University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Publications, Agencies and Staff of the U.S.
Department of Commerce

U.S. Department of Commerce

2011

Evaluating the utility of *cox1* for cetacean species identification

A. Viricel

National Marine Fisheries Service

P. E. Rosel

National Marine Fisheries Service, Patricia.Rosel@noaa.gov

Follow this and additional works at: <https://digitalcommons.unl.edu/usdeptcommercepub>

Part of the [Environmental Sciences Commons](#)

Viricel, A. and Rosel, P. E., "Evaluating the utility of *cox1* for cetacean species identification" (2011).

Publications, Agencies and Staff of the U.S. Department of Commerce. 323.

<https://digitalcommons.unl.edu/usdeptcommercepub/323>

This Article is brought to you for free and open access by the U.S. Department of Commerce at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications, Agencies and Staff of the U.S. Department of Commerce by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

MARINE MAMMAL SCIENCE, **(*): ***_*** (** 2011)

2011 by the Society for Marine Mammalogy

Published 2011. This article is a US Government work and is in the public domain in the USA.

DOI: 10.1111/j.1748-7692.2010.00460.x

Evaluating the utility of *cox1* for cetacean species identification

A. VIRICEL

National Marine Fisheries Service,
646 Cajundome Boulevard, Suite 234,
Lafayette, Louisiana 70506, U.S.A.
and

Department of Biology,
Billeaud Hall, Room 108,
300 E. St. Mary Boulevard,
University of Louisiana at Lafayette,
Lafayette, Louisiana 70504, U.S.A.

P. E. ROSEL

National Marine Fisheries Service,
646 Cajundome Boulevard, Suite 234,
Lafayette, Louisiana 70506, U.S.A.
E-mail: patricia.rosel@noaa.gov

ABSTRACT

The mitochondrial cytochrome *c* oxidase I (*cox1*) gene has been promoted as a universal reference gene, or barcode, to identify organisms to the species level. We evaluated whether *cox1* would be appropriate to diagnose cetacean species. The 5' end of *cox1* (686 base pairs, bp) was sequenced for 46 of 86 recognized species of cetaceans. In addition, we included 105 sequences from GenBank, increasing our taxonomic coverage to 61 species. Particular focus was placed on sampling two subfamilies that contain closely related taxa: the Delphininae and the Globicephalinae. Species-specific sequences were observed for all but three taxa (*Delphinus delphis*, *D. capensis*, and *Stenella coeruleoalba*). Although correct assignment was seen for most species, significant overlap between intra- and interspecific variation makes *cox1* an imperfect barcode for cetaceans. The efficacy of *cox1* was compared to the 5' end of the cytochrome *b* (*cytb*) gene, a mitochondrial region routinely used for cetacean species identification. Although *cytb* performed better than *cox1* for some species, this marker could not differentiate other closely related taxa (*Eubalaena* spp.). Species identification for taxa not reliably identified using *cox1* or *cytb* might be best addressed through use of multiple mitochondrial DNA fragments or other newly developed markers.

Key words: DNA barcoding, species identification, mitochondrial DNA, forensics, whales, dolphins, conservation genetics.

Although the field of molecular taxonomy predates DNA barcoding (*e.g.*, Dizon *et al.* 2000, Valentini *et al.* 2009), the effort to produce DNA sequences for

species-level identification of a large array of taxa has vastly expanded in the last 10 yr and has become known as DNA barcoding (*e.g.*, Barrett and Hebert 2005, Meier *et al.* 2006, Clare *et al.* 2007, Cesari *et al.* 2009). This field generated such groups as The Barcode of Life Initiative (BOLI: <http://www.dnabarcodes.org>), an international consortium started in 2003 whose goal is to develop “DNA barcoding as a global standard for identifying species.” DNA barcoding aims to develop a universal DNA-based species diagnosis technique in which a single reference gene is used to classify a specimen to species through DNA sequence comparisons (Hebert *et al.* 2003*a*, 2004). The mitochondrial cytochrome *c* oxidase I (*cox1*) gene has been most heavily promoted for this purpose (Hebert *et al.* 2003*b*). Benefits of DNA barcoding include rapid biodiversity surveys and accelerated discovery of new taxa (*e.g.*, Bradford *et al.* 2010), species identification from early life stages (*e.g.*, Richardson *et al.* 2006), and forensic applications. However, the theoretical basis and universal applicability of this diagnosis technique has been widely debated (DeSalle *et al.* 2005, Will *et al.* 2005, Nielsen and Matz 2006, Rubinoff *et al.* 2006). One of the methods of species identification most commonly used in DNA barcoding employs distance-based phylogenetic trees (using the neighbor-joining, NJ, method: Saitou and Nei 1987) incorporating reference sequences from voucher specimens and examining where sequences from unknown specimens fall on a tree (Ratnasingham and Hebert 2007). This approach has led to several critiques. First, single gene trees do not necessarily equate to species trees (Maddison 1997) and basing diagnosis on a portion of a single gene can therefore yield inaccurate taxonomic identification due to incomplete lineage sorting or introgression, among other possible scenarios (*e.g.*, McGuire *et al.* 2007). Furthermore, there is disagreement concerning the assumption inherent in this approach that species should and do form monophyletic clusters (Funk and Omland 2003, Meier *et al.* 2006). Amplification of nontarget regions in the form of nuclear pseudogenes of mitochondrial origin can also bias molecular identification (*e.g.*, Thalmann *et al.* 2004). Other critiques of DNA barcoding relate to the use of *cox1* as a “universal” barcode to be applied to all taxa, as certain taxa do not display sufficient variability at this gene to provide phylogenetic resolution at the species level (*e.g.*, McFadden *et al.* 2010). After comparing 12 mitochondrial protein-coding genes in two mammalian species (*Bos* spp.), Santamaria *et al.* (2007) concluded that *cox1* is in fact one of the least suitable barcodes among these genes, due to overlap in intra- and interspecific distances. In addition, mitochondrial gene evolution is particularly slow in some groups, resulting in shared sequences among currently recognized taxa (*e.g.*, McFadden *et al.* 2010). In anthozoans, for example, mutation rates in mitochondrial DNA are very low (McFadden *et al.* 2000, France and Hoover 2002) and *cox1* is therefore not likely to be a suitable barcoding gene for this group due to low interspecific divergence (Neigel *et al.* 2007, Shearer and Coffroth 2008, McFadden *et al.* 2010).

Despite these criticisms, the *cox1* gene has been shown to identify species of a variety of taxa successfully (Hebert *et al.* 2003*a*, 2004; Cywinska *et al.* 2006; Hajibabaei *et al.* 2006; Kelly *et al.* 2007) and use of this gene for DNA barcoding has gained wide popularity. *Cox1* profiles from 203 arachnid species allowed correct identification of 100% of the specimens analyzed (Barrett and Hebert 2005) and this gene has proven quite useful for identifying bird species as well (Hebert *et al.* 2004). The success of these studies reflects, in part, the lack of overlap between intra- and interspecific nucleotide divergence (referred to as the “barcoding gap”) in the investigated taxa. Overlap between intra- and interspecific genetic variability results in lower identification success rates in other taxonomic groups (cowries

~83%, Meyer and Paulay 2005; blue butterflies ~82%, Wiemers and Fiedler 2007).

Despite the popularity of the *cox1* gene for barcoding of other species, DNA-based identification of cetaceans is currently primarily based on two other mitochondrial regions, namely the gene coding for cytochrome *b* (*cytb*) and the control region (Ross *et al.* 2003, Dalebout *et al.* 2004, Sholl *et al.* 2008, Jayasankar *et al.* 2008). A first assessment of the usefulness of the “primary barcode” sequence, that is, *cox1*, for cetacean species diagnosis has been conducted on four dolphin species (*Tursiops truncatus*, *Delphinus delphis*, *Stenella coeruleoalba*, and *Stenella frontalis*: Amaral *et al.* 2007). Their findings suggested that *cox1* is not appropriate to distinguish between these closely related taxa. The goal of the present study is to investigate whether *cox1* can be used to diagnose species of cetaceans using a data set with larger taxonomic coverage ($n = 61$ species). Modern cetaceans originated during the Late Eocene (~34 mya) and hence are considered a relatively recent group (Fordyce 1980). Some families date back only to the mid-Miocene 10–15 mya (*e.g.*, Barnes 1976), yet contain a majority of the species within the group. The subfamily Delphininae, in particular, is thought to be of very recent origin and to have arisen from a recent, rapid radiation event (Perrin and Reeves 2004). Thus, this subfamily provides an excellent empirical test of the efficacy of *cox1* to diagnose species under these conditions. We compare the efficacy of *cox1* to the more commonly used *cytb* gene for species diagnosis in cetaceans by analyzing alignments of newly generated sequences with augmentation of species coverage through inclusion of GenBank sequences to determine whether *cox1* performs better than *cytb*. This comparison would inform the marine mammal community whether it should focus on creating an entirely new DNA sequence database based on *cox1* sequences to replace those currently and commonly in use.

METHODS

Taxonomic Sample, DNA Extraction and Sequencing

For *cox1*, samples from 46 of the 86 currently recognized extant species of cetaceans (Committee on Taxonomy 2009) were sequenced for this study. Where possible, multiple individuals were chosen for each species to account for intraspecific variation and specimens were selected to encompass geographical variation within species (Table S1). Additionally, samples associated with voucher information were preferentially used to ensure the assignment of *cox1* sequences to taxonomically verified specimens. Finally, particular focus was placed on sampling two Delphinidae subfamilies that contain closely related and/or recently radiated taxa: the Delphininae (11 species sampled) and the Globicephalinae (6 species sampled). Multiple individuals of each species, including samples from regions throughout the species geographic distribution, were included to examine identification success rates among closely related taxa.

For 16 samples (Table S1), total genomic DNA was extracted from approximately 15 mg of tissue following protocols described in Rosel and Block (1996). DNA had been previously extracted from 81 additional samples included in the study following the same protocol. New primers targeting 731 base pairs (bp) of the 5' end of the cetacean mitochondrial *cox1* gene were designed using sequences from GenBank (Table S2): L5357 5'-CAGTCTAATGCCTACTCGGCC-3' and H6090 5'-GAAGTBTATATYYTAATTCTACC-3' (position numbers were

determined using the *Phocoena phocoena* complete mitochondrial genome, accession NC_005280). For samples of poor quality DNA (Table S1), two internal primers were designed to allow amplification of the entire 731-bp piece in two smaller overlapping fragments: H5730 5'-GCAGGYACAGGYTGAACYG-3' and L5652 5'-GTMTAAAYAAAYATRAGCTTCTG-3'. The 25 μ L reactions included 20 mM Tris-HCl, pH 8.4; 50 mM KCl, 1.5-mM MgCl₂, 0.3 μ M for primers L5357, H6090, and H5730, 0.6 μ M for primer L5652, 0.15 mM dNTPs, 2.5 U *Taq* polymerase (Invitrogen), and 0.1–25 ng DNA. When amplification was poor, 2–8 μ g of bovine serum albumin (Sigma Chemical Company, St Louis, MO) were added and the final concentration of primer H6090 was doubled. The PCR profile for primers L5357-H6090 was: 94°C for 30 s, followed by 32 cycles of 20 s at 94°C, 20 s at 45°C, and 40 s at 72°C and a final extension of 10 min at 72°C. Profiles for the internal primers were: 94°C for 30 s, followed by 30 cycles of 20 s at 94°C, 20 s at 45°C for H6090-L5652 or 55°C for L5357-H5730, and 30 s at 72°C and a final extension of 7 min at 72°C.

For *cytb*, although a large number of sequences were available in GenBank, the 5' end of the *cytb* gene was sequenced for 27 additional specimens (Table S1) to obtain the same number of sequences per species as in the *cox1* data set. Whenever possible, sequences from the same specimens were either generated or obtained from GenBank for both genes (Table 1). Sixteen samples (Table S1) were amplified and sequenced using primers L14724 5'-TGACTTGAARAACCAAYCGTTG-3' (Palumbi *et al.* 1991) and H15387 5'-GAATGGGATTATGTCTATGT-3' (Rosel *et al.* 1994), which generates a 704-bp product. For 11 samples with lower DNA quality (Table S1), the internal primer H15149 5'-CAGAATGATATTTGTCCTCA-3' (Kocher *et al.* 1989) was used with primer L14524 to amplify a shorter (465 bp) portion of the *cytb* gene. Reaction conditions were the same as described above with 0.3 μ M of each primer. The PCR profile was: 95°C for 30 s, followed by 35 cycles of 30 s at 94°C, 30 s at 45°C, 30 s at 72°C, and a final extension of 7 min at 72°C. For both genes, sequences were obtained in both directions using an ABI 3130 Genetic Analyzer and BigDye Terminator v.1.1 kit (Applied Biosystems Inc., Foster City, CA). Sequences for each gene were edited using SEQUENCHER v. 4.7 (Gene Codes Corp., Ann Arbor, MI) and aligned by eye using SEQPUP v. 0.6f (Gilbert 1996).

We created and analyzed four different alignments. For *cox1*, we first made a 686-bp alignment from the new sequences generated for this study. We then created a second 602-bp alignment by adding 105 cetacean sequences from GenBank (Table S2) to the first alignment. These two distinct *cox1* alignments (excluding and including sequences from GenBank) were used to assess the impact of potentially misidentified GenBank sequences on analyses (see Ross and Murugan 2006). A 347-bp *cytb* alignment was made using 27 sequences generated for this study and 163 sequences downloaded from GenBank (Table S1 and S2). A second, longer (571 bp) *cytb* alignment was also created by removing 43 shorter sequences from the first alignment (Table S1 and S2), to be more comparable in length to the *cox1* alignments. As most *cytb* sequences in our data set were obtained from GenBank, possible misidentifications were examined using the tree-based method (see below).

Analyses

Genetic variation—First, we characterized patterns of sequence variation and the degree of overlap in distance measures within and among species and genera for both genes. The numbers of variable sites and of phylogenetically informative sites were

Table 1. Number of sequences and unique haplotypes analyzed at *cox1* and *cytb*. “Overlap” identifies the number of individuals for which both *cox1* and *cytb* sequences were obtained from the same individuals. Family (F) and subfamily (Sf) designations follow Rice (1998) except for the following: subfamily designations within the Delphinidae are *sensu* LeDuc *et al.* (1999) and *Eschrichtius robustus* is included here in the Balaenopteridae family due to its placement in the phylogenies by Sasaki *et al.* (2005) and Agnarsson and May-Collado (2008). *Balaenoptera brydei* is no longer accepted as a species (Perrin 2009), however, the nomenclature was kept to reflect the way these sequences were labeled in GenBank. Geographical coverage is given for each species; however, this list is not exhaustive as location was not available for all sequences from GenBank. Abbreviated locations: Atlantic Ocean (AO); northeastern, northwestern, southeastern and southwestern Atlantic (NEA, NWA, SEA, SWA), North Atlantic Ocean (NA), Pacific Ocean (PO), northwestern and northeastern Pacific (NWP, NEP), central North Pacific (CNP), eastern tropical Pacific (ETP), Gulf of Mexico (GOM), Mediterranean Sea (MS), Black Sea (BS), Indian Ocean (IO) and eastern Indian Ocean (EIO).

F	Sf	Species	Nb sequences (Nb haplotypes)		Overlap	Geographic coverage	
			<i>cox1</i>	<i>cytb</i>		<i>cox1</i>	<i>cytb</i>
Balaenopteridae		<i>Balaenoptera bonaerensis</i>	1 (1)	1 (1)	1	Antarctic Ocean	Antarctic Ocean
		<i>Balaenoptera borealis</i>	2 (2)	2 (2)	1	NWA, Antarctic Ocean	Antarctic Ocean
		<i>Balaenoptera acutorostrata</i>	3 (3)	2 (2)	2	NWA, NA	NA
		<i>Balaenoptera physalus</i>	3 (3)	4 (2)	1	NWA, NEP	NWA
		<i>Balaenoptera edeni</i>	2 (2)	2 (2)	1	NWA, NWP	NWP
		<i>Balaenoptera musculus</i>	1 (1)	1 (1)	1		
		<i>Balaenoptera omurai</i>	2 (2)	2 (1)	2	NWP	NWP
		<i>Balaenoptera brydei</i>	2 (2)	2 (1)	2	NWP	NWP
		<i>Eschrichtius robustus</i>	5 (2)	3 (1)	2	NEP	NEP
		<i>Megaptera notaeangliae</i>	4 (2)	3 (2)	2	NWA, Antarctic Ocean, NEP	Antarctic Ocean
Balaenidae		<i>Eubalaena japonica</i>	1 (1)	1 (1)	1	Bering Sea	Bering Sea
		<i>Eubalaena glacialis</i>	2 (1)	2 (2)	1	NWA	Bering Sea, NWA
		<i>Eubalaena australis</i>	3 (3)	3 (1)	3	SWA, SEA, Antarctic Ocean	SWA, SEA, Antarctic Ocean
		<i>Balaena mysticetus</i>	6 (1)	6 (5)	2	Arctic Ocean, NWP, NA	NWP, NA

(Continued)

Table 1. (Continued)

F	Sf	Species	Nb sequences (Nb haplotypes)		Overlap	Geographic coverage	
			<i>cox1</i>	<i>cytb</i>		<i>cox1</i>	<i>cytb</i>
Kogiidae		<i>Kogia breviceps</i>	2 (2)	2 (2)	1	NWA	NWA
		<i>Kogia sima</i>	1 (1)	1 (1)	0	GOM	NWA
Physeteridae		<i>Physeter macrocephalus</i>	2 (2)	2 (2)	1	NWA, NA	NA
Neobalaenidae		<i>Caperea marginata</i>	2 (2)	2 (2)	2	SWP	SWP
Ziphiidae		<i>Mesoplodon bidens</i>	1 (1)	1 (1)	0	NWA	NWA
		<i>Mesoplodon mirus</i>	1 (1)	1 (1)	0	NWA	NEP
		<i>Mesoplodon carlbubbsi</i>	1 (1)	1 (1)	0	NEP	
		<i>Mesoplodon europaeus</i>	1 (1)	1 (1)	0	NWA	
		<i>Mesoplodon densirostris</i>	1 (1)	1 (1)	0	NWA	
		<i>Mesoplodon stejnegeri</i>	1 (1)	1 (1)	0	CNP	CNP
		<i>Berardius bairdii</i>	1 (1)	1 (1)	1		
		<i>Ziphius cavirostris</i>	3 (3)	4 (3)	0	NWA, NEP	NWA, NEP
		<i>Hyperoodon ampullatus</i>	1 (1)	1 (1)	1		
Monodontidae		<i>Monodon moneros</i>	1 (1)	1 (1)	1		
Delphinidae		<i>Delphinapterus leucas</i>	3 (2)	3 (2)	0	Arctic Ocean, NEP	
		Globicephalinae					
		<i>Pseudorca crassidens</i>	4 (2)	4 (2)	3	NWA, GOM	NWA, GOM, CNP
		<i>Globicephala melas</i>	5 (2)	5 (2)	1	NWA	NWA, SWP
		<i>Globicephala macrorhynchus</i>	4 (1)	4 (2)	0	NWA, GOM	AO, SWP
		<i>Grampus griseus</i>	5 (2)	5 (3)	4	NWA, GOM, NWP	NWA, NWP, AO

(Continued)

Table 1. (Continued)

F	Sf	Species	Nb sequences (Nb haplotypes)		Overlap	Geographic coverage	
			<i>cox1</i>	<i>cytb</i>		<i>cox1</i>	<i>cytb</i>
		<i>Peponocephala electra</i>	4 (2)	4 (1)	1	NWA, GOM	GOM, NWA, CNP
		<i>Feresa attenuata</i>	2 (1)	2 (2)	1	NWA	NWA, NWP
	Orcininae	<i>Orcinus orca</i>	3 (1)	3 (2)	0	GOM, CNP, NEP	PO
	Stenoninae	<i>Steno bredanensis</i>	2 (2)	2 (2)	0	NWA	AO, PO
		<i>Sotalia fluviatilis</i>	1 (1)	1 (1)	0	Brazilian Amazon	Brazilian Amazon
	Delphininae	<i>Sousa chinensis</i>	3 (2)	3 (3)	1	NWP	SEA, NWP
		<i>Stenella frontalis</i>	8 (4)	5 (4)	2	NWA, GOM, NEA	NWA
		<i>Stenella clymene</i>	3 (3)	3 (3)	2	NWA, GOM	NWA
		<i>Stenella coerulealba</i>	20 (17)	13 (8)	11	NWA, GOM, NEA, NWP	NWA, NEA, PO, MS, NWP
		<i>Stenella longirostris</i>	5 (4)	5 (3)	2	NWA, GOM, ETP	NWA, EIO
		<i>Stenella attenuata</i>	6 (4)	6 (5)	1	NWA, ETP, NWP	AO, PO, NWP
		<i>Lagenodelphis hosei</i>	1 (1)	1 (1)	0	NWA	AO
		<i>Tursiops aduncus</i>	2 (2)	2 (2)	1	IO, NWA	IO, NWA
		<i>Tursiops truncatus</i>	14 (10)	14 (8)	7	NWA, GOM, NEA, NWP	NWA, NEA, GOM, PO, NWP
		<i>Delphinus capensis</i>	8 (3)	9 (3)	8	NEP, SEP, NWP	NEP, SEP, NWP
		<i>Delphinus delphis</i>	20 (14)	19 (10)	17	NWA, NEA, NEP, ETP	NWA, NEA, PO, BS
	Incertae sedis	<i>Lagenorhynchus albirostris</i>	2 (2)	2 (2)	1	NWA	NWA
	Incertae sedis	<i>Lagenorhynchus acutus</i>	1 (1)	1 (1)	0	NWA	NEA

(Continued)

Table 1. (Continued)

F	Sf	Species	Nb sequences (Nb haplotypes)		Overlap	Geographic coverage	
			<i>cox1</i>	<i>cytb</i>		<i>cox1</i>	<i>cytb</i>
	Lissodelphininae						
		<i>Lagenorhynchus obliquidens</i>	3 (3)	3 (2)	0	NEP	NEP
		<i>Lissodelphis borealis</i>	1 (1)	1 (1)	0	NEP	NEP
Pontoporidae							
		<i>Pontoporia blainvillei</i>	2 (2)	2 (2)	1	SWA	
Iniidae							
		<i>Iniia geoffrensis</i>	2 (2)	2 (2)	1	Brazilian Amazon	Peruvian Amazon
Phocoenidae							
		<i>Phocoena phocoena</i>	5 (4)	5 (3)	1	NWA, CNP	BS
		<i>Neophocaena phocaenoides</i>	1 (1)	1 (1)	0	Yangtze River	
		<i>Phocoena sinus</i>	1 (1)	1 (1)	0	Sea of Cortez	Sea of Cortez
		<i>Phocoena spinipinnis</i>	1 (1)	1 (1)	0	SEP	SEP
		<i>Phocaenoides dalli</i>	3 (3)	3 (2)	0	NEP, CNP	NWA, CNP
Platanistidae							
		<i>Platanista minor</i>	1 (1)	1 (1)	1		

determined using DnaSP v. 4.10 (Rozas *et al.* 2003). To examine patterns of intra- and interspecific diversity, histograms of pairwise distance distributions were constructed using the Kimura-2-Parameter (K2P) (Kimura 1980) model of substitution as suggested for barcoding by Barrett and Hebert (2005) using R v. 2.10.1 (R Development Core Team 2005). Weighted mean intraspecific and intrageneric distances were calculated (weighing the mean by the number of sequences in each species). The weighted mean was chosen over the arithmetic mean to avoid the potential bias arising from species being represented by different number of sequences. To assess whether our sampling was representative, we investigated the relationship between sample size and mean intraspecific distance by conducting a linear regression analysis using R: we calculated the adjusted r^2 (adjusted for the number of predictors and sample size) between the mean pairwise intraspecific distance of each species and the number of individuals analyzed for that species. Finally, all alignments were examined using MacClade v. 4.0 (Maddison and Maddison 2000) to see whether any species shared haplotypes.

Tree-based identification—Early barcoding studies relied heavily on tree-based methods to assign unknown sequences to species (*e.g.*, Hebert *et al.* 2003a) and this method is still commonly used. More recently, alternative methods, including distance-based (Ratnasingham and Hebert 2007) and character-based (Rach *et al.* 2008) have gained approval and popularity. For direct comparison to other barcoding studies, we examined all three methods to test the efficacy of both *cox1* and *cytb* as barcodes for cetaceans using criteria similar to those used in other taxa (*e.g.*, Barrett and Hebert 2005, Meier *et al.* 2006, Huang *et al.* 2008, McFadden *et al.* 2010). NJ trees of unique haplotypes were generated for the purpose of identifying samples to species. Trees were reconstructed for both *cox1* and both *cytb* alignments using the package Analyses of Phylogenetics and Evolution (APE) (Paradis *et al.* 2004) in R. Genetic distances between haplotypes were estimated using the K2P model of evolution and 1,000 bootstrap replicates were performed. We evaluated other substitution models using jModelTest (Guindon and Gascuel 2003, Posada 2008). Using the model of substitution chosen with the Akaike Information Criterion (*cox1*: TPM2uf+I+G; *cytb*: TVM+I+G) did not alter the resulting trees and bootstrap support (data not shown). All trees were rooted with sequences of three ungulate species: *Bos taurus* (accession V00654), *Ovis aries* (accession NC_001941), and *Hippopotamus amphibius* (accession NC_000889).

Trees were examined for species that did not form species-specific clusters of all sequences, excluding those species represented by a single individual. To quantitatively assess the efficacy of each mitochondrial region using a tree-based approach, we determined the percentage of species (those represented by at least two sequences) for which all specimens would be correctly identified to species. We followed the “liberal” approach described by Ross *et al.* (2008), where monophyly of all conspecific sequences is not required for successful identifications. In other words, a sequence sister to a conspecific clade is considered to belong to the species forming this clade. Using simulated sequences, Ross *et al.* (2008) showed that this liberal method provides a larger proportion of correct identifications than a “strict” (monophyly required) approach. For comparison, we investigated the impact of using the strict approach on the number of successfully identified species.

Distance-based identification—The success of the distance-based identification method was tested on all alignments as implemented in the function Species-Identifier v. 1 within the program TAXONDNA (Meier *et al.* 2006). Using K2P pairwise distances, each sequence of the data set (the “query”) is removed

sequentially and compared back against the data set to test for assignment to the correct species. There are four possible outcomes for each query sequence: identification is either successful, unsuccessful (*i.e.*, cannot be identified to any species), ambiguous (*i.e.*, could be assigned to more than one species), or incorrect (*i.e.*, misidentified). Using the “best close match” criterion, the identification is considered successful when the sequences in the data set with the lowest genetic distance (*i.e.*, the closest matches) are conspecifics with the query sequence and the genetic distance between the query and the closest matches falls below a specified threshold (see below). The query sequence remains unidentified when none of the pairwise distances between it and the other sequences in the data set are below the threshold. Identification is considered ambiguous when the closest matches have pairwise distances below the threshold but include sequences of both the correct species and other species. Misidentification occurs when the closest match with pairwise distances below the threshold does not include a conspecific. Distance thresholds were calculated following Meier *et al.* (2006) in which the threshold represents the “distance below which 95% of all intraspecific [pairwise] distances are found” and is calculated from each data set (Meier *et al.* 2006). This resulted in 0.7% and 1.8% thresholds for the *cox1* data sets (with and without sequences from GenBank, respectively). Using the same method, the thresholds were set at 2% and 1.9% for the 347-bp and 571-bp *cytb* alignments, respectively. For these calculations, singletons (species represented by only one sequence) were removed from the alignments.

Character-based identification—We searched for character states diagnostic to a species (DeSalle *et al.* 2005) using DnaSP within three taxonomic groups containing multiple closely related species for which we had sequences: the genera *Balaenoptera* and *Mesoplodon* and the *Stenella*–*Tursiops*–*Delphinus* complex. We considered a character state to be diagnostic to a species if it was shared by all members of the species, but absent for all other taxa (*i.e.*, “pure characteristic attributes” *sensu* Rach *et al.* 2008). A character key was constructed for the *Stenella*–*Tursiops*–*Delphinus* complex using TeXshade v. 1.21 (Beitz 2000).

RESULTS

After examining sequences from GenBank in the generated NJ trees, we investigated four potential misidentifications. Three *cox1* sequences (accession EF090643, DQ465997, and DQ466015) that did not cluster with conspecifics could not be conclusively attributed to misidentifications and were kept in the second *cox1* alignment for all analyses. The fourth sequence we investigated, a *S. coeruleoalba cytb* sequence from GenBank (accession DQ466019) was confirmed by the author to be in fact a common dolphin (*D. delphis*) sequence. This sequence was removed from our data set.

Genetic Variation

New *cox1* sequences were obtained from 97 samples representing 10 families of cetaceans (Table S1). Except for three taxa (*Delphinus capensis*, *D. delphis*, and *S. coeruleoalba*), each species possessed unique *cox1* sequences. Adding *cox1* GenBank sequences reduced the alignment to 602 bp. The number of polymorphic and phylogenetically informative sites are reported in Table 2. The weighted mean pairwise intraspecific distance was larger when including sequences from GenBank (Table 2), mainly due to the inclusion of species with apparently high intraspecific variation

Table 2. Number of polymorphic and phylogenetically informative sites, and weighted mean percent intraspecific and intragenetic divergence (K2P) for each data set.

Data set	Nb polymorphic sites	Nb phylogenetically informative sites	Mean% intraspecific divergence	Mean% intrageneric divergence
Cytochrome <i>c</i> oxidase I				
686 bp	272	249	0.27 (± 0.08 SE)	6.40 (± 0.67 SE)
602 bp	248	236	0.50 (± 0.13 SE)	5.1 (± 1.30 SE)
Cytochrome <i>b</i>				
347 bp	170	155	0.62 (± 0.11 SE)	6.30 (± 1.60 SE)
571 bp	274	250	0.68 (± 0.20 SE)	6.20 (± 0.79 SE)

(e.g., *Balaenoptera edeni*, *Kogia breviceps*, and *Inia geoffrensis*; see “Discussion”). There was no significant relationship between the number of individuals analyzed per species and mean intraspecific variation (regression analysis: adjusted $r^2 = 0.027$, $P = 0.16$). There was considerable overlap between intraspecific and interspecific genetic variation in the subfamily Delphininae but not in the Globicephalinae (Fig. 1).

The 347-bp and 571-bp *cytb* alignments included 190 and 148 sequences representing 61 and 56 species, respectively. Each species possessed unique *cytb* sequences except for two taxa: *Eubalaena glacialis* ($n = 1$) and all *E. australis* ($n = 3$) specimens shared the same haplotype. Although their sequences were identical at the 5' portion examined in this study, the complete *cytb* gene sequences for these specimens are available in GenBank and 1–2 bp differences at the 3' end of the *cytb* gene differentiate the three *E. australis* sequences from the *E. glacialis* sequence. The number of polymorphic and phylogenetically informative sites was similar across the longer (571 bp) *cytb* alignment and the *cox1* alignments (Table 2). Intraspecific and intragenetic mean pairwise distances across the shorter (347 bp) *cytb* fragment were similar to the longer fragment (Table 2). There was no significant relationship between the number of individuals analyzed per species and mean intraspecific variation (regression analysis: adjusted $r^2 = -0.025$, $P = 0.85$). As with *cox1*, there was overlap between intra- and interspecific genetic variation at *cytb* in the subfamily Delphininae for both *cytb* alignments (Fig. 1). Contrary to *cox1*, an overlap between intra- and interspecific distance was also observed in the Globicephalinae (Fig. 1).

Tree-based Identification—*cox1*

Adding sequences from GenBank did not modify the tree topology except for a few problematic sequences discussed below. Therefore, the tree from the second alignment (including sequences from GenBank) is presented here as it provides broader taxonomic coverage (Fig. 2). For 35 (i.e., 87.5%) of the 40 species represented by at least two sequences, all specimens could be correctly identified using the “liberal” approach recommended by Ross *et al.* 2008. Applying a “strict” (monophyly required) approach decreased the number of species that were correctly identified ($n = 27$) as shown in Ross *et al.* (2008), because some sequences fall sister to a single-species clade. It is noteworthy that the number of species that would be misidentified remains the same. Sequences from four species formed a polytomy in addition to clusters of conspecific sequences: both *Delphinus* species, *S. coeruleoalba*, and *S. frontalis*. One *D. delphis* sequence from GenBank (accession DQ465997) was nested within one of two

Figure 1. Patterns of intra- and interspecific divergence using newly generated and GenBank (A) *cox1* (602 bp) sequences and (B) *cytb* (347 bp) sequences.

clusters of sequences from *S. coeruleoalba* and one *Tursiops* sequence from GenBank (accession DQ466015) did not cluster with others in the tree (Fig. 2). Moreover, two clades of conspecific sequences (*Stenella clymene* and *E. australis*) showed low bootstrap support (<70%) suggesting species identification of these two taxa using *cox1* should be treated with caution.

Tree-based Identification—*cytb*

Except for one clade (*D. delphis*, see below), the two *cytb* alignments produced similar tree topologies. As a result, only the tree reconstructed from the shorter (347 bp) alignment is presented as it provides broader taxonomic coverage (Fig. 3). For 31 (*i.e.*, 77.5%) of the 40 species represented by at least two sequences, all specimens could be correctly identified. As seen with *cox1*, applying a strict approach decreased the number of species that were correctly identified ($n = 24$). The number of species that would be misidentified remained the same using this approach. In addition to *E. australis* and *E. glacialis* sharing a haplotype, sequences from seven species did not form a conspecific cluster: *D. delphis*, *D. capensis*, *B. edeni*, *S. frontalis*, *S. coeruleoalba*,

Figure 3. Neighbor-joining tree (50% majority rule consensus) reconstructed using 133 unique 347-bp *cytb* sequences from 61 species of cetaceans and three outgroups. Clades of conspecific sequences were collapsed into triangles. The number of sequences is given in brackets for collapsed clades and for haplotypes shared by multiple specimens. Numbers above or left of nodes denote bootstrap values.

Globicephala macrorhynchus, and *G. melas*. The 571-bp *cytb* alignment improved the resolution for one of these species: the longer *D. delphis* sequences formed a conspecific cluster (data not shown). One species (*S. clymene*) that did form a cluster of conspecific sequences showed low bootstrap support providing low confidence in identifying this species using the 347-bp *cytb* fragment.

Distance-based Identification

The number of misidentified and unidentified sequences using the “best close match” criterion was similar for the 5' end of the *cox1* and *cytb* genes even though the *cytb* alignments were shorter compared to *cox1* (Table 3). Moreover, the longer *cytb* alignment did not result in lower numbers of misidentified and unidentified sequences than the shorter *cytb* alignment (Table 3).

Overall for *cox1*, successful identification of all specimens occurred for 33 (*i.e.*, 82.5%) of the 40 species represented by at least two sequences in the second alignment (including sequences from GenBank). Identification of a large number of sequences ($n = 27$) was ambiguous mainly due to the two *Delphinus* species sharing a haplotype. Three sequences were misidentified: a *D. delphis* sequence (accession DQ465997) for which the closest match was a *S. coeruleoalba*, a *S. coeruleoalba* sequence (accession DQ466007) for which the closest match was a *S. clymene* and a *S. coeruleoalba* sequence (accession EF090643) sharing a haplotype with *D. delphis* sequences. Five sequences remained unidentified: a divergent *T. truncatus* sequence (accession DQ466015), and two species with intraspecific pairwise distances above the threshold: the two *I. geoffrensis* sequences (pairwise distance: 3.2%) and the two *B. edeni* sequences (pairwise distance: 1.82%) (see “Discussion”).

Comparatively, ambiguous identifications using the 5' end of *cytb* were due to *E. glacialis* and *E. australis* sharing a haplotype, and low pairwise distances between several *Stenella* spp. sequences and between the two *Globicephala* species. Using the best close match method, four sequences were misidentified: a *G. melas* sequence (accession GMU13132) for which the closest match was a *G. macrorhynchus*, a *B. edeni* sequence (accession X75583) for which the closest match was a *Balaenoptera brydei* (but see “Discussion”), a *D. capensis* sequence (accession NC_012061) for which the closest match was a *D. delphis* and the *E. glacialis* sequence sharing a haplotype

Table 3. Number of successful (S), ambiguous (A), misidentified (M) and unidentified (U) sequences determined using SpeciesIdentifier. The corresponding percentages are given in brackets for each category. The first data set utilizes only new *cox1* sequences, while the second *cox1* data set is augmented with GenBank sequences. The length of the alignment for each data set is given in brackets; n = number of species in data set.

		Best close match			
Data set	n	S	A	M	U
Cytochrome <i>c</i> oxidase I					
65 seq. (686 bp)	14	50 (76.9%)	12 (18.5%)	0 (0.0%)	3 (4.6%)
181 seq. (602 bp)	40	146 (80.7%)	27 (14.9%)	3 (1.7%)	5 (2.8%)
Cytochrome <i>b</i>					
169 seq. (347 bp)	40	153 (90.5%)	9 (5.3%)	4 (2.4%)	3 (1.8%)
128 seq. (571 bp)	37	114 (89.1%)	7 (5.5%)	4 (3.1%)	3 (2.3%)

Table 4. Number of pure characteristic attributes identified for each species within three taxonomic groups.

Group	Species	Nb pure characteristic attributes	
		<i>cox1</i> (602 bp)	<i>cytb</i> (347 bp)
<i>Balaenoptera</i> spp.			
	<i>B. acutorostrata</i>	6	8
	<i>B. bonaerensis</i>	3	3
	<i>B. borealis</i>	5	0
	<i>B. brydei</i>	5	2
	<i>B. edeni</i>	0	0
	<i>B. musculus</i>	10	11
	<i>B. omurai</i>	10	5
	<i>B. physalus</i>	10	6
<i>Mesoplodon</i> spp.			
	<i>M. bidens</i>	21	4
	<i>M. carlhubbsi</i>	14	4
	<i>M. densirostris</i>	10	7
	<i>M. europaeus</i>	15	6
	<i>M. mirus</i>	12	8
	<i>M. stejnegeri</i>	7	8
<i>Stenella–Tursiops–Delphinus</i> complex			
	<i>D. capensis</i>	0	0
	<i>D. delphis</i>	0	0
	<i>S. attenuata</i>	2	1
	<i>S. clymene</i>	0	0
	<i>S. coeruleoalba</i>	0	0
	<i>S. frontalis</i>	1	0
	<i>S. longirostris</i>	6	5
	<i>T. aduncus</i>	3	3
	<i>T. truncatus</i>	0	4

with *E. australis* sequences. As with *cox1*, the two *I. geoffrensis* sequences remained unidentified due to their high pairwise distance (3.5%). A *B. edeni* sequence also remained unidentified due to high pairwise distance with its conspecific (2.3%). Overall, successful identification of all specimens occurred for 31 (*i.e.*, 77.5%) of the 40 species in the shorter *cytb* alignment and for 29 (*i.e.*, 78.4%) of the 37 species in the longer *cytb* alignment.

Character-based Identification

Across the 602 bp *cox1* and 347 bp *cytb* alignments, we identified diagnostic characters (*i.e.*, “pure characteristic attributes”) for all *Mesoplodon* species (Table 4). In comparison, “pure characteristic attributes” were found for all *Balaenoptera* species except *B. edeni* using the *cox1* alignment and for all but two *Balaenoptera* species across the 347 bp *cytb* alignment (Table 4). For the *Stenella–Tursiops–Delphinus* complex, the number of diagnostic characters was comparable between the 602 bp *cox1* alignment and the 347 bp *cytb* alignment, but only four of the nine species had diagnostic characters for either gene (and not the same four taxa for each gene) (Table 4, Fig. 4).

	<u>cox1</u>		<u>cytb</u>	
<i>S. attenuata</i>	CCTTTCAACATC	6 (4)	TTTCATCTACTAC	6 (5)
<i>S. coeruleoalba</i>	t..c.....	20 (17)	...t.....	13 (8)
<i>S. clymene</i>	t..c.....	3 (3)	...t.....	3 (3)
<i>S. frontalis</i>	t..c.....c.	8 (4)	...t.....	5 (4)
<i>S. longirostris</i>	t.ccctg.t..t	5 (4)	.c.tt.t.g.c..	5 (3)
<i>T. aduncus</i>	tt.c...g.g..	2 (2)	..ct.....gt	2 (2)
<i>T. truncatus</i>	t..c.....	14 (10)	c..t.c.c.t...	14 (8)
<i>D. capensis</i>	t..c.....	8 (3)	...t.....	9 (3)
<i>D. delphis</i>	t..c.....	20 (14)	...t.....	19 (10)

Figure 4. Pure characteristic attributes identified within the *Stenella–Tursiops–Delphinus* complex across the 602-bp *cox1* alignment and across the 347-bp *cytb* alignment. One representative sequence is shown for each species. The total number of individuals analyzed for each species is given to the right of each string of diagnostic characters, with the number of haplotypes in parentheses.

DISCUSSION

Using 97 sequences generated for this study and 105 sequences retrieved from GenBank representing a total of 61 species, we assessed the efficacy of *cox1* for species identification within the group Cetacea. Furthermore, we compared the performance of *cox1* to one of the commonly used markers for cetaceans, the 5' end of the *cytb* gene. Although the efficacy of the mitochondrial control region was not addressed here, it is important to note that its usefulness to distinguish between closely related cetaceans has been widely acknowledged (Rosel *et al.* 1994, Dalebout *et al.* 2004, Chivers *et al.* 2005, Gaines *et al.* 2005) and that this marker is currently the primary means of cetacean identification in DNA surveillance (Ross *et al.* 2003). However, this marker also has weaknesses, particularly in identifying members of the Delphininae (Bero 2001, Dizon *et al.* 2000, Kingston *et al.* 2009) because levels of intraspecific variation exceed interspecific variation for some pairwise species comparisons.

Variability of *cox1* in Cetaceans

Mean intraspecific and intrageneric pairwise distances observed at *cox1* were within the range reported in other taxa (Hebert *et al.* 2003b, Barrett and Hebert 2005, Yoo *et al.* 2006, Clare *et al.* 2007). Even at shorter sequence length, mean intraspecific pairwise distances were larger at *cytb* compared to *cox1*, which likely reflects the different rates of evolution of the two genes (Lopez *et al.* 1997). Due to limitations in sample availability, we could not sample across the entire geographic range of these widely distributed species, and limited geographic sampling may decrease the level of detected intraspecific variability (Meyer and Paulay 2005). However, results from regression analyses suggested there was no significant relationship between the number of individuals analyzed per species and mean intraspecific divergence in our data sets. Additionally, it is noteworthy that cetaceans show relatively low levels of genetic variability (Shimura and Numachi 1987, Wada 1988, Schlötterer *et al.* 1991, Kingston and Rosel 2004). Jackson *et al.* (2009) attributed low rates of molecular evolution in mysticetes to low metabolic rates, large body size, and/or long generation times. Lower rates of molecular evolution will limit the amount of genetic divergence that can accumulate within and between species within a given time frame. Furthermore, for many taxa, conspecific variation at *cox1* is very limited

(Clare *et al.* 2007, Ward *et al.* 2008, Ward 2009). Thus, the combination of a relatively low level of genetic variability in cetaceans coupled with the broader taxonomic trend of low intraspecific variability at *cox1* sequences will lessen the impacts of the limited geographic sampling. The low interspecific distances we found between some closely related species will greatly impede the efficacy of *cox1* for species identifications for these cetacean taxa and this problem will continue even with broader geographic coverage.

Impact of Potential Misidentifications and Imperfect Taxonomy on Identification Success

Factors that may have impacted the success rate in our study include misidentification of samples and unresolved or imperfect taxonomy. Further examination of the three *cox1* GenBank sequences (DQ466015, EF090643, and DQ465997) that were incorrectly identified by both the tree-based and distance-based methods suggested misidentification of these specimens was unlikely. For example, although the *D. delphis* *cox1* sequence (accession DQ465997) grouped with *S. coeruleoalba* sequences, the *cytb* and control region sequences for the same animal clustered with other *D. delphis*. And, although the *S. coeruleoalba* sequence (accession EF090643) shared the same haplotype as three *D. delphis* specimens across the 602 bp of *cox1* sequence used here, this specimen does have a single base pair difference with *D. delphis* sequences at the 5' end of *cox1* in the longer alignment from Amaral *et al.* (2007). For the *cytb* sequences however, it is not possible to rule out misidentification of specimens for three of them. In the tree-based method, the *G. melas* sequence (U13132) is found nested within the *G. macrorhynchus* clade (Fig. 3). This placement contradicts what is known about pilot whale mtDNA variability. Oremus *et al.* (2009) surveyed the highly variable mtDNA control region in pilot whales worldwide and found diversity to be very low in both species. Thus, it is reasonable to conclude that this sample may be misidentified or mislabeled in GenBank. If we remove this particular sequence, then there is no overlap between intra- and interspecies differences for the Globicephalinae in *cytb*. The control region sequence for the *D. capensis* sequence NC_012061, also available on GenBank, groups with *D. delphis* sequences when tested using DNA Surveillance-Witness for the Whales v. 4.3 (Ross *et al.* 2003), suggesting the possibility that this specimen was also misidentified, but given the uncertain worldwide taxonomy for this genus one cannot be sure. Finally, the geographic location (Bering Sea) of the *E. glacialis* voucher specimen (AY398662) sharing a *cytb* haplotype with three southern right whales (*E. australis*) suggests this specimen is in fact a North Pacific right whale, *Eubalaena japonica*. North Pacific and southern right whales have been shown to be more closely related to each other than to the North Atlantic right whale, *E. glacialis* (Rosenbaum *et al.* 2000), which is consistent with our findings. Although *E. japonica* and *E. australis* can be differentiated using the mitochondrial control region (Rosenbaum *et al.* 2000), these closely related taxa simply have not accumulated fixed differences in the section of *cytb* assayed.

Imperfect or unresolved taxonomy seems a likely explanation for the inability to unambiguously identify *Inia* and Bryde's whale sequences using the distance-based method. Both these groups exhibited high levels of intraspecific diversity for both genes. As a result, these specimens remained unidentified using the distance-based approach with a threshold set at the 95th percentile of all intraspecific distances. Using the tree-based method, however, *Inia* sequences were successfully identified for both genes, but this was not the case for Bryde's whale sequences at *cytb*. High

levels of intraspecific variation within the genus *Inia* have been reported at *cytb* and the control region by Banguera-Hinestroza *et al.* (2002) who suggested the existence of at least two evolutionary significant units that could represent distinct species. The number of species within the Bryde's whale complex that includes *B. edeni* and *B. brydei* is still unclear (Kato and Perrin 2009) and in fact currently only one Bryde's whale species, *B. edeni*, is officially recognized. However, sequences in GenBank are labeled with the various species names and we chose to maintain those identifications for the purpose of testing these genes. It is clear that the identification problem seen for these taxa does not represent a failure of the genetic markers in this case, but is rather the result of unresolved taxonomy.

Lack of Barcoding Gap in Cetaceans

Overlap between intra- and interspecific variation was observed for both genes, particularly in the subfamily Delphininae. This lack of a "barcoding gap" can diminish success rates for species identification (*e.g.*, Wiemers and Fiedler 2007), as mean rates of success tend to decrease with an increase in overlap (Ross *et al.* 2008). In cetaceans, this overlap does appear to impact our ability to distinguish between closely related species. Indeed, all methods of species diagnosis we tested produced a large number of ambiguous identifications and some incorrect identifications. While the barcoding gap was inferred early on to be a feature of *cox1* sequences, newer studies illustrate that it is not a universal feature (Hickerson *et al.* 2006, Ward 2009) and may more likely be an artifact of insufficient taxonomic sampling (Wiemers and Fiedler 2007). A variety of studies have now shown overlap between intra- and interspecific variation in cetaceans and have suggested a recent and rapid radiation event as the cause (Agnarsson and May-Collado 2008, Dalebout *et al.* 2008, Jackson *et al.* 2009, Kingston *et al.* 2009, Steeman *et al.* 2009). Rapid radiation events result in low internodal divergences between species, complicating species identification *via* barcoding methods. Shared ancestral polymorphisms and incomplete lineage sorting can result in shared haplotypes between species and apparent paraphyly/polyphyly in phylogenetic trees. Both these factors will impact success rates of tree-based, character-based, and distance-based methods and have likely played a role in the misidentifications seen here for some cetacean taxa, particularly for members of the subfamily Delphininae.

Difficulties in Barcoding Recently Diverged Taxa: the Case of the Delphininae

Here, 9 of the 11 species of Delphininae sampled were represented by three or more individuals and using *cox1* (when excluding sequences from GenBank), four of those nine species (*D. delphis*, *D. capensis*, *S. coeruleoalba*, and *S. frontalis*) did not form the species-specific clusters necessary for tree-based identifications. Furthermore, pure character attributes were only found for four of nine species within the *Stenella-Tursiops-Delphinus* complex (Fig. 4). In fact, *D. delphis* and *D. capensis* shared a *cox1* haplotype. Interestingly, mitochondrial DNA control region and *cytb* sequences were not identical between the individuals of these two species that shared *cox1* sequences here, indicating that misidentification of samples is not the cause of the shared sequences. As seen previously in the study by Amaral *et al.* (2007), the species *S. coeruleoalba* also did not form a single cluster in the *cox1* trees. This species was also problematic in the *cytb* tree. Overall, three mitochondrial

genes (*cox1*, *cytb*, control region) tested have failed to provide a reliable barcode for identifying *S. coeruleoalba*. Both *S. coeruleoalba* and *D. delphis* are globally distributed species with large abundances. These characteristics may contribute to the high levels of intraspecific variation, but do not necessarily explain why the two species tend to be confused specifically with each other using these barcoding methods. Interspecific hybridization is well known in delphinids (Zornetzer and Duffield 2003, Acevedo-Gutierrez *et al.* 2005, Silva *et al.* 2005, Caballero and Baker 2010, Kingston *et al.* 2009), but we are unaware of documented hybrids between these two species.

We also encountered one *T. truncatus cox1* sequence from GenBank (accession DQ466015) that did not cluster with the rest of the individuals of this species. The first 10 hits obtained by blasting this sequence in GenBank (searching for most similar sequences to the query) include *D. delphis*, *S. coeruleoalba*, and *T. truncatus* sequences, all of which showed a minimum of 47-bp differences to the query—an enormous number of differences when average intraspecific variation in this gene is 0.27% (± 0.08 SE). It is of interest to note that this haplotype was shared by three specimens and clustered with its conspecifics in the Amaral *et al.* (2007) *cox1* study. This might be due to the fact that sequences from other Delphinidae were not represented in the Amaral *et al.* (2007) phylogenetic tree, illustrating the importance of complete taxon sampling in groups of closely related species when performing DNA barcoding tests.

Several additional genera could prove problematic for *cox1* barcoding due to closely related species: *Orcaella*, *Sousa*, *Orcinus*, and *Kogia*. For *Orcaella*, a new species has only recently been recognized due to extremely low levels of morphological differences between species (Beasley *et al.* 2005) and species delineation in *Sousa* remains unclear (Jefferson and Waerebeek 2004). Furthermore, genetic diversity consistent with interspecific levels of variation has been found within the nominal species *Orcinus orca* (LeDuc *et al.* 2008) and *Kogia sima* (Chivers *et al.* 2005) suggesting the presence of additional species within these genera. Future analysis of *cox1* and *cytb* genes for these taxa will further our understanding of the ability of these genes to distinguish closely related cetacean taxa.

Conclusions

For most but not all cetacean species examined, *cox1* sequences were unique to a species. However, while *cox1* can be successful as a barcode for many cetacean species, it is not a comprehensive solution to the difficulties of identifying all members of the order Cetacea and does not outperform the *cytb* gene that has been commonly used for cetacean genetic identification. In cetaceans, the 5' end of *cytb* performed better than the 5' end of *cox1* for some species (*i.e.*, *D. delphis*, using the longer fragment) but did offer poor resolution to differentiate some closely related taxa (*e.g.*, *Eubalaena* spp.). Although we have not assessed the effectiveness of combining the *cox1* and *cytb* fragments, this could possibly improve the resolution of problematic taxa. Overall, the shorter *cytb* fragment (347 bp) provides similar efficacy to the longer *cytb* fragment (571 bp), using both tree-based and distance-based approaches. The shorter *cytb* fragment also contains an equivalent number of diagnostic characters for members of the *Stenella–Tursiops–Delphinus* complex when compared to a *cox1* fragment nearly twice as long. Thus, the shorter *cytb* fragment appears to be a better choice over *cox1* for most species diagnosis. This is of particular importance as it is not uncommon that cetacean species identifications involve decomposed carcasses or

highly processed samples (*e.g.*, Baker *et al.* 1996, 2006, 2007) where degradation of DNA forces the use of smaller fragments in DNA-based assays. Similar performance of *cox1* and *cytb* for species identification has been previously reported for mammals (Pfundner *et al.* 2004, Hajibabaei *et al.* 2007). Species identification for those taxa not reliably identified using one of the three mitochondrial DNA fragments (*cox1*, *cytb*, and control region), might best be addressed through use of multiple mitochondrial DNA fragments routinely used for identification purposes (*e.g.*, Ross *et al.* 2003) or other newly developed markers such as amplified fragment length polymorphisms (AFLPs) (*e.g.*, Kingston and Rosel 2004, Kingston *et al.* 2009).

ACKNOWLEDGMENTS

This project was funded by the NOAA Fisheries Southeast Fisheries Science Center, a grant from the Graduate Student Organization of the University of Louisiana at Lafayette (ULL), and a ULL Fellowship to Amélia Viricel. We would like to acknowledge Sarah Kingston, Stephanie Blochowiak, and Kristine Hiltunen for generating some of the *cytb* sequences and Eric Pante, Michael Schwartz, and several anonymous reviewers for constructive comments on the manuscript. We thank the many people and institutions who provided samples, including K. Williams (AFSC), B. Sima (Aquarium of the Americas), J. Reyes (CEPEC, Peru), V. Ridoux (CRMM, France), A. Read (Duke University), A. Costidis (FFWCC), M. C. Pinedo (Fundação Universidade do Rio Grande), L. Youngner (Georgia DNR), V. da Silva and T. Martin (INPA, Brazil), J. Heyning (LA County Museum), J. Brush (Maryland DNR), N. Barros (Mote Marine Lab), R. Bailey (MMSC), F. Wenzel (NEFSC-Woods Hole), W. McFee (NOS-Charleston), G. Anli (NNU, China), S. Ridgway (NOSC), K. Durham (Riverhead Foundation), A. Hohn, R. Baird and P. Martone (SEFSC-Beaufort), S. Gomez, R. Ewing, B. Mase and J. Litz (SEFSC-Miami), W. Jones and S. Bowen (SEFSC-Panama City), K. Mullin and W. Hoggard (SEFSC-Pascagoula), C. Potter (Smithsonian Institution), T. Jefferson (SWFSC), E. Rogan (UCC, Ireland), W. McLellan (UNCW), and S. Barco (Virginia Marine Science Museum). Skin biopsy samples were collected under U.S. MMPA permit numbers 779-1339-00, 779-1633-00 and 775-1600-10.

LITERATURE CITED

- Acevedo-Gutierrez, A., A. DiBerardinis, S. Larkin, L. Larkin and P. Forestell. 2005. Social interactions between tucuxi and bottlenose dolphins in Gandoca-Manzanillo, Costa Rica. *Latin American Journal of Aquatic Mammals* 4:49–54.
- Agnarsson, I., and L. J. May-Collado. 2008. The phylogeny of Cetartiodactyla: The importance of dense taxon sampling, missing data, and the remarkable promise of cytochrome *b* to provide reliable species-level phylogenies. *Molecular Phylogenetics and Evolution* 48:964–985.
- Amaral, A. R., M. Sequeira and M. M. Coelho. 2007. A first approach to the usefulness of cytochrome *c* oxidase I barcodes in the identification of closely related delphinid cetacean species. *Marine and Freshwater Research* 58:505–510.
- Arnason, U., and A. Gullberg. 1993. Comparison between the complete mtDNA sequences of the blue and the fin whale, two species that can hybridize in nature. *Journal of Molecular Evolution* 37:312–322.
- Baker, C. S., F. Cipriano and S. R. Palumbi. 1996. Molecular genetic identification of whale and dolphin products from commercial markets in Korea and Japan. *Molecular Ecology* 5:671–685.
- Baker, C. S., J. G. Cooke, S. Lavery, *et al.* 2007. Estimating the number of whales entering trade using DNA profiling and capture-recapture analysis of market products. *Molecular Ecology* 16:2617–2626.

- Baker, C. S., V. Lukoschek, S. Lavery, M. L. Dalebout, M. Yong-un, T. Endo and N. Funahashi. 2006. Incomplete reporting of whale, dolphin and porpoise 'bycatch' revealed by molecular monitoring of Korean markets. *Animal Conservation* 9:474–482.
- Banguera-Hinestroza, E., H. Cárdenas, M. Ruiz-García, M. Marmontel, E. Gaitán, R. Vázquez and F. García-Vallejo. 2002. Molecular identification of evolutionary significant units in the Amazon river dolphin *Inia* sp. (Cetacea: Iniidae). *The Journal of Heredity* 93:312–322.
- Barnes, L. G. 1976. Outline of eastern North Pacific fossil cetacean assemblages. *Systematic Zoology* 25:321–343.
- Barrett, R. D. H., and P. D. N. Hebert. 2005. Identifying spiders through DNA barcodes. *Canadian Journal of Zoology* 83:481–491.
- Beasley, I., K. M., Robertson and P. Arnold. 2005. Description of a new dolphin, the Australian snubfin dolphin *Orcaella heinsobnsi* sp. n. (Cetacea, Delphinidae). *Marine Mammal Science* 21:365–400.
- Beitz, E. 2000. TeXshade: Shading and labeling of multiple sequence alignments using LaTeX2e. *Bioinformatics* 16:135–139.
- Bero, D. 2001. Population structure of the Atlantic spotted dolphin (*Stenella frontalis*) in the Gulf of Mexico and western North Atlantic. M.S. thesis, College of Charleston, Charleston, SC. 86 pp.
- Bradford, T., M. Adams, W. F. Humphreys, A. D. Austin and S. J. B. Cooper. 2010. DNA barcoding of stygofauna uncovers cryptic amphipod diversity in a calcrete aquifer in Western Australia's arid zone. *Molecular Ecology Resources* 10:41–50.
- Caballero, S., and C. S. Baker. 2010. Captive-born intergeneric hybrid of a Guiana and bottlenose dolphin: *Sotalia guianensis* × *Tursiops truncatus*. *Zoo Biology* 29:1–11.
- Cesari, M., R. Bertolani, L. Rebecchi and R. Guidetti. 2009. DNA barcoding in Tardigrada: the first case study on *Macrobiotus macrocalix* Bertolani & Rebecchi 1993 (Eutardigrada, Macrobiotidae). *Molecular Ecology Resources* 9:699–706.
- Chivers, S. J., R. G. Leduc, K. M. Robertson, N. B. Barros and A. E. Dizon. 2005. Genetic variation of *Kogia* spp. with preliminary evidence for two species of *Kogia sima*. *Marine Mammal Science* 21:619–634.
- Clare, E. L., B. K. Lim, M. D. Engstrom, J. L. Eger and P. D. N. Hebert. 2007. DNA barcoding of Neotropical bats: Species identification and discovery within Guyana. *Molecular Ecology Notes* 7:184–190.
- Committee on Taxonomy. 2009. List of marine mammal species and subspecies. Society for Marine Mammalogy. Available at <http://www.marinemammalscience.org> (accessed 6 March 2010).
- Cywinska, A., F. F. Hunter and P. D. N. Hebert. 2006. Identifying Canadian mosquito species through DNA barcodes. *Medical and Veterinary Entomology* 20:213–424.
- Dalebout, M. L., C. S. Baker, J. G. Mead, V. G. Cockcroft and T. K. Yamada. 2004. A comprehensive and validated molecular taxonomy of beaked whales, family Ziphiidae. *Journal of Heredity* 95:459–473.
- Dalebout, M. L., D. Steel and C. S. Baker. 2008. Phylogeny of the beaked whale genus *Mesoplodon* (Ziphiidae: Cetacea) revealed by nuclear introns: Implications for the evolution of male tusks. *Systematic Biology* 57:857–875.
- DeSalle R., M. G. Egan and M. Siddall. 2005. The unholy trinity: Taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B* 360:1905–1916.
- Dizon, A., C. S. Baker, F. Cipriano, G. Lento, P. Palsbøll and R. Reeves. 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. NOAA Technical Memorandum NMFS-SWFSC-286, Washington, DC. 51 pp.
- Fordyce, R. E. 1980. Whale evolution and Oligocene Southern Ocean environments. *Palaeogeography, Palaeoclimatology, Palaeoecology* 31:319–336.

- France, S. C., and L. L. Hoover. 2002. DNA sequences of the mitochondrial *COI* gene have low levels of divergence among deep-sea octocorals (Cnidaria: Anthozoa). *Hydrobiologia* 471:149–155.
- Funk, D. J., and K. E. Omland. 2003. Species-level paraphyly and polyphyly: Frequency, causes and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution and Systematics* 34:397–423.
- Gaines, C. A., M. P. Hare, S. E. Beck and H. C. Rosenbaum. 2005. Nuclear markers confirm taxonomic status and relationships among highly endangered and closely related right whale species. *Proceedings of the Royal Society B* 272:533–542.
- Gilbert, D. G. 1996. SeqPup: A biosequence editor and analysis platform, version 0.6f. Bionet Software. Available at <ftp://iubio.bio.indiana.edu/molbio/seqpup/>.
- Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52:696–604.
- Hajibabaei, M., D. H. Janzen, J. M. Burns, W. Hallwachs and P. D. N. Hebert. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences* 103:968–971.
- Hajibabaei, M., G. Singer, E. Clare and P. Hebert. 2007. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biology* 5:24.
- Hebert, P. D. N., A. Cywinska and S. L. Ball. 2003a. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B* 270:313–321.
- Hebert, P. D. N., S. Ratnasingham and J. R. deWaard. 2003b. Barcoding animal life: Cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B* 270:S96–S99.
- Hebert P. N., M. Y. Stoeckle, T. S. Zemlak and C. M. Francis. 2004. Identification of birds through DNA barcodes. *Public Library of Science, Biology* 2:1657–1663.
- Hickerson, M. J., C. P. Meyer and C. Moritz. 2006. DNA barcoding will often fail to discover new animal species over broad parameter apace. *Systematic Biology* 55:729–739.
- Huang, D., R. Meier, P. A. Todd and L. M. Chou. 2008. Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. *Journal of Molecular Evolution* 66:167–174.
- Jackson, J. A., C. S. Baker, M. Vant, D. J. Steel, L. Medrano-Gonzales and S. R. Palumbi. 2009. Big and slow: Phylogenetic estimates of molecular evolution in baleen whales (suborder Mysticeti). *Molecular Biology and Evolution* 26:2427–2440.
- Jayasankar, P., B. Anoop, Vivekanandan, *et al.* 2008. Molecular identification of delphinids and finless porpoise (Cetacea) from the Arabian Sea and Bay of Bengal. *Zootaxa* 1853:57–67.
- Jefferson, T. A., and K. Van Waerebeek. 2004. Geographic variation in skull morphology of humpback dolphins (*Sousa* spp.). *Aquatic Mammals* 30:3–17.
- Kato, H., and W. F. Perrin. 2009. Bryde's whales *Balaenoptera edeni/brydei*. Pages 158–163 in W. F. Perrin, B. Würsig and J. G. M. Thewissen, eds. *Encyclopedia of marine mammals*. Academic Press, Amsterdam, The Netherlands.
- Kelly, R. P., I. N. Sarkar, D. J. Eernisse and R. DeSalle. 2007. DNA barcoding using chitons (genus *Mopalia*). *Molecular Ecology Notes* 7:177–183.
- Kimura, P. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111–120.
- Kingston, S. E., L. D. Adams and P. E. Rosel. 2009. Testing mitochondrial sequences and anonymous nuclear markers for phylogeny reconstruction in a rapidly radiating group: Molecular systematics of the Delphininae (Cetacea: Odontoceti: Delphinidae). *BMC Evolutionary Biology* 9:245.
- Kingston, S. E., and P. E. Rosel. 2004. Genetic differentiation among recently diverged delphinid taxa determined using AFLP markers. *Journal of Heredity* 95:1–10.
- Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca and A. C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the USA* 86:6196–6200.

- LeDuc, R. G., W. F. Perrin and A. E. Dizon. 1999. Phylogenetic relationships among the delphinid cetaceans based on full cytochrome *b* sequences. *Marine Mammal Science* 15:619–648.
- LeDuc, R. G., K. M. Robertson and R. L. Pitman. 2008. Mitochondrial sequence divergence among Antarctic killer whale ecotypes is consistent with multiple species. *Biology Letters* 23:426–429.
- Lopez, J. V., M. Culver, J. C. Stephens, W. E. Johnson and S. J. O'Brien. 1997. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. *Molecular Biology and Evolution* 14:277–286.
- Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46:523–536.
- Maddison, D. R., and W. P. Maddison. 2000. *MacClade version 4: Analysis of phylogeny and character evolution*. Sinauer Associates, Sunderland, MA.
- McFadden, C. S., Y. Benayahu, E. Pante, J. N. Thoma, P. A. Nevarez and S. C. France. 2010. Limitations of mitochondrial gene barcoding in Octocorallia. *Molecular Ecology Resources*. doi: 10.1111/j.1755-0998.2010.02875.x.
- McFadden, C. S., I. Tullis, M. B. Hutchinson and K. Winner. 2000. Rates of evolution of cnidarian mitochondrial genes. *American Zoologist* 40:1124–1124.
- McGuire, J. A., C. W. Linkem, M. S. Koo, *et al.* 2007. Mitochondrial introgression and incomplete lineage sorting through space and time: Phylogenetics of crotaphytid lizards. *Evolution* 61:2879–2897.
- Meier, R., K. Shiyang, G. Vaidya and P. K. L. Ng. 2006. DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology* 55:715–728.
- Meyer, C. P., and G. Paulay. 2005. DNA barcoding: Error rates based on comprehensive sampling. *Public Library of Science, Biology* 3:2229–2238.
- Neigel, J., A. Domingo and J. Stake. 2007. DNA barcoding as a tool for coral reef conservation. *Coral Reefs* 26:487–499.
- Nielsen, R., and M. Matz. 2006. Statistical approaches for DNA barcoding. *Systematic Biology* 55:162–169.
- Oremus, M., R. Gales, M. L. Dalebout, *et al.* 2009. Worldwide mitochondrial DNA diversity and phylogeography of pilot whales (*Globicephala* spp.). *Biological Journal of the Linnean Society* 98:729–744.
- Palumbi, S., A. Martin, S. Romano, W. O. McMillan, L. Stice and G. Grabowski. 1991. The simple fool's guide to PCR. University of Hawaii, Honolulu, HI.
- Paradis, E., J. Claude and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Perrin, W. F. 2009. World Cetacea database. Available online at <http://www.marinespecies.org/cetacea> (accessed 22 February 2010).
- Perrin, W. F., and R. R. Reeves. 2004. Electronic appendix 5: Report of the working group on species- and subspecies-level taxonomy. Pages 25–60 in R. R. Reeves, W. F. Perrin, B. L. Taylor, C. S. Baker and S. L. Mesnick, eds. Report of the workshop on shortcomings of cetacean taxonomy in relation to needs of conservation and management. NOAA-TM-NMFS-SWFSC-363, La Jolla, CA. 94 pp.
- Pfunder, M., O. Holzgang and J. E. Frey. 2004. Development of microarray-based diagnostics of voles and shrews for use in biodiversity monitoring studies, and evaluation of mitochondrial cytochrome oxidase I vs. cytochrome *b* as genetic markers. *Molecular Ecology* 13:1277–1286.
- Posada, D. 2008. jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* 25:1253–1256.
- R Development Core Team. 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rach, J., R. DeSalle, I. N. Sarkar, B. Schierwater and H. Hadrys. 2008. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proceedings of the Royal Society B* 275:237–247.

- Ratnasingham, S., and P. D. N. Hebert. 2007. bold: The barcode of life data system (<http://www.barcodinglife.org>). *Molecular Ecology Notes* 7:355–364.
- Rice, D. W. 1998. Marine mammals of the world: Systematics and distribution. Special Publication Number 4, The Society of Marine Mammalogy.
- Richardson, D. E., J. D. Vanwyke, A. M. Exum, R. K. Cowen and D. L. Crawford. 2006. High-throughput species identification: From DNA isolation to bioinformatics. *Molecular Ecology Notes* 7:199–207.
- Rosel, P. E., and B. A. Block. 1996. Mitochondrial control region variability and global population structure in the swordfish, *Xiphius gladius*. *Marine Biology* 125:11–22.
- Rosel, P. E., A. E. Dizon and J. E. Heyning. 1994. Genetic analysis of sympatric morphotypes of common dolphins (genus *Delphinus*). *Marine Biology* 119:159–167.
- Rosenbaum, H. C., R. L. Brownell, Jr., M. W. Brown, *et al.* 2000. World-wide genetic differentiation of *Eubalaena*: Questioning the number of right whale species. *Molecular Ecology* 9:1793–1802.
- Ross, H. A., G. M. Lento, M. L. Dalebout, *et al.* 2003. DNA surveillance: Web-based molecular identification of whales, dolphins and porpoises. *Journal of Heredity* 94:111–114.
- Ross, H. A., and S. Murugan. 2006. Using phylogenetic analyses and reference datasets to validate the species identities of cetacean sequences in GenBank. *Molecular Phylogenetics and Evolution* 40:866–871.
- Ross, H. A., S. Murugan and W. L. S. Li. 2008. Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology* 57:216–230.
- Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
- Rubioff, D., S. Cameron and K. Will. 2006. A genomic perspective on the shortcomings of mitochondrial DNA for “barcoding” identification. *Journal of Heredity* 97:581–594.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406–425.
- Santamaria, M., C. Lanave, S. Vicario and C. Saccone. 2007. Variability of the mitochondrial genome in mammals at the inter-species/intra-species boundary. *Biological Chemistry* 388:943–946.
- Sasaki, T., M. Nikaido, H. Hamilton, *et al.* 2005. Mitochondrial phylogenetics and evolution of mysticete whales. *Systematic Biology* 54:77–90.
- Schlötterer, C., B. Amos and D. Tautz. 1991. Conservation of polymorphic simple sequence loci in cetacean species. *Nature* 354:63–65.
- Shearer, T. L., and M. A. Coffroth. 2008. Barcoding corals: Limited by interspecific divergence, not intraspecific variation. *Molecular Ecology Notes* 8:247–255.
- Shimura, E., and K. I. Numachi. 1987. Genetic variability and differentiation in the toothed whales. *Scientific Reports of the Whales Research Institute, Tokyo* 38:141–163.
- Sholl, T. G. C., F. F. Nascimento, O. Leoncini, C. R. Bonvicino and S. Siciliano. 2008. Taxonomic identification of dolphin love charms commercialized in the Amazonian region through the analysis of cytochrome *b* DNA. *Journal of the Marine Biological Association of the United Kingdom* 88:1207–1210.
- Silva, J. M., F. J. L. Silva and I. Sazima. 2005. Two presumed interspecific hybrids in the genus *Stenella* (Delphinidae) in the Tropical West Atlantic. *Aquatic Mammals* 31:468–472.
- Steeman, M. E., M. B. Hebsgaard, R. E. Fordyce, *et al.* 2009. Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic Biology* 58:573–585.
- Thalmann, O., J. Hebler, H. N. Poinar, S. Pääbo and L. Vigilant. 2004. Unreliable mtDNA data due to nuclear insertions: A cautionary tale from analysis of humans and other great apes. *Molecular Ecology* 13:321–335.
- Valentini, A., F. Pompanon and P. Taberlet. 2009. DNA barcoding for ecologists. *Trends in Ecology and Evolution* 24:110–117.
- Wada, S. 1988. Genetic differentiation between the two forms of short-finned pilot whales off the Pacific coast of Japan. *Scientific Reports of the Whales Research Institute, Tokyo* 39:91–101.

- Ward, R. D. 2009. DNA barcode divergence among species and genera of birds and fishes. *Molecular Ecology Resources* 9:1077–1085.
- Ward, R. D., B. H. Holmes, W. T. White and P. R. Last. 2008. DNA barcoding Australasian chondrichthyans: results and potential uses in conservation. *Marine and Freshwater Research* 59:57–71.
- Wiemers, N., and K. Fiedler. 2007. Does the DNA barcoding gap exist? A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology* 4:8.
- Will, K. W., B. D. Mishler and Q. D. Wheeler. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology* 54:844–851.
- Yoo, H. S., J. Y. Eah, J. S. Kim, M. S. Min, W. K. Paek, H. Lee and C. B. Kim. 2006. DNA barcoding Korean birds. *Molecules and Cells* 22:323–327.
- Zornetzer, H. R., and D. A. Duffield. 2003. Captive-born bottlenose dolphin $\times$ common dolphin (*Tursiops truncatus* $\times$ *Delphinus capensis*) intergeneric hybrids. *Canadian Journal of Zoology* 81:1755–1762.

Received: 26 March 2010

Accepted: 27 November 2010

SUPPORTING INFORMATION

The following supporting information is available for this article online:

Table S1. List of 97 specimens from 46 cetacean species sequenced at *cox1* for this study and 27 specimens sequenced at cytochrome *b* (for this study or previous projects in our lab).

Table S2. Accession numbers of 105 *cox1* and 156 cytochrome *b* sequences from 60 species obtained from GenBank.