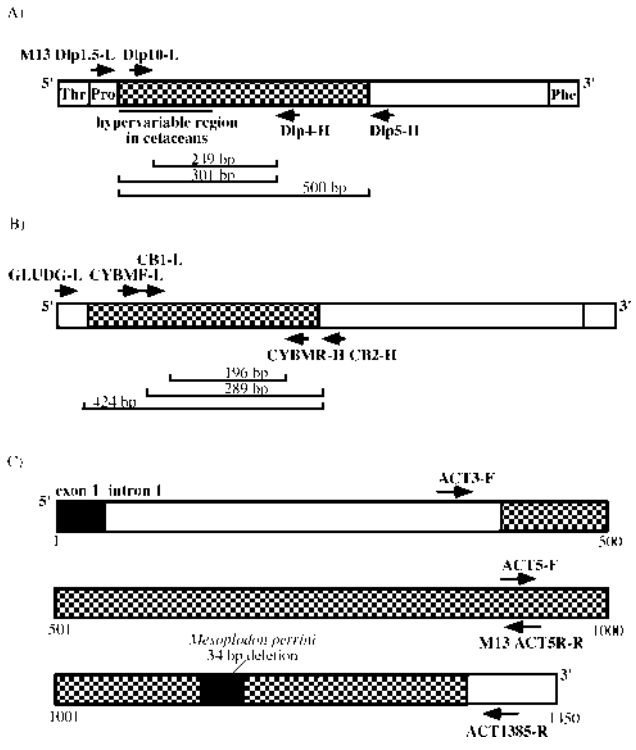


Figure 1. Primer map for (a) mtDNA control region, (b) mtDNA cytochrome *b*, and (c) nuclear actin intron 1. See Table 2 for primer sequences.

Phylogenetic Analyses and Identification of Test Specimens

For identification of test specimens, control region and cytochrome *b* datasets were analyzed separately. Inclusion of representatives from other cetacean families did not affect the results of phylogenetic reconstruction for the mtDNA or nDNA datasets at any nodes relevant to this article (Dalebout 2002), and as such, were not included in the analyses presented here. Instead, trees were rooted with Baird's beaked whale (*Berardius bairdii*), which appears to represent the ancestral lineage among the Ziphiidae based on morphological and molecular data (Arnason and Gullberg 1996; Dalebout 2002; Messenger and McQuire 1998).

For all datasets, appropriate models of molecular evolution were selected using the program ModelTest 3.06 (Posada and Crandall 1998). This program uses a series of hierarchical likelihood ratio tests to compare the fit of the nested general-time-reversible (GTR) family of nucleotide substitution models. Output from ModelTest was used to provide the starting parameters for maximum-likelihood (ML) reconstruction of phylogenetic relationships among sequences as implemented in the program PAUP* (Swofford 2000). ML reconstructions used the heuristic search option with random sequence addition (10 replicates) and subtree pruning-regrafting branch swapping. The statistical consistency of groupings was evaluated by ML bootstrap resamplings of the data (control region, 200; cytochrome *b*, 200; actin, 500). The number of bootstrap replicates differs between datasets due to limitations of time and availability of computers. Trees obtained from maximum parsimony (MP) analyses (heuristic search, random sequence addition-100 replicates, tree bisection-reconnection branch swapping,

insertion-deletions coded as missing data) did not differ from the ML trees at any nodes relevant to this article.

Consistency and Sensitivity of Species Identification

To investigate the consistency and sensitivity of the mtDNA loci for beaked whale species identification, aligned files of reference sequences were subjected to a series of deletion subsamplings. Starting with full-length sequences, the files were reduced 50 bp at a time to find the minimum sequence length required for reliable assignment of test specimens to species. Deletion series were performed from both the 5′ and 3′ ends of the alignment to determine the effect of sequence position within the fragment. To allow identification of beaked whale-specific insertion-deletions (indels) and other sequence landmarks, files used for these analyses included representatives of other cetacean species (length of aligned control region file with outgroups, 450 bp). A bootstrap score of $\geq$80% was chosen as the cutoff for positive "species identification" in this simulation. Although the choice of cutoff value was arbitrary, a score of $\geq$80% was considered a sufficient indication of robust support for the node uniting the reference sequences for any one species. In a similar application, Hillis and Bull (1993) found that a bootstrap score of $\geq$70% corresponded to a probability of $\geq$95% that the corresponding clade was valid. At each stage of deletion subsampling, the robustness of species-specific groupings was assessed by 1000 neighbor-joining bootstrap resamplings of the data using PAUP*, with appropriate models of evolution selected by ModelTest 3.06. The distribution of variable sites and inferred number of mutations at each site (steps) were calculated using the program MacClade 4.0 (Maddison and Maddison 1992) based on the relationships among ziphiids suggested by the ML trees constructed previously.

Results

Reference Database of mtDNA Sequences

The mtDNA control region and cytochrome *b* fragments were amplified and sequenced to compile a comprehensive database for the 20 beaked whale species recognized at the start of this study (Rice 1998). However, sequences derived from putative specimens of Hector's beaked whale (*Mesoplodon hectori*) from the North Pacific (Henshaw et al. 1997) and the Southern Hemisphere (Dalebout et al. 1998) did not group together in phylogenetic analyses as would be expected if they represented the same species. Additional analyses confirmed that the north Pacific specimens represented a new species of beaked whale, *Mesoplodon perrini*, bringing the total in this family to 21 (Dalebout et al. 2002) (see Discussion).

The mtDNA reference database consists of two sets of aligned sequences for all 21 beaked whale species; the 5′ end of the control region (consensus length 437 bp), and the 5′ end of the cytochrome *b* gene (consensus length 384 bp). The consensus length of each alignment was defined by the longest sequences available for each species (except those known to date only from osteological material from which full-length mtDNA sequences could not be amplified). Where possible, two representatives of each species were included, obtained from different geographic regions within their range (Table 1A,B). In four cases, sequences were derived from holotype specimens (*Indopacetus pacificus*, *Mesoplodon ginkgodens*, *M. perrini*, and *M. traversii*).

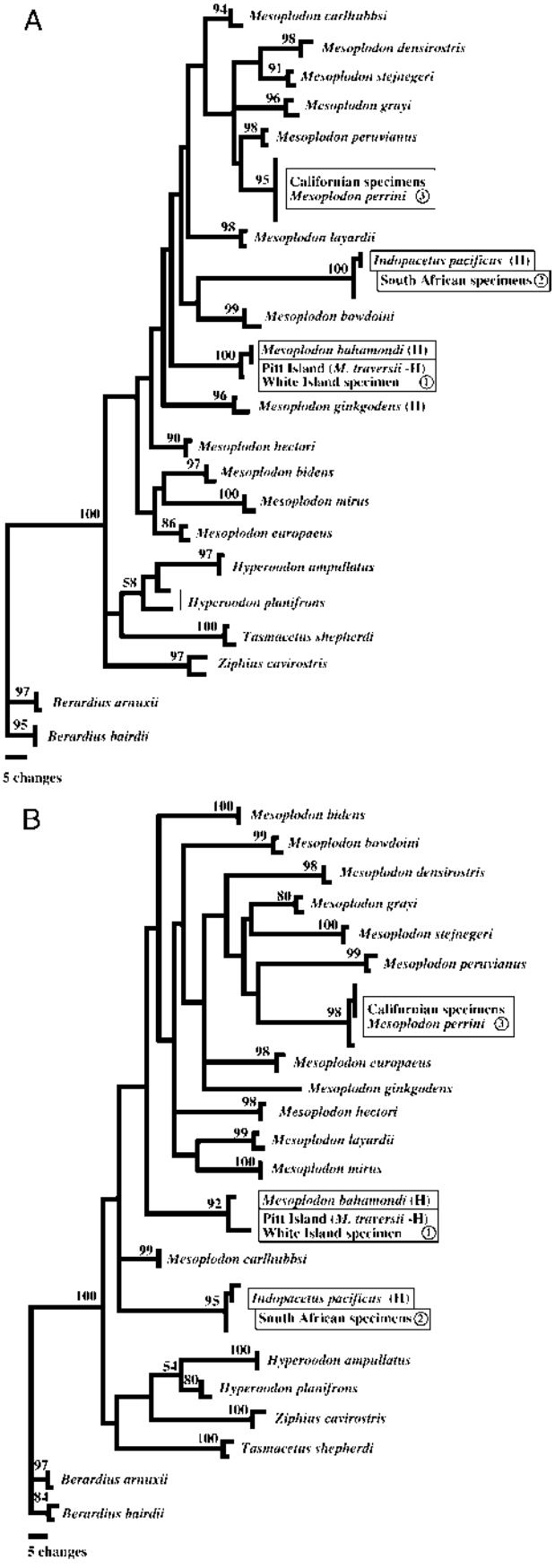
Mitochondrial Loci: Genetic Diversity and Species Distinctiveness

Among the beaked whales, the control region alignment required only minor adjustments for indels. No indels were found in the cytochrome *b* alignment, as expected for sequences coding for a functional protein. There was no evidence of stop codons or other interruptions to the reading frame indicative of possible nuclear pseudogene sequences (e.g., Lopez et al. 1994). For both loci, patterns of nucleotide substitution were consistent with those reported by other researchers (e.g., Arnason and Gullberg 1996; Arnason et al. 1993; Baker and Medrano-Gonzalez 2002), and sequences obtained from different PCRs using different primer pairs were unambiguous. Contrary to what has been reported for other mammals (Lopez et al. 1997), interspecific divergence among beaked whales was greater for cytochrome *b* (average 13.19%; range 2.90–19.43%) than for the control region (average 8.57%; range 3.37–20.49%). For intraspecific diversity, this relationship was reversed (control region: average 0.85%; range 0.23–3.10%; cytochrome *b*: average 0.51%; range 0–1.15%).

Phylogenetic analyses of the mtDNA reference sequences confirmed the alpha taxonomy of all 21 beaked whale species described to date (Figure 2a,b). Although the validity of some beaked whale species has been questioned due to the subtlety of diagnostic morphological features (e.g., McCann 1962a,b; Moore 1968), each was found to be genetically distinct. Reference sequences representing each species grouped together with high bootstrap support, species-specific lineages were distinguished from other lineages by a series of synapomorphic nucleotide substitutions (as reflected by branch lengths), and sister taxa were reciprocally monophyletic. Only the control region sequences of the southern bottlenose whale (*Hyperoodon planifrons*) were weakly paraphyletic with those of its congener, the northern bottlenose whale (*H. ampullatus*) (Figure 2a) due to a deep intraspecific divergence at this locus among specimens of the former species (Dalebout 2002). A reciprocally monophyletic relationship was suggested by cytochrome *b* (Fig. 2b) and by analyses combining sequences from both loci (not shown).

Mitochondrial Loci: Consistency and Sensitivity of Species Identification

Mapping of variable sites inferred from the ML tree showed that the first 160 bp of the 5′ end of the beaked whale mtDNA control region forms a hypervariable section (Figure 3a, top panel). The remaining 3′ portion of this

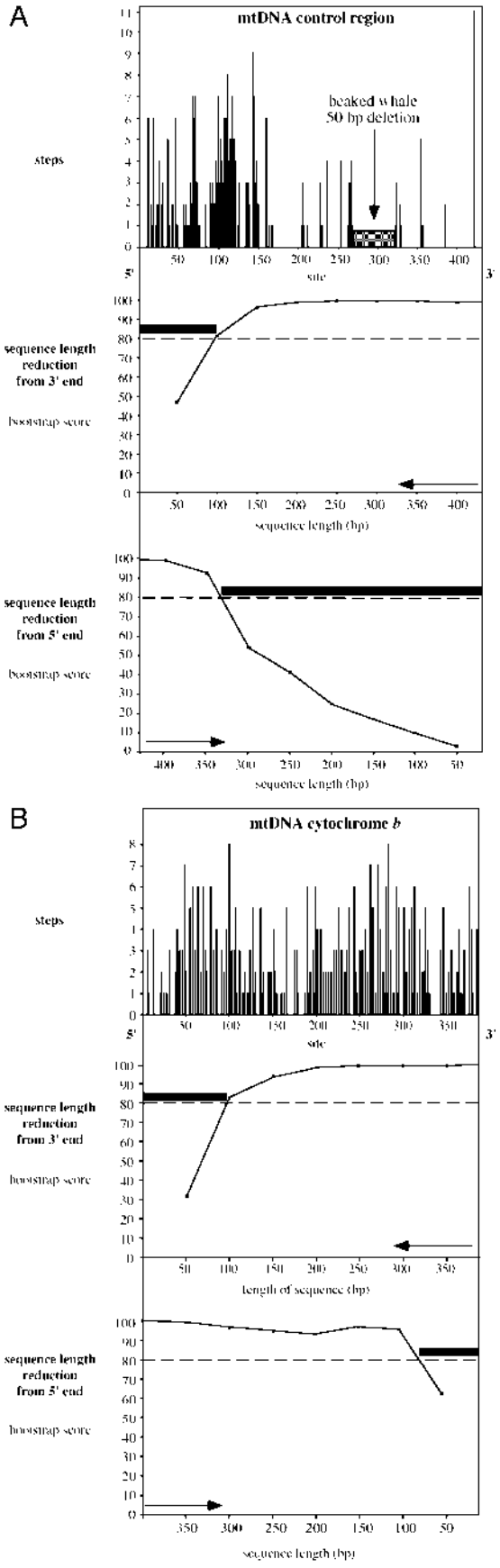


fragment was far more conserved and included a diagnostic 50 bp deletion (Dalebout et al. 1998). A similar, albeit less extreme distribution of variable sites has been documented for this locus in other cetaceans (Baker and Medrano 2002; Hoelzel et al. 1991). The smaller sequence fragments obtained from tooth and bone specimens either spanned the hypervariable portion of this locus or a significant portion of it. Results from deletion subsampling indicated that much of the phylogenetic signal was contained in the 5′ hypervariable segment. For beaked whale "species identification" (i.e., robust assignment of "test" specimens to species), a minimum of 100 bp from the 5′ end of the control region would be required (bootstrap score $\geq$80%) (Figure 3a, middle panel). For the 3′ end, approximately 350 bp would be required to obtain a similar consistency (Figure 3a, bottom panel).

In comparison, variable sites were more evenly distributed in cytochrome *b* (Figure 3b, top panel). The wider distribution of variable and informative sites at this locus was reflected in the results of deletion subsampling. A 100 bp segment from the 5′ end would be sufficient for reliable beaked whale species identification (Figure 3b, middle panel), while approximately 75 bp of the 3′ end would be required for similar consistency (Figure 3b, bottom panel). The majority of changes in cytochrome *b* were at the third position (85.37%), followed by the first position (10.26%), with second position sites least likely to experience a substitution (4.37%). Similar patterns of cytochrome *b* polymorphism have been found in other cetacean families (Arnason and Gullberg 1994; LeDuc et al. 1999).

←

Figure 2. Phylogenetic relationships among the 21 described beaked whale species (Ziphiidae), based on (a) mtDNA control region sequences (437 bp) and (b) cytochrome *b* sequences (384 bp) from ML analyses. Numbers above internal nodes indicate bootstrap values $\geq$50%. All described species are represented by two reference specimens where possible. Although well suited to answering questions of species identity, these rapidly evolving mitochondrial loci offered little resolution of higher-level relationships in this group. Boxed taxon names highlight cases demonstrating the utility and power of the beaked whale DNA database (see Discussion): (1) partial specimens of the species known previously as *Mesoplodon bahamondi*, linked to the only known specimen of *M. traversii* through mtDNA analysis (van Helden et al. 2002); (2) mtDNA identification of new specimens of *Indopacetus pacificus* as a result of which the external appearance of this species was finally revealed (Dalebout et al. 2003); and (3) specimens of *M. perrini*, a new species of living beaked whale discovered through phylogenetic analysis of mtDNA sequences (Dalebout et al. 2002). H, holotype included among specimens analyzed. ML scores (−ln L), control region = 2318.40; cytochrome *b* = 2467.00. Models of evolution; control region − HKY + I + G, Pinvar = 0.6388, α = 0.8746; cytochrome *b* − TrN + I + G, Pinvar = 0.5309; α = 1.9451.



Reference Database of nDNA Sequences

Actin intron fragments were amplified and sequenced for 17 of the 21 species (Table 1C). As with the mtDNA datasets, two representatives of each species were included where possible. Nuclear sequences were obtained only from species for which fresh tissue samples were available. The nDNA reference database consists of aligned sequences (consensus length 925 bp), with each beaked whale species represented by two (and in some cases three) alleles. Where both reference specimens for a species were homozygous for the same allele, two copies of that allele were included in the database to reflect sampling effort.

nDNA: Genetic Diversity and Species Distinctiveness

The overall structure and low number of indels observed among the beaked whale actin intron sequences were similar to those reported for baleen whales (Palumbi and Baker 1994). Average interspecific pairwise sequence divergence was 1.05% (range 0.11–2.01%) and average intraspecific diversity was 0.24% (range 0% [i.e., all individuals screened were homozygous for the same allele] to 0.66%). All alleles obtained were species specific and, for the majority of species, grouped together to the exclusion of alleles representing other species (i.e., monophyly of alleles within species) (Figure 4). While bootstrap scores for some species-specific clades were low, the overall consistency index was high (as obtained through both MP analyses and MP evaluation of the ML tree). Several internal branches received additional support through the presence of unique indels, the distribution of which was mapped onto the ML tree after construction (Figure 4, white crossbars). For some closely

←

Figure 3. Consistency and sensitivity of mtDNA control region and cytochrome *b* sequences for beaked whale species identification: (a) 450 bp of the 5′ end of the control region (as aligned to sequences from other cetacean taxa), three panels; and (b) 384 bp of the 5′ end of cytochrome *b*, three panels. Top panels show the distribution of variable sites for each locus based on ML trees. Middle panels show the effect of sequence length reduction from the 3′ end. Bottom panels show the effect of sequence length reduction from the 5′ end. The dashed line in these panels indicates the level of consistency considered acceptable for species identification (≥80% bootstrap score; averaged over all beaked whale species). Gray bars highlight the minimum length of sequence required to attain this level of consistency. Site is the nucleotide position in the alignment. Steps is the number of mutations occurring at that site across all beaked whale taxa in the alignment. Gray box in the top panel of control region figure (a) indicates the position of the 50 bp deletion specific to beaked whales. Bootstrap scores obtained from neighbor-joining analyses using the following models of evolution: control region, HKY + G, $\alpha = 0.4203$; cytochrome *b*, GTR + I + G, Pinvar = 0.4775, $\alpha = 1.2411$.

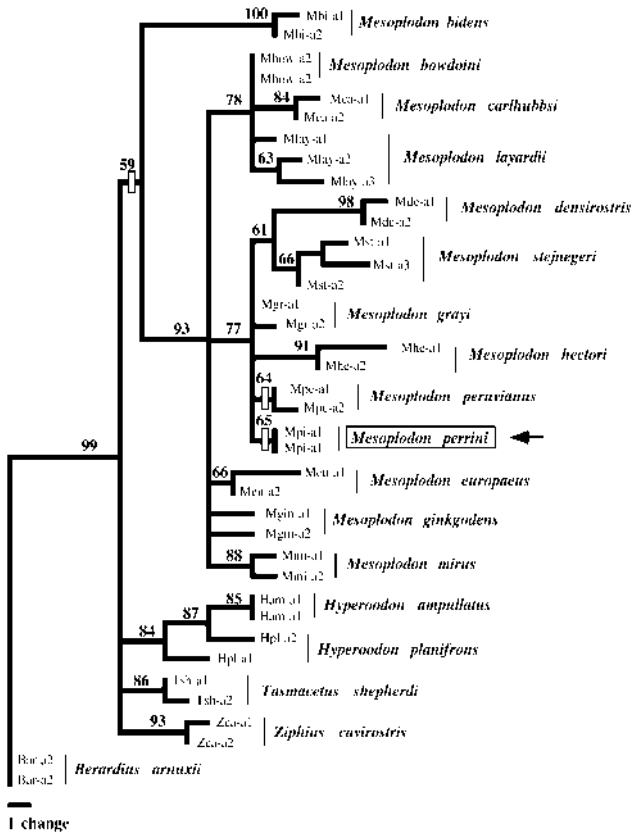


Figure 4. Phylogenetic relationships among beaked whale alleles from the nuclear actin intron (925 bp) from ML analyses (17 of 21 species represented). Branch termini represent alleles labeled as in Table 1C. Numbers above internal nodes indicate bootstrap scores ≥50%. White crossbars represent diagnostic indels. *Mesoplodon perrini* (arrow) is distinguished from all other beaked whales by a unique 34 bp deletion. A single base pair deletion unites all species in the genus *Mesoplodon* and another distinguishes *M. peruvianus*. ML score (–ln L) = 1854.76. Model of evolution, HKY. Consistency index (CI) = 0.936, retention index (RI) = 0.962 (based on MP evaluation of ML tree).

related species, alleles did not form species-specific clades due to a lack of synapomorphies (e.g., Gray's beaked whale [*Mesoplodon grayi*]). A paraphyletic relationship was observed among alleles from the two *Hyperoodon* species.

The utility of a parallel nDNA database for beaked whale species identification is twofold. First, hybrid animals cannot be detected unless biparentally inherited loci are screened. Second, in the absence of female-mediated gene flow, populations may appear monophyletic at mtDNA loci even if male-mediated gene flow continues. Evaluation of phylogenetic patterns at nDNA loci can therefore help assess levels of genetic isolation among populations and species. With the cessation of all gene flow, lineage sorting in a population will result eventually in all alleles tracing back to a single lineage (coalescence). Neutral theory predicts that the evolution of monophyly will be four times slower in nuclear than in mitochondrial genes. This is due largely to the slower rate of genetic drift in autosomal nuclear loci because of the fourfold difference in the effective population size. As such, recently diverged species may be monophyletic for mtDNA haplotypes but not nDNA alleles.

Although generally similar in external morphology, sufficient evolutionary time appears to have elapsed since the divergence of most beaked whale species for these lineages to have attained monophyly at both mitochondrial and nuclear loci. Similar findings have been reported for several baleen whale species (family Balaenopteridae) (Palumbi et al. 2001). In contrast, the north Pacific white-sided dolphin (*Lagenorhynchus obliquidens*) and southern hemisphere dusky dolphin (*L. obscurus* [family Delphinidae]) were found to be paraphyletic in analyses of nuclear butyrophilin intron sequences (Hare et al. 2002). Given the patterns of comparative intraspecific diversity and interspecific divergence observed among beaked whales at the mtDNA [see also Dalebout (2002)], our findings of nuclear actin monophyly for most species are consistent with the predictions of the "three-times rule." This rule states that, on average, most nuclear loci will be monophyletic when the branch length leading to the mtDNA sequences of a species is three times longer than the average mtDNA sequence diversity observed within that species (Hare 2001; Palumbi et al. 2001).

Discussion

DNA taxonomy, together with the often-associated but less contentious concept of Web-based taxonomy, has recently become a topic of major discussion. Bisby (2000) and Godfray (2002) put forward a plea that Web-based technology be used to provide wider access to taxonomic information for biodiversity research. In response, Tautz et al. (2002, 2003) suggested that many of the problems currently faced by traditional morphology-based taxonomy could be further resolved by using DNA sequences as a universal reference standard. Extending this idea, Hebert et al. (2003) proposed that DNA "barcodes" (sequence profiles based on a single genetic locus) could be used as a global bioidentification system for all animals. Recommendations such as these, suggesting that "DNA taxonomy can provide a new scaffold for our accumulated taxonomic knowledge and a reliable tool for species identification and description" (Tautz et al. 2002) have resulted in vociferous debate (e.g., Dunn 2003; Lipscomb et al. 2003; Seberg et al. 2003). However, many of these objections appear to be fueled by the mistaken notion that adoption of a molecular taxonomic approach will lead to all morphological information in species descriptions being discarded.

Components of a Molecular Taxonomy

Given our experience with the family Ziphiidae, and that of others and ourselves with the order Cetacea as a whole (Baker et al. 2003; Ross et al. 2003), we offer the following components for consideration as the basis of a molecular taxonomy for all organisms.

Comprehensiveness

All described species in a group should be represented. Multiple representatives from different geographic locations should be included for each species to reflect the full range of genetic diversity. The number of specimens required for this will differ among taxa, depending on the levels of intraspecific genetic diversity and divergence from other closely related species. Where databases are to be established for several loci, the same suite of reference specimens should ideally be used to generate all sequences. Unfortunately this was not possible for the beaked whale databases presented here (Table 1) due opportunistic methods of data collection and the nature of some source material.

Validation

DNA sequence data should be obtained from holotype specimens wherever possible. Otherwise DNA sequences should be obtained only from validated specimens; those examined by experts for that group and from which diagnostic skeletal material or photographs have been collected (Dizon et al. 2000). Native DNA may no longer be present in holotype specimens due to age, museum preparation methods, and storage conditions. In these cases we suggest that validated specimens from the type locality should be used to stand in for the holotype (e.g., Dalebout et al. 2002; Tautz et al. 2003; Winston 1999). Under the International Code of Zoological Nomenclature (ICZN), a neotype can be designated only if the holotype specimen has been destroyed or lost. We propose that where native DNA cannot be obtained from a holotype specimen, an official "DNA neotype" should be formally designated for the purposes of molecular taxonomy.

Locus Sensitivity

The locus or loci used for such analyses should be appropriate to the group of interest. Although use of a single, universal genetic marker might be desirable (e.g., cytochrome oxidase I; Hebert et al. 2003), we feel that such an approach is unlikely to be sufficiently sensitive for all organisms. Given the vast differences in molecular rates and patterns, and the lack of a single definition of what constitutes a species agreeable to all biologists, there is no single molecular marker suitable for the identification of all organisms. The mtDNA control region or cytochrome *b* is likely to be more suitable for most mammalian groups, but not all taxa will conform to the same patterns of molecular evolution. For example, in a recent comparison of divergence rates among mammalian mtDNA loci, cytochrome *b* was ranked 13th out of 18 on average (where a rank of 1 was the most divergent; Lopez et al. 1997). The 5′ end of the control region was ranked second on average in these same comparisons. For the Ziphiidae, however, this pattern of divergence appears to be reversed. Minimum sequence length required for robust species identification should also be considered. It may not be possible to obtain full-length sequences from historic tooth and bone material (e.g., holotype specimens), but short fragments may still be sufficient depending on locus sensitivity. Deletion subsampling analyses performed here allowed us to determine the minimum length of mtDNA sequences required for robust beaked whale species identification. With such issues in mind, we suggest that molecular taxonomists work closely with species specialists to select and develop appropriate loci and sampling programs. Ultimately a truly universal molecular taxonomy will require a hierarchical suite of markers to fully resolve the links between the roots and branch tips in the tree of life.

Genetic Distinctiveness and Exclusivity

Sequences from specimens assumed to represent a given taxon should form monophyletic lineages (Milinkovitch et al. 2002). Lineages representing the same species are then expected to group together to the exclusion of lineages representing other described species, with synapomorphic nucleotide substitutions distinguishing these lineages from one another. Bootstrap scores and other measures of robustness are expected to be high for species-specific clades in most cases.

Concordance

Phylogenetic analyses of multiple loci, together with assessment of morphological features, behavior, and geographic distribution, should yield concordant results. Within this framework, a molecular taxonomic approach could also lead to the discovery of new species. If all known species in a group are represented in a molecular phylogeny, the discovery of a new lineage, represented by one or more specimens, must be considered worthy of further scrutiny (Figure 5). This scrutiny would involve tests of concordance to determine the appropriate level for taxonomic classification (i.e., species, subspecies, evolutionarily significant units) as follows (Avise 2000; Avise and Ball 1990): (1) concordance across sequence characters within a genetic locus (putative gene-tree clades should be robust, as reflected in high bootstrap scores) leading to conclusive exclusion; (2) concordance in these genealogical patterns across multiple loci, both mitochondrial and nuclear (i.e., gene-tree partitions should reflect phylogenetic partitions at the species level); (3) concordance with biogeographical patterns; and (4) reexamination of the morphology of the specimen(s) in question.

Universal Accessibility and Curation

Databases of DNA reference sequences and information on validated source specimens should be easily updateable and accessible to everyone, together with standardized phylogenetic programs for species identification (Ross et al. 2003). Such universal access can be facilitated through the Worldwide Web, with molecular and morphological data curated by species specialists for each group (e.g., Bisby 2000; Godfray 2002). Following this recommendation, the beaked whale mtDNA reference database presented here, together with associated information, will be available

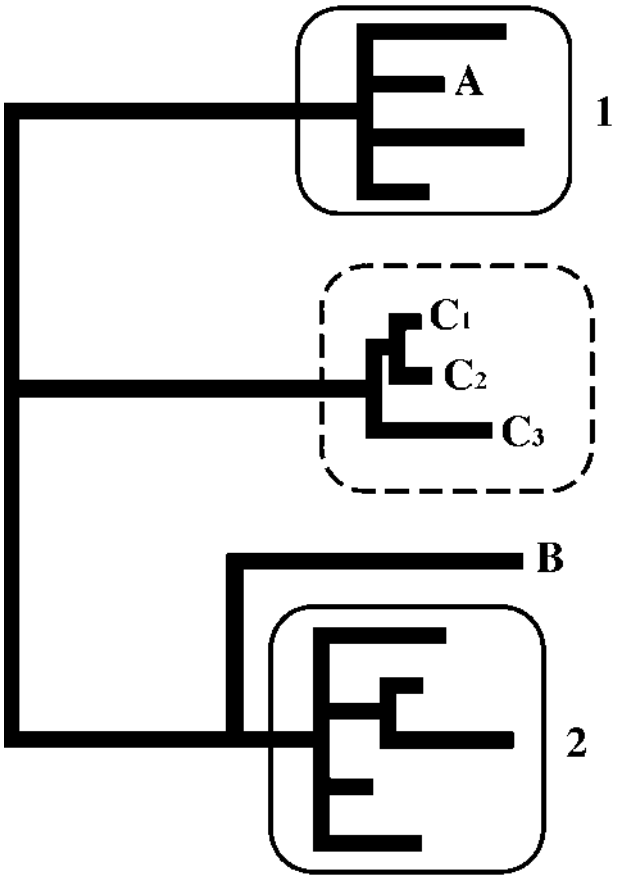


Figure 5. In a hypothetical phylogenetic tree, the circled clades, 1 and 2, would be considered species according to branching patterns predicted under the phylogenetic species concept (e.g., Simpson 1961). Test specimen A would be identified as a member of species 1 because the sequence nests within the known "reference" sequences for that species. Test specimen B could be considered a possible member of species 2 with an as yet unseen level of genetic divergence, perhaps the result of geographic isolation or variation. The group of test specimens labeled $C_{1–3}$ (dashed-line box) falls outside the groupings of any known species and shows a similar level and pattern of divergence. As such, these specimens would warrant further examination as possible representatives of new, previously unrecognized species.

through the Web-based species identification program, DNA Surveillance (http://www.dna-surveillance.auckland.ac.nz) (Ross et al. 2003). This program implements a phylogenetic approach to species identification of whales, dolphins, and porpoises using a hierarchical set of aligned sequences from the mtDNA control region and cytochrome *b* for the majority of recognized species in this order (Baker et al. 2003; Ross et al. 2003). Using this approach, we aim to bridge the gap between the nonarchived international genetic database, GenBank, and new Web-based archiving initiatives of museums and other institutions (e.g., Bisby 2000; Bisby et al. 2002; Godfray 2002; Knapp et al. 2002).

Applications and Discoveries

The following three discoveries, discussed in detail elsewhere (Dalebout et al. 2002, 2003; van Helden et al. 2002), are summarized here as examples of the utility of a molecular taxonomic approach for beaked whales.

Fragmented Evidence and Orderly Nomenclature

The first discovery concerns three partial specimens: a beach-cast skull collected from Robinson Crusoe Island, Chile, in 1986, described as a new species (*Mesoplodon bahamondi*) by Reyes et al. (1996); a beach-cast skull collected from White Island, New Zealand, in the 1950's, identified first as *M. ginkgodens* (Baker and van Helden 1999) and later as *M. bahamondi* (Baker 2001) from morphology; and a lower jaw with teeth collected from Pitt Island, New Zealand, in 1872, registered as *M. layardii* in the Museum of New Zealand Te Papa Tongarewa. No lower jaw or teeth were collected for the two beach-cast skulls. These three partial specimens were linked and shown to represent the same species through phylogenetic analyses of mtDNA sequences (Figure 2a,b, number 1) (van Helden et al. 2002). A subsequent literature review revealed that the lower jaw had in fact been described as a new species by Gray (1874), but had been synonymized with *M. layardii* by subsequent researchers. With all three specimens now recognized as the same species, the original name, *M. traversii*, takes precedence over *M. bahamondi* (van Helden et al. 2002). Known to date only from these three specimens and with no records of its external appearance, *M. traversii* is the rarest of all living cetaceans. Sequences from the holotype (Pitt Island) and Chilean specimen are now included in the DNA reference database (Table 1).

Correct Identification and Revelation of External Appearance

Until recently, Longman's beaked whale (*Indopacetus pacificus*) was known only from two beach-cast skulls collected in Australia and Somalia in 1882 and 1956, respectively (Azzaroli 1968; Longman 1926). In 1976 and 1992, two juvenile beaked whales stranded in South Africa were identified as *H. planifrons* from external morphology (Ross 1984; Port Elizabeth Museum records). However, mtDNA typing of skeletal material from these specimens indicated instead that they represented *I. pacificus* (Figure 2a,b; number 2) (Dalebout et al. 2003). Photographs taken of the South African animals at the time of stranding revealed the external appearance of this species for the first time (Dalebout et al. 2003). Further, comparisons to images of unidentified "tropical bottlenose whales" observed at sea in the tropical Indian and Pacific Oceans (e.g., Mörzer Bruyns 1971; Pitman et al. 1999) confirmed that these animals also represented *I. pacificus* (Dalebout et al. 2003). Overall, what was once considered to be the rarest of whales may be widely distributed and relatively common compared to many other ziphiids.

Discovery and Diagnosis of New Taxa

The third discovery concerns five beaked whales that stranded in southern California between 1975 and 1997 (Dalebout et al. 2002). Four of these animals were identified

as *M. hectori* from morphology (Mead 1981), while the fifth was identified as *Ziphius cavirostris* (National Marine Fisheries Service Southwest Fisheries Science Center records). While there are numerous records of the latter species from this region (Jefferson et al. 1993), *M. hectori* had been recorded previously only in the southern hemisphere. mtDNA sequences from the Californian specimens grouped together to the exclusion of all other recognized beaked whale species in phylogenetic analyses, including *M. hectori* and *Z. cavirostris*, suggesting that they could represent a new species (Figure 2a,b; number 3) (Dalebout et al. 2002). Morphological reevaluation supported this conclusion, although osteological features distinguishing this species from *M. hectori* are subtle, and only molecular characters are considered fully diagnostic (Dalebout et al. 2002). A new species, Perrin's beaked whale (*M. perrini*), was described and named on the basis of these results (Dalebout et al. 2002). Sequences from two of these specimens, including the holotype, are now included in the DNA reference database (Table 1).

The validity of this discovery was further confirmed by the results from phylogenetic analyses of nuclear actin sequences presented here. A unique 34 bp deletion distinguished the allele representing the two specimens of *M. perrini* screened for this locus (both homozygotes) from the alleles of all other beaked whales; a pattern reflected in the reconstruction of phylogenetic relationships among the actin intron sequences (Figure 4, arrow).

Conclusion

While recognizing the power of molecular or DNA taxonomy, we do not suggest, however, that taxonomy should be "impoverished" to reliance solely on genetic data for species descriptions (Lipscomb et al. 2003). Instead, we see DNA-based taxonomy as a natural extension of the modern synthesis of Darwinian evolution and Mendelian genetics. Species are the result of descent with modification, and molecular characters are the direct archive of this history (Zuckerkandl and Pauling 1965). If necessary, a molecular taxonomy could operate in a morphological vacuum (as is currently the case with many discoveries of new microorganisms) (e.g., Nee 2003), but only once a database of reference sequences derived from validated physical specimens has been established. For most organisms, however, such an extreme reductionism is clearly not required or desirable and would deny much of what is most interesting about the wealth of biological diversity with which we share the Earth. Nonetheless, we feel that a DNA-based approach, operating within a universal code of practice, can help overcome many of the problems facing taxonomic identification today.

As a result of our findings, the Smithsonian Institution's National Museum of Natural History (United States) has proposed allowing DNA typing of all cetacean holotypes in its collection. Given this precedent, we recommend that in addition to material collected traditionally for museums (osteological specimens, morphological descriptions, and life-history samples), the collection of genetic data should become standard for all cetaceans. Recognizing the pivotal role of stranding networks in cetacean research, we also stress the importance of maintaining and expanding such networks wherever possible.

Acknowledgments

We thank the people and institutions listed in Table 1 for collection and access to the specimens used to establish the beaked whale DNA reference database. We also thank the following for laboratory support, access to specimens, and/or valuable discussion: A. E. Dizon, W. F. Perrin, and K. Robertson, National Marine Fisheries Service, Southwest Fisheries Science Center; S. Van Dyck, Queensland Museum; J. E. Heyning, Los Angeles County Museum of Natural History; G. M. Lento, M. Lewis, E. Matisoo-Smith, J. Robins, A. G. Rodrigo, and D. J. Saul, University of Auckland; T. Kakuda and I. Nishiumi, National Science Museum, Tokyo; S. J. O'Brien and N. Yuhki and colleagues, Laboratory of Genomic Diversity, National Cancer Institute; V. Peddemors and the Sodwana Bay staff, KwaZulu-Natal Nature Conservation Services and Natal Sharks Board; C. W. Potter, Smithsonian Institution National Museum of Natural History; G. J. B. Ross, Environment Australia; G. Watson and staff, Port Elizabeth Museum, South Africa; K. van Waerebeek, Peruvian Center for Cetacean Research, Lima, Peru; J. C. Reyes, ACOREMA, Pisco, Peru; A. van Helden, Museum of New Zealand Te Papa Tongarewa; and A. N. Baker, New Zealand Department of Conservation. Special thanks to the many devoted field staff of the New Zealand Department of Conservation for collection of beach-cast and stranded specimens. Comments from three anonymous reviewers helped improve the manuscript. M. L. Dalebout was supported by a New Zealand Vice-Chancellors' Committee (NZVCC) William Georgetti scholarship, a University of Auckland doctoral scholarship, a Smithsonian Institution short-term visitor's fellowship, and a Killam postdoctoral fellowship. Financial support for the project was provided to C. S. Baker by the University of Auckland Research Council, New Zealand Lottery Board, International Fund for Animal Welfare, and New Zealand Marsden Fund.

References

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

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

Arnason U, Gullberg A, and Widegren B, 1993. Cetacean mitochondrial DNA control region: sequences of all extant baleen whales and two sperm whale species. Mol Biol Evol 10:960–970.

Avise JC, 2000. Phylogeography: the history and formation of species. Cambridge, MA: Harvard University Press.

Avise JC and Ball RM, 1990. Principles of genealogical concordance in species concepts in a biological taxonomy. Oxf Surv Evol Biol 7:45–67.

Azzaroli ML, 1968. Second specimen of *Mesoplodon pacificus*, the rarest living beaked whale. Monit Zool Ital (NS) 2:67–79.

Baker AN, 2001. Status, relationships and distribution of *Mesoplodon bowdoini* Andrews, 1908 (Cetacea: Ziphiidae). Mar Mamm Sci 17:473–493.

Baker AN and van Helden A, 1999. New records of beaked whales, genus *Mesoplodon*, from New Zealand (Cetacea: Ziphiidae). J R Soc N Z 29:235–244.

Baker AN, Smith ANH, and Pichler FB, 2002. Geographical variation in Hector's dolphin: recognition of a new subspecies of *Cephalorhynchus hectori*. J R Soc N Z 32:713–727.

Baker CS, Cipriano F, and Palumbi SR, 1996. Molecular genetic identification of whale and dolphin products from commercial markets in Korea and Japan. Mol Ecol 5:671–685.

Baker CS, Dalebout ML, Lavery S, and Ross HA, 2003. www.DNA-surveillance: applied molecular taxonomy for species conservation and discovery. Trends Ecol Evol 18:271–272.

Baker CS and Medrano-Gonzalez L, 2002. World-wide distribution and diversity of humpback whale mitochondrial DNA lineages. In: Cell and molecular biology of marine mammals (Pfeiffer CJ, ed). Malabar, FL: Krieger; 84–99.

Baker CS, Medrano-Gonzalez L, Calambokidis J, Perry A, Pichler F, Rosenbaum H, Straley JM, Urban-Ramirez J, Yamaguchi M, and Von Ziegesar O, 1998. Population structure of nuclear and mitochondrial DNA variation among humpback whales in the north Pacific. Mol Ecol 7:695–707.

Baker CS and Palumbi SR, 1994. Which whales are hunted? A molecular genetic approach to monitoring whaling. Science 265:1538–1539.

Baker CS, Perry A, Bannister JL, Weinrich MT, Abernethy RB, Calambokidis J, Lien J, Lambertsen RH, Urban Ramirez J, Vasquez O, et al., 1993. Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. Proc Natl Acad Sci USA 90:8239–8243.

Baker CS, Slade RW, Bannister JL, Abernethy RB, Weinrich MT, Lien J, Urban J, Corkeron P, Calmabokidis J, Vasquez O, et al., 1994. Hierarchical structure of mitochondrial DNA gene flow among humpback whales *Megaptera novaeangliae*, world-wide. Mol Ecol 3:313–327.

Bisby FA, 2000. The quiet revolution: biodiversity informatics and the Internet. Science 289:2309–2312.

Bisby FA, Shimura J, Ruggiero M, Edwards J, and Haeuser C, 2002. Taxonomy, at the click of a mouse. Nature 418:367.

Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheim-van Dillen PME, and van der Noorda J, 1990. Rapid and simple method for purification of nucleic acids. J Clin Microbiol 28:495–503.

Dalebout ML, 2002. Species identity, genetic diversity and molecular systematic relationships among the Ziphiidae (beaked whales) (PhD dissertation). Auckland, New Zealand: University of Auckland.

Dalebout ML and Baker CS, 2002. *Mesoplodon perrini* sp. n. Dalebout, Mead, Baker, Baker and van Helden 2002: notification of new species of living beaked whale (Ziphiidae) from the eastern north Pacific. Bull Zool Nom 59:284–286.

Dalebout ML, Hooker SK, and Christensen I. 2001. Genetic diversity and population structure among northern bottlenose whales, *Hyperoodon ampullatus*, in the western North Atlantic Ocean. Can J Zool 79:478–484.

Dalebout ML, Mead JG, Baker CS, Baker AN, and van Helden AL, 2002. A new species of beaked whale *Mesoplodon perrini* sp. n. (Cetacea: Ziphiidae) discovered through phylogenetic analyses of mitochondrial DNA sequences. Mar Mamm Sci 18:1–32.

Dalebout ML, Ross GJB, Baker CS, Anderson RC, Best PB, Cockcroft VG, Hinsz HL, Peddemors V, and Pitman RL, 2003. Appearance, distribution and genetic distinctiveness of Longman's beaked whale, *Indopacetus pacificus*. Mar Mamm Sci 19:421–461.

Dalebout ML, van Helden A, Van Waerebeek K, and Baker CS, 1998. Molecular genetic identification of southern hemisphere beaked whales (Cetacea: Ziphiidae). Mol Ecol 7:687–694.

Davis LG, Dibner MD, and Batteny JF, 1986. Basic methods in molecular biology. Amsterdam: Elsevier.

DeSalle R and Birstein VJ, 1996. PCR identification of black caviar. Nature 381:197–198.

Deveraux JP, Haeberli P, and Smithies O, 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395.

Dizon A, Baker CS, Cipriano F, Lento G, Palsboll P, and Reeves R, 2000. Molecular genetic identification of whales, dolphins and porpoises: proceedings of a workshop on the forensic use of molecular techniques to identify wildlife products in the marketplace. Technical Memorandum NMFS NOAA-TM-NMFS-SWFSC-286. La Jolla, CA: National Oceanic and Atmospheric Administration.

Dunn C, 2003. Keeping taxonomy based in morphology. Trends Ecol Evol 18:270–271.

Fuhrman JA, McCallum K, and Davis AA, 1992. Novel major archaebacterial group from marine plankton. Nature 356:148–149.

Giao PM, Tuoc D, Dung VV, Wikramanayake ED, Amato G, and MacKinnon JR, 1998. Description of *Muntiacus truongsonensis*, a new species of muntjac (Artiodactyla: Muntiacidae) from central Vietnam, and implications for conservation. Ann Conserv 1:61–68.

Godfray HCJ, 2002. Challenges for taxonomy. Nature 417:17–19.

Gray JE, 1874. Notes on Dr. Hector's paper on the whales and dolphins of the New Zealand seas. Trans N Z Inst VI:93–97.

Hare M, 2001. Prospects for nuclear gene phylogeography. Trends Ecol Evol 16:700–706.

Hare MP, Cipriano F, and Palumbi SR, 2002. Genetic evidence on the demography of speciation in allopatric dolphin species. Evolution 56:804–816.

Hebert PDN, Cywinska A, Ball SL, and deWaard JR, 2003. Biological identifications through DNA barcodes. Proc R Soc Lond Ser B 270:313–321.

Hennig W, 1966. Phylogenetic systematics. Urbana: University of Illinois Press.

Henshaw MD, LeDuc RG, Chivers SJ, and Dizon AE, 1997. Identification of beaked whales (family Ziphiidae) using mtDNA sequences. Mar Mamm Sci 13:487–495.

Hillis DM and Bull JJ, 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol 42:182–192.

Hoelzel RA, Hancock JM, and Dover GA, 1991. Evolution of the cetacean mitochondrial D-loop region. Mol Biol Evol 8:475–493.

Höss M and Pääbo S, 1993. DNA purification from Pleistocene bones by a silica-based purification method. Nucleic Acids Res 21:3913–3914.

Jefferson TA, Leatherwood S, and Webber MA, 1993. FAO species identification guide: marine mammals of the world. Rome: United States Environment Programme, Food and Agriculture Organization of the United Nations (FAO).

Knapp S, Bateman RM, Chalmers NR, Humphries CJ, Rainbow PS, Smith AB, Taylor PD, Vane-Wright RI, and Wilkinson M, 2002. Taxonomy needs evolution, not revolution. Nature 419:559.

LeDuc RG, Perrin WF, and Dizon AE, 1999. Phylogenetic relationships among the delphinid cetaceans based on full cytochrome *b* sequences. Mar Mamm Sci 15:619–648.

Lento GM, Mattlin RH, Chambers GK, and Baker CS, 1994. Geographic distribution of mitochondrial cytochrome *b* DNA haplotypes in New Zealand fur seals (*Arctocephalus forsteri*). Can J Zool 72:293–299.

Lipscomb D, Platnick N, and Wheeler Q, 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. Trends Ecol Evol 18:65–66.

Longman HA, 1926. New records of Cetacea, with a list of Queensland species. Mem Queensland Mus 8:266–278.

Lopez JV, Culver M, Stephens JC, Johnson WE, and O'Brien SJ, 1997. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. Mol Biol Evol 14:277–286.

Lopez JV, Yuhki N, Masuda R, Modi W, and O'Brien SJ, 1994. *Numt*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J Mol Evol 39:174–190.

Maddison WP and Maddison DR, 1992. MacClade (version 4.0): analysis of phylogeny and character evolution. Sunderland, MA: Sinauer Associates.

Malik S, Wilson PJ, Smith RJ, Lavigne DM, and White BN, 1997. Pinniped penises in trade: a molecular genetic investigation. Conserv Biol 11:1365–1374.

Matisoo-Smith E, Allen JS, Ladefoged TN, Roberts RM, and Lambert DM, 1997. Ancient DNA from Polynesian rats: extraction, amplification and sequence from single small bones. Electrophoresis 18:1534–1537.

McCann C, 1962a. The taxonomic status of the beaked whale *Mesoplodon hectori* (Gray)–Cetacea. Rec Dom Mus 4:83–94.

McCann C, 1962b. The taxonomic status of the beaked whale *Mesoplodon pacificus* Longman–Cetacea. Rec Dom Mus 4:95–100.

Mead JG, 1981. First records of *Mesoplodon hectori* (Ziphiidae) from the northern hemisphere and a description of the adult male. J Mamm 62:430–432.

Mead JG, 1989. Beaked whales of the genus *Mesoplodon*. In: Handbook of marine mammals, vol. 4 (Ridgway SH and Harrison R, eds). London: Academic Press; 349–430.

Messenger SL and McQuire JA, 1998. Morphology, molecules and the phylogenetics of cetaceans. Syst Biol 47:90–124.

Milinkovitch M, LeDuc R, Tiedemann R, and Dizon AE, 2002. Applications of molecular data in cetacean taxonomy and population genetics with species emphasis on defining species boundaries. In: Marine mammals: biology and conservation (Evans PGH and Raga JA, eds). Amsterdam: Kluwer Academic/Plenum; 325–359.

Milinkovitch M, Meyer A, and Powell JR, 1994. Phylogeny of all major groups of cetaceans based on DNA sequences from three mitochondrial genes. Mol Biol Evol 11:939–948.

Moon-van der Staay SY, De Wachter R, and Vaulot D, 2001. Oceanic 18S rRNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409:607–610.

Moore JC, 1968. Relationships among the living genera of beaked whales. Field Zool 53:209–298.

Mörzer Bruyns WFJ, 1971. Field guide of whales and dolphins. Amsterdam: C. A. Mees.

Nee S, 2003. Unveiling prokaryotic diversity. Trends Ecol Evol 18:62–63.

Pääbo S, 1990. Amplifying ancient DNA. In: PCR protocols: a guide to methods and applications (Innis MA, Gelfand DH, Sninsky JJ, and White TJ, eds). San Diego: Academic Press; 159–166.

Pace NR, 1997. A molecular view of microbial diversity and the biosphere. Science 276:734–740.

Palumbi SR, 1996. Nucleic acids II: the polymerase chain reaction. In: Molecular systematics (Hillis DM, Moritz C, and Mable BK, eds). Sunderland, MA: Sinauer Associates; 205–247.

Palumbi SR and Baker CS, 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol Biol Evol 11:426–435.

Palumbi SR, Cipriano F, and Hare MP, 2001. Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. Evolution 55:859–868.

Pichler FB, Dalebout ML, and Baker CS, 2001. Non-destructive DNA extraction from sperm whale teeth and scrimshaw. Mol Ecol Notes 1:106–109.

Pitman RL, Palacios DM, Brennan PL, Balcomb KC, and Miyashita T, 1999. Sightings and possible identity of a bottlenose whale in the tropical Indo-Pacific: *Indopacetus pacificus*?. Mar Mamm Sci 15:531–549.

Posada D and Crandall KA, 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.

Reyes JC, Van Waerebeek K, Cárdenas JC, and Yañez JL, 1996. *Mesoplodon bahamondi* sp. n. (Cetacea, Ziphiidae), a living beaked whale from the Juan Fernandez Archipelago, Chile. Bol Mus Nac Hist Nat Chile 45:31–44.

Rice DW, 1998. Marine mammals of the world: systematics and distribution. Special publication no. 4. Lawrence, KS: Society for Marine Mammalogy.

Roman J and Bowen BW, 2000. The mock turtle syndrome: genetic identification of turtle meat purchased in the south-eastern United States of America. Ann Conserv 3:61–65.

Ross HA, Lento GM, Dalebout ML, Goode M, Ewing G, McLaren P, Rodrigo AG, Lavery S, and Baker CS, 2003. DNA surveillance: molecular identification of whales, dolphins and porpoises. J Hered 94:111–114.

Ross GJB, 1984. The smaller cetaceans of the south east coast of southern Africa. Ann Cape Prov Mus (Nat Hist) 15:173–410.

Seberg O, Humphries CJ, Knapp S, Stevenson DW, Petersen G, Scharff N, and Andersen NM, 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy. Trends Ecol Evol 18:63–65.

Simpson GG, 1961. Principles of animal taxonomy. New York: Columbia University Press.

Smith EFG, Arctander P, Fjeldsa J, and Amir OG, 1991. A new species of shrike (Laniidae: *Laniarius*) from Somalia, verified by DNA sequence data from the only known individual. IBIS 133:227–235.

Swofford DL, 2000. PAUP*: phylogenetic analysis using parsimony and other phylogenetic reconstruction methods (version 4.0b10). Sunderland, MA: Sinauer Associates.

Tautz D, Arctander P, Minelli A, Thomas RH, and Vogler AP, 2002. DNA points the way ahead in taxonomy. Nature 418:479.

Tautz D, Arctander P, Minelli A, Thomas RH, and Vogler AP, 2003. A plea for DNA taxonomy. Trends Ecol Evol 18:70–74.

van Helden AL, Baker AN, Dalebout ML, Reyes JC, Van Waerebeek K, and Baker CS, 2002. Resurrection of *Mesoplodon traversii* (Gray, 1874), senior synonym of *M. bahamondi* Reyes, Van Waerebeek, Cárdenas and Yañez, 1995 (Cetacea: Ziphiidae). Mar Mamm Sci 18:609–621.

Wakeham-Dawson A, Morris S, and Tubbs P, 2002. Type specimens: dead or alive?. Bull Zool Nom 59:282–284.

Wayne RK, Meyer A, Lehman N, Van Valkenburgh B, Kat PW, Fuller TK, Girman D, and O'Brien SJ, 1990. Large sequence divergence among mitochondrial DNA genotypes within populations of eastern African black-backed jackals. Proc Natl Acad Sci USA 87:1772–1776.

Wheeler QD and Meier R (eds) 2000. Species concepts and phylogenetic theory. New York: Columbia University Press.

Wilson EO, 1992. Diversity of life. Cambridge, MA: Harvard University Press.

Winston JE, 1999. Describing species: practical taxonomic procedure for biologists. New York: Columbia University Press.

Zuckerkandl E and Pauling L, 1965. Evolutionary divergence and convergence in proteins. In: Evolving genes and proteins (Bryson V and Vogel HJ, eds). New York: Academic Press; 97–166.

Received September 15, 2003
Accepted March 16, 2004

Corresponding Editor: Stephen J. O'Brien