View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by College of William & Mary: W&M Publish

W&M ScholarWorks

VIMS Articles

Virginia Institute of Marine Science

3-2007

Specific identification of western Atlantic Ocean scombrids using mitochondrial DNA cytochrome C oxidase subunit I (COI) gene region sequences

MA Paine
Virginia Institute of Marine Science

Jan McDowell
Virginia Institute of Marine Science

John E. Graves
Virginia Institute of Marine Science

Follow this and additional works at: https://scholarworks.wm.edu/vimsarticles

Part of the Aquaculture and Fisheries Commons, and the Marine Biology Commons

Recommended Citation

Paine, MA; McDowell, Jan; and Graves, John E., "Specific identification of western Atlantic Ocean scombrids using mitochondrial DNA cytochrome C oxidase subunit I (COI) gene region sequences" (2007). *VIMS Articles*. 1505.
https://scholarworks.wm.edu/vimsarticles/1505

This Article is brought to you for free and open access by the Virginia Institute of Marine Science at W&M ScholarWorks. It has been accepted for inclusion in VIMS Articles by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.

BULLETIN OF MARINE SCIENCE, 80(2): 353–367, 2007

SPECIFIC IDENTIFICATION OF WESTERN ATLANTIC OCEAN SCOMBRIDS USING MITOCHONDRIAL DNA CYTOCHROME *C* OXIDASE SUBUNIT I (COI) GENE REGION SEQUENCES

Melissa A. Paine, Jan R. McDowell, and John E. Graves

ABSTRACT

Identification of scombrids (tunas, mackerels, bonitos, etc.) is difficult when morphological characters are ambiguous or missing, such as with early life history stages or tissues found in the stomachs of predators. The mitochondrial cytochrome *c* oxidase subunit I (COI) gene region was evaluated as a molecular marker for the specific identification of the 17 members of the family Scombridae common to the western Atlantic Ocean. A 950 base pair region in the COI gene was sequenced from up to 20 individuals of each species, and suites of nucleotide polymorphisms that unambiguously distinguish among these scombrid species were identified. A shorter 250 base pair fragment of COI proved to be sufficient for species identification and was better suited for analyzing degraded tissue samples. Scombrid larvae collected in the Florida Straits and scombrid remains in the stomachs of large pelagic predators were used to demonstrate the utility of both the long and short COI fragments.

Members of the family Scombridae (tunas, mackerels, bonitos, etc.) are important components of pelagic ecosystems, with several species supporting large commercial and recreational fisheries throughout the world's oceans. Proper identification of these species at all life stages and in various conditions, even as degraded stomach contents, is essential to better understand early life history characteristics and ecological relationships in the pelagic ecosystem, and to enable effective management. In addition, specific identification of processed tissues or fillets is necessary for enforcement of fisheries management regulations.

While specific identification of adult scombrids is essentially unambiguous (Collette and Nauen, 1983), identification is problematic in situations where morphological characters are difficult to interpret (early life history stages) or missing (fillets, digested stomach contents). Identification of early life history stages of scombrids has been especially challenging. Scombrid eggs, larvae (especially those of the genus *Thunnus*), and juveniles generally cannot be distinguished unambiguously based solely on morphology (Richards et al., 1990).

Molecular markers can provide a means for positive identification when morphological identification is uncertain. Various molecular markers have been used to identify fish eggs and larvae including allozymes (Morgan, 1975), polymerase chain reaction (PCR)/restriction fragment length polymorphism (RFLP) analysis (Daniel and Graves, 1994; McDowell and Graves, 2002), multiplex PCR (Rocha-Olivares, 1998; Hyde et al., 2005) and sequencing (Hare et al., 1994; Kirby and Reid, 2001; Perez et al., 2005). Many of these techniques have been used to identify scombrids. Allozymes have been successfully used to discriminate between early juveniles of bigeye tuna *Thunnus obesus* (Lowe, 1839) and yellowfin tuna *Thunnus albacares* (Bonnaterre, 1788) (Graves et al., 1988) as well as between adult Pacific northern bluefin tuna *Thunnus thynnus orientalis* Serventy, 1956 and southern bluefin tuna *Thunnus maccoyii* Castelnau, 1872 (Ward, 1995). Several studies have used PCR/

Bulletin of Marine Science
© 2007 Rosenstiel School of Marine and Atmospheric Science of the University of Miami

RFLP analysis to identify species of the scombrid tribes Thunnini and Sardini (Chow et al., 2003) as well as eight species of the genus *Thunnus* (Chow and Inoue, 1993). In addition, sequencing of a mitochondrial gene region has been used to identify *Thunnus* species (Bartlett and Davidson, 1991; Ram et al., 1996; Quintero et al., 1998; Terol et al., 2002; Ward et al., 2005). While each of these techniques has advantages and disadvantages, sequencing provides the highest level of resolution as it shows genetic differences at the nucleotide level. A few studies have used sequence analysis to identify scombrids, but these investigations were limited as they only distinguished between a few species, used a region that revealed considerable intraspecific variation, had limited sample sizes, or encountered problems with non-specific amplification (Bartlett and Davidson, 1991; Ram et al., 1996; Quintero et al., 1998; Terol et al., 2002; Ward et al., 2005). Additionally, molecular techniques have only recently appeared promising as an answer for the difficulties associated with identifying degraded remains in stomachs (Harper et al., 2005; Gorokhova, 2006), and sequencing has been used to identify gut contents of marine invertebrates (Blankenship and Yayanos, 2005) and large pelagic fishes (Smith et al., 2005).

The mitochondrial genome has been preferred for analysis in many genetic studies as it has a high number of copies per cell, which facilitates PCR amplification, and because the presence of a single allele makes it possible to sequence products directly (Avise, 1994). Many mitochondrial gene regions (cytochrome *b,* ND4, 16S, COI) have been successfully used for fish identification (Bartlett and Davidson, 1991; McDowell and Graves, 2002; Hyde et al., 2005; Lopez and Pardo, 2005). These gene regions display different levels of genetic variation as a result of different evolutionary rates. While variation is necessary to highlight interspecific differences, too much variation can hinder primer design. Because of this, the use of a conserved region is advantageous for effective amplification across many species.

One of the most conserved protein coding genes in the mitochondrial (mt) genome is cytochrome *c* oxidase subunit I (COI) (Brown, 1985). COI is critical for cellular energy production and this functional importance constrains its evolution (Rawson and Burton, 2002). The high level of conservation of COI allows for the design of a unique primer pair that successfully amplifies the same fragment across the diverse members of the Scombridae. Previous work has taken advantage of COI for broad taxonomic studies (11 invertebrate phyla, Folmer et al., 1994; 11 animal phyla, Hebert et al., 2003), but COI has also been useful to distinguish closely related genera in species identification (three copepod genera, Bucklin et al., 1999). Efforts to use a segment of DNA as a barcode of identity have successfully employed COI to identify various taxa including fish species (Steinke et al., 2005; Ward et al., 2005). Because COI is informative for distinguishing species across and within many different taxa, it is well suited for identification across a family as diverse as the Scombridae. In this study, a molecular key is developed based on the mitochondrial COI region for the specific identification of the 17 scombrids present in the western Atlantic Ocean.

Materials and Methods

Tissue samples were obtained from up to 20 specimens of each of the 17 scombrid species common to the western Atlantic Ocean: *Acanthocybium solandri* (Cuvier, 1831); *Auxis rochei* (Risso, 1810); *Auxis thazard* (Lacépède, 1800); *Euthynnus alletteratus* (Rafinesque, 1810); *Katsuwonus pelamis* (Linnaeus, 1758); *Sarda sarda* (Bloch, 1793); *Scomber colias* Gme-

lin, 1789; *Scomber scombrus* Linnaeus, 1758; *Scomberomorus brasiliensis* Collette, Russo and Zavalla-Camin, 1978; *Scomberomorus cavalla* (Cuvier, 1829); *Scomberomorus maculatus* (Mitchill, 1815); *Scomberomorus regalis* (Bloch, 1793); *Thunnus alalunga* (Bonnaterre, 1788); *Thunnus albacares* (Bonnaterre, 1788); *Thunnus atlanticus* (Lesson, 1830); *Thunnus obesus* (LOWE, 1834) and *Thunnus thynnus* (Linnaeus, 1758). All specimens were identified based on morphological characters (Collette and Nauen, 1983). Tissue samples were either stored in DMSO buffer (Seutin et al., 1991) or frozen. Published COI sequences of *A. thazard* and *A. rochei* (Infante et al., 2004) were used to supplement the number of samples for these species. Collection information is provided in Table 1.

To evaluate the efficacy of COI as a marker to identify scombrids, specimens of larval scombrids stored in ethanol were obtained from D. Richardson and R. Cowen, Rosenstiel School of Marine and Atmospheric Science, University of Miami. In addition, stomach content samples containing putative scombrids were collected from blue marlin and white marlin caught in recreational fishing operations out of Cape May, NJ, USA and La Guaira, Venezuela. Putative scombrids were removed from the marlin stomachs dockside and rinsed with water. Either a muscle sample was removed and placed in DMSO buffer (Seutin et al., 1991) or the whole fish was frozen until analysis.

Total genomic DNA was extracted from adult tissues of known scombrid species using a standard phenol/chloroform isolation protocol (modified from Sambrook and Russell, 2001). A series of extractions was performed on each sample using equilibrated phenol, followed by phenol: chloroform: isoamyl alcohol (25:24:1) and finally, chloroform: isoamyl alcohol (24:1). Following extraction, DNA was precipitated with ethanol. For larval fishes, one eyeball (right eyeball when available) was removed and rinsed with distilled water. DNA was extracted from this tissue using proteinase-K and Chelex beads (Bio-Rad Laboratories, Hercules, CA) (Estoup et al., 1996). Each larva was photographed using a digital camera attached to a stereomicroscope via a phototube, capturing as much detail as possible for future morphological or meristic analysis.

Primers that amplify the COI gene region across the scombrid family were designed using conserved regions of seven scombrid COI sequences (*A. rochei, A. thazard, E. alletteratus, K. pelamis, S. scombrus, T. alalunga, and T. thynnus)* available through GenBank (National Center for Biotechnology Information). Two sets of primers were developed that amplify a ~950 base pair (bp) fragment (long fragment) of the COI gene and a ~250 bp fragment (short fragment) located within the 950 bp fragment:

950 bp fragment:	LCOI 121	CTA AGC CAA CCA GGT GCC CTT CT
	HCOI 1199	AAT AGT GGG AAT CAG TGT ACG A
250 bp fragment:	LCOI 646	AAT ACA ACC TTC TTC GAC C
	HCOI 947	GTT GGA ATT GCG ATA ATC

(The number in the primer name designates the position of the 5′ end within the COI gene (1550 bp)). All primers were ordered from Invitrogen Corporation (Carlsbad, California).

Polymerase chain reactions were performed on each known or putative scombrid sample. Each 25 μl reaction consisted of 0.25 μl template DNA, 2.5 μl 10× PCR Buffer plus magnesium (QIAGEN, Inc., Valencia, CA), 0.5 μl 10 mM dNTP (QIAGEN), 0.25 μl forward primer (100 pm μl^{-1}), 0.25 μl reverse primer (100 pm μl^{-1}), 0.125 μl *Taq* DNA polymerase (QIAGEN), 5.0 μl BSA (bovine serum albumin) (1 mg mL^{-1}) and 16.125 μl sterile water (modified from McDowell and Graves, 2002). Amplifications using Chelex extractions contained 2.5 μl DNA template. Reactions were carried out in an MJ Research Corporation PTC-200 Peltier thermal cycler (Watertown, MA) under the following conditions: initial denaturation at 94 °C for 4 min, followed by 35 cycles of 94 °C for 1 min, 57 °C for 1 min, 72 °C for 2 min, a final extension at 72 °C for 5 min, and final hold at 4 °C. Amplifications done with the LCOI 646/ HCOI 1085 primers used an annealing temperature of 54 °C, but were otherwise run using the same conditions.

Table 1. Scombrid reference sample collection location and sample size.

Species name	Common name	Abbreviation	Catch location (no. of samples), date collected	Sample size
Acanthocybium solandri	Wahoo	ASOL	Cape May, New Jersey (12), 2004; Gulf of Mexico (5); Isla Mujeres, Mexico (1)	18
Auxis rochei	Bullet tuna	AUXR	Virginia (4), 2004; Philippines (1), 2003; Pacific (3) and Atlantic/Mediterranean (8) from Infante et al., 2004.	16
Auxis thazard	Frigate tuna	AUXT	Panama, Pacific (5), 2004; Philippines (1), 2003; Seychelles (1), 2002; Spain (3) from Infante et al., 2004.	10
Euthynnus alletteratus	Little tunny	EUTH	Isla Mujeres, Mexico (10)	10
Katsuwonus pelamis	Skipjack tuna	SKJT	Virginia (16), 2004; Hawaii (3), 2004	19
Sarda sarda	Atlantic bonito	SARD	Mid-Atlantic Bight (1), 1994; Mediterranean (2), 2002	3
Scomber colias	Chub mackerel	CHMK	Gulf of Mexico (10); Virginia (4), 2004; Argentina (3); Ivory Coast (3), 1992	20
Scomber scombrus	Atlantic mackerel	ATLM	Massachusetts (10); Plymouth, England (10)	20
Scomberomorus brasiliensis	Serra Spanish mackerel	SCBR	Brazil (2); Trinidad (1); Saba (1), 2006	4
Scomberomorus cavalla	King mackerel	KMCK	Panama City, Florida (2), 1994; South Carolina (2), 1996; Virginia (3), 2005; Maryland (1), 2005; Florida Keys (1), 2006	9
Scomberomorus maculatus	Spanish mackerel	SPMK	South Carolina (13), 1997; Chesapeake Bay (5), 1994; Gulf of Mexico (2)	20
Scomberomorus regalis	Cero	CERO	Bahamas (2); Colombia (1); Florida (1)	4
Thunnus alalunga	Albacore	ALBC	Mid-Atlantic Bight (3), 2004; Atlantic (12); North Carolina (2), 2005	17
Thunnus albacares	Yellowfin tuna	YFT	Mid-Atlantic Bight (14), 2004; Hawaii (4), 2004	18
Thunnus atlanticus	Blackfin tuna	BLKF	Isla Mujeres (6), 2004; Florida Straits (5), 2004; North Carolina (7) 2005	18
Thunnus obesus	Bigeye tuna	BET	Florida Straits (14), 2004; Hawaii (4), 2004	18
Thunnus thynnus	Atlantic bluefin tuna	BLFT	Ocean City, Maryland (14), 2003; Spain (2), 1998; Mediterranean (2), 1998; New Jersey (2)	18

Sequencing was performed on either gel-based or capillary-based automated sequencers. For gel-based sequencing, purified PCR (using ExoSAP; USB Corporation) products were cycle sequenced using a Thermo Sequenase Primer Cycle Sequencing Kit (Amersham Biosciences, Piscataway, NJ) and loaded onto a Li-Cor NEN IR^2 4200 global sequencing system (Li-Cor, Lincoln, NE). The sequencing program eSeq version 2.0 was used to read sequences and to check base calls. For capillary-based sequencing, purified PCR products were cycle sequenced using a 1/8 dilution of the manufacturer's (Applied Biosystems BigDye) sequencing reaction protocol for a 5 µl reaction: 0.25 µl BigDye reagent, 0.875 µl 5× BigDye Buffer, 0.32 µl primer, 1.0 µl template (10–40 ng for 1000 bp product), 2.55 µl water. The sequencing reaction products were loaded onto an ABI 3100 capillary sequencer (Applied Biosystems, Foster City, CA) and analyzed using the program Sequencing Analysis 5.1.1.

Both primer pairs successfully amplified samples taken from known scombrids. The longer fragment was used to generate information on the reference samples to identify those sites that discriminated between species. The internal primer pair was designed after the sequence of the long fragment was known, in order to amplify the shortest possible fragment that included informative sites. For the "unknown" samples (larvae, stomach contents), PCR was performed using the shorter primer pair. In cases where the shorter primer pair did not generate a PCR product, universal COI primers were used as a positive control. If the universal primers generated an amplicon, the sample was inferred to be a non-scombrid. If the universal primers did not result in a successful amplification, the sample was considered too degraded for analysis. The universal primers were designed using Primaclade (Gadberry et al., 2005) with known COI sequences of several marine fish species in GenBank: *Mola mola* Linnaeus, 1758, *Trachurus trachurus* Linnaeus, 1758, *Hoplostethus japonicus* Hilgendorf, 1879, *Antigonia capros* Lowe, 1843, *Emmelichthys struhsakeri* Heemstra and Randall, 1977, *A. thazard*, *E. alletteratus*, *T. alalunga*, *K. pelamis*, *S. scombrus*, *and T. thynnus*. The universal COI primers used were: COIuniv113Fdeg (5′GRG CNG ARC TAA GYC AAC C3′) and COIuniv697R (5′CCR AAG AAT CAG AAB AGR TG3′).

All sequences were edited using Sequencher version 4.2.2 (Gene Codes Corp.). Edited reference sequences (long fragment) of each species were aligned using the ClustalW program in MacVector version 7.2 (Accelyrs Inc.) to assess intraspecific variation. A consensus sequence of all haplotypes was generated for each species and these representative sequences were aligned to visualize informative interspecific differences using the program MEGA version 3.0 (Kumar et al., 2004). MacClade v. 4.07 (Maddison and Maddison, 2005) was used to assess variability at each base position. For unknown samples, the species identity was inferred by noting where the sample sequence clustered in a UPGMA tree using absolute number of differences between the consensus sequences.

Preliminary analyses suggested either misidentification or introgression in one sample identified as *S. regalis* based on morphology. To discriminate between misidentification and introgression between *S. maculatus* and *S. regalis*, the nuclear ITS-1 region was analyzed in four samples of each of these two species using the primers F-ITS-1 (5′GAG GAA GTA AAA GTC GTA ACA AGG3′) and 5.8SR2 (5′GTG CGT TCG AAR KGT CGA TGA TCA AT3′) (K. Johnson, Virginia Institute of Marine Science, unpubl.). PCR products were cloned into the pCR4-TOPO vector (Invitrogen Corporation, Carlsbad, California) and three clones from each sample were sequenced. This fragment was amplified and sequenced as previously described for the long COI fragment, using the capillary sequencer except 5 µl of Q solution (Qiagen, Valencia, California) was used in the 25 µl reaction and the annealing temperature was 45 °C. The restriction enzyme *Sma* I (New England BioLabs, Ipswich, Massachusetts) was found to distinguish between *S. maculatus* and *S. regalis* based on sequence alignment information. Samples were subsequently distinguished using a 15 µl restriction digestion reaction using 0.5 µl *Sma* I, 1.5 µl 10× NEB4 10× reaction buffer, 11 µl deionized sterile water, and 2 µl PCR product, digested at 25 °C for 1 hr, then 65 °C for 20 min.

Results

Two amplicons were generated in this study, a long (945 bp) and a short (264 bp) fragment of the COI gene. Within the long fragment there were 279 (30%) variable sites and in the short fragment there were 64 (24%) variable sites. The vast majority (93.7%) of substitutions occurred at the third codon position, while only 5.7% occurred at the first position and 0.7% at the second position. There were no insertions or deletions within the COI regions analyzed.

The long fragment of COI exhibited a wide range of differences between the 17 species analyzed (GenBank Accession nos DQ835818–DQ835957; EF379124). The number of nucleotide differences between consensus sequences of each species ranged from 2 base changes (between *T. obesus* and *T. albacares* or *T. atlanticus*) to 152 (between *S. scombrus* and *S. cavalla*) (Table 2). The differences between species in the short fragment ranged from 1 base change (between *T. obesus* and *T. atlanticus* and between *S. maculatus* and *S. regalis*) to 48 (between *S. colias* and *A. thazard*). Within species, variation of the long fragment ranged from 0 in *S. brasiliensis* and *S. regalis* to 26 variable sites within *S. sarda*. Reference samples of *K. pelamis*, *T. albacares*, *T. obesus*, *A. rochei*, and *A. thazard* included Atlantic and Pacific individuals, and thus the intraspecific variation observed in these species encompassed any inter-oceanic differences.

The positively identified samples of a species consistently grouped together in a UPGMA tree of all COI sequences in this study. A consensus sequence was generated for each species to serve as a representative of that species in a reference UPGMA tree (Fig. 1). A single *S. regalis* sample had a COI sequence that was more similar to *S. maculatus* than *S. regalis*. To distinguish between misidentification and introgression, the nuclear ITS-1 region was sequenced (GenBank Accession nos EF379101–EF379123). The ITS-1 sequence data indicated that the anomalous sample was indeed *S. regalis*, suggesting introgression. The COI sequence of this sample was not included in the consensus sequence of *S. regalis*. The differences between these two species in the ITS-1 region are shown in Figure 2. From the ITS-1 sequence information, a restriction fragment length polymorphism was found between *S. regalis* and *S. maculatus* whereby the restriction enzyme *Sma*I would create 250 bp and 600 bp bands for *S. regalis* and 280 bp and 600 bp bands for *S. maculatus* (Fig. 3).

From the consensus sequences, an unambiguous molecular key was developed that allows identification of all 17 western Atlantic scombrids. Positions at which a species has a consistent, unique combination of nucleotide base pairs are indicated in Figure 4. The shorter COI fragment also provided dependable species identification as it includes diagnostic sites with the added advantage of the ability to amplify smaller, degraded DNA fragments. Clustering of an unidentified sample in the UPGMA tree was the quickest method of identification, but in cases where an individual did not clearly group with one species, discriminatory base positions were located in the unknown sequence and compared with the molecular key for identification.

Scombrid larvae from the Florida Straits in the western North Atlantic Ocean were used to test the efficacy of this marker. These individuals were sufficiently large (4.5–12 mm) to be identified morphologically to genus; however, some were damaged, making specific identification based on morphological characters difficult if not impossible. Fifty-two scombrid larvae were identified based on DNA sequence and, when possible, using morphological characters following Richards (2006) and

Table 2. Pairwise distances between western Atlantic scombrids based upon number of nucleotide differences for consensus haplotype COI sequences of each species. Differences in the longer fragment (945 bp) are shown in the lower left portion of the matrix and differences in the shorter fragment (260 bp) are shown in the upper right. The column labeled "Intra" contains the number of variable sites in the long fragment within each of the species in that row (MEGA version 3.0 Kumar, Tamura, Nei, 2004). Species abbreviations are given in Table 1.

	ALBC	BLFT	BET	YFT	BLKF	AUXT	AUXR	EUTH	SKJT	SARD	ASOL	CHMK	ATLM	KMCK	SPMK	CERO	SCBR	Intra
ALBC		7	5	6	4	29	25	23	17	23	22	43	36	35	35	35	36	7
BLFT	17		2	4	3	29	24	23	17	22	20	42	34	37	33	33	34	15
BET	14	6		2	1	27	23	23	16	23	17	44	36	36	32	32	34	15
YFT	14	7	2		3	29	24	23	17	23	18	43	37	35	33	33	34	17
BLKF	12	8	2	4		28	24	26	16	24	19	46	38	36	36	36	37	8
AUXT	98	97	95	96	95		6	26	18	22	22	48	44	28	31	31	33	8
AUXR	88	87	86	88	87	23		26	16	20	20	44	42	26	31	31	31	14
EUTH	89	89	86	85	90	89	92		18	19	23	39	40	28	27	26	28	8
SKJT	80	85	82	84	80	65	58	73		15	18	43	36	26	31	31	32	17
SARD	82	77	77	80	79	88	84	83	85		21	35	34	26	27	27	28	26
ASOL	100	99	95	97	96	107	100	108	102	86		38	35	27	28	27	27	16
CHMK	130	125	127	126	130	126	118	126	120	115	138		25	43	37	37	38	16
ATLM	119	120	123	122	126	134	126	133	124	117	134	82		45	33	33	34	15
KMCK	132	129	132	128	132	137	130	134	124	107	118	143	152		23	24	23	9
SPMK	115	114	113	110	118	131	127	123	124	105	118	133	133	107		1	2	14
CERO	118	118	117	114	122	132	126	124	126	107	119	135	134	108	4		3	0
SCBR	124	121	120	116	124	135	128	127	126	110	115	133	138	114	18	20		0

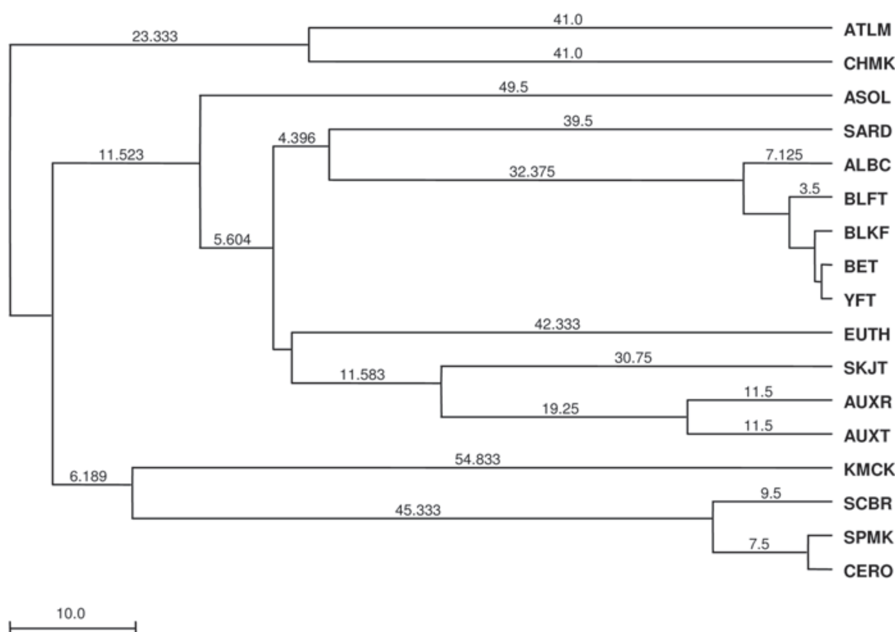

Figure 1. UPGMA tree based on absolute number of nucleotide differences between reference COI sequences. Each species group is a consensus sequence of all haplotypes of that given species. Species abbreviations are given in Table 1.

Nishikawa and Rimmer (1987). From these guides, useful morphological characters were: forebrain pigment and ventral pigment spot in *K. pelamis*, lower jaw pigmentation in *E. alletteratus*, and lateral tail pigmentation in *A. thazard*. Morphological identification to species level was possible for 18 larvae, and in each case, morphological assignment supported genetic identification. The remaining 34 unknown larvae were identified to species solely by noting their placement with known samples in a UPGMA tree of long fragment consensus sequences (Fig. 5).

To test the efficacy of the marker on degraded tissues, the short fragment was amplified from putative scombrids found in billfish stomach contents. The shorter fragment of COI was analyzed in the stomach contents as these tissues are generally deteriorated and therefore the DNA may also be degraded. When the sequences of these COI fragments were aligned to the known reference samples, nine samples clustered with *A. rochei* in the UPGMA tree (Fig. 6). Two samples from stomachs of billfishes caught in Hawaii did not cluster with any of the scombrid species. The search engine Basic Local Alignment Search Tool (BLAST) (National Center for Biotechnology Information) was used to find the closest match between these samples and species in GenBank (National Center for Biotechnology Information). One sample shared 84% identity with the COI gene of *Myripristis berndti* (blotcheye soldierfish) (Jordan and Evermann, 1903) and the other sample had 83% identity with the COI gene of *H. japonicus* (flintperch).

Taxon/Node	**24**	122	**144**	**152**	**164**	**187**	**270**	**576**	**577**	**578**	**579**	**591**	598	**602**	617	632
SPMK1ITS	T	G	T	C	C	T	C	T	C	T	C	C	G	C	G	A
SPMK3AITS	T	G	T	C	C	T	C	T	C	T	C	C	G	C	G	A
SPMK1AITS	T	A	T	C	C	T	C	T	C	T	C	C	G	C	G	A
SPMK3BITS	T	G	T	C	C	T	C	T	C	T	C	C	G	C	G	M
SPMK3ITS	T	G	T	C	C	T	C	T	C	T	C	C	G	C	G	A
SPMK2ITS	T	G	T	C	C	T	C	T	C	T	C	C	G	C	G	A
SPMK3CITS	T	G	T	C	C	T	C	T	C	T	C	C	G	C	G	A
SPMK1BITS	T	G	T	C	C	T	C	T	C	T	C	C	G	C	G	A
CERO1ITS	C	G	C	T	G	C	T	C	T	A	A	G	A	T	T	C
CERO4ITS	C	G	C	T	G	C	T	C	T	A	A	G	G	T	T	A
CERO3ITS	C	G	C	T	G	C	T	C	T	A	A	C	A	T	G	C
CERO1AITS	C	G	C	T	G	C	T	C	T	A	A	C	G	T	G	A
CERO3AITS	C	G	C	T	G	C	T	C	T	A	A	C	G	T	G	A
CERO2CITS	C	G	C	T	G	C	T	C	T	A	A	C	G	T	G	A
CERO3BITS	C	G	C	T	G	C	T	C	T	A	A	C	G	T	G	A
CERO1CITS	C	A	C	T	G	C	T	C	T	A	A	C	G	T	G	A
CERO2BITS	C	G	C	T	G	C	T	C	T	A	A	C	G	T	G	A
CERO2AITS	C	G	C	T	G	C	T	C	T	A	A	C	G	T	G	A
CERO4BITS	C	G	C	T	G	C	T	C	T	A	A	C	G	T	G	A
CERO3CITS	C	G	C	T	G	C	T	C	T	A	A	C	G	T	G	A
CERO1BITS	C	G	C	T	G	C	T	C	T	A	A	C	G	T	G	A
CERO4CITS	C	G	C	T	G	C	T	C	T	A	A	C	G	T	G	A
CERO4AITS	C	G	C	T	G	C	T	C	T	A	A	C	G	T	G	A

Figure 2. ITS-1 sequence alignment of *Scomberomorus maculatus* and *Scomberomorus regalis* showing interspecific nucleotide differences. Only variable sites are displayed, with diagnostic sites in bold. Sequence names with the letters A, B, or C denote cloned PCR product and the other sequences are direct sequence of PCR product.

Discussion

Both the long and short COI fragments met the two requirements of a good molecular marker for scombrid identification: consistent interspecific differences and minimal intraspecific variation. This molecular key was developed to unambiguously identify all scombrids occurring in the western Atlantic Ocean, several of which have a circumtropical distribution. To evaluate the applicability of the marker outside the Atlantic, several Pacific conspecifics of some of the circumglobal species covered in this study were sequenced to determine if intraspecific differences exist between conspecifics from different ocean basins. Previous studies have shown evidence of inter-oceanic differences in bigeye tuna and albacore based on other gene regions (Chow and Ushiama, 1995; Alvarado Bremer et al., 1998; Chow et al., 2000). The diagnostic sites of the COI region still allowed for unambiguous identification of

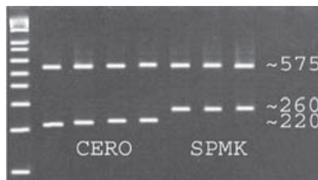

Figure 3. Restriction digestion pattern of *Scomberomorus regalis* and *Scomberomorus maculatus* samples using the restriction enzyme *Sma*I on ITS-1 fragments. First lane is 1KB+ ladder, lanes two through five are digests of *S. regalis* (CERO) samples and lanes six through eight are digests of *S. maculatus* (SPMK) samples. Bands have been accentuated using Adobe Photoshop and numbers associated with bands denote the size of the band in number of base pairs.

```
     6666666666666666666666666666666666677777777777777777777777777777777777777777777777777777777
     6666677777777778888888888999999999900000000001111111111222222222233333333334444444444555555
     5678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
ALBC CTGCAGGAGGGGGAGACCCAATCCTTTACCAGCATCTATTCTGATTCTTTGGACATCCAGAAGTCTACATTCTTATTCTTCCCGGATTCGG
BLFT ......N.......R................R..C.......K.......................................T..G.....
YFT  ...............................A..C.......................R................................
BET  ...............................A..C....................Y...................................
BLKF ...............................A..C........................................................
SKJT .......T..A...........T........A..C..............C..C.....R..............A...A..A..........
SARD .......C..R..T..T..C...........R..CT.............Y..C..C........A....................C.....
AUXR ......................T........A..C..............Y..T..................T.A.....G..A..C.....
AUXT ......................T........A..C..............C.....................T.A.....G..A..C.....
EUTH .......C........T..............A..C..............C..C..C........A......T.A.....A..A........
ASOL .C........A..T.................A..C.................Y..C..C.....A..T...........C..A.....T..
KMCK .A.....G..A....................A..C.................C..S..C........T........C........R.....
CERO .C.......................G..T.....C..............C..T..C...........T........C...........T..
SPMK .C.....N.................G..T.....CY.............C..T..C..C.R......T........C...........T..
SCBR .C.......................G..T..A..C..............C..G..C..C........T........C...........T..
ATLM .C.G......A.....T..T..T........A.................C.....C....................C..C..A.....Y..
CHMK ...GG..............C..T........A..C..C...........C..C..C..C....................K..A........
                                                                                     *
     7777777777777777777777777777777777777777777788888888888888888888888888888888888888888888888
     5555666666666677777777778888888888999999999900000000001111111111222222222233333333334444444
     6789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456
ALBC AATGATTTCCCACATTGTTGCCTACTACTCAGGTAAAAAAGAACCTTTCGGCTACATGGGTATGGTATGAGCCATGATGGCCATCGGCCTA
BLFT ...K..C....................................................................................
YFT  ......C.................................................................................Y..
BET  ...R..C....................................................................................
BLKF ......C..........................................................................T.........
SKJT ......Y........C............G.C.......................T..........................T.........
SARD ...A..Y.....T..C.....M........C....................T........Y.....R.....A...........Y.....W
AUXR ............T..C...........YG.C.........................................................Y..
AUXT ......C..Y..T..C...........TG.C.....................................................T......
EUTH ......C..T..................G.C.......................T........R..G........................
ASOL ...A.....T..T.....C........T.....C.................T..T.....Y....................T.........
KMCK ...A...........C..............T.................Y..R..Y........A..G........................
CERO ...A..........................C....................A........A..............................
SPMK ...A..Y.......................C....................A........A..............................
SCBR ...A..........................C....................A.......................................
ATLM ...............C............G.C..............C..............C..........................N...
CHMK ...A..C.....T..C........Y...G.C..............C..............Y........R.....................
                            *     *     *
     88888888888888888888888888888888888888888888888888888999999999999999999999999999999
     44455555555556666666666777777777788888888889999999999000000000011111111112222222222
     78901234567890123456789012345678901234567890123456789012345678901234567890123456789
ALBC CTAGGGTTCATTGTATGAGCCCAYCACATGTTCACGGTAGGAATGGACGTAGACACACGGGCATACTTTACATCCGCAACTAT
BLFT ...........C.....R.....C............................................Y..............
YFT  ...........C...........T...........A...............................................
BET  ...........C...........C.................W.........................................
BLKF ...........C...........C...........................................................
SKJT ...........C...........C...........A.......................A..C.................Y..
SARD ...........C.....R..............T..A.......................A..C..Y..C..............
AUXR ...........C...........C...........A..........................C.....C...........A..
AUXT ...........C..T..R.....C...........A.......................A..C.....C...........A..
EUTH .......................T...........A........R..............A..C.....C...........A..
ASOL ...........C...........C........T..A........A..............A..C.................A..
KMCK .....T........Y........T........T..A......W.....T..........A..C..Y..C...........A..
CERO .....A..T.....T........T........T..A.....T..A..............A........C..............
SPMK .....A..T.....T........T....A..T..A.....T..R..............A........C..............
SCBR .....A..T.....T........T....A..T..A.....T..A..............A........C..............
ATLM .....C..T.....C........C........T..A..T....................A.......................
CHMK .....C..T........R.....T........T..A..C.....A..............A..G..T..C..............
```

Figure 4. Molecular key of interspecific differences in the shorter COI fragment between consensus sequences. The sites that are useful in distinguishing very closely related species (i.e., *Thunnus albacares*/*Thunnus obesus*, *Thunnus atlanticus*/*Thunnus obesus*, and *Scomberomorus regalis*/*Scomberomorus maculatus*) have an asterisk and are bolded. Species abbreviations are given in Table 1.

circumglobal species between ocean basins. This was not unexpected given the high level of conservation in COI (Brown, 1985). The success of this marker in other ocean basins suggests that its utility may easily be extended to other scombrid species that occur elsewhere.

While the COI marker has proven to be effective for species identification of scombrids it, like other molecular markers has limitations. Sequencing has drawbacks including limitations imposed by cost and time, but this technology is being improved

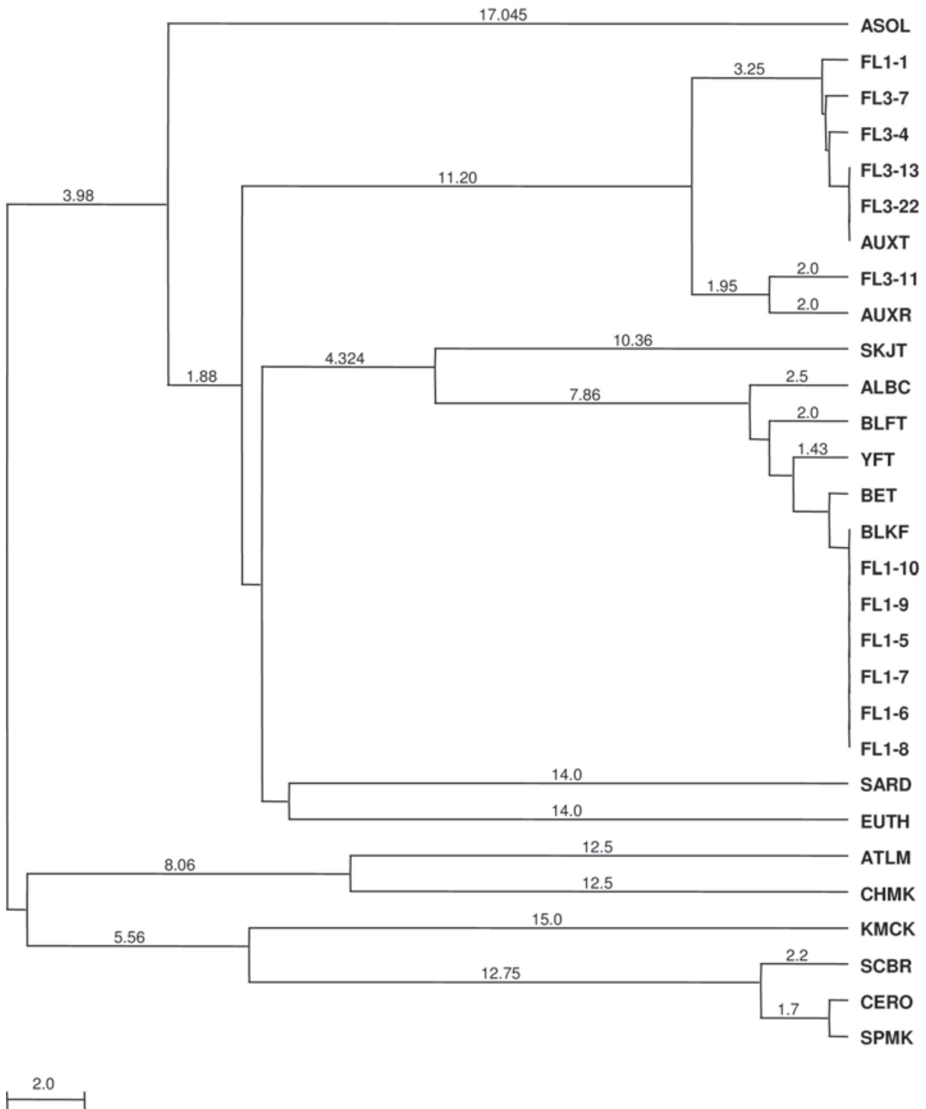

Figure 5. UPGMA tree including long fragment consensus COI sequences and a select number of unknown larval scombrid samples from Florida clustered with their respective species. Unknown samples are designated by a FL prefix. Species abbreviations are given in Table 1.

upon continually, making it quite an attractive high resolution technique for species identification. An alternative method using PCR/RFLP analysis of an amplified gene region (Chow and Inoue, 1993; Daniel and Graves, 1994; McDowell and Graves, 2002; Chow et al., 2003) is practical, but it becomes more difficult with increasing number of species to find unique or unambiguous fragment patterns that will distinguish each species. Similarly, a multiplex assay increases speed and decreases cost of analysis, but requires the design of many species-specific primers, which would be challenging given the number of species in this study.

Another concern with using only a mitochondrial marker is the possibility that introgression may lead to the misidentification of samples. Mitochondrial introgres-

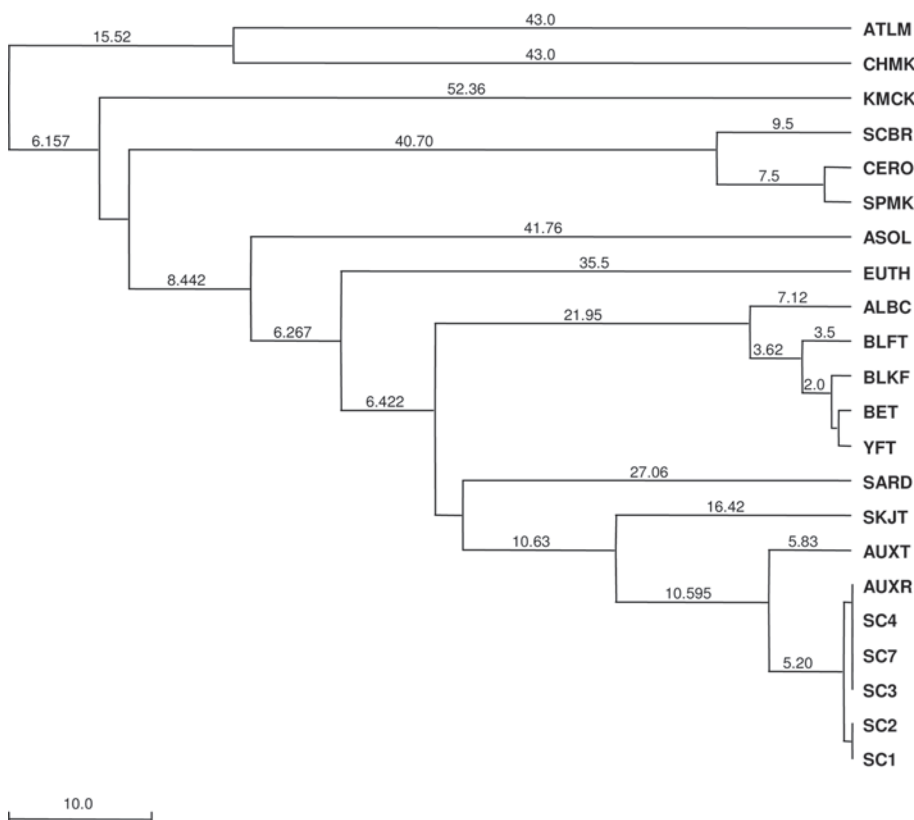

Figure 6. UPGMA tree of consensus reference COI sequences and unknown billfish stomach content samples. Unknown samples are denoted by a number with a SC prefix. Five of the nine stomach content samples analyzed are shown here. All nine samples cluster with *Auxis rochei*. Species abbreviations are given in Table 1.

sion has been previously reported in scombrids. The mitochondrial genome of the albacore, *T. alalunga,* has introgressed onto the Pacific bluefin tuna *T. orientalis* genetic background within the Pacific at a high frequency (98%) (Chow and Kishino, 1995) and at a low frequency (5%; 6.8%) in the eastern Atlantic/Mediterranean in Atlantic bluefin tuna *T. thynnus* (Vinas et al., 2003; Carlsson et al., 2004). Additionally, mitochondrial introgression has also been reported in the genus *Scomberomorus*. Banford et al. (1999) posited that the *S. regalis* mitochondrial genome has introgressed into *S. maculatus*. In the present study, one *S. regalis* sample clustered with *S. maculatus*, suggesting either misidentification or introgression. Subsequent analysis of the nuclear ITS region demonstrated that introgression may be bidirectional, as the observed introgression is in the opposite direction as that seen in the study by Banford et al. (1999). Clearly there is a need for further analysis, including more *S. regalis* samples, to adequately resolve this issue. Until then, a nuclear marker should be employed in addition to the COI marker to verify species identity of any putative *S. regalis* or *S. maculatus* samples. The RFLP in ITS-1 found in this study, which generated different sized bands between these species, could be used for future studies as a quick diagnostic for introgression.

The demonstrated ability of this key to provide species identifications of scombrid larvae and scombrid remains in stomach contents indicates its potential for use

in population studies, forensic analyses, and early life history investigations. This marker has numerous applications, from verifying that samples are indeed the correct species in population studies employing analysis of nuclear microsatellite loci, to providing species level identification of fillets that are being sold illegally, which is critical for management enforcement (Lopez and Pardo, 2005). While DNA barcoding has been proven successful for species identification (Hebert et al., 2003; Ward et al., 2005), the present study demonstrates the potential of using a molecular marker to identify degraded tissue, a notoriously difficult task. In cases where morphological identification is not possible, a molecular key provides a reliable and relatively rapid means of unambiguously identifying scombrid species.

Acknowledgments

Many thanks to all those who generously provided tissue or DNA samples: D. Richardson, R. Cowen, B. Collette, D. Kerstetter, K. Neill, J. Walter, O. Sanjur, T. Orrell, A. Mahon, K. Carpenter, P. D'Antoni, and H. Banford. A. Pao assisted with stomach content analyses and E. Brasseur facilitated taking pictures of the larvae. Funding was provided through Cooperative Marine Education and Research (CMER) and NOAA Highly Migratory Species awards. Virginia Institute of Marine Science contribution no. 2823.

Literature Cited

Alvarado Bremer, J. R., B. Stequert, N. W. Robertson, and B. Ely. 1998. Genetic evidence for inter-oceanic subdivision of bigeye tuna (*Thunnus obesus*) populations. Mar. Biol. 132: 547–557.

Avise, J. C. 1994. Molecular markers, natural history, and evolution. Chapman & Hall. New York.

Banford, H. M., E. Bermingham, B. B. Collette, and S. S. McCafferty. 1999. Phylogenetic systematics of the *Scomberomorus regalis* (Teleostei: Scombridae) species group: molecules, morphology and biogeography of Spanish mackerels. Copeia 1999: 596–613.

Bartlett, S. E. and W. S. Davidson. 1991. Identification of *Thunnus* tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome *b* genes. Can. J. Fish. Aquat. Sci. 48: 309–317.

Blankenship, L. E. and A. A. Yayanos. 2005. Universal primers and PCR of gut contents to study marine invertebrate diets. Mol. Ecol. 14: 891–899.

Brown, W. M. 1985. The mitochondrial genome of animals. Pages 95–130 *in* R. J. MacIntyre, ed. Molecular Evolutionary Genetics. Plenum Press, New York.

Bucklin, A., M. Guarnieri, R. S. Hill, A. M. Bentley, and S. Kaartvedt. 1999. Taxonomic and systematic assessment of planktonic copepods using mitochondrial COI sequence variation and competitive, species-specific PCR. Hydrobiologia 401: 239–254.

Carlsson, J., J. R. McDowell, P. Diaz-Jaimes, J. E. L. Carlsson, S. B. Boles, J. R. Gold, and J. E. Graves. 2004. Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) population structure in the Mediterranean Sea. Mol. Ecol. 13: 3345–3356.

Chow, S. and S. Inoue. 1993. Intra- and interspecific restriction fragment length polymorphism in mitochondrial genes of *Thunnus* tuna species. Bull. Natl. Res. Inst. Fish. 30: 207–224.

_______ and H. Kishino. 1995. Phylogenetic relationships between tuna species of the genus *Thunnus* (Scombridae: Teleostei): Inconsistent implications from morphology, nuclear and mitochondrial genomes. J. Mol. Evol. 41: 741–748.

_______ and H. Ushiama. 1995. Global population structure of albacore (*Thunnus alalunga*) inferred by RFLP analysis of the mitochondrial ATPase gene. Mar. Biol. 123: 39–45.

_______, H. Okamoto, N. Miyabe, K. Hiramatsu, and N. Barut. 2000. Genetic divergence between Atlantic and Indo-Pacific stocks of bigeye tuna (*Thunnus obesus*) and admixture around South Africa. Mol. Ecol. 9: 221–227.

_______, K. Nohara, R. Tanabe, T. Itoh, S. Tsuji, Y. Nishikawa, S. Uyeyanagi, and K. Uchikawa. 2003. Genetic and morphological identification of larval and small juvenile tunas (Pisces: Scombridae) caught by a mid-water trawl in the western Pacific. Bull. Fish. Res. Agen. 8: 1–14.

Collette, B. B. and C. E. Nauen. 1983. FAO species catalogue. Vol. 2: Scombrids of the world: an annotated and illustrated catalogue of tunas, mackerels, bonitos and related species known to date. FAO Fisheries Synopsis No. 125: 137 p.

Daniel, L. B., III and J. E. Graves. 1994. Morphometric and genetic identification of eggs of spring-spawning sciaenids in lower Chesapeake Bay. Fish. Bull. 92: 254–261.

Estoup, A., C. R. Largiader, E. Perrot, and D. Chourrout. 1996. Rapid one-tube DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes. Mol. Mar. Biol. Biotech. 5: 295–298.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotech. 3: 294–299.

Gadberry, M. D., S. T. Malcomber, A. N. Doust, and E. A. Kellogg. 2005. Primaclade–a flexible tool to find conserved PCR primers across multiple species. Bioinformatics 21: 1263–1264.

Gorokhova, E. 2006. Molecular identification of the invasive cladoceran *Cercopagis pengoi* (Cladocera: Onychopoda) in stomachs of predators. Limnol. Oceanogr. Methods 4: 1–6.

Graves, J. E., M. A. Simovich, and K. M. Schaefer. 1988. Electrophoretic identification of early juvenile yellowfin tuna, *Thunnus albacares*. Fish. Bull. 86: 835–838.

Hare, J. A., R. K. Cowen, J. P. Zehr, F. Jeanes, and K. H. Day. 1994. Biological and oceanographic insights from larval labrid (Pisces: Labridae) identification using mtDNA sequences. Mar. Biol. 118: 17–24.

Harper, J. L., R. A. King, C. S. Dodd, J. D. Harwood, D. M. Glen, M. W. Bruford, and W. O. C. Symondson. 2005. Rapid screening of invertebrate predators for multiple prey DNA targets. Mol. Ecol. 14: 819–827.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. deWaard. 2003. Biological identifications through DNA barcodes. Proc. Roy. Soc. Lond., B 270: 313–321.

Hyde, J. R., E. Lynn, R. Humphreys Jr., M. Musyl, A. P. West, and R. Vetter. 2005. Shipboard identification of fish eggs and larvae by multiplex PCR, and description of fertilized eggs of blue marlin, shortbill spearfish, and wahoo. Mar. Ecol. Prog. Ser. 286: 269–277.

Infante, C., G. Catanese, M. Ponce, and M. Manchado. 2004. Novel method for the authentication of frigate tunas (*Auxis thazard* and *Auxis rochei*) in commercial canned products. J. Agric. Food Chem. 52: 7435–7443.

Kirby, R. R. and P. C. Reid. 2001. PCR from the CPR offers a historical perspective on marine population ecology. J. Mar. Biol. Assoc. U. K. 81: 539–540.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinfo. 5: 150–163.

Lopez, I. and M. A. Pardo. 2005. Application of relative quantification TaqMan real-time polymerase chain reaction technology for the identification and quantification of *Thunnus alalunga* and *Thunnus albacares*. J. Agric. Food Chem. 53: 4554–4560.

Maddison, D. R. and W. P. Maddison. 2005. MacClade v. 4.07. Sinauer Associates, Inc. Sunderland, MA.

McDowell, J. R. and J. E. Graves. 2002. Nuclear and mitochondrial DNA markers for specific identification of istiophorid and xiphiid billfishes. Fish. Bull. 100: 537–544.

Morgan, R. P. 1975. Distinguishing larval white perch and striped bass by electrophoresis. Chesapeake Sci. 16: 68–70.

Nishikawa, Y. and D. W. Rimmer. 1987. Identification of larval tunas, billfishes and other scombroid fishes (suborder Scombroidei): an illustrated guide. CSIRO Marine Laboratories no. 186: 26 p.

Perez, J., P. Alvarez, J. L. Martinez, and E. Garcia-Vazquez. 2005. Genetic identification of hake and megrim eggs in formaldehyde-fixed plankton samples. ICES J. Mar. Sci. 62: 908–914.

Quintero, J., C. G. Sotelo, H. Rehbein, S. E. Pryde, I. Medina, R. I. Perez-Marin, M. Rey-Mendez, and I. M. Mackie. 1998. Use of mtDNA direct polymerase chain reaction (PCR) sequencing and PCR-Restriction fragment length polymorphism methodologies in species identification of canned tuna. J. Agric. Food Chem. 46: 1662–1669.

Ram, J. L., M. L. Ram, and F. F. Baidoun. 1996. Authentication of canned tuna and bonito by sequence and restriction site analysis of polymerase chain reaction products of mitochondrial DNA. J. Agric. Food Chem. 44: 2460–2467.

Rawson, P. D. and R. S. Burton. 2002. Functional coadaptation between cytochrome *c* and cytochrome *c* oxidase within allopatric populations of a marine copepod. PNAS 99: 12955–12958.

Richards, W. J. 2006. Scombridae: Mackerels and Tunas. Pages 2187–2227 *in* W. J. Richards, ed. Early stages of Atlantic fishes: an identification guide for the western central North Atlantic. Boca Raton.

____________, T. Potthoff, and J. -M. Kim. 1990. Problems identifying tuna larvae to species (Pisces: Scombridae: Thunnus) from the Gulf of Mexico. Fish. Bull. 88: 607–609.

Rocha-Olivares, A. 1998. Multiplex haplotype-specific PCR: a new approach for species identification of the early life stages of rockfishes of the species-rich genus *Sebastes* Cuvier. J. Exp. Mar. Biol. Ecol. 231: 279–290.

Sambrook, J. and D. W. Russell. 2001. Molecular cloning: A laboratory manual. Third edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

Seutin, G., B. N. White, and P. T. Boag. 1991. Preservation of avian blood and tissue samples for DNA analysis. Can. J. Zool. 69: 82–90.

Smith, P. J., S. M. McVeagh, V. Allain, and C. Sanchez. 2005. DNA identification of gut contents of large pelagic fishes. J. Fish Biol. 67: 1178–1183.

Steinke, D., M. Vences. W. Saltzburg, and A. Meyer. 2005. DNA barcoding Australia's fish species. Phil. Trans. R. Soc., B 360: 1975–1980.

Terol, J., R. Mascarell, V. Fernandez-Pedrosa, and M. Perez-Alonso. 2002. Statistical validation of the identification of tuna species: Bootstrap analysis of mitochondrial DNA sequences. J. Agric. Food Chem. 50: 963–969.

Vinas, J., C. Pla, M. Y. Tawil, A. Hattour, A. F. Farrugia, and J. M. de la Serna. 2003. Mitochondrial genetic characterization of bluefin tuna (*Thunnus thynnus*) from three Mediterranean (Libya, Malta, Tunisia); and one Atlantic locations (Gulf of Cadiz). ICCAT 55: 1282–1288.

Ward, R. D. 1995. Population genetics of tunas. J. Fish Biol. 47 (Supplement A): 259–280.

__________, T. S. Zemlak, B. H. Innes, P. R. Last, and P. D. N. Hebert. 2005. DNA barcoding Australia's fish species. Phil. Trans. R. Soc., B 360: 1847–57.

Date Submitted: 10 July, 2006. Date Accepted: 14 February, 2007.

Address: (M.A.P., J.R.M., J.E.G.) *Virginia Institute of Marine Science, School of Marine Science, College of William and Mary, Rte 1208 Greate Rd, Gloucester Point, Virginia 23062.* Corresponding Author: (M.A.P.) *E-mail: <melpaine@vims.edu>.*