

# Linking adults and immatures of South African marine fishes<sup>1</sup>

Dirk Steinke, Allan D. Connell, and Paul D.N. Hebert

**Abstract:** The early life-history stages of fishes are poorly known, impeding acquisition of the identifications needed to monitor larval recruitment and year-class strength. A comprehensive database of COI sequences, linked to authoritatively identified voucher specimens, promises to change this situation, representing a significant advance for fisheries science. Barcode records were obtained from 2526 early larvae and pelagic eggs of fishes collected on the inshore shelf within 5 km of the KwaZulu-Natal coast, about 50 km south of Durban, South Africa. Barcodes were also obtained from 3215 adults, representing 946 South African fish species. Using the COI reference library on BOLD based on adults, 89% of the immature fishes could be identified to a species level; they represented 450 species. Most of the uncertain sequences could be assigned to a genus, family, or order; only 92 specimens (4%) were unassigned. Accumulation curves based on inference of phylogenetic diversity indicate near-completeness of the collecting effort. The entire set of adult and larval fishes included 1006 species, representing 43% of all fish species known from South African waters. However, this total included 189 species not previously recorded from this region. The fact that almost 90% of the immatures gained a species identification demonstrates the power and completeness of the DNA barcode reference library for fishes generated during the 10 years of FishBOL.

*Key words:* DNA barcoding, fish, larval fish, fish eggs.

**Résumé :** Les premiers stades de développement des poissons sont peu connus, ce qui limite l'acquisition des identifications requises pour suivre le recrutement larvaire et l'abondance d'une cohorte annuelle. Une base de données extensives des séquences COI, associée à des spécimens de référence identifiés par des autorités en la matière, promet de changer cette situation et représente une avancée significative pour les sciences aquatiques. Des codes à barres ont été obtenus pour 2526 stades larvaires précoces ou d'œufs pélagiques d'espèces de poissons échantillonnés sur le plateau continental à l'intérieur d'une zone de 5 km de la côte du KwaZulu-Natal, à environ 50 km au sud de Durban, en Afrique du Sud. Des codes à barres ont également été obtenus pour 3215 adultes représentant 946 espèces de poissons d'Afrique du Sud. À l'aide d'une collection de référence COI sur BOLD, basée sur des poissons adultes, 89 % des poissons immatures ont été identifiés au niveau de l'espèce; ils représentaient au-delà de 450 espèces. La plupart des séquences d'origine incertaine ont été assignées à un genre, une famille ou un ordre et seuls 92 spécimens (4 %) sont demeurés non-assignés. Les courbes d'accumulation basées sur une inférence de la diversité phylogénétique suggèrent que l'effort d'échantillonnage s'est avéré presque complet. La collection entière de poissons adultes et immatures comprenait 1006 espèces, ce que représente 43 % de toutes les espèces connues dans les eaux de l'Afrique du Sud. Cependant, ce total comprend 189 espèces qui n'avaient pas encore été recensées dans cette région. Le fait que près de 90 % des échantillons immatures aient été assignés à une espèce démontre la puissance et l'intégralité de la collection de référence de codes à barres de l'ADN pour les poissons qui a été produite au cours des 10 années du projet FishBOL. [Traduit par la Rédaction]

*Mots-clés :* codage à barre, poisson, larves de poisson, œufs de poisson.

Received 10 December 2015. Accepted 4 March 2016.

Corresponding Editor: Martha Valdez-Moreno.

**D. Steinke and P.D.N. Hebert.** Canadian Centre for DNA Barcoding, Biodiversity Institute of Ontario, University of Guelph, Guelph, ON N1G 2W1, Canada.

**A.D. Connell.\*** South African Institute of Aquatic Biodiversity, Private Bag 1015, Grahamstown, South Africa 6140.

**Corresponding author:** Dirk Steinke (email: [dsteinke@uoguelph.ca](mailto:dsteinke@uoguelph.ca)).

\*Allan D. Connell died on 18 March 2016 in a dive accident. Without his enthusiasm and tireless efforts this study would have not been possible. This is a profound loss not only to his family and friends but also to the world of marine biology. Allan left this world doing what he loved most. This study is his legacy, a testimony for 10 years of diligent work.

<sup>1</sup>This paper is part of a special issue entitled Barcodes to Biomes.

Copyright remains with the author(s) or their institution(s). Permission for reuse (free in most cases) can be obtained from [RightsLink](http://RightsLink).

## Introduction

Although it is widely accepted (Olson et al. 1991; Ward 2000) that fisheries research would greatly benefit from a system enabling accurate, rapid species identifications for eggs and larvae (hereafter immatures), it is estimated that only 8%–10% of fish species can currently be identified at all life stages (Fahay 2007; Froese and Pauly 2015). This fact reflects the past reliance on morphological characters, meristic counts, and morphometric analyses to identify immature fishes. Because of the dramatic morphological reconfigurations during development, the diagnosis of immatures requires the synthesis of keys based on entirely different characters (e.g., melanophores, myomeres, head spines) than those used for the identification of adults. The limited number of characters, the multiple developmental stages, and the small size of immatures complicate taxonomic assignments (Busby et al. 2000; Baldwin and Smith 2003, Richards 2006), making it unlikely that any comprehensive morphological identification system will be developed (Leis and Hay 2004). Molecular methods have gained increasing recognition as an alternative tool for species identification (Garland and Zimmer 2002), and several studies (e.g., Hubert et al. 2010; Ko et al. 2013; Valdez-Moreno et al. 2010) have confirmed that one of these approaches, DNA barcoding, is a powerful tool for identifying larval fishes.

DNA barcoding employs sequence diversity in a standardized gene region as a basis for species identification. The effectiveness of a 650-bp segment of the cytochrome *c* oxidase I (COI) gene for species identification in fishes has now been heavily validated (e.g., Hubert et al. 2008; Steinke et al. 2009; Ward et al. 2005, 2008). A comprehensive database of COI sequences, linked to authoritatively identified voucher specimens for all fishes, would represent a major advance for fisheries science (Becker et al. 2011; Costa and Carvalho 2011; Ward et al. 2009). Aside from providing identifications for adults, this reference library would allow the identification of all life stages (Côté et al. 2013; Smith et al. 2005).

This study employs DNA barcoding to identify immature fishes collected during a 28 year survey of the inshore shelf along a short section of the KwaZulu-Natal coastline, about 50 km south of Durban, South Africa (Connell 2010). The study was initially motivated by concerns relating to the effects of effluent discharge into the marine environment on the early life stages of marine fishes, especially those with pelagic eggs. To assess the importance of the area as a spawning ground for coastal marine fishes, plankton tows were conducted on a regular basis for more than a decade to assess the diversity of fish species spawning in the area, and the intensity and seasonality of their spawning. For 16 years, the identification of immatures relied solely on morphological identifications. However, many specimens collected after 2004 were also DNA barcoded to increase the fraction of

the immatures that could be connected to adults. To strengthen the prospects for identification, an effort was also made to develop a comprehensive reference library of DNA barcodes based upon expertly identified adult fishes. The overall program represents the largest effort to develop an identification system for a regional fish fauna based on DNA barcoding.

## Materials and methods

### Sampling

This study examined 3215 adults, representing 946 fish species from South African waters. Whenever possible, at least two adults were analyzed per species. All specimens are new collections with vouchers deposited in the South African Institute of Aquatic Biodiversity (Grahamstown, South Africa), except for the largest species. A detailed overview of the samples is provided in the public dataset, *Adult fishes of South Africa* [DS-SAAF] (<http://dx.doi.org/10.5883/DS-SAAF>), on the Barcode of Life Datasystems (BOLD, Ratnasingham and Hebert 2007). It contains all specimen and collection data, sequences, specimen images, and trace files. All sequences in this study have also been deposited in GenBank. Accession numbers are available through the BOLD dataset. In addition, 2526 immatures were analyzed from wild-collected eggs and larvae, especially in the flexion or post-flexion stage. Detailed specimen data are provided in the public dataset *Fish larvae of Park Rynie, South Africa* [DS-SALF] (<http://dx.doi.org/10.5883/DS-SALF>). These samples were collected September 2003 – February 2015 at both inshore (0.5 km) and offshore (5 km) sites near Park Rynie, KwaZulu-Natal, by towing the upper 2 m of the water column at two knots for 10 min using a 300  $\mu\text{m}$  mesh plankton net. During the course of the study, it became apparent that certain important species were excluded, as they were attracted to estuarine systems. Thus, a series of samples was collected from the entrance to Durban harbour. Some of these specimens were rarely or never encountered in the coastal samples, revealing the importance of Durban Harbour to estuarine fishes in the region, and yielding data on species that otherwise would not have been documented. On the KwaZulu-Natal coast, the daytime spring high tide occurs at about 16:00 local time, every 2 weeks. These spring tide periods are peak spawning times for estuarine fishes. South African Navy tide- and sunset-tables were used to sample the harbor entrance at dusk on the day on which the afternoon high tide coincided with sunset. A single sample was collected every 2 weeks.

The contents of each tow were placed in a separate 15 L container of seawater, for transport to the laboratory, where they were concentrated into a glass bowl, and sorted, visually, with the aid of a low-powered binocular microscope. Single fish eggs were isolated into containers of clean seawater where they developed for 3–5 days. Undeveloped eggs were discarded. All larvae were anesthetized

using MS-222, photographed, and subsequently transferred to 95% ethanol.

#### DNA analysis

DNA was extracted from muscle tissue, eyes, or in the case of larvae, the whole larva, using an automated Glass Fiber protocol (Ivanova et al. 2006). The 652-bp barcode region of COI was subsequently amplified under the following thermal conditions: 2 min at 95 °C; 35 cycles of 0.5 min at 94 °C, 0.5 min at 52 °C, and 1 min at 72 °C; 10 min at 72 °C; held at 4 °C. The 12.5 µL PCR reaction mixes included 6.25 µL of 10% trehalose, 2.00 µL of ultra-pure water, 1.25 µL 10X PCR buffer (200 mmol/L Tris-HCl (pH 8.4), 500 mmol/L KCl), 0.625 µL MgCl<sub>2</sub> (50 mmol/L), 0.125 µL of each primer cocktail (0.01 mmol/L, using primer cocktails C\_FishF1t1 and C\_FishR1t1 or C\_VF1LFt1 and C\_VR1LRt1, alternatively, both from Ivanova et al. 2007), 0.062 µL of each dNTP (10 mmol/L), 0.060 µL of Platinum® Taq Polymerase (Invitrogen), and 2.0 µL of DNA template. PCR amplicons were visualized on a 1.2% agarose gel E-Gel® (Invitrogen) and bidirectionally sequenced using sequencing primers M13F or M13R (Messing 1983) and the BigDye® Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Inc.) on an ABI 3730xl capillary sequencer following manufacturer's instructions. Bi-directional sequences were assembled and edited using either SEQSCAPE v.2.1.1 (Applied Biosystems) or CodonCode Aligner software (CodonCode Corporation, USA).

#### Sequence analysis

The analysis tools in BOLD (Ratnasingham and Hebert 2007) were used to calculate the nucleotide composition of the sequences and the distribution of Kimura 2-parameter (K2P) distances (Kimura 1980) within and between species. Relationships between all adults were visualized with a NJ tree based on K2P distances. In addition, all adult barcodes were assigned a Barcode Index Number (BIN) as implemented in BOLD (Ratnasingham and Hebert 2013). Sequences are assigned to BINs through the Refined Single Linkage algorithm which employs a three-step process. Sequence clusters are initially delineated by single linkage analysis employing a 2.2% sequence divergence as a minimum distance between clusters. The resulting operational taxonomic units are then refined by Markov clustering followed by selection of the optimal partition based upon the Silhouette criterion (Ratnasingham and Hebert 2013). BIN assignments on BOLD are constantly updated as new sequences are added so a particular BIN can be split or merged as new data are obtained (Ratnasingham and Hebert 2013). BIN assignments were used for data curation, interpretation of species boundaries, and as an indication of potential cryptic species.

Larval identifications were based upon COI sequence data using the “species”-level identification function of the BOLD ID Engine (version August 2015). Specimens

that showed a sequence match of 99% or higher were assigned a species-level identification. Sequences for which the ID engine did not return matches meeting this threshold were examined for closely similar taxa based upon NJ analysis. In cases where immatures showed a sequence similarity score of 95%–99% to reference taxa in the BOLD database, an effort was made to place the taxon in a genus, family, and order if the query sequence fell within a monophyletic cluster of sequences assigned to this family or genus. Placements were always made to the lowest level that did not have taxonomic conflict. When similarity scores dropped below 95%, the immature was classed as “currently unidentifiable”.

To gauge the completeness of the sampling effort for the larval dataset, barcode accumulation curves were calculated based on phylogenetic diversity (Smith et al. 2009) and BIN assignments. For the former, NJ trees were constructed using MEGA 6 (Tamura et al. 2013) and the resultant tree was input into Conserve (Agapow and Crozier 2008). Sampling units of 10 randomly chosen sequences were added to the total phylogenetic tree at each iteration, and the total phylogenetic diversity (Faith 1994; Crozier 1997) was recalculated. This iterative and additive process was used to analyze the barcode phylogenetic diversity for all larval specimens collected. The BIN-based accumulation curve was calculated in BOLD using a randomized sampling order and 100 iterations.

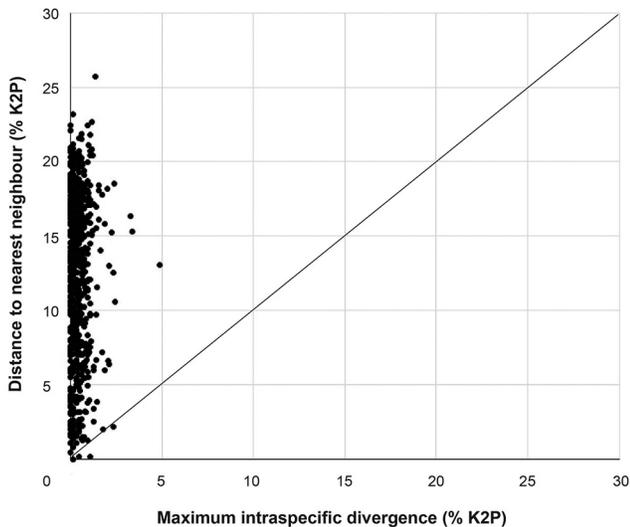
A list of all fish species ( $N = 1919$ ) known from South African marine waters was obtained from FishBase in August 2015 (Froese and Pauly 2015). It was used to determine the fraction of known taxa with barcode coverage and to recognize species that were newly recorded during the present study.

All barcode sequences, and relevant collection data (localities, GPS coordinates, dates, etc.) are available in BOLD, along with voucher repository details, sequence trace files, and photographs of the adults and larvae. All these data are combined into two datasets, one for all adults (<http://dx.doi.org/10.5883/DS-SAAF>) and one for larvae (<http://dx.doi.org/10.5883/DS-SALF>). All barcode sequences are also available in GenBank (see individual datasets for accession numbers hyperlinked to GenBank flatfiles).

## Results

The adult fish dataset included 3215 sequence records that derived from 916 named species and another 30 species that have been recognized but not yet formally described. Among these species, 703 were represented by two or more individuals. They possessed intraspecific divergences averaging 0.25% (with a mean of the maximum intraspecific divergences of 4.64%), while the mean distance to the NN taxon was 48-fold higher, averaging 12.09%. As a consequence, there was a clear barcode gap for most species (Fig. 1).

**Fig. 1.** Plot of maximum intraspecific divergence at COI versus nearest-neighbour distance for 703 species of adult marine fishes from South Africa represented by two or more individuals. Points above the diagonal line indicate species with a barcode gap.



Most (3167/3215) of the sequences from adult fishes met the quality (<1% n) and length (>500 bp) criteria for BIN assignment, and they were assigned to 949 BINs. There was a perfect correspondence between the specimens assigned to a particular BIN and the members of a particular species in nearly all cases (936 of 949). The 14 exceptions involved five cases of a BIN merger; four of these cases involved two species while the other involved three (Table 1). The other three exceptions each involved a BIN split with the members of a particular species assigned to two BINs (Table 1).

To evaluate if the 2526 records in the larval dataset represented a complete larval inventory for the sampling approach and site, barcode accumulation curves were determined (Fig. 2). Since both curves approach an asymptote, the sampling program has captured most of fish species breeding in the area.

Based upon the entire library of COI sequences for fishes on BOLD, 2240 larval fishes (89%) could be identified to a species level (supplementary data, Table S1<sup>2</sup>). These individuals were assigned to 347 species. The 286 specimens that failed to gain a species identification belonged to 127 BINs (= presumptive species). Members of 101 of these BINs could be assigned to a genus, family, or order (Fig. 3). The remaining 92 specimens belonged to 26 BINs and failed to show a close sequence match to any records on BOLD, meaning that their taxonomic status remains uncertain, even at an ordinal level.

The species detected in the present study were compared to the list of all marine fishes ( $N = 1919$ ) known

**Table 1.** Discordances between BIN and species assignments.

Species sharing a BIN	
BOLD:AAB2317	<i>Engraulis encrasicolus</i> / <i>E. japonicus</i>
BOLD:AAB6371	<i>Etmopterus brachyurus</i> / <i>E. lucifer</i>
BOLD:AAC4045	<i>Poroderma africanum</i> / <i>P. pantherinum</i>
BOLD:AAD1027	<i>Serranus knysnaensis</i> / <i>S. novemcinctus</i>
BOLD:AAA8614	<i>Trachurus delagoa</i> / <i>T. murphyi</i> / <i>T. trachurus</i>
Species assigned to two BINs	
<i>Owstonia weberi</i>	BOLD:AAE5219/BOLD:ACF0579
<i>Paraceasio xanthura</i>	BOLD:AAI6821/BOLD:ABZ2393
<i>Selar crumenophthalmus</i>	BOLD:AAB0870/BOLD:0871

from South Africa (Froese and Pauly 2015). Thirty recognized, but undescribed species in the adult dataset were excluded from this analysis as prior occurrence data for these taxa were not available. This analysis showed that 15% of the species collected as adults and 26% of those collected as immatures represented new records for South Africa (Fig. 4). In total, 189 species were newly reported from South Africa from the 1006 species that were collected (Table S2)<sup>2</sup>. The other 817 species have been previously reported from South African waters and represent 43% of its ichthyofauna.

## Discussion

This study assembled DNA barcode sequences for 1006 species of marine fishes that occur in South Africa. This total includes about 43% of the known marine ichthyofauna of the country (Froese and Pauly 2015), as well as another 189 species that are newly recorded. This coverage is remarkable as the collecting effort for this study largely relied upon a single researcher (A.D. Connell).

### Adult reference material

Among the 946 species whose adults were barcoded in this study, only three possessed sufficiently deep sequence divergence to be assigned to two BINs (Table 1). This incidence of BIN splits (0.3%) is lower than the mean frequency (2%) of BIN splits in prior work on other fish faunas (O'Dor et al. 2012 for an overview). The small sample size for each species examined in this study constrained the number of localities that were examined across the range of each species, undoubtedly lowering the prospects of detecting BIN splits. In addition, 30 provisional taxa were excluded prior to analysis because they had already been flagged as overlooked species as a result of earlier barcode studies (Becker et al. 2011; Ward et al. 2009). Although formal descriptions for these taxa are pending, informal names were used to separate these species-in-waiting from the other species where BIN boundaries were discordant with current species boundaries.

<sup>2</sup>Supplementary data are available with the article through the journal Web site at <http://nrcresearchpress.com/doi/suppl/10.1139/gen-2015-0212>.

Fig. 2. DNA barcode accumulation curves using (a) phylogenetic diversity and (b) BINs for randomly picked barcode sequences of 2526 larval marine fish samples.

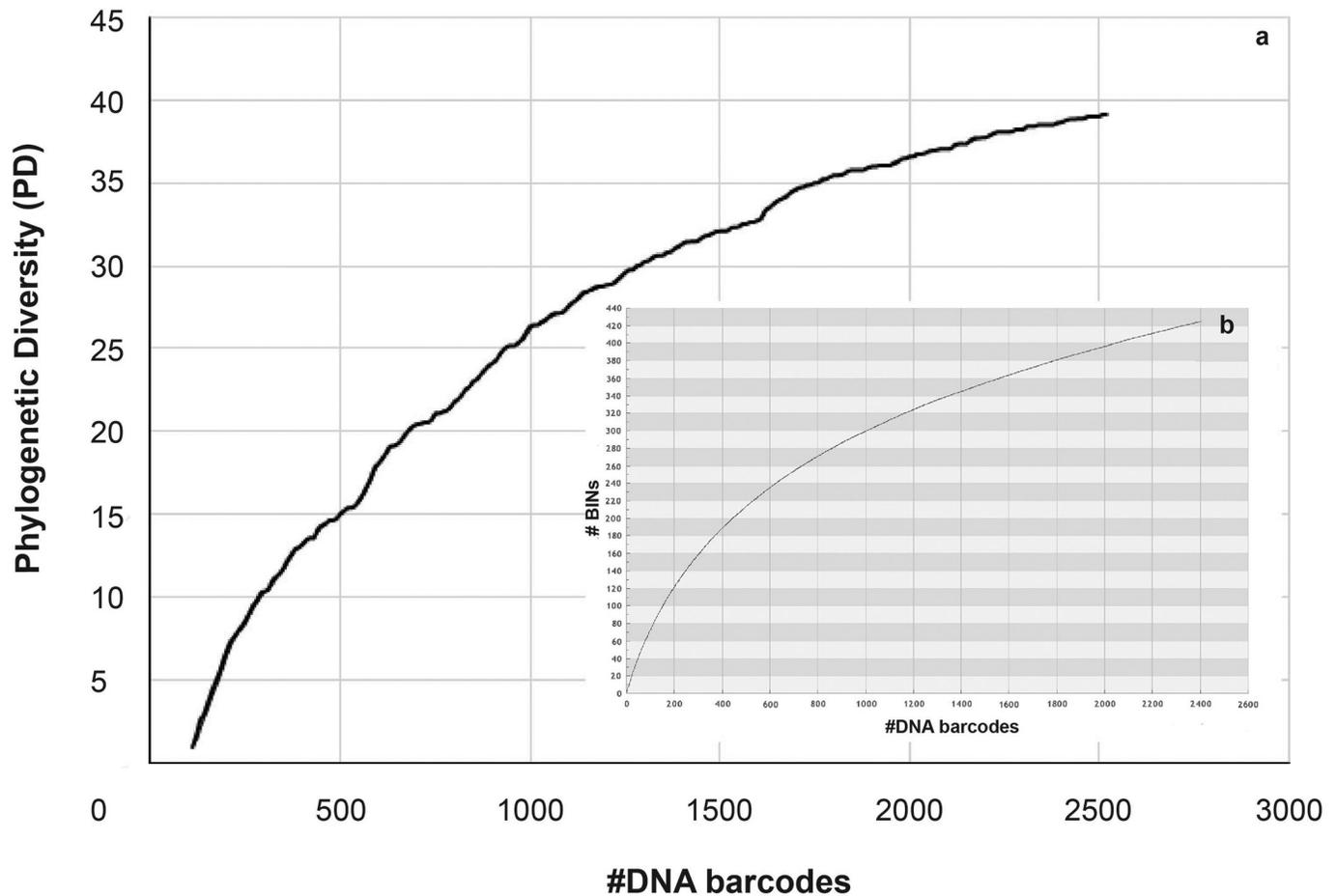
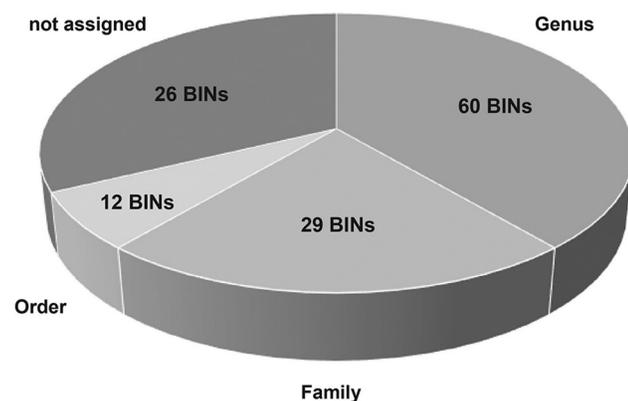


Fig. 3. Pie chart depicting the taxonomic status for the 127 BINs of immature specimens that could not be assigned to a species. All but 26 could be assigned to at least an order.



The present study revealed three new cases of BIN splits involving the following taxa:

1. *Owstonia weberi*

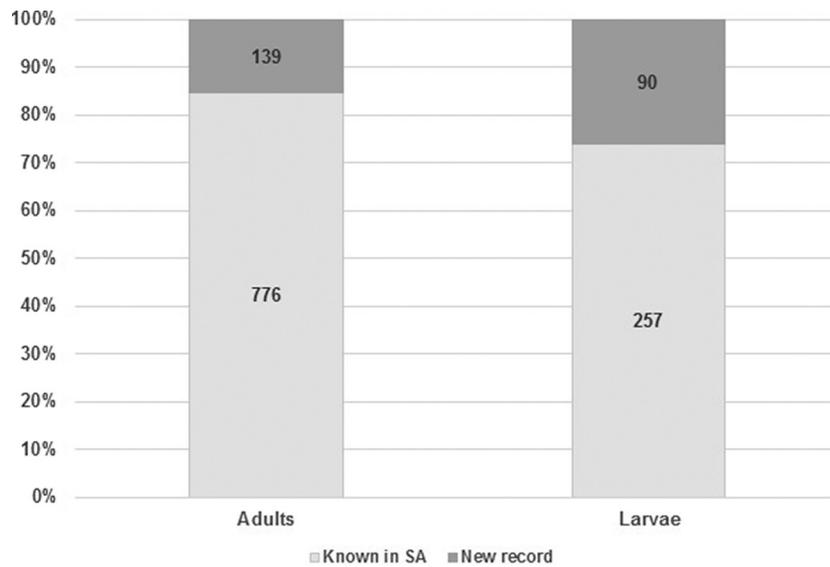
Although COI divergence was quite low (1.28%), *O. weberi* was placed in two BINs represented by individuals collected from a single location (KwaZulu-Natal coast north of Durban). *Parasphenanthias microlepis*

was described from this region just 12 years after the description of *O. weberi* (Gilchrist 1922) from the same locality and was later synonymized with it (Böhlke 1984) as no diagnostic morphological characters were apparent. Although further sampling and genetic analysis of type material is required, the present results suggest that *P. microlepis* may be valid.

2. *Paracaesio xanthura*

A similar situation occurs in *P. xanthura* whose specimens were represented by two quite divergent (2.9%) allopatric BINs, which might also reflect a case of unjustified synonymization. One BIN contained a single specimen from Pomene, Mozambique, while the other was represented by five specimens from Park Rynie. Bleeker (1869) described this species from specimens collected at Nossi-Bé, Madagascar. Another species, *Paracaesio cantharoides* (Barnard 1937), was described from KwaZulu-Natal, but Allen (1985) synonymized it with *P. xanthura* based on their morphological similarity and overlapping distributions. However, Carpenter and Niem (2001) suggested that *P. xanthura* might actually include at least two species, a view supported by our results. The marked difference between the two BINs suggests that *P. cantharoides* is a valid species.

**Fig. 4.** The numbers of new and known fish species from South African waters based on the barcode analysis of both adults and immatures.



### 3. *Selar crumenophthalmus*

A recent study indicated that specimens of bigeye scad (*S. crumenophthalmus*) from Indonesia included two COI lineages with 2.6% divergence that occurred sympatrically (Mat Jaafar et al. 2012). Our study indicates the presence of the same two lineages, each assigned to a different BIN in South African waters.

Our study detected five instances of BIN sharing (Table 1). Such cases can result from two different patterns of sequence variation. Species involved in BIN sharing may belong to separate monophyletic clusters separated by very low divergence or the species involved may have identical or intermingled sequences. In the latter case, DNA barcodes cannot be used to discriminate the species, but they can in the former case. Three of the five cases of BIN sharing detected in this study involved species with low divergences:

#### 1. *Serranus knysnaensis*/*Serranus novemcinctus*

These two species of dwarf sea basses were assigned to one BIN together with a third species (*Serranus cabrilla*) that was not collected in this study. *Serranus novemcinctus* has been recognized as a valid species since its description (Kner 1864), but *S. knysnaensis* was first synonymized with *S. cabrilla* and subsequently resurrected (Heemstra and Heemstra 2004). All three taxa appear to represent valid taxa as they form monophyletic clusters separated by 1.8%–2.25% sequence divergence, values which are not exceptionally low for closely related species of ray finned fishes (Bucklin et al. 2011).

#### 2. Elasmobranchs

Both South African species of lantern shark (*Etmopterus brachyurus*/*Etmopterus lucifer*) were assigned to the same BIN, and two catsharks (*Poroderma africanum*/*Poroderma pantherinum*) also shared a BIN (Table 1). However, in

each case, sequences fell into two distinct groups separated by 1.2% and 1.1% sequence divergence, respectively. Prior barcoding studies on elasmobranchs have shown that COI divergences are lower than in most other chordates (Spies et al. 2006; Smith et al. 2008; Ward et al. 2008), perhaps reflecting their lower rates of mtDNA evolution (Martin and Palumbi 1993; Johns and Avise 1998).

The last two cases of BIN sharing involved specimens assigned to different species with intermingled or identical barcodes. These taxa need to be tested for sequence divergence at nuclear loci, ideally at localities where they are sympatric, to ascertain if their component species are truly reproductively isolated. If no differences are detected, the taxa involved may reflect cases of oversplitting while if differences are present, their shared barcodes may indicate incomplete lineage sorting or introgressive hybridization.

#### 1. *Engraulis encrasicolus*/*Engraulis japonicus*

Earlier studies (Grant et al. 2005, Whitehead et al. 1988) have indicated that all five species of Old World anchovies (*E. australis*, *E. capensis*, *E. encrasicolus*, *E. eurystole*, *E. japonicus*) lack unique evolutionary histories and show little or no morphological divergence. As a consequence, it has been suggested that they actually represent just a single species (Whitehead et al. 1988), a conclusion supported by our findings and other records in BOLD that show that individuals identified as all five members of the Old World anchovies fall into a single BIN (BOLD:AAB2317).

#### 2. *Trachurus delagoa*/*Trachurus murphyi*/*Trachurus trachurus*

Low levels of mtDNA sequence divergence have been reported among the 11 species of *Trachurus*, with mean congeneric divergences of  $3.43\% \pm 1\%$  for cytochrome *b* and  $3.50\% \pm 1\%$  for D-loop, values that are among the lowest reported for any genus of marine fishes (Cárdenas

et al. 2005). A recent study (Neira et al. 2015) indicated that COI barcodes could not discriminate *T. declivis* and *T. novaehollandiae* and recommended the use of cytochrome *b* and species-specific single-nucleotide polymorphisms (SNPs). In fact, most *Trachurus* barcodes available on BOLD, including those for three of the four species examined in this study (*T. capensis* possessed a different BIN), belong to a single BIN.

### Larval identification

The early life-history stages of fishes are poorly known, impeding acquisition of the identifications needed to monitor larval recruitment and year-class strength. By contrast, 2240 of the 2526 immatures collected in this study gained a species identification, reflecting the fact that reference barcodes were available for 74% of all marine fishes known from South Africa. The identified larvae represented 347 species, and the barcode results suggested that the remaining individuals belonged to 127 additional species. Most of the latter species could be placed in a genus, family, or order (Fig. 3). It is important to note that those taxa that currently lack a species identification will gain one as the barcode library moves toward completion.

Smith et al. (2009) showed that accumulation curve based on estimates of phylogenetic diversity (PD) derived from barcode data can be calculated independently from the assignment of sequences to a species or BIN. In our case, this approach allowed inclusion of all larval records, including those without a species assignment as well as those that failed to meet the quality requirements to gain a BIN assignment (Ratnasingham and Hebert 2013). The PD accumulation curve for all larval fish (Fig. 2) approached an asymptote, indicating that sampling is comprehensive for assessments employing diurnal plankton tows. The BIN-based accumulation curve (Fig. 2) shows a similar but not as pronounced shape that can be explained by the fact that 121 larval sequences did not gain a BIN assignment due to the quality requirements described above.

The adult and larval fishes examined in this study represent 43% of all fishes known from South African waters and additionally includes 189 newly reported species (Fig. 4; Table S2<sup>2</sup>). The incidence of new species discovery is particularly striking for the larval dataset that included 90 species new to South Africa, especially in light of the fact that our sampling methods excluded marine fishes that do not have pelagic eggs (~600 further species). Half of these new taxa are tropical reef fishes that have a pelagic larval stage before they settle on a suitable reef (Richards 2006). Understanding their early life histories is important to comprehend population dynamics and, more broadly, the biogeography, ecology, and evolution of coral reef fishes. Despite its undoubted importance, the settlement transition between coastal and pelagic waters has not been comprehensively studied, and one of the reasons for this is the lack of information

on larval stages. The fact that more than a quarter of larvae assigned to species in our study are new records for South Africa demonstrates this clearly, but it also shows the utility of DNA barcoding for studies that seek insights into larval recruitment and horizontal migration. Another 28% (25 species) of the new records represent fish that only occur in deep waters as adults, species which initially live in shallower water and descend as they mature (Randall and Farrell 1997). Again, little is known about this process because of the difficulty in identifying juvenile stages. Although further parametrization of the barcode reference library is desirable, the past decade of work (Becker et al. 2011; Ward et al. 2009) has delivered a library that is already highly effective.

In summary, this study has delivered DNA barcodes for over 5000 individuals representing more than 1000 species of South African fishes. The fact that these specimens represent about 3% of all known fish species and 10% of the fish species barcoded to date is startling as they derive from the efforts of a single researcher. This effort took essentially 10 years, but it clearly shows that it does not require a very large workforce to accomplish completion of DNA barcode reference libraries for major species assemblages. This study also demonstrates that DNA barcodes can regularly deliver species-level identifications for marine fishes, further validating the effectiveness of COI as the core of a DNA-based identification system for fish.

### Acknowledgements

The authors are grateful to a network of collectors around the South African coast, who willingly, and often generously, contributed to the collection of adult fish. Knut Sorensen, the skipper and owner of the trawler Ocean Spray, operating out of Durban, kept boxes of bycatch, which were always interesting and challenging. Rob Cooper, of Marine & Coastal Management, provided many trawled specimens from the fisheries research vessel RV Africana, operating in the Western Cape. Along the coast of KwaZulu-Natal, many anglers, with good knowledge of local species, kept specimens they were unsure of. Among these anglers are Alan Fraser, Peter Snyman, Gerald Buckle, Patrick Eaton, Archie Viviers, and Jeremy Williams. To these and many others who contributed the occasional fish, we are extremely grateful. The Natal Sharks Board contributed tissues of many elasmobranchs, the effort co-ordinated through Sheldon Dudley and Jeremy Cliff. Staff at the Oceanographic Research Institute and uShaka Marine World complex also contributed, especially Bruce Mann, Simon Chater, and Sean Fenessey. Philip and Elaine Heemstra provided material during collecting trips to the Aliwal Shoal. Steven Weerts and Alan Whitfield collected some interesting estuarine species. On the skiboating front, hundreds of surf launches were made with the late Pieter van Niekerk, then Gerald Smith and George Halekas. More recently Mike and Valda Fraser have shared many

launches and diving experiences, Mike as skipper/diver and Valda as underwater photographer. We thank Emily Berzitis, Alex Borisenko, Erin Corstorphine, Natalya Ivanova, Christa Maitland, Paola Pierossi, and Tyler Zemlak for their invaluable help with specimen handling and logistics as well as DNA analyses. We also thank Benjamin Victor, Jack Randall, Tyson Roberts, Yukio Iwatsuki, Sophie van der Heyden, Phil Heemstra, Wouter Holleman, Brent Chiazzari, Helen Larson, and Rick Winterbottom for their help in identifying a number of taxa. Funding for this work was provided by the Gordon and Betty Moore Foundation, NSERC, the Canada Research Chairs program and the government of Canada through Genome Canada and the Ontario Genomics Institute in support of the International Barcode of Life Project.

## References

- Agapow, P.M., and Crozier, R.H. 2008. Conserve IV. Available from <http://www.agapow.net/software/consolve>.
- Allen, G.R. 1985. FAO species catalogue. Snappers of the world. An annotated and illustrated catalogue of lutjanid species known to date. FAO Fisheries Synopsis, 6: 1–208.
- Baldwin, C.C., and Smith, D.G. 2003. Larval Gobiidae (Teleostei: Perciformes) of Carrie Bow Cay, Belize, Central America. *Bull. Mar. Sci.* 72: 639–674.
- Barnard, K.H. 1937. Further notes on South African marine fishes. *Ann. South Afr. Mus.* 32: 41–67.
- Becker, S., Hanner, R., and Steinke, D. 2011. Five years of FISH-BOL: Brief status report. *Mitochondrial DNA*, 22(S1): 3–9. PMID:21271850.
- Bleeker, P. 1869. Description d'une espèce inédite de *Caesio* de l'île de Nossibé. *Verslagen en Mededeelingen der Koninklijke Akademie van Wetenschappen. Afdeling Natuurkunde*, 3: 78–79.
- Böhlke, E.B. 1984. Catalog of type specimens in the ichthyological collection of the Academy of Natural Sciences of Philadelphia. *Spec. Publ. Acad. Nat. Sci. Phila.* 14: 1–216.
- Bucklin, A., Steinke, D., and Blanco-Bercial, L. 2011. DNA barcoding of marine Metazoa. *Ann. Rev. Mar. Sci.* 3: 471–508. doi:10.1146/annurev-marine-120308-080950. PMID:21329214.
- Busby, M.S., Matarese, A.C., Blood, D.M., and Konieczna, M. 2000. Advances in ichthyoplankton taxonomy in the Northeastern Pacific Ocean and Bering Sea: Research conducted by the Alaska Fisheries Science Center 1965–1999. *Bull. Sea Fish. Inst.* 3: 11–20.
- Cárdenas, L., Hernández, C.E., Poulin, E., Magoulas, A., Kornfield, I., and Ojeda, F.P. 2005. Origin, diversification, and historical biogeography of the genus *Trachurus* (Perciformes: Carangidae). *Mol. Phylogenet. Evol.* 35: 496–507. doi:10.1016/j.ympev.2005.01.011. PMID:15804418.
- Carpenter, K.E., and Niem, V.H. 2001. FAO species identification guide for fishery purposes. The living marine resources of the Western Central Pacific. Vol. 5. Bony fishes. Part 3: (Menidae to Pomacentridae). Rome, FAO. pp. 2791–3380.
- Connell, A.D. 2010. A 21-year ichthyoplankton collection confirms sardine spawning in KwaZulu-Natal waters. *Afr. J. Mar. Sci.* 32: 331–336. doi:10.2989/1814232X.2010.502638.
- Costa, F.O., and Carvalho, G.R. 2007. The Barcode of Life Initiative: synopsis and prospective societal impacts of DNA barcoding of fish. *Genomics Soc. Policy*, 3: 29–40. doi:10.1186/1746-5354-3-2-29.
- Côté, I.M., Green, S.J., Morris, J.A., Jr., Akins, J.L., and Steinke, D. 2013. Diet richness of invasive Indo-Pacific lionfish revealed by DNA barcoding. *Mar. Ecol. Progr. Ser.* 72: 249–256.
- Crozier, R.H. 1997. Preserving the information content of species: genetic diversity, phylogeny and conservation worth. *Annu. Rev. Ecol. Syst.* 28: 243–268. doi:10.1146/annurev.ecolsys.28.1.243.
- Fahay, M.P. 2007. Early stages of fishes in the western north Atlantic Ocean (Davis Strait, Southern Greenland and Flemish Cap to Cape Hatteras). Northwest Atlantic Fisheries Organization, Dartmouth, N.S.
- Faith, D.P. 1994. Phylogenetic pattern and the quantification of organismal biodiversity. *Philos. Trans. R. Soc. B Biol. Sci.* 345: 45–58. doi:10.1098/rstb.1994.0085.
- Froese, R., and Pauly, D. (Editors). 2015. Fishbase. World Wide Web electronic publication Available from <http://www.fishbase.org>.
- Garland, E.D., and Zimmer, C.A. 2002. Techniques for the identification of bivalve larvae. *Mar. Ecol. Progr. Ser.* 225: 299–310. doi:10.3354/meps225299.
- Gilchrist, J.D.F. 1922. Deep-sea fishes procured by the S.S. "Pickle" (Part I). *Rep. Fish. Mar. Biol. Surv. Union S. Afr. Rep.* 2: 41–79.
- Grant, W.S., Leslie, R.W., and Bowen, B.W. 2005. Molecular genetic assessment of bipolarity in the anchovy genus *Engraulis*. *J. Fish Biol.* 67: 1242–1265. doi:10.1111/j.1095-8649.2005.00820.x.
- Heemstra, P.C., and Heemstra, E. 2004. Coastal fishes of southern Africa. NISC and SAIAB. i-xxiv + 1–488.
- Hubert, N., Hanner, R., Holm, E., Mandrak, N.E., Taylor, E., Burrige, M., et al. 2008. Identifying Canadian freshwater fishes through DNA barcodes. *PLOS ONE*, 3: e2490. doi:10.1371/journal.pone.0002490. PMID:22423312.
- Hubert, N., Delrieu-Trottin, E., Irissou, J.O., Meyer, C., and Planes, S. 2010. Identifying early stages of coral reef fishes through DNA barcoding: a test case with the families Acanthuridae and Holocentridae. *Mol. Phylogenet. Evol.* 55: 1195–1203. doi:10.1016/j.ympev.2010.02.023. PMID:20188843.
- Ivanova, N.V., deWaard, J.R., and Hebert, P.D.N. 2006. An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Mol. Ecol. Notes*, 6: 998–1002. doi:10.1111/j.1471-8286.2006.01428.x.
- Ivanova, N.V., Zemlak, T.S., Hanner, R.H., and Hebert, P.D.N. 2007. Universal primer cocktails for fish DNA barcoding. *Mol. Ecol. Notes*, 7: 544–548. doi:10.1111/j.1471-8286.2007.01748.x.
- Johns, G., and Avise, J.C. 1998. A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome *b* gene. *Mol. Biol. Evol.* 15: 1481–1490. doi:10.1093/oxfordjournals.molbev.a025875. PMID:12572611.
- Kimura, M. 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111–120. doi:10.1007/BF01731581.
- Kner, R. 1864. Specielles Verzeichniss der während der Reise der kaiserlichen Fregatte "Novara" gesammelten Fische. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften. Mathematisch-Naturwissenschaftliche Classe v.* 49: 481–486.
- Ko, H.L., Wang, Y.T., Chiu, T.S., Lee, M.A., Leu, M.Y., Chang, K.Z., et al. 2013. Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. *PLOS ONE*, 8: e53451. doi:10.1371/journal.pone.0053451. PMID:23382845.
- Leis, J.M., and Hay, A.C. 2004. Larval development of *Achoerodus viridis* (Pisces: Labridae), the Australian eastern blue grouper. *Ichthyol. Res.* 51: 46–51. doi:10.1007/s10228-003-0192-4.
- Martin, A.P., and Palumbi, S.R. 1993. Protein evolution in different cellular environments: cytochrome *b* in sharks and mammals. *Mol. Biol. Evol.* 10: 873–891. PMID:8355604.
- Mat Jaafar, T.N.A., Taylor, M.I., Mohd Nor, S.A., de Bruyn, M., and Carvalho, G.R. 2012. DNA barcoding reveals cryptic diversity within commercially exploited Indo-Malay Carangidae (Teleostei: Perciformes). *PLOS ONE*, 7: e49623. doi:10.1371/journal.pone.0049623. PMID:23209586.

- Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* **101**: 20–78. PMID:6310323.
- Neira, F.J., Perry, R.A., Burrridge, C.P., Lyle, J.M., and Keane, J.P. 2015. Molecular discrimination of shelf-spawned eggs of two co-occurring *Trachurus* spp. (Carangidae) in southeastern Australia: a key step to future egg-based biomass estimates. *ICES J. Mar. Sci.* **72**: 614. doi:10.1093/icesjms/fsu151.
- O'Dor, R., Boustany, A.M., Chittenden, C.M., Costello, M.J., Moustahfid, H., Payne, J., et al. 2012. A census of fishes and everything they eat: How the Census of Marine Life advanced fisheries science. *Fisheries*, **37**: 398–409. doi:10.1080/03632415.2012.714323.
- Olson, R.R., Runstadler, J.A., and Kocher, T.D. 1991. Whose larvae? *Nature*, **351**: 357–358. PMID:2034285.
- Randall, D.J., and Farrell, A.P. 1997. *Deep-sea fishes*. Academic, San Diego, Calif.
- Ratnasingham, S., and Hebert, P.D.N. 2007. BOLD: The Barcode of Life Data System ([www.barcodinglife.org](http://www.barcodinglife.org)). *Mol. Ecol. Notes*, **7**: 355–364. PMID:18784790.
- Ratnasingham, S., and Hebert, P.D.N. 2013. A DNA-based registry for all animal species: The Barcode Index Number (BIN) system. *PLOS ONE*, **8**: e66213. doi:10.1371/journal.pone.0066213. PMID:23861743.
- Richards, W.J. 2006. *Early stages of Atlantic fishes: an identification guide for the western central north Atlantic*. Taylor & Francis, Boca Raton, Fla.
- Smith, M.A., Fernandez-Triana, J., Roughley, R., and Hebert, P.D.N. 2009. DNA barcode accumulation curves for understudied taxa and areas. *Mol. Ecol. Resour.* **9**: 208–216. doi:10.1111/j.1755-0998.2009.02646.x. PMID:21564980.
- Smith, P.J., McVeagh, S.M., Allain, V., and Sanchez, C. 2005. DNA identification of gut contents of large pelagic fishes. *J. Fish Biol.* **67**: 1178–1183. doi:10.1111/j.0022-1112.2005.00804.x.
- Smith, P.J., Steinke, D., McVeagh, M.S., Stewart, A.L., Struthers, C.D., and Roberts, C.D. 2008. Molecular analysis of Southern Ocean skates (*Bathyraja*) reveals a new species of Antarctic skate. *J. Fish Biol.* **73**: 1170–1182. doi:10.1111/j.1095-8649.2008.01957.x.
- Spies, I.B., Gaichas, S., Stevenson, D.E., Orr, J.W., and Canino, M.F. 2006. DNA-based identification of Alaska skates (*Amblyraja*, *Bathyraja* and *Raja*: Rajidae) using cytochrome c oxidase subunit I (COI) variation. *J. Fish Biol.* **69**: 283–292. doi:10.1111/j.1095-8649.2006.01286.x.
- Steinke, D., Zemplak, T.S., and Hebert, P.D.N. 2009. Barcoding Nemo: DNA-based identifications for the ornamental fish trade. *PLoS ONE*, **4**: e6300. doi:10.1371/journal.pone.0006300.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**: 2725–2729. doi:10.1093/molbev/mst197. PMID:24132122.
- Valdez-Moreno, M., Vásquez-Yeomans, L., Elías-Gutiérrez, M., Ivanova, N.V., and Hebert, P.D.N. 2010. Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management. *Mar. Freshw. Res.* **61**: 665–671. doi:10.1071/MF09222.
- Ward, R.D. 2000. Genetics in fisheries management. *Hydrobiologia*, **420**: 191–201. doi:10.1023/A:1003928327503.
- Ward, R.D., Zemplak, T.S., Innes, B.H., Last, P.R., and Hebert, P.D.N. 2005. DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. B Biol. Sci.* **360**: 1847–1857. doi:10.1098/rstb.2005.1716.
- Ward, R.D., Holmes, B.H., White, W.T., and Last, P.R. 2008. DNA barcoding Australasian chondrichthyans: results and potential uses in conservation. *Mar. Freshw. Res.* **59**: 57–71. doi:10.1071/MF07148.
- Ward, R.D., Hanner, R., and Hebert, P.D.N. 2009. The campaign to DNA barcode all fishes, FISH-BOL. *J. Fish Biol.* **74**: 329–356. doi:10.1111/j.1095-8649.2008.02080.x. PMID:20735564.
- Whitehead, P.J.P., Nelson, G.J., and Wongratana, T. 1988. *FAO species catalogue. Clupeoid fishes of the world (Suborder Clupeoidei). An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, anchovies and wolf-herrings. Part 2. Engraulididae*. *FAO Fish. Synop.* **7**: 305–579.