MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) doi: 10.1111/j.1755-0998.2011.03039.x

Establishment of a mitochondrial DNA sequence database for the identification of fish species commercially available in South Africa

DONNA-MAREÈ CAWTHORN,* HARRIS ANDREW STEINMAN† and R. CORLI WITTHUHN*
*Department of Food Science, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa, †Food & Allergy Consulting & Testing Services (F.A.C.T.S), PO Box 565, Milnerton 7435, South Africa

Abstract

The limitations intrinsic to morphology-based identification systems have created an urgent need for reliable genetic methods that enable the unequivocal recognition of fish species, particularly those that are prone to overexploitation and/or market substitution. The aim of this study was to develop a comprehensive reference library of DNA sequence data to allow the explicit identification of 53 commercially available fish species in South Africa, most of which were locally caught marine species. Sequences of approximately 655 base pairs were generated for all species from the cytochrome *c* oxidase I (COI) gene, the region widely adopted for DNA barcoding. Specimens of the genus *Thunnus* were examined in further detail, employing additional mitochondrial DNA control region sequencing. Cumulative analysis of the sequences from the COI region revealed mean conspecific, congeneric and confamilial Kimura 2-parameter distances of 0.10%, 4.58% and 15.43%, respectively. The results showed that the vast majority (98%) of fish species examined could be readily differentiated by their COI barcodes, but that supplementary control region sequencing was more useful for the discrimination of three *Thunnus* species. Additionally, the analysis of COI data raised the prospect that *Thyrsites atun* (snoek) could constitute a species pair. The present study has established the necessary genetic information to permit the unambiguous identification of 53 commonly marketed fish species in South Africa, the applications of which hold a plethora of benefits relating to ecology research, fisheries management and control of commercial practices.

Keywords: COI region, DNA barcoding, DNA-based identification, mitochondrial DNA, South African fish species

Received 18 February 2011; revision received 13 May 2011; accepted 17 May 2011

Introduction

The rapidly declining state of the world's fish stocks and the escalating incidence of fraud and species substitutions on global markets have culminated in an urgent need to reassess the manner in which fish species are identified (Di Finizio *et al.* 2007; Jacquet & Pauly 2008). Over 30 000 fish species exist worldwide (Froese & Pauly 2010) and while these can often be discriminated by knowledgeable individuals when in their whole state, morphological differences may be subtle in some cases or not easily recognized when foreign fish are traded over international borders (Teletchea 2009). Furthermore, when the industrial processing of fish removes distinguishing external features, identification by means of visual inspection becomes difficult (Gil 2007).

Correspondence: Donna-Mareè Cawthorn, Fax: +27 21 808 3510; E-mail: 13946455@sun.ac.za

The sequencing of specific DNA fragments amplified by the polymerase chain reaction (PCR) is considered to be the most accurate and informative method for the identification of fresh, frozen and processed fish species (Lockley & Bardsley 2000; Gil 2007; Hubalkova *et al.* 2007). Nonetheless, the success of such authentication techniques relies on the establishment of databases containing reference DNA sequences from a large variety of expertly identified fish specimens (Martinez *et al.* 2005; Steinke *et al.* 2009a; Ward *et al.* 2009). The sequences of unknown samples can then be compared and potentially matched with those of reference samples deposited in international, open-access sequence databases, such as GenBank (http://www.ncbi.nlm.nih.gov) and the Barcode of Life Database (BOLD) (http://www.boldsystems.org).

The ability to make unambiguous species identifications using sequencing methods is highly dependent on the specific DNA marker chosen for analysis. For such purposes, it is imperative that the inter-species sequence

© 2011 Blackwell Publishing Ltd

variation presented in the region of interest is substantially greater than the intra-species variation (Ward *et al.* 2005). Although both the nuclear DNA (nDNA) and mitochondrial DNA (mtDNA) may be targeted for species identification, the analysis of mtDNA offers numerous advantages over the nDNA for fish species identification (Gil 2007). The maternal inheritance and lack of recombination of vertebrate mtDNA means that it is conserved across extended evolutionary distances, and the high copy number and mutation rate makes it suitable for discriminating between a large number of species (Rokas *et al.* 2003; Teletchea 2009). In particular, the mitochondrial ribosomal RNA genes, cytochrome *b* gene and the control region have been used in the past for the identification of different fish species and for the detection of commercial substitution of fish products (Lockley & Bardsley 2000; Rasmussen & Morrisey 2009).

To promote rapid and automated sample processing, it would be beneficial if the sequence diversity in a single standardized gene region could be globally utilized for species identification (Stoeckle 2003; Blaxter 2004). Recently it was proposed, and subsequently demonstrated, that an approximately 650 base pair (bp) region of the mitochondrial cytochrome *c* oxidase I (COI) gene has sufficient discriminatory power to serve as a unique 'barcode' for the identification of the large majority of animal species (Hebert *et al.* 2003a,b; Ward *et al.* 2005; Dawnay *et al.* 2007). The acceptance of the COI gene by the Consortium for the Barcode of Life for DNA barcoding purposes has provoked a number of international collaborative research efforts, including the Fish Barcode of Life Initiative (FISH-BOL), which aims to barcode all of the world's fish species (Ward *et al.* 2009).

South Africa comprises one of the most well established and lucrative fishing nations on the African continent (FAO 2010), owing largely to the wide variety of fish species found within its national fishing zone. In 2007, capture production for fisheries in South Africa was estimated at 670 571 metric tons, with the bulk of this production being consumed domestically (FAO 2009). In spite of the commercial importance of the fishing industry in this country and the fact that species substitutions have been suspected on the local market (Smith & Smith 1966; Von der Heyden *et al.* 2010), the public availability of reference DNA sequence data for many fish species commonly encountered in South Africa is limited or nonexistent. The information that is available is fragmented, representing sequences from varying gene regions for different taxonomic groups. This presents a major challenge for the identification of fish species in South Africa using DNA-based methods. The aim of this study was to construct a mtDNA database for 53 commercially significant fish species in South Africa, containing sufficient sequence data to permit the unambiguous identification of all reference specimens. Fulfillment of this aim entailed generation of sequence data from two different mtDNA loci (the COI for all fishes, as well as the control region for *Thunnus* species), evaluation of the utility of these DNA markers to allow explicit species resolutions, and the submission of reference sequences to public databases to allow remote comparisons for future fish authentication studies.

Materials and methods

Samples

Fifty-three putative fish species, representing 42 genera within 23 families and nine orders of ray-finned fishes (*Actinopterygii*), were collected from major trawling companies, fish processing facilities and the Department of Environmental Affairs and Tourism (Marine and Coastal Management Division) in South Africa (Table 1). The large majority of the fish in this collection were marine species, 47 of which were caught locally. Seven species were imported from outside African borders, but were included as these are commonly available on the South African market (Cawthorn *et al.* 2011). At least three specimens of each species were analysed, totaling 194 individuals. The numbers (*N*) of specimens per species ranged from 3 to 12, with a mean of 3.66 specimens per species. All specimens were morphologically identified at the species level by fish taxonomists based on established meristic, morphometric and colouration criteria. Vouchers specimens for all samples are currently stored at the Department of Food Science, Stellenbosch University, South Africa.

Extraction of DNA

Tissue subsamples were excised from the lateral muscle on the right-hand side of each fish specimen using a sterile scalpel and forceps. DNA was extracted using the SureFood® PREP allergen kit (r-Biopharm; supplied by AEC-Amersham, Cape Town, South Africa) according to the manufacturer's instruction. This method was found to extract the highest yields of DNA from fish muscle in a previous comparative study (Cawthorn *et al.* 2010). The concentration and purity of the extracted DNA was estimated at 260 and 280 nm in a spectrophotometer (Beckman Coulter DU530; Beckman Instruments, Fullerton, CA, USA). Extracted DNA was stored at −20 °C until further use.

PCR amplification

The extracted DNA from all specimens was used as a template for PCR amplification with oligonucleotide primers targeting the COI gene (Table 2). A 450-bp

© 2011 Blackwell Publishing Ltd

Table 1 Species, sample sizes (N), localities and GenBank accession numbers of all specimens analysed in this study, with grey shading representing the first entries into GenBank for the given gene region

Order	Family	Species	Common name* (Local name)	*N*	Country of origin	COI gene GenBank accession no.
Clupeiformes	Clupeidae	*Clupea harengus*	Atlantic herring	3	North sea	HQ611120–HQ611122
		Sardinops sagax	South American pilchard (pilchard/sardine)	3	South Africa	HQ611132–HQ611134
	Engraulidae	*Engraulis encrasicolus*	European anchovy (anchovy)	3	South Africa	HM007778–HM007780
		Engraulis japonicus	Japanese anchovy (Cape anchovy)	3	South Africa	HM007796–HM007798
Gadiformes	Merlucciidae	*Merluccius capensis*	Shallow-water Cape hake	4	South Africa	HM007690–HM007692, HQ611082
		Merluccius paradoxus	Deep-water Cape hake	7	South Africa	HM007683–HM007689
Lophiiformes	Lophiidae	*Lophius vomerinus*	Devil anglerfish (Cape monk)	3	South Africa	HM007765–HM007767
Ophidiiformes	Ophidiidae	*Genypterus blacodes*	Pink cusk-eel (ling)	3	New Zealand / Argentina	HQ611135, HQ611136, HQ611137
		Genypterus capensis	Kingklip	12	South Africa	HM007735–HM007746
Perciformes	Bramidae	*Brama brama*	Atlantic pomfret (angelfish)	4	South Africa	HQ611083–HQ611086
	Carangidae	*Seriola lalandi*	Yellowtail amberjack	4	South Africa	HM007727–HM007730
		Seriola quinqueradiata	Japanese amberjack	3	China	HQ641665–HQ641667
		Trachurus capensis	Cape horse mackerel (maasbanker)	3	South Africa	HM007775–HM007777
Perciformes	Coryphaenidae	*Coryphaena hippurus*	Common dolphinfish (dorado)	4	South Africa	HM007704–HM007707
	Centrolophidae	*Centrolophus niger*	Rudderfish (black ruff)	3	South Africa	HM007793–HM007795
		Schedophilus velaini	Violet warehou (black butterfish)	3	South Africa	HQ611129–HQ611131
		Seriolella brama	Common warehou	3	Australia / New Zealand	HM007731–HM007733, HM007734
	Epigonidae	*Epigonus telescopus*	Black cardinal fish (cardinal)	3	South Africa	HM007701–HM007703
	Gempylidae	*Lepidocybium flavobrunneum*	Escolar (butterfish)	3	South Africa	HM007724–HM007726
		Ruvettus pretiosus	Oilfish (butterfish)	3	South Africa	HM007721–HM007723
Perciformes	Gempylidae	*Thyrsites atun*	Snoek	6	South Africa / New zealand	HQ611106–HQ611108, HQ611109, HQ611110, HQ641670
	Istiophoridae	*Makaira nigricans*	Blue marlin	3	South Africa	HQ611114–HQ611116
		Tetrapturus angustirostris	Shortbill spearfish	3	South Africa	HQ611111–HQ611113
	Sciaenidae	*Argyrosomus inodorus*	Mild meagre (silver kob)	7	South Africa	HM007711–HM007717
		Argyrosomus japonicus	Japanese meagre (dusky kob)	3	South Africa	HM007718–HM007720
		Atractoscion aequidens	Geelbek croaker (geelbek/Cape salmon)	5	South Africa	HM007696–HM007700
	Scombridae	*Gasterochisma melampus*	Butterfly kingfish (gastora)	3	South Africa	HM007708–HM007710
		Katsuwonus pelamis	Skipjack tuna	3	South Africa	HQ611090–HQ611092
		Scomber japonicus	Chub mackerel	3	South Africa	HQ611117–HQ611119
		Scomberomorus commerson	Spanish mackerel (king mackerel/couta)	3	South Africa	HM007790–HM007792
		Thunnus alalunga	Albacore (longfin tuna)	3	South Africa	HM007772–HM007774
		Thunnus albacares	Yellowfin tuna	4	South Africa	HM007768–HM007771
		Thunnus obesus	Bigeye tuna	3	South Africa	HQ611138–HQ611140

© 2011 Blackwell Publishing Ltd

Table 1 Continued

Order	Family	Species	Common name* (Local name)	*N*	Country of origin	COI gene GenBank accession no.
	Serranidae	*Cephalopholis sonnerati*	Tomato hind (tomato rockcod)	3	South Africa	HQ611096–HQ611098
		Epinephelus marginatus	Dusky gouper (yellowbelly rockcod)	3	South Africa	HQ611093–HQ611095
	Sparidae	*Argyrozona argyrozona*	Carpenter seabream (silverfish)	3	South Africa	HM007753–HM007755
		Cheimerius nufar	Santer seabream	4	South Africa	HQ611102–HQ611105
		Chrysoblephus anglicus	Englishman seabream	3	South Africa	HQ611099–HQ611101
		Chrysoblephus laticeps	Roman seabream	3	South Africa	HM007750–HM007752
		Chrysoblephus puniceus	Slinger seabram	3	South Africa	HQ611087–HQ611089
Perciformes	Sparidae	*Pachymetopon blochii*	Hottentot seabream	3	South Africa	HM007693–HM007695
		Pterogymnus laniarius	Panga seabream	3	South Africa	HM007781–HM007783
		Rhabdosargus globiceps	White stumpnose	3	South Africa	HM007759–HM007761
	Trichiuridae	*Lepidopus caudatus*	Silver scabbardfish (buttersnoek/ ribbon fish)	3	South Africa	HM007784–HM007786
	Xiphiidae	*Xiphias gladius*	Swordfish	3	South Africa	HM007787–HM007789
Pleuronectiformes	Soleidae	*Austroglossus microlepis*	West coast sole	6	South Africa	HM007674–HM007678, HQ641664
		Austroglossus pectoralis	Mud sole (East coast sole)	4	South Africa	HM007679–HM007682
Salmoniformes	Salmonidae	*Oncorhynchus mykiss*	Rainbow trout	3	South Africa	HQ611123–HQ611125
		Oncorhynchus keta	Chum salmon (Alaskan salmon)	3	USA	HQ611126–HQ611128
		Salmo salar	Atlantic salmon	5	Norway / Canada	HM007799–HM007801, HQ641668–HQ641669
Scorpaeniformes	Sebastidae	*Helicolenus dactylopterus*	Blackbelly rosefish (jacopever)	3	South Africa	HM007747–HM007749
	Triglidae	*Chelidonichthys capensis*	Cape gurnard	3	South Africa	HM007756–HM007758
Zeiformes	Zeidae	*Zeus capensis*	Cape dory	3	South Africa	HM007762–HM007764

*Common names are from FishBase (http://www.fishbase.org).

Table 2 PCR primers used for the generation of mitochondrial DNA sequence data for fish species commercially available in South Africa

Primer	Primer sequence (5′ to 3′)	mtDNA target	Amplicon size (bp)	Reference
Fish DNA barcoding primer cocktail (C_FishF1t1/C_FishR1t1)				
VF2_t1	TGTAAAACGACGGCCAGTCAACCAACCACAAAGACATTGGCAC	Cytochrome *c* oxidase 1 (COI) gene	652	Ivanova *et al.* (2007)
FishF2_t1	TGTAAAACGACGGCCAGTCGACTAATCATAAAGATATCGGCAC			
FishR2_t1	CAGGAAACAGCTATGACACTTCAGGGTGACCGAAGAATCAGAA			
FR1d_t1	CAGGAAACAGCTATGACACCTCAGGGTGTCCGAARAAYCARAA			
M13F*	TGTAAAACGACGGCCAGT			Messing (1983)
M13R*	CAGGAAACAGCTATGAC			
Control region primers				
L15998	TAC CCC AAA CTC CCA AAG CTA	Control region (D-loop)	450	Alvarado Bremer (1994)
CSBDH	TgA ATT AGG AAC CAG ATG CCA G			

*Sequencing primers for M13-tailed PCR products.

fragment of the mtDNA control region (Table 2) was also amplified in cases where the amplification of the aforementioned gene region failed to resolve the identification of closely related species. PCR amplifications were carried out in a Mastercycler Personal (Eppendorf, Germany) utilizing the reaction mixtures and thermal

© 2011 Blackwell Publishing Ltd

cycling regimes shown in Table 3. PCR products were resolved by electrophoresis (90 V, 45 min) in 1.5% (m/v) agarose (Sigma-Aldrich, Gauteng, South Africa) gels and were visualized under an ultraviolet light.

Sequencing and sequence analysis

Purification of PCR products was performed using the NucleoFast 96 PCR Clean-up Kit (Macherey-Nagel; supplied by Separations, Gauteng, South Africa) according to the manufacturer's instruction. The purified PCR products were sequenced using BigDye chemistry and were analysed on an ABI 3100 Genetic Analyser (Applied Biosystems, Foster City, CA, USA). Primers for M13-tailed PCR products (Messing 1983) (Table 2) were used for the sequencing of COI amplicons, while the PCR amplification primers were used as sequencing primers for the control region amplicons (Table 2). The COI sequences were aligned using the complete alignment application in CLUSTAL X version 2.0 (Larkin *et al.* 2007) and the same procedure was used for the alignment of those sequences generated from the mtDNA control region. All generated DNA sequences were submitted to GenBank. The GenBank accession nos for COI sequences are given in Table 1, while the accession nos for the control region sequences generated from three *Thunnus* species are HQ853210–HQ853219. COI sequences, along with specimen and collection details, were also submitted to BOLD under the project name 'Barcoding marine fish species of South Africa'.

The mean overall nucleotide base frequencies for the COI sequences (Table 4) were computed in BOLD. COI sequence divergence values, hereafter referred to as distance or D, were computed using the Kimura 2-parameter (K2P) distance model (Kimura 1980). The mean K2P distances in percentages within different taxonomic levels are provided in Table 5, the distributions of which are represented in Fig. 1. The neighbour-joining (NJ) tree based on COI sequences was compiled in MEGA version 4.0 (Tamura *et al.* 2007) using the pairwise deletion of missing nucleotide data option and the robustness of topology nodes was evaluated using the nonparametric bootstrap method with 1000 replications (Felsenstein 1985). The full COI K2P/NJ tree has been lodged as an Electronic appendix (Fig. S1) and is presented in Fig. 2 in summary form.

Additional GenBank sequences

The validity of COI and control region sequencing for *Thunnus* species identification was further investigated by supplementing the sequences generated in this study for South African *Thunnus* specimens with additional sequences from GenBank for all eight recognized species within this genus (Collette *et al.* 2001). The number of

Table 3 Reaction mixtures and thermal cycling parameters used for PCR amplifications

Primers	Total volume	Primer 1 (10 μM)	Primer 2 (10 μM)	Primer 3 (10 μM)	Primer 4 (10 μM)	*Taq* DNA polymerase* (5 U/μL)	Reaction buffer* (10X)	$MgCl_2$* (25 mM)	dNTPs† (10 mM)	DNA template	PCR cycling conditions
C_FishF1t1/ C_FishR1t1	25 μL	0.25 μL (0.10 μM)	0.25 μL (0.10 μM)	0.25 μL (0.10 μM)	0.25 μL (0.10 μM)	0.125 μL (0.625 U)	2.5 μL (1X)	2.5 μL (2.5 mM)	0.5 μL (0.2 mM)	2.0 μL (*c.* 2 μg)	94 °C – 2 min; 35× (94 °C – 30 s, 52 °C – 40 s, 72 °C – 60 s); 72 °C – 10 min
L15998/ CSBDH	25 μL	1.0 μL (0.40 μM)	1.0 μL (0.40 μM)			0.10 μL (0.5 U)	2.5 μL (1X)	2.0 μL (2.0 mM)	2.0 μL (0.8 mM)	1.0 μL (*c.* 1 μg)	94 °C – 5 min; 35× (94 °C – 45 s, 54 °C – 45 s, 72 °C – 60 s); 72 °C – 10 min

*Super-Therm, supplied by Southern Cross Biotechnologies, Cape Town, South Africa.
†AB gene, supplied by Southern Cross Biotechnologies.

© 2011 Blackwell Publishing Ltd

Table 4 Mean percentage base compositions (with standard errors) of COI gene sequences

Gene	Number of species	G%	C%	A%	T%	GC%	GC% Codon position 1	GC% Codon position 2	GC% Codon position 3
COI	53	18.63 ± 0.07	28.37 ± 0.12	23.43 ± 0.08	29.58 ± 0.11	47.00 ± 0.15	56.24 ± 0.11	42.87 ± 0.03	41.60 ± 0.40

Table 5 Summary of genetic divergences in the COI region calculated for different taxonomic levels using K2P distances (%). Data are represented for 53 fish species within 42 genera and 23 families

Gene	Comparisons within	No. sequences	Taxa	No. comparisons	Minimum distance (%)	Mean distance (%)	Maximum distance (%)	Standard error distance (%)
COI	Species	194	53	320	0.00	0.10	1.54	0.020
	Genus	194	42	199	0.46	4.58	11.47	0.216
	Family	194	23	626	3.33	15.50	23.06	0.139
	Order	194	9	6705	14.49	22.75	29.89	0.033
	Class	194	1	10871	16.06	23.86	30.87	0.020

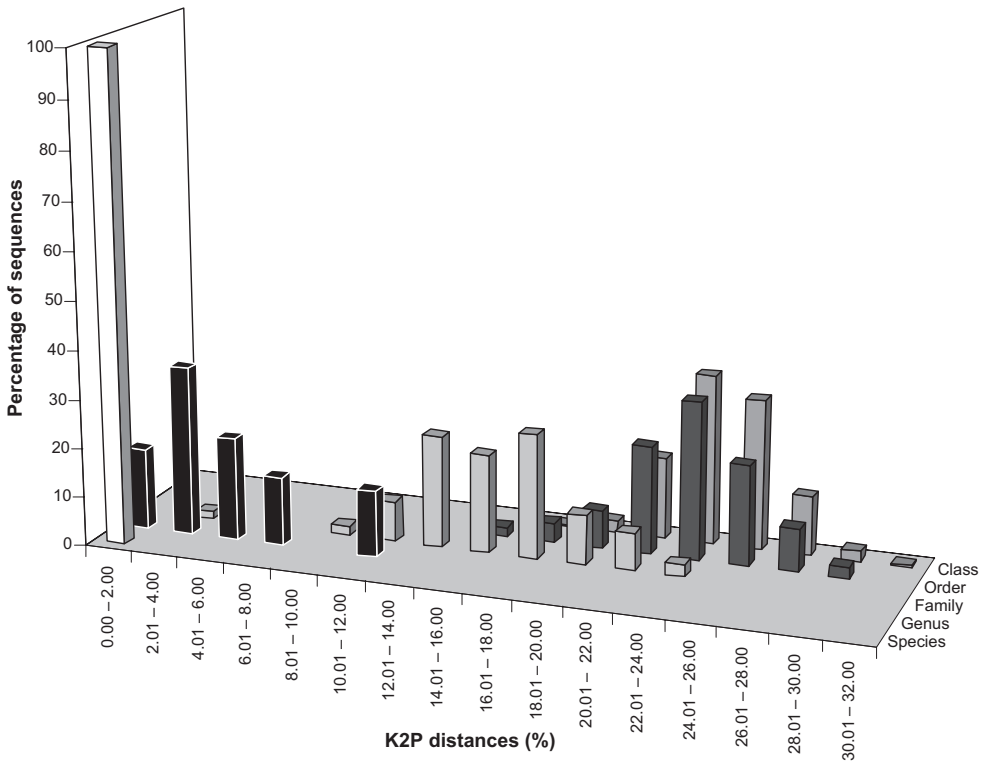


Fig. 1 The distribution of Kimura 2-parameter (K2P) distances (in percentages) for the cytochrome *c* oxidase I region within various taxonomic levels.

sequences acquired per species was based on their availability in GenBank. For the COI gene, 38 *Thunnus* sequences were analysed in total. These included six sequences from *Thunnus alalunga* (albacore), six from *Thunnus obesus* (bigeye tuna), five sequences each from *Thunnus albacares* (yellowfin tuna), *Thunnus orientalis* (northern bluefin tuna), *Thunnus maccoyii* (southern bluefin tuna) and *Thunnus atlanticus* (Atlantic blackfin tuna), as well as four from *Thunnus tonggol* (longtail tuna). COI data for *Thunnus thynnus* (Atlantic bluefin tuna) were limited in

© 2011 Blackwell Publishing Ltd

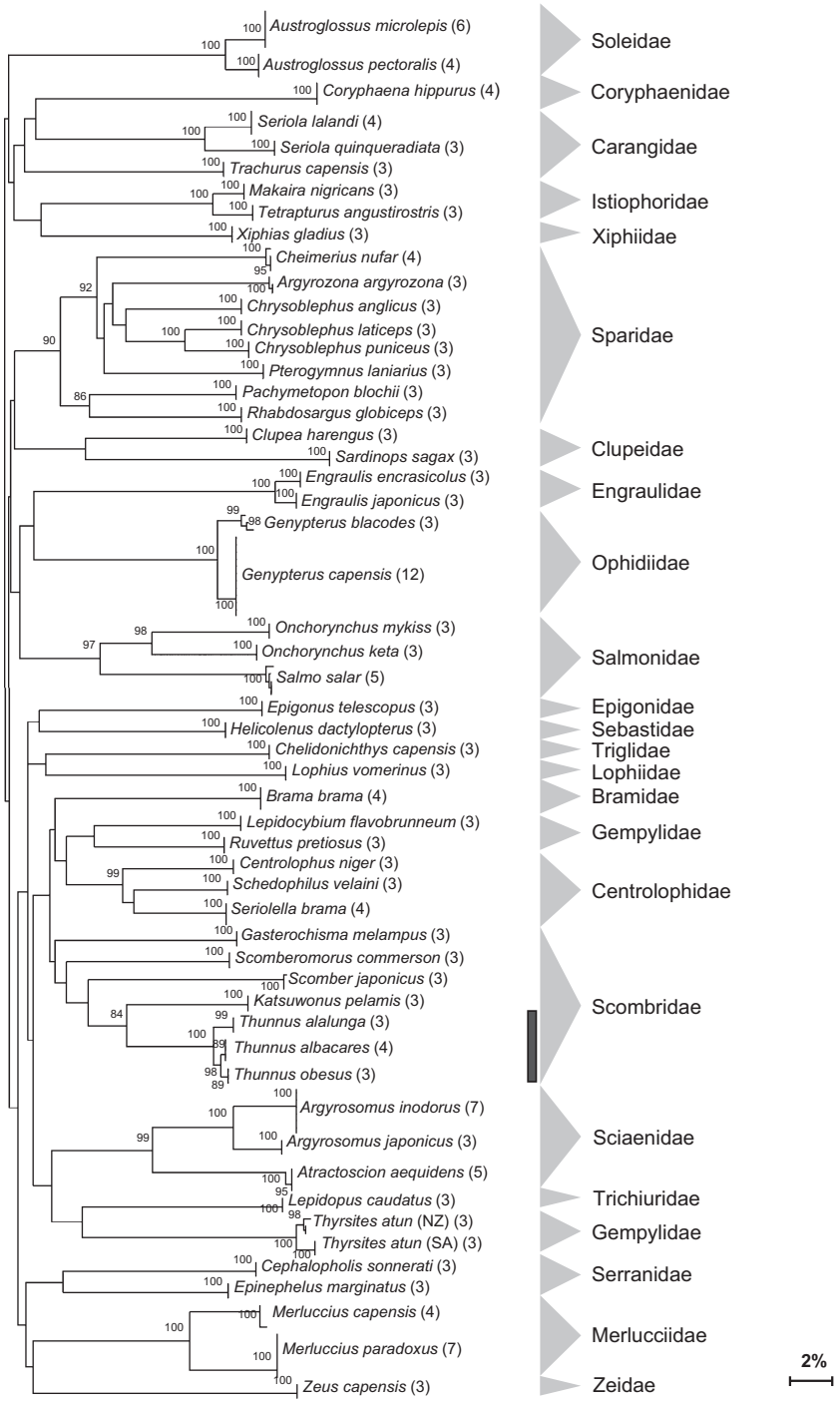


Fig. 2 Kimura 2-parameter distance neighbour-joining tree of cytochrome *c* oxidase I (COI) sequences generated from 53 fish species, with the numbers of specimens analysed per species in parenthesis. Numbers at nodes indicate bootstrap values (values higher than 80 are given). Species showing <2% COI divergence are indicated by a black bar.

GenBank and only two sequences from this species were included. The geographical origins and accession nos of all investigated COI sequences are provided in Fig. 3.

For the mtDNA control region, 27 sequences from the eight members of the genus *Thunnus* were analysed, which included six sequences each from *T. obesus* and *T. albacares*, five from *T. alalunga*, four from *T. thynnus* and three from *T. orientalis*. The single control region sequences available in GenBank for *T. atlanticus*, *T. maccoyii* and *T. tonggol* were also included. The accession nos

© 2011 Blackwell Publishing Ltd

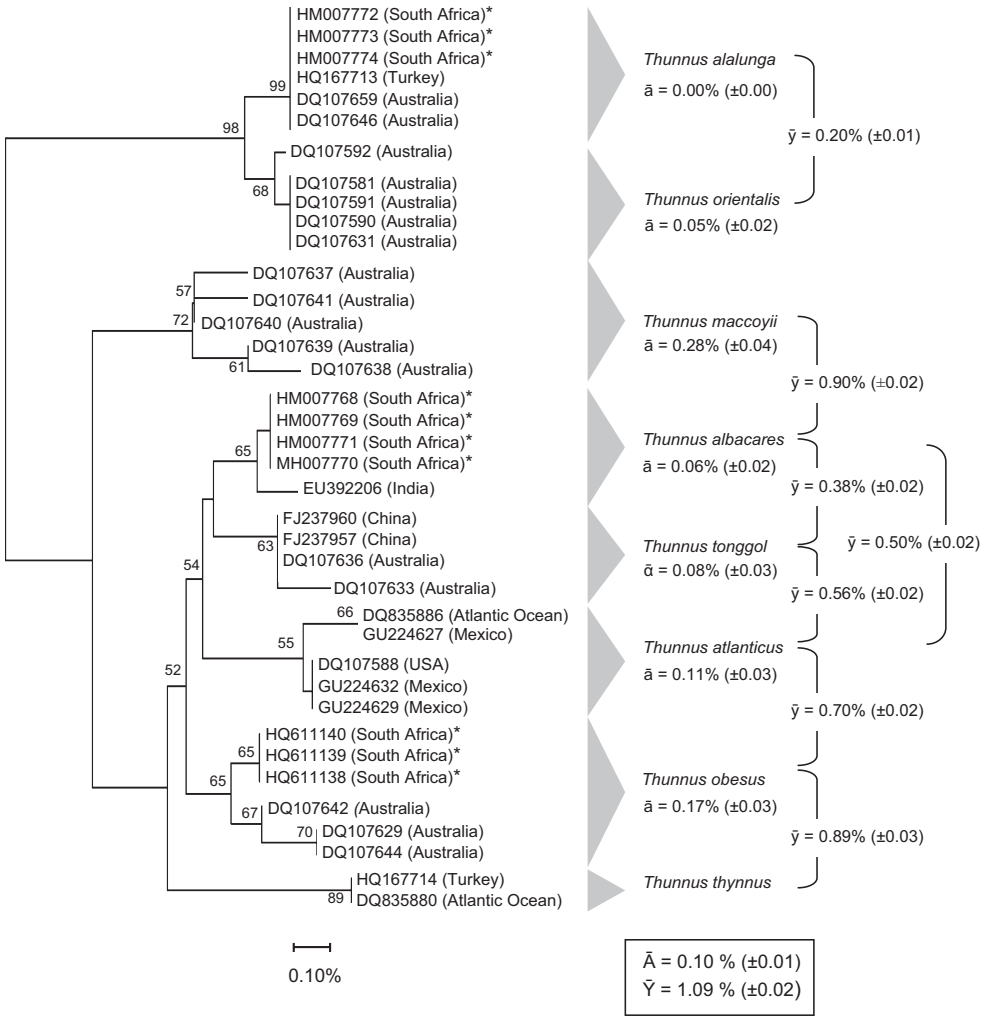


Fig. 3 Kimura 2-parameter distance neighbour-joining tree of 38 cytochrome *c* oxidase I sequences from the eight species of tuna within the genus *Thunnus*. GenBank accession nos and geographical origins of all specimens are provided, with asterisks indicating members of the three *Thunnus* species examined in this study. Bootstrap values higher than 50 are shown. Mean distances between members of each species ($\bar{a}$) are provided, as is the mean conspecific distance found overall for all eight species ($\bar{A}$). The mean congeneric distance between species and their nearest neighbours are indicated ($\bar{\bar{y}}$), as well as the overall mean congeneric distance found for all eight species ($\bar{\bar{Y}}$).

of all evaluated control region sequences of the *Thunnus* species are provided in Fig. 4. K2P distances and NJ trees for the COI and control region sequences of *Thunnus* species were computed in MEGA 4.

Results and discussion

Common names, species and family details, as well as GenBank accession nos for all fish specimens sequenced for the COI gene region are presented in Table 1. The DNA obtained from all 53 examined fish species was readily amplified using the barcoding primer cocktail described by Ivanova *et al*. (2007) (Table 2), generating a total of 194 sequences owing to the analysis of multiple specimens of each species. Mitochondrial control region amplicons were also successfully recovered from 10 specimens representing three species of the genus *Thunnus*. In this study, new DNA sequence data were generated for many commercially important fish species in South Africa that were previously unavailable in GenBank. For 23% of the fish species evaluated, this was the first time that DNA sequence data of any kind were submitted to GenBank. COI sequences were deposited in GenBank for the first time for 30% of the species (Table 1).

Cytochrome c oxidase I (barcoding) sequences

Read lengths of the 194 COI sequences were approximately 655 bp, with <2% of these being <600 bp in length. No indels, stop codons or sequences indicative of NUMTs (nuclear DNA sequences originating from mtDNA sequences) were encountered in any of the COI

© 2011 Blackwell Publishing Ltd

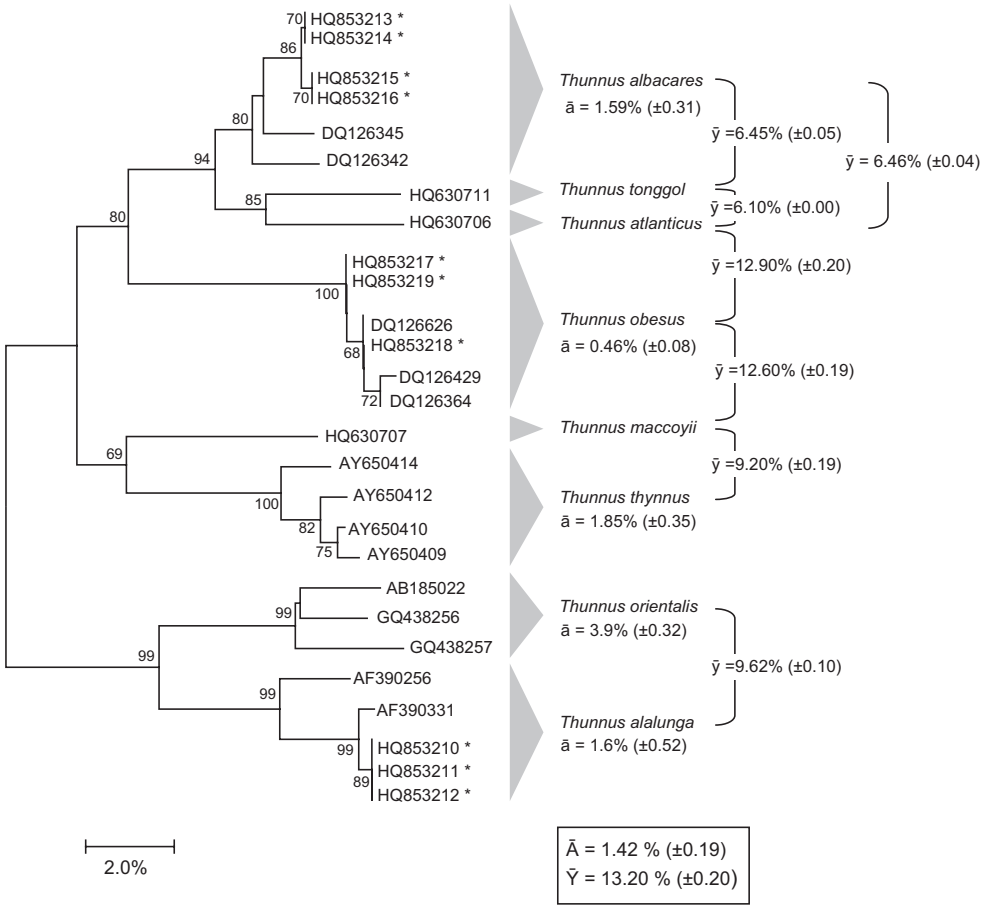


Fig. 4 Kimura 2-parameter distance neighbour-joining tree of 27 mtDNA control region sequences from the eight species of tuna within the genus *Thunnus*. GenBank accession nos of all specimens are provided, with asterisks indicating members of the three *Thunnus* species examined in this study. Bootstrap values higher than 50 are shown. Mean distances between members of each species ($\bar{a}$) are provided, as is the mean conspecific distance found overall for all eight species ($\bar{A}$). The mean congeneric distance between species and their nearest neighbours are indicated ($\bar{y}$), as well as the overall mean congeneric distance found for all eight species ($\bar{Y}$).

sequences, a result in agreement with previous fish barcoding reports (Steinke *et al.* 2009b; Lakra *et al.* 2011). The mean overall nucleotide base frequencies observed for the COI sequences from the 53 fish species analysed (Table 4) corresponded with those COI nucleotide frequencies reported by Ward *et al.* (2005) for 143 Australian teleost species. Overall, most nucleotide changes were seen to occur at the third codon base position, as illustrated by observation of the standard errors of the GC percentages of 0.40, 0.11 and 0.03 at the third, first and second base positions, respectively (Table 4). Such a finding is concurrent with the fact that in protein coding DNA regions, the degenerate nature of the genetic code results in the third codon base evolving the most rapidly, with most of these changes being synonymous and not leading to amino acid changes. It is reportedly this third base variability that affords COI barcoding with its power to discriminate closely related species (Ward & Holmes 2007).

Although the primary aim of DNA barcoding is to delineate species boundaries and to establish identification systems, phylogenetic relationships were apparent in the COI sequence data. Species within genera consistently clustered together, as did genera within families (Fig. 2). As anticipated, the mean K2P distances within species, genera and families increased with taxonomic rank (Table 5, Fig. 1). The mean distance between genera within families (D = 15.50%) was more than threefold greater than that seen within species of genera (4.58%). In turn, the mean distance between species within genera was *c.* 45-fold greater than the mean distance between individuals within species (D = 0.10%), demonstrating the general discriminative ability of COI barcoding.

The average within-species distance observed for the COI region in this study (D = 0.10%) was somewhat lower than the corresponding values of 0.25% and 0.30% reported in other barcoding studies on marine fish (Steinke *et al.* 2009a; Lakra *et al.* 2011). Previous studies

© 2011 Blackwell Publishing Ltd

on several hundred marine fish species (Ward *et al.* 2005, 2008; Rock *et al.* 2008; Steinke *et al.* 2009a,b) revealed that more than 95% of intra-species comparisons showed an average COI divergence of <2%. In this study, all (100%) of the intra-species values showed <2% divergence, with 97% of these exhibiting divergence values below 1% (98% of species within the study) (Fig. 1). Zero intra-species sequence divergence was found for 45 of 53 (85%) fish species examined, and these 45 species all clustered with 100% bootstrap support in the NJ tree (Fig. 2). Nonetheless, as sample sizes were generally small, further sampling could uncover some intra-specific variability for these species. The maximum intra-species COI variation was observed between specimens of *Thyrsites atun* (snoek), which was one of few species for which both locally caught and imported specimens were analysed. Clear geographic differentiation was apparent for *T. atun*, with one clade represented by individuals from South Africa (SA) and another from New Zealand (NZ), and no shared haplotypes between the SA and NZ specimens. The extent of SA-NZ differentiation, an average K2P distance of 1.54% (Table 5), far exceeded that found within the SA (0.00%) or NZ (0.30%) groups. The three SA specimens of *T. atun* separated at a 100% bootstrap value from the three NZ specimens (Fig. 2). Hebert *et al.* (2004) have proposed that genetically divergent specimens may be flagged as provisional species when they show a 10-fold greater mean intra-specific variation than the group under study. It is therefore possible that the locally caught *T. atun* and the *T. atun* imported from NZ could potentially refer to two distinct biological species. However, such a distinction would require further clarification, employing extensive morphological and meristic investigations, as well as additional genetic analyses.

For species within genera, the mean COI distance of 4.58% in this study was comparable with that of 3.75% recorded among congeneric fish species from Canadian Pacific waters (Steinke *et al.* 2009a) and that of 6.67% for fish species of the Scotia sea (Rock *et al.* 2008), but was considerably lower than the 9.93% recorded for Australian marine fish (Ward *et al.* 2005). For 50 of 53 (94%) fish species in the data set, inter-species distance values exceeding 2% were calculated, suggesting that these species could be readily discriminated by their COI barcodes. The remaining three species analysed, all members of the genus *Thunnus*, exhibited congeneric distances below 2%. Examination of the K2P/NJ tree (Fig. 2) revealed that, although each of these species clustered into a separate grouping with no specimens being misplaced, the genetic distances between the three *Thunnus* species were small. The mean congeneric distance between specimens of *Thunnus alalunga* and *T. albacares* was 1.51%, and between *T. alalunga* and *T. obseus* this value was 1.53%. Overall, the minimum inter-species distance in the study (0.46%) was found to be between individuals of *T. albacares* and *T. obesus*. Therefore, while DNA barcoding was shown to be a highly effective tool for the identification of the vast majority of the fish species analysed, these results appear to indicate that the unambiguous identification of *Thunnus* species using COI sequencing may not be possible.

Thunnus species identification

Members of the genus *Thunnus*, commonly referred to as tunas, are large, migratory fish belonging to the family Scombridae. Although the genus is believed to be monophyletic, the phylogenetic relationships between the individual species within the genus remains controversial and classifications based on morphology, mtDNA and nDNA have not been entirely concordant (Chow & Kishino 1995; Elliot & Ward 1995; Alvarado Bremer *et al.* 1997; Lowenstein *et al.* 2009). Of the eight *Thunnus* species recognized by Collette *et al.* (2001), five of these are found in South African waters: *T. obesus*, *T. albacares*, *T. alalunga*, *T. maccoyii* and *T. thynnus* (Froese & Pauly 2010). Certain members of this genus are extremely commercially significant, which makes them prone to both overexploitation (Safina 2001; Allen *et al.* 2010) and fraudulent trading practices (FSA 2000; Jacquet & Pauly 2008; Lowenstein *et al.* 2009). As tunas are often presented on the market in their processed forms, the development of reliable genetic methodologies is critically important to ensure accurate species identifications. Nonetheless, Viñas & Tudela (2009) reported that a number of *Thunnus* species are genetically very similar and that confounded identifications may result if DNA markers with low genetic variation are employed.

DNA barcoding has been proposed as a single universal tool for animal species identifications (Dasmahapatra & Mallet 2006) and a recent study examining the molecular evolutionary behaviour of the COI region in fishes implied that barcoding should be applicable for the identification of all marine species (Ward & Holmes 2007). However, it has increasingly been recognized that recently speciated taxa could pose difficulties for the identification by DNA barcoding (Elias *et al.* 2007; Lowenstein *et al.* 2009). As the genus *Thunnus* is recently derived, believed to have evolved during periods of rapid speciation, the analysis of faster-evolving genes rather than slowly evolving ones is preferable for identification of these species, as the latter generally do not show sufficient mutations to allow clear distinction between the different members of the genus (Collette *et al.* 2001). While Lowenstein *et al.* (2009) showed that a character-based analysis of COI sequences allowed the discrimination of members of the *Thunnus* genus, reports by the same authors, as well as by Wong & Hanner

© 2011 Blackwell Publishing Ltd

(2008), have suggested that the genetic-distance methods employed by BOLD and the BLASTn algorithm in GenBank may not permit the clear assignment of *Thunnus* specimens to the species level. Indeed, the preliminary results from the barcoding of three *Thunnus* species in this study (*T. alalunga*, *T. albacares* and *T. obesus*) showed that the unambiguous identification of members of this genus may be problematic based on the COI region. Comparison of the generated COI sequences for these *Thunnus* specimens with those sequences already available in GenBank and in BOLD provided additional support for this finding, where in all cases the sequences of putative species showed equal percentage similarities with those of one or more congeneric species. In order to further investigate the utility of COI barcoding for the unambiguous identification of *Thunnus* species, the generated COI sequences for these three *Thunnus* species were consequently analysed together with representative sequences from GenBank for all eight members of the genus. Figure 3 shows the resulting COI phenogram for 38 specimens, along with the GenBank accession no and geographical origin of each specimen.

Genetic differences within species of *Thunnus* were small at the COI level, with a mean K2P distance of 0.11% (Fig. 3). While members of the same *Thunnus* species clustered under the same nodes, the bootstrap percentage values for most of the species separations were relatively low (60–70% level). The mean COI distance between *Thunnus* species was 1.09%, a value more than four times smaller than that found for all 53 fish species analysed in this study (D = 4.58%). *Thunnus alalunga* and *T. orientalis*, the only clade for which high bootstrap support (98%) was obtained in Fig. 3, showed the minimum inter-species distance (D = 0.20%) within the *Thunnus* genus. Only a single nucleotide variation was observed in the COI sequences of the two aforementioned species. Interestingly, the sequence of one member of *T. orientalis* (GenBank accession no DQ107592) also differed by a single nucleotide from the sequences of the other members of the species, indicating the problems inherent with making identifications based on single base polymorphisms. Observation of the COI K2P/NJ tree (Fig. 3) revealed that the three species comprising the *Neothunnus* subgenus (*T. albacares*, *T. atlanticus* and *T. tonggol*) (Collette 1978) assembled into a loosely defined clade with 54% bootstrap support. This finding is in contrast to that of Viñas & Tudela (2009), who reported that members of *Neothunnus* did not group together based on their COI sequences. The mean congeneric distance found between members of *Neothunnus* (D = 0.47%) appears to indicate that these species show very close COI sequence congruence. In particular, the COI sequences of *T. tonggol* differed by only two nucleotides from those of *T. albacares* and by only three nucleotides from those of *T. atlanticus*.

Fundamental to the success of DNA barcoding for species-level identifications is the requirement for members within a species to show substantially less COI variation than that variation seen amongst different species (Ward *et al*. 2005). This, however, was not seen to be the case for all species of *Thunnus* examined. In fact, the maximum intra-species COI variation observed among specimens of *T. maccoyii* (0.46%) was more than threefold higher than the minimum inter-species variation observed between certain specimens of *T. alalunga* and *T. orientalis* (0.15%). The maximum intra-species variation (0.46%) was also higher than the mean inter-species distance found between *T. albacares and T. tonggol* (0.38%). Overall, this study has highlighted that *Thunnus* species are likely to pose a challenge for the barcoding system due to the low COI sequence divergence amongst certain congeners and the use of the COI region as a single genetic marker for *Thunnus* species identification has been brought into question.

It has been suggested that a more variable gene region, such as the mitochondrial control region, could hold considerable promise for confirming the identity of the closely related members belonging to the *Thunnus* genus (Viñas & Tudela 2009). Thus, sequences of the 5′ region of the control region were generated for the specimens of *T. alalunga*, *T. albacares* and *T. obesus* examined in this study (accession nos HQ853210–HQ853212, HQ853213–HQ853216 and HQ853217–HQ853219, respectively). The average read length of the control region sequences was 398 bp. The mean K2P distance found between members of the aforementioned three *Thunnus* species was 12.22%, while the mean within-species distance was 0.18%. As with the COI region, the discriminatory power of the control region for *Thunnus* species identification was further validated by comparing the generated sequences with additional sequences from GenBank. The resulting phylogenetic reconstruction based on 27 control region sequences representing all eight *Thunnus* species is presented in Fig. 4.

While it is accepted that the samples sizes of sequences for certain species were small due to their limited availability in GenBank, the control region phenogram (Fig. 4) appears to be considerably more consistent than the corresponding one based on the COI region (Fig. 3), with species separating into distinct clusters with high bootstrap support (80–100%). The mean K2P distance between the eight *Thunnus* species (D = 13.20%) was *c.* ninefold higher than the within-species value (D = 1.42%) based on the control region sequences. Both of the aforementioned values were also more than 10-fold greater than the corresponding values found for the eight species based on the COI region, emphasizing the greater nucleotide diversity exhibited in the control region of *Thunnus* species. *Thunnus alalunga* and *T orientalis*, the

© 2011 Blackwell Publishing Ltd

two species found to differ by only one nucleotide in their COI sequences (D = 0.20%), were easily identifiable by their control region sequences, which showed a mean distance of 9.62%. The three members of the subgenus *Neothunnus* grouped together in the control region phenogram, showing a higher level of bootstrap support (86%) than that seen in the COI tree for these species (54%). In contrast to the findings based on the COI region, the maximum conspecific distance found for all eight species (4.4%) was lower than the minimum congeneric distance (6.1%) seen between species.

The discriminatory power of the control region has been shown to be superior to the COI region for the species of *Thunnus* examined in this study. Nonetheless, introgression between several members of this genus is known to occur (Chow & Kishino 1995; Alvarado Bremer *et al.* 1997, 2005; Viñas & Tudela 2009) and Chow & Kishino (1995) proposed that mtDNA sequence data alone would not be sufficient for differentiating *Thunnus* species if horizontal transfer of mtDNA exists. In this context, Viñas & Tudela (2009) showed that mtDNA control region sequencing could not discriminate between *T. alalunga* and introgressed members of *T. thynnus* showing albacore-like sequences. These authors recommended that the sequencing of a nuclear fragment, the ribosomal DNA first internal transcribed spacer (ITS1), was more suitable for differentiating the aforementioned *Thunnus* specimens. Therefore, it appears that while sequencing of the mtDNA control region can serve as an extremely useful tool for *Thunnus* species identification, this will probably require supplementary sequencing of a nDNA locus (such as the ITS1) if potentially introgressed members of this genus are to be revealed.

Conclusions

To our knowledge, this is the first study that has sought to establish a comprehensive DNA sequence database, including COI barcoding records, for those fish species commonly encountered on the South African market. In particular, the utility of DNA barcoding has been demonstrated and the results suggest that the COI region could serve as a single genetic marker to discriminate the vast majority of the fish species included in this study, with only members of the genus *Thunnus* likely requiring further confirmation through the use of mitochondrial control region sequencing and possibly also nDNA analysis. The patterning of divergence in the mtDNA regions examined has revealed close correspondence with those species assigned through prior morphological analyses, emphasizing the value of integrating DNA-based methodologies with classical identification approaches to validate existing taxonomic systems. In addition, the uncovering of a potentially cryptic species of *T. atun* in this work has shown the potential of DNA barcoding in revealing overlooked diversity and for highlighting taxa which require additional investigation and resolution.

Overall, the results and the DNA sequences generated from this research have provided the necessary genetic information to allow the explicit identification of 53 commercially important fish species in South Africa. Aside from permitting identification of whole fish specimens, these developed systems should also be extendable to the identification of fish at any developmental stage, from eggs to adulthood, and for body fragments. A realm of applications has consequently been opened relating to conservation, ecology research, control of commercial practices and detection of retail fraud, from which the entire fisheries supply chain, regulators and inevitably consumers will undoubtedly benefit. Such work should prove pertinent within an environment where both natural and human-induced activities are rapidly accelerating modifications in the abundance and distribution of fish species.

References

Allen R, Joseph J, Squires D (2010) *Conservation and Management of Transnational Tuna Fisheries*, pp. 3–87. John Wiley & Sons Inc., New Jersey, USA.

Alvarado Bremer JR (1994) *Assessment of morphological and genetic variation of the swordfish (*Xiphias gladius *Linnaeus): evolutionary patterns of nucleotide substitution in the mitochondrial genome*. PhD Thesis, University of Toronto, Canada (As cited by Viñas & Tudela, 2009).

Alvarado Bremer JR, Naseri I, Ely B (1997) Orthodox and unorthodox phylogenetic relationships among tunas revealed by the nucleotide sequence analysis of the mitochondrial control region. *Journal of Fish Biology*, **50**, 540–554.

Alvarado Bremer JR, Viñas J, Mejuto J, Ely B, Pla C (2005) Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. *Molecular Phylogenetics and Evolution*, **36**, 169–187.

Blaxter ML (2004) The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **359**, 669–679.

Cawthorn DM, Steinman HA, Witthuhn RC (2010) Comparative study of different methods for the extraction of DNA from fish species commercially available in South Africa. *Food Control*, **22**, 231–244.

Cawthorn DM, Steinman HA, Witthuhn RC (2011) Evaluating the availability of fish species on the South African market and the factors undermining sustainability and consumer choice. *Food Control*, **22**, 1748–1759.

Chow S, Kishino H (1995) Phylogenetic relationships between tuna species of the genus *Thunnus* (Scombridae: Teleostei): inconsistent implications from morphology, nuclear and mitochondrial genomes. *Journal of Molecular Evolution*, **41**, 741–748.

Collette BB (1978) Adaptations and systematics of the mackerels and tunas. In: *The Physiological Ecology of Tunas* (eds Sharp GD, Dizon AE), pp. 7–39. Academic Press, London, UK.

Collette BB, Reeb C, Block BA (2001) Systematics of the tunas and mackerels (Scombridae). In: *Tuna: Physiology, Ecology and Evolution* (eds Block BA, Stevens ED), pp. 5–30. Academic Press, London, UK.

Dasmahapatra KK, Mallet J (2006) DNA barcodes: recent successes and future prospects. *Heredity*, **97**, 254–255.

Dawnay N, Ogden R, McEwing R, Carvalho GR, Thorpe RS (2007) Validation of the barcoding gene COI for use in forensic genetic species identification. *Forensic Science International*, **173**, 1–6.

© 2011 Blackwell Publishing Ltd

Di Finizio A, Guerriero G, Russo GL, Ciarcia G (2007) Identification of gadoid species (Pisces, Gadidae) by sequencing and PCR–RFLP analysis of mitochondrial 12S and 16S rRNA gene fragments. *European Food Research and Technology*, **225**, 337–344.

Elias M, Hill RI, Willmott KR *et al.* (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 2881–2889.

Elliot NG, Ward RD (1995) Genetic relationships of eight species of Pacific tunas (Teleostei: Scombridae) inferred from allozyme analysis. *Marine and Freshwater Research*, **46**, 1021–1032.

FAO (Food and Agriculture Organization of the United Nations) (2009) *FAO Yearbook of Fishery and Aquaculture Statistics 2007*, pp. 1–72. FAO, Rome, Italy.

FAO (Food and Agriculture Organization of the United Nations) (2010) Fishery and aquaculture country profiles: South Africa national fishery sector overview. Available at: http://www.fao.org/fishery/country-sector/FI-CP_ZA/en. Accessed 12 July 2010.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.

Froese R, Pauly D (2010) FishBase. World Wide Web electronic publication, version 01/2010. Available at: http://www.fishbase.org. Accessed 21 September 2010.

FSA (Food Standards Agency) (2000) Survey of misdescription of tuna species in tuna products. Available at: http://www.food.gov.uk/science/surveillance/fsis2000/01tuna. Accessed 10 December 2010.

Gil LA (2007) PCR-based methods for fish and fishery products authentication. *Trends in Food Science and Technology*, **18**, 558–566.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, **270**, S96–S99.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. *PLoS Biology*, **2**, 1657–1663.

Hubalkova Z, Kralik P, Tremlova B, Rencova E (2007) Methods of gadoid fish species identification in food and their economic impact in the Czech Republic: a review. *Veterinarni Medicina*, **52**, 273–292.

Ivanova NV, Zemlak TS, Hanner R, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, **7**, 544–548.

Jacquet JL, Pauly D (2008) Trade secrets: renaming and mislabelling of seafood. *Marine Policy*, **32**, 309–318.

Kimura M (1980) A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Lakra WS, Verma MS, Goswani M *et al.* (2011) DNA barcoding Indian marine fishes. *Molecular Ecology Resources*, **11**, 60–71.

Larkin MA, Blackshields G, Brown NP *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

Lockley AK, Bardsley RG (2000) DNA-based methods for food authentication. *Trends in Food Science and Technology*, **11**, 67–77.

Lowenstein JH, Amato G, Kolokotronis SO (2009) The real *maccoyii*: identifying tuna sushi with DNA Barcodes – contrasting characteristic attributes and genetic distances. *PLoS ONE*, **4**, e7866.

Martinez I, James D, Loréal H (2005) *Application of modern analytical techniques to ensure seafood safety and authenticity. FAO Fisheries Technical Paper No. 455*, pp. 1–73. FAO, Rome, Italy.

Messing J (1983) New M13 vectors for cloning. *Methods in Enzymology*, **101**, 20–78.

Rasmussen RS, Morrisey MT (2009) Application of DNA-based methods to identify fish and seafood substitution on the commercial market. *Comprehensive Reviews in Food Science and Food Safety*, **8**, 118–154.

Rock J, Costa FO, Walker DI, North AW, Hutchinson WF, Carvalho GR (2008) DNA barcodes of fish of the Scotia Sea, Antarctica indicate priority groups for taxonomic and systematics focus. *Antarctic Science*, **20**, 253–262.

Rokas A, Ladoukakis E, Zouros E (2003) Animal mitochondrial DNA recombination revisited. *Trends in Ecology and Evolution*, **18**, 411–417.

Safina C (2001) Tuna conservation. In: *Tuna: Physiology, Ecology and Evolution* (eds Block BA, Stevens ED), pp. 414–461. Academic Press, London, UK.

Smith JLB, Smith MM (1966) *Fishes of the Tsitsikamma Coastal National Park*, pp. 102. National Parks Board Trust, Pretoria, South Africa.

Steinke D, Zemlak TS, Hebert PDN, Adamowicz SA (2009a) DNA barcoding of Pacific Canada's fishes. *Marine Biology*, **156**, 2641–2647.

Steinke D, Zemlak TS, Hebert PDN (2009b) Barcoding Nemo: DNA-based identifications for the ornamental fish trade. *PLoS ONE*, **4**, e6300.

Stoeckle M (2003) Taxonomy, DNA, and the bar code of life. *BioScience*, **53**, 796–797.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA 4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Teletchea F (2009) Molecular identification methods of fish species: reassessment and possible applications. *Reviews in Fish Biology and Fisheries*, **19**, 265–293.

Viñas J, Tudela S (2009) A validated methodology for genetic identification of tuna species (genus *Thunnus*). *PLoS ONE*, **4**, e7606.

Von der Heyden S, Barendse J, Seebregts AJ, Matthee CA (2010) Misleading the masses: detection of mislabelled and substituted frozen fish products in South Africa. *ICES Journal of Marine Science*, **67**, 176–185.

Ward RD, Holmes B (2007) An analysis of nucleotide and amino acid variability in the barcode region of cytochrome *C* oxidase I (*cox1*) in fishes. *Molecular Ecology Notes*, **7**, 899–907.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1847–1857.

Ward RD, Holmes BH, White WT, Last PR (2008) DNA barcoding Australasian chondrichthyans: results and potential uses in conservation. *Marine and Freshwater Research*, **59**, 57–71.

Ward RD, Hanner R, Herbert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, **74**, 329–356.

Wong EHK, Hanner RH (2008) DNA barcoding detects market substitution in North American seafood. *Food Research International*, **41**, 828–837.

Data Accessibility

DNA sequences: GenBank accessions HM007674–HM007801; HQ611082–HQ611140; HQ641664–HQ641670; HQ853210–HQ853219. BOLD accessions BARFS001-09; BARFS002-10–BARFS194-10.

Final DNA sequence assembly uploaded as online supplemental material.

Phylogenetic data: TreeBASE study accession no. S11583. Study accession URL: http://purl.org/phylo/treebase/phylows/study/TB2:S1158.

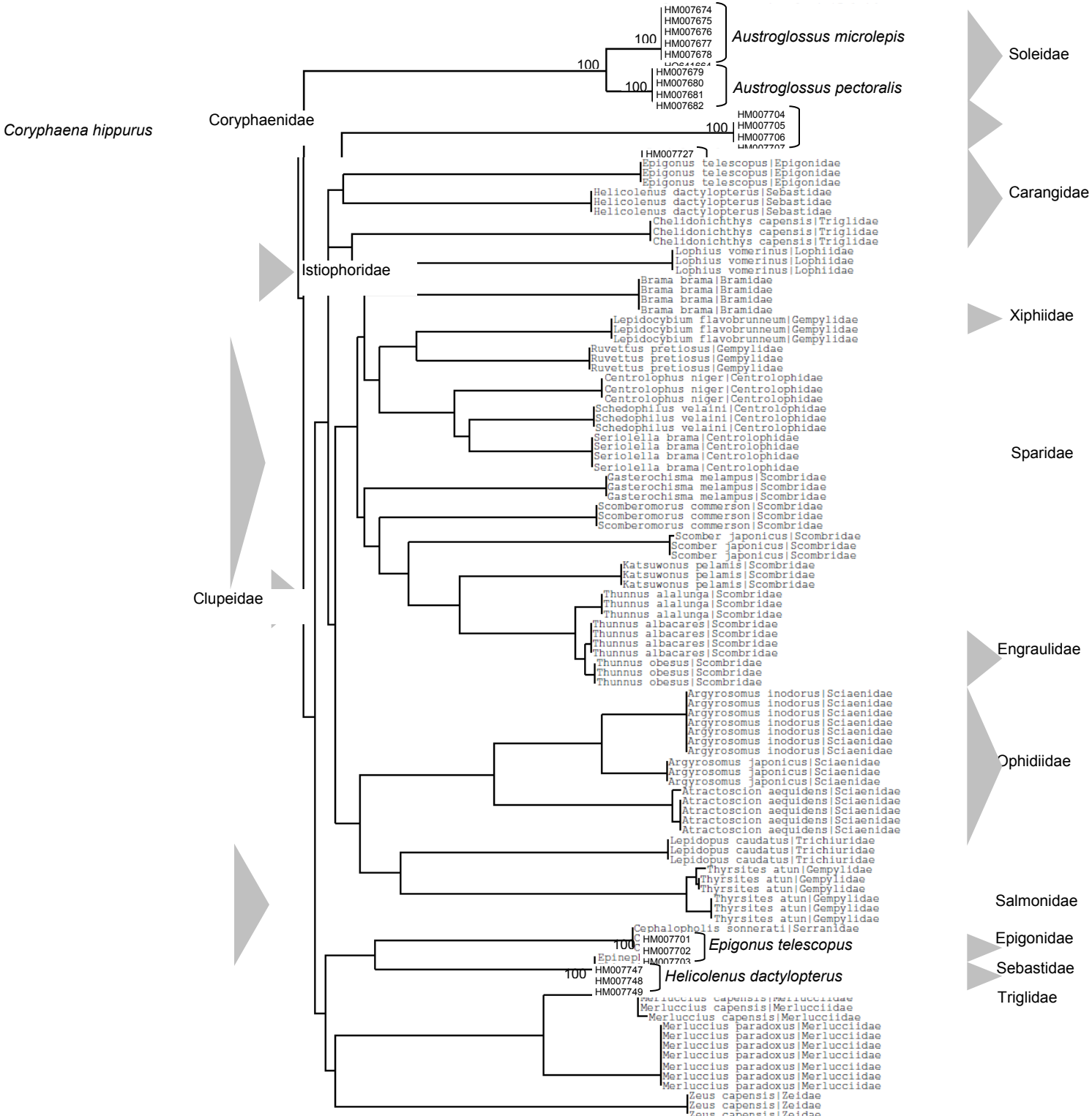
Supporting Information

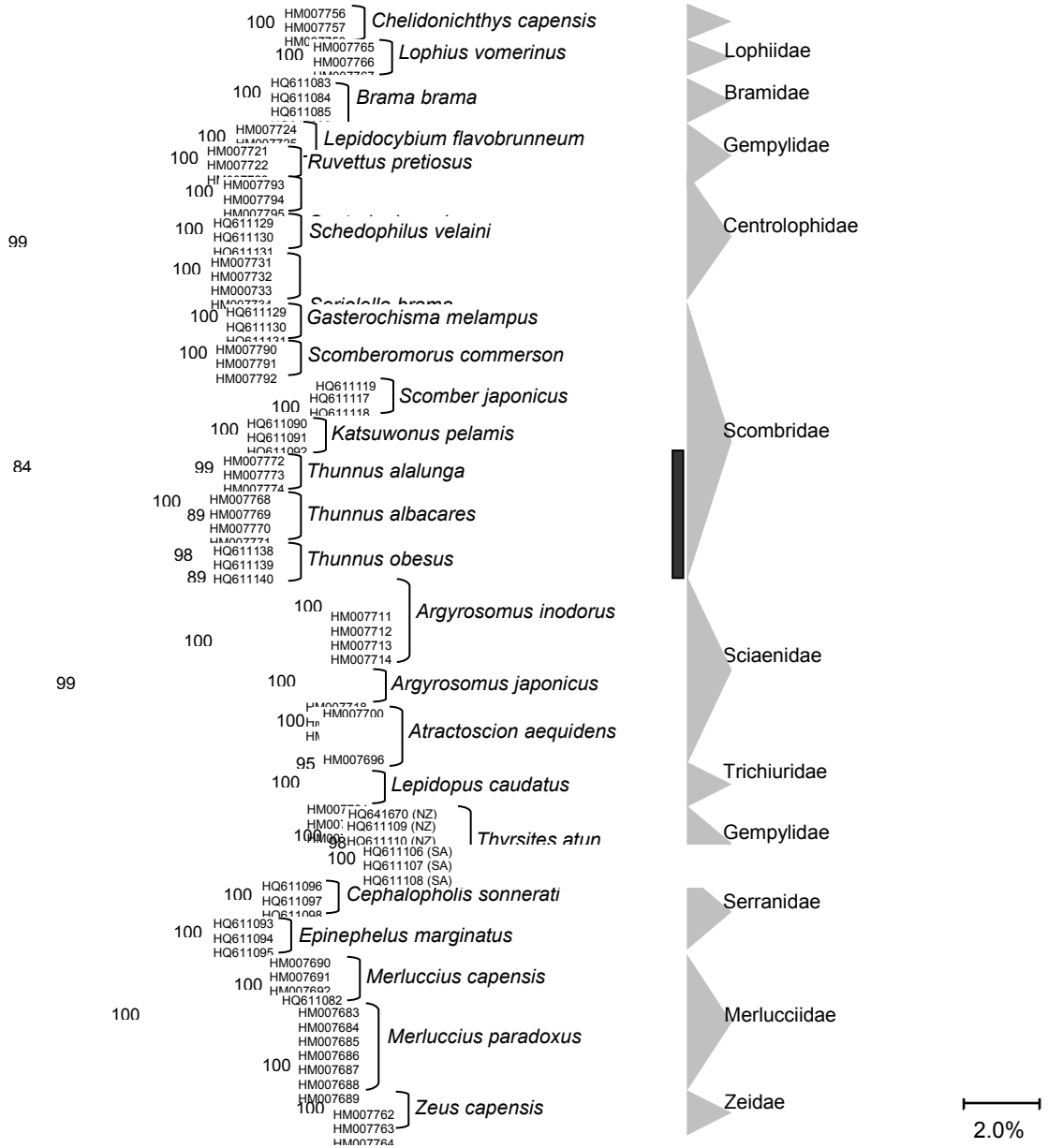
Additional supporting information may be found in the online version of this article.

Fig. S1 K2P distance neighbour-joining tree of 194 COI sequences from 53 fish species, with GenBank accession nos for each specimen.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2011 Blackwell Publishing Ltd


Coryphaena hippurus
Coryphaenidae
Istiophoridae
Clupeidae
HM007674
HM007675
HM007676
HM007677
HM007678
Austroglossus microlepis
HM007679
HM007680
HM007681
HM007682
Austroglossus pectoralis
HM007704
HM007705
HM007706
HM007727
Epigonus telescopus|Epigonidae
Helicolenus dactylopterus|Sebastidae
Chelidonichthys capensis|Triglidae
Lophius vomerinus|Lophiidae
Brama brama|Bramidae
Lepidocybium flavobrunneum|Gempylidae
Ruvettus pretiosus|Gempylidae
Centrolophus niger|Centrolophidae
Schedophilus velaini|Centrolophidae
Seriolella brama|Centrolophidae
Gasterochisma melampus|Scombridae
Scomberomorus commerson|Scombridae
Scomber japonicus|Scombridae
Katsuwonus pelamis|Scombridae
Thunnus alalunga|Scombridae
Thunnus albacares|Scombridae
Thunnus obesus|Scombridae
Argyrosomus inodorus|Sciaenidae
Argyrosomus japonicus|Sciaenidae
Atractoscion aequidens|Sciaenidae
Lepidopus caudatus|Trichiuridae
Thyrsites atun|Gempylidae
Cephalopholis sonnerati|Serranidae
HM007701
HM007702
Epigonus telescopus
HM007747
HM007748
HM007749
Helicolenus dactylopterus
Merluccius capensis|Merlucciidae
Merluccius paradoxus|Merlucciidae
Zeus capensis|Zeidae
100
Soleidae
Carangidae
Xiphiidae
Sparidae
Engraulidae
Ophidiidae
Salmonidae
Epigonidae
Sebastidae
Triglidae



Supplement 1 K2P distance neighbour-joining tree of 194 COI sequences from 53 fish species, with GenBank accession numbers for each specimen. Numbers at nodes indicate bootstrap values (values higher than 80 are given). Species showing less than 2% COI divergence are indicated by a black bar.