Food Control 118 (2020) 107348

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Difficulties in DNA barcoding-based authentication of snapper products due to ambiguous nucleotide sequences in public databases

Kristina Kappel[a,*], Ute Schröder[b]

[a] National Reference Centre for Authentic Food, Max Rubner-Institut, Hermann-Weigmann-Str. 1, 24103, Kiel, Germany
[b] Department of Safety and Quality of Milk and Fish Products, Max Rubner-Institut, Palmaille 9, 22767, Hamburg, Germany

ARTICLE INFO

Keywords:
Food control
Fishery product
Snapper
Fish species
DNA barcoding
Lutjanidae

ABSTRACT

Fishery products are traded worldwide, have long and branched supply chains and are among the most counterfeit foods. Food control authorities are responsible of controlling food labelling in order to ensure the safety and quality of products. This is particularly important for tropical and subtropical fish species, such as snappers (Lutjanidae), because certain species are known to be frequently contaminated with ciguatoxins that cause harmful Ciguatera Fish Poisoning outbreaks. The analytical method of choice for the identification of a sample at species level is 'DNA barcoding', the sequencing of specific genetic markers and the comparison of the sequences with international nucleotide databases, such as GenBank and BOLD. However, the results of these analyses can be severely impaired by an insufficient data basis and especially by database entries with incorrect species annotations. In this work, the available nucleotide sequences for common genetic markers (COI, *cytb*, 16S rDNA, 12S rDNA, *rag1* and ITS) of snapper species in GenBank were subjected to careful examination with regard to their unambiguousness and clarity. Phylogenetic neighbour joining trees were prepared and checked for ambiguous and contradictory placement of species' sequences with special emphasis on the Malabar blood snapper (*Lutjanus malabaricus*), the two-spot red snapper (*L. bohar*) and the crimson snapper (*L. erythropterus*). The results indicate that ambiguous and contradictory database nucleotide sequences impede the DNA barcoding-based authentication of snapper products at species level. A species assignment for the Malabar snapper and the crimson snapper based solely on database queries seems questionable.

1. Introduction

Fishery products feature long and branched supply chains and are particularly susceptible to mislabelling and substitution. Unintentional mislabelling can occur, for example due to misidentification of similar looking species at the catch level or because of the existence of ambiguous and mismatched trade names between countries. But also common is the deliberate substitution due to the high demand for these goods and substantial price differences between species (Reilly, 2018). By implementing the regulation on the common organisation of the markets in fishery and aquaculture products ('REGULATION (EU) No 1379/2013'), the European Union (EU) has laid down detailed labelling rules to be followed when selling fishery products to consumers. According to this regulation, unprepared products have to be labelled not only with the commercial designation but also with the scientific (Latin) name of the particular species. For this purpose, all EU member states are obliged to publish a list specifying commercial designations for all marketed fish and seafood species.

In many cases, mislabelled fishery products affect the quality of the products and therefore only have a financial impact on the consumers. However, in some cases the substitutes may have considerable negative health effects. This is especially true for tropical and subtropical reef fishes, like snappers, parrotfishes, groupers, barracudas, mackerels, jacks and others, as they may contain ciguatoxins which cause Ciguatera Fish Poisoning (CFP). CFP is a serious marine toxin related illness with sometimes protracted neurological symptoms (Friedman et al., 2008). To date, a validated, rapid and cost-effective test concerning the ciguatoxin (CTX) content in fish specimens is not available (Friedman et al., 2017). Thus, the prevention of CFP is mainly based on avoiding medium and high-risk species from CFP endemic areas and depends primarily on the knowledge of local fishermen (Friedman et al., 2017). However, it is extremely difficult for wholesalers and retailers from importing countries to identify exotic fish specimens at species level, especially if the fish has already been filleted in the country of origin.

Particular species known to accumulate large amounts of

* Corresponding author.
E-mail addresses: kristina.kappel@mri.bund.de (K. Kappel), ute.schroeder@mri.bund.de (U. Schröder).

https://doi.org/10.1016/j.foodcont.2020.107348
Received 27 February 2020; Received in revised form 6 May 2020; Accepted 7 May 2020
Available online 03 June 2020

ciguatoxins, for example, are the closely related, long-lived and distinctly red snapper species two-spot red snapper (*Lutjanus bohar*) and humpback red snapper (*L. gibbus*). These species are therefore banned from sale in some countries (Allen & Talbot, 1985; Edwards, Zammit, & Farrell, 2019; Laurent, Yeeting, Labrosse, & Gaudechoux, 2005). According to Friedemann (2019), there have been six documented outbreaks of CFP in Germany since 2012, all of which could be attributed to imported snappers (Lutjanidae) from India, Indonesia and Vietnam. In at least five of these instances, the two-spot red snapper was not only the source of the poisoning, but the products consumed had also been labelled as 'Red Snapper'. However, in Germany the commercial designation 'Red Snapper' is exclusively allowed for the Malabar blood snapper (*Lutjanus malabaricus*) according to the German list of trade descriptions for fishery and aquaculture products (see https://www.ble.de). Because of this Friedemann recommends to not import the two-spot red snapper from the Indo-Pacific Ocean. Yet she also states, that snappers are often insufficiently labelled (e.g. *Lutjanus* spp.) at import. Thus, on the one hand it is paramount that the imported goods are tested by the importing companies in order to identify them at species level before they are traded to consumers or restaurants. On the other hand, national food control authorities should regularly verify that the labelling of snapper products is correct in order to ensure that snapper species posing high CFP risks are neither traded on markets nor served in restaurants.

The current method of choice for identifying an individual fish at species level is the sequencing of common genetic markers, such as cytochrome *b* gene (*cytb*) or cytochrome *c* oxidase I gene (COI, also abbreviated as *cox1*) segments. The determined sequences are compared to sequences in international nucleotide databases, like GenBank (https://www.ncbi.nlm.nih.gov/nucleotide) and the Barcode of Life Data System (BOLD, http://www.boldsystems.org). This strategy is commonly known as DNA barcoding. Whereas GenBank constitutes an annotated collection of all sorts of publicly available DNA sequences (Benson et al., 2013), BOLD was especially designed to serve as an online workbench and database for DNA barcode data and even features an identification engine to identify a sample at species level (Ratnasingham & Hebert, 2007). While GenBank can be searched with all kinds of genetic markers, BOLD can currently only be searched with COI sequences when analysing animal samples. However, the use of these databases for the species authentication or the species determination of a sample requires that the data for the respective taxa are correct and unambiguous. Single entries with incorrect species annotations may be recognized as such if there are sufficient entries with correct annotations. Yet, sometimes there are no or only very few entries for a respective species or the data for the taxon are ambiguous and overall confusing. This was recently demonstrated by Li et al. (2018), who had carried out an analysis of intraspecific and interspecific divergences of *cytb* sequences from fish GenBank entries and out of 35,130 sequences had identified 1303 problematic sequences that they believed needed further inspection.

The goal of the present study was to evaluate the validity and unambiguousness of the DNA sequences for several genetic markers (i.e. COI, *cytb*, 16S RNA gene (16S rDNA), 12S RNA gene (12S rDNA), recombination activating gene 1 (*rag1*) and the internal transcribed spacer region (ITS)), which are deposited in GenBank with the annotation of Lutjanidae species. For this purpose, respective nucleotide sequences were downloaded from GenBank and used for generating neighbour joining (NJ) trees. Subsequently, the NJ trees were inspected for ambiguities with special emphasis on the placement of sequences annotated as Malabar blood snapper, two-spot red snapper and crimson snapper (*L. erythropterus*). BOLD BINs including sequences from these species were checked for clarity, too. Additional COI sequences for selected snapper species were extracted from the Reference Standard Sequence Library (RSSL) for Seafood Identification of the U.S. Food and Drug Administration (FDA). This library constitutes a curated database with COI sequences from reference specimens with authoritative taxonomic authentication and was explicitly established to authenticate seafood products (Deeds et al., 2014). To analyse the feasibility of DNA barcoding-based authentication of snapper products as an example, a fish specimen labelled as 'Red Snapper (*Lutjanus malabaricus*)' was bought at a local supermarket in Germany. The results of this study should support food control agencies as well as service laboratories to better judge GenBank and BOLD query results when authenticating snapper products at species level. However, this study did not aim to analyse the phylogenetic relationships of snapper species.

2. Materials and methods

2.1. GenBank sequences

COI, *cytb*, 16S rDNA, 12S rDNA, *rag1* and ITS DNA sequences for Lutjanidae species were downloaded in FASTA format from GenBank with an in-house Python script on January 29th, 2019. The search terms were 'txid308050 [Organism] AND (cox1 [All Fields] OR COI [All Fields])', 'txid308050 [Organism] AND cytb [All Fields]', 'txid308050 [Organism] AND 16S [All Fields]', 'txid308050 [Organism] AND 12S [All Fields]', 'txid308050 [Organism] AND rag1 [All Fields]', and 'txid308050 [Organism] AND internal transcribed spacer [All Fields]', respectively. The sequences' names were changed in the FASTA-files with an in-house C++ script, so that the specified taxon was displayed at the beginning, followed by the accession number (e.g. '>L_malabaricus_EU502677'). This served for a better legibility during subsequent analyses. The sequences belonging to the different genetic markers were aligned with ClustalX 2.1 (Larkin et al., 2007) in separate alignments and the alignments were visualised in AliView 1.18 (Larsson, 2014). If individual sequences could not be aligned with the main part of the sequences, it was checked whether they were displayed in reverse-complement format. In that case, the sequences were transferred into the required direction and were again aligned to the other sequences. If that was not the case, the sequences were aligned to a complete mitochondrial genome of a randomly selected Lutjanidae species (in this case *L. malabaricus* (NC_012736)) in order to check, whether they exhibited overlapping regions with the main part of the sequences. If so, the sequences were moved manually to the correct position of the alignment. If no overlapping regions could be identified, the sequences were discarded. If large subsets of sequences represented completely different regions of the marker genes, they were subsequently treated in separate alignments. In order to obtain alignments suitable for the generation of NJ trees, short sequences and sequences with bad qualities (e.g. sequences displaying a lot of ambiguous bases; protein coding genes with insertions or deletions) were deleted and the alignments were truncated to reasonable lengths. During this reduction process as many sequences as possible were retained without losing too much nucleotide positions of the alignment. The final alignments were saved as FASTA files. The protein-coding genes were translated into amino acid sequences with the Vertebrate Code (*cytb*, COI) or the Standard Code (*rag1*) and checked for stop codons. Neighbour-joining trees (Nei & Saitou, 1987) were prepared with MEGA 7.0.14 (Kumar, Stecher, & Tamura, 2016) from FASTA files. The evolutionary distances were computed using the Kimura 2-parameter (K2P) method (Kimura, 1980) and are in the units of the number of base substitutions per site. For the presentation of the trees, clusters were condensed and different taxa present within the individual clusters were displayed with the number of their frequencies in parentheses.

2.2. FDA sequences

All available COI sequences of Lutjanidae species found in mixed clusters with *L. malabaricus*, *L. bohar* or *L. erythropterus* sequences in the NJ trees (see 2.1) were manually downloaded from the FDA Reference Standard Sequence Library (RSSL) for Seafood Identification on February 25th, 2019. The sequences were integrated in the COI

alignment and preparation of the phylogenetic COI tree was repeated as described above.

2.3. BOLD sequences

BOLD was searched for available Barcode Index Numbers (BINs) containing entries with *L. malabaricus*, *L. erythropterus* and *L. bohar* annotations as well as *L. fulvus* annotations on September 10th, 2019. The latter species was also included in the search because the COI sequence of the market sample resembled the COI sequences from *L. fulvus* entries (see below). The BINs were recorded together with the different species annotations and the number of occurrences.

2.4. Analysis of a 'Red Snapper' market sample

A fish specimen labelled with the commercial designation 'Red Snapper' and the scientific name '*Lutjanus malabaricus*' was bought at a local supermarket in Kiel, Germany, in March 2019. A small piece of muscle tissue (about 50 mg) was dissected and total DNA was isolated with a modified CTAB-method (Iwobi, Huber, Hauner, Miller, & Busch, 2011). DNA purity and concentration were measured spectrophotometrically.

The genetic markers COI, *cytb*, 16S rDNA, 12S rDNA and *rag1* were amplified in separate PCRs with the primer systems specified in Table 1. The PCR reactions were carried out in volumes of 20 µl consisting of 10 µl AccuStart II PCR Supermix (Quantabio, Beverly, MA, USA), 125 nM (in case of COI PCR) or 500 nM of each primer (*cytb*, 16S rDNA, 12S rDNA and *rag1* PCRs) (synthesized by Biomers, Ulm, Germany) and 20 ng isolated sample DNA. The PCR cycling conditions are indicated in Table 2. All PCRs were conducted with 35 cycles of denaturation, primer annealing and elongation.

The PCR products were inspected by agarose gel electrophoresis. Successful reactions were diluted one to ten with molecular biology grade water and were sent without prior purification to LGC Genomics (Berlin, Germany) for sequencing in both directions.

The electropherograms were examined in Chromas Lite 2.6.6 (Technelysium Pty Ltd, South Brisbane, Queensland, Australia), the base calling was verified and primers were deleted. Consensus/assembled sequences were inserted into the alignments of the sequences downloaded from GenBank. The NJ trees were generated again in order to assess the placement of the market sample sequences within the trees (in Figs. 1–5 indicated with the symbol of a small fish). A nucleotide BLAST search in GenBank (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was performed with the identified DNA sequences from the market sample as query sequences on September 10th, 2019 and on October 30th (only *rag1* sequences). The maximum number of aligned sequences to display was set at 1000 and the hits were sorted by identity. All species were recorded, which produced hits with more than or equal to 98% identities.

Table 2
PCR cycling conditions. All PCRs were performed with 35 cycles of denaturation, primer annealing and elongation.

Genetic marker	Initial denaturation	Denaturation	Primer annealing	Elongation	Final elongation
COI	3 min, 95 °C	60 s, 94 °C	60 s, 54 °C	60 s, 72 °C	–
cytb	3 min, 95 °C	30 s, 94 °C	30 s, 50 °C	45 s, 72 °C	–
16S rDNA	3 min, 95 °C	30 s, 94 °C	30 s, 50 °C	45 s, 72 °C	–
12S rDNA	3 min, 95 °C	30 s, 94 °C	30 s, 55 °C	45 s, 72 °C	–
rag1	2 min, 95 °C	45 s, 94 °C	45 s, 62 °C	60 s, 72 °C	5 min, 72 °C

An attempt was also made to investigate the internal spacer 1 (ITS1), the first part of the ITS region. Unfortunately, the ITS1 sequence of the market sample could not be determined, neither with published primers (Chow, Nakagawa, Suzuki, Takeyama, & Matsunaga, 2006) nor with in-house primers designed based on the published snapper ITS sequences. Although several PCR conditions had been tested (different annealing temperatures, cycle numbers, $MgCl_2$ concentrations) resulting in different numbers and sizes of amplicons, an amplicon with the expected length could not be generated (data not shown).

3. Results

A dataset with FASTA files for the alignments of all considered sequences and files for the complete uncondensed and the condensed NJ trees (in the.mts format from MEGA) for all investigated genetic markers in this study is available at Mendeley Data (Kappel & Schröder, 2020).

3.1. COI sequences

1961 COI sequences belonging to Lutjanidae species were found in GenBank with a combination of the search terms 'cox1' and 'COI' (see above). Since there is an inconsistency in naming of genes in GenBank, it is likely that some entries were missed, but *cox1* and COI seem to be the most common annotations for cytochrome *c* oxidase subunit I gene in fishes. After removal of short sequences, sequences of bad quality and sequences from different parts of the gene, the final alignment comprised 1737 sequences and 480 nucleotide positions. No stop codons were present.

After a first inspection of the resulting COI NJ tree (data not shown), additional COI sequences belonging to species which were found in mixed clusters with *L. malabaricus*, *L. bohar* or *L. erythropterus* were downloaded from the FDA RSSL for Seafood Identification, in particular: two *L. bohar* sequences, one *L. malabaricus* sequence, three *L. lutjanus* sequences, two *L. rivulatus* sequences, one *L. fulvus* sequence, one *L. gibbus* sequence and three *L. argentimaculatus* sequences. No FDA

Table 1
PCR primers used for the analysis of the 'Red Snapper' market sample. Lower case letters indicate M13-tails, which served as binding sites for the sequencing primers in the subsequent sequencing reaction.

Genetic marker	Primer	Sequence (5'- > 3')	Fragment length	Reference
COI	VF2_t1	tgtaaaacgacggccagTCAACCAACCACAAAGACATTGGCAC	655 bp	CEN (2019)
	FishF2_t1	tgtaaaacgacggccagTCGACTAATCATAAAGATATCGGCAC		
	FishR2_t1	caggaaacagctatgacACTTCAGGGTGACCGAAGAATCAGAA		
	FR1d_t1	caggaaacagctatgacACCTCAGGGTGTCCGAARAAYCARAA		
cytb	L14735	ccagggttttcccagtcacgAAAAACCACCGTTGTTATTCAACTA	415 bp	CEN (2019)
	H15149ad	cggataacaatttcacacaggGCICCTCARAATGAYATTTGTCCTCA		
16S rDNA	16sar L	ccagggttttcccagtcacgCGCCTGTTTATCAAAAACAT	576 bp	Palumbi et al. (2002)
	16sbr H	cggataacaatttcacacaggCCGGTCTGAACTCAGATCACGT		
12S rDNA	MiFish-U-F teleo_R	GTCGGTAAAACTCGTGCCAG	665 bp	Miya et al. (2015)
		CTTCCGGTACACTTACCATG		Valentini et al. (2016)
rag1	RAG1F	AGCTGTAGTCAGTAYCACAARATG	~ 1600 bp	Quenouille, Bermingham, and Planes (2004)
	RAG9R	GTGTAGAGCCAGTGRTGYTT		

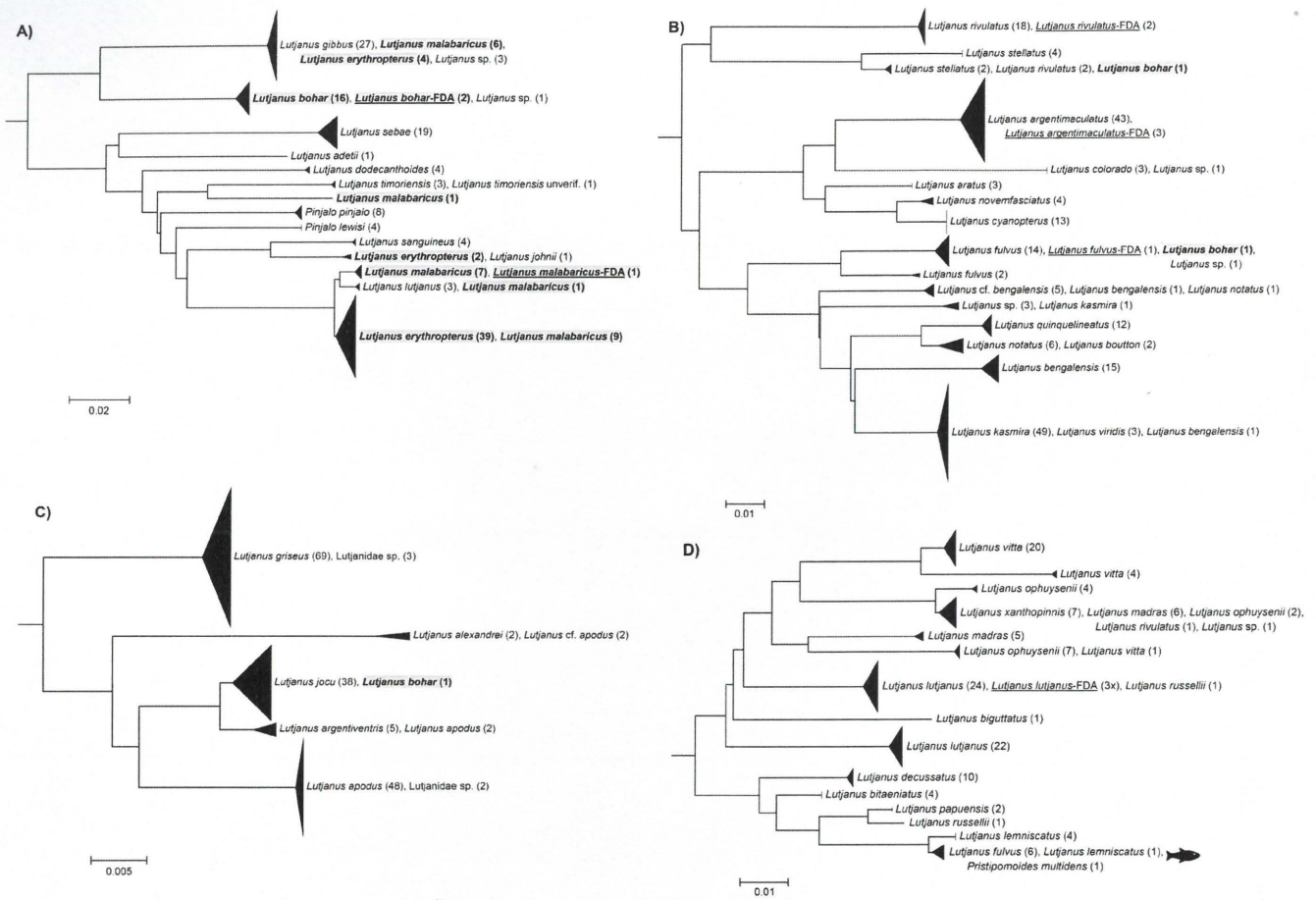

Fig. 1. All COI NJ subtrees containing *L. malabaricus* (A), *L. erythropterus* (A), *L. bohar* (A–C) and *L. fulvus* (B, E) sequences. The subtrees were extracted from the NJ tree containing 1737 snapper (Lutjanidae) sequences from GenBank and selected sequences from the FDA RSSL for Seafood Identification. The K2P distances (Kimura, 1980) were calculated in MEGA7 (Kumar et al., 2016) for a 480 bp segment and are displayed in the units of the number of base substitutions per site. The numbers of the sequences for the given species are specified in parentheses. The Malabar blood snapper, the crimson snapper, and the two-spot red snapper sequences are highlighted in grey. The sequences from the FDA RSSL for Seafood Identification are underlined. The placement of the 'Red Snapper' market sample is indicated with the symbol of a small fish. The complete tree can be accessed in the Mendeley Data (Kappel & Schröder, 2020).

entries existed for the following species: *L. erythropterus*, *L. jocu*, *L. stellatus*, *L. sanguineus* and *L. johnii*. The downloaded FDA sequences were integrated in the phylogenetic tree. As the NJ tree was too large for presentation on one page, only the subtrees which include *L. malabaricus*, *L. bohar*, *L. erythropterus* and *L. fulvus* sequences are displayed in Fig. 1 A-D.

In the complete COI NJ tree, 112 separate clusters were identified by visual inspection, which could not be further divided into individual clusters representing separate species. Of all these 112 separate clusters, 28 (25%) clusters were mixed and contained more than one species (see condensed COI tree in the Mendeley Data). Higher taxa (e.g. Lutjanidae sp. or *Lutjanus* sp.) or taxa indicated as uncertain taxonomic assignments (e.g. *Lutjanus russellii* unverified, *Caesio* cf. *caerulaurea*, *Lutjanus* aff. *johnii*) were not considered.

The 24 Malabar blood snapper COI sequences from GenBank were all retained in the data set when the alignment was reduced to the chosen length and are present in the subtree in Fig. 1 A. They are distributed over five branches and share clusters with sequences annotated as *L. gibbus*, *L. erythropterus* (two clusters) or *L. lutjanus*. One cluster exclusively features *L. malabaricus* sequences including the reference sequence for *L. malabaricus* from the FDA. The 19 two-spot red snapper COI sequences from GenBank are represented in four clusters in the COI NJ trees (see Fig. 1A–C). In addition to three single sequences grouped with other *Lutjanus* species, one cluster only represents *L. bohar* sequences and this cluster also features the FDA *L. bohar* reference sequence. The distribution of crimson snapper sequences within the COI NJ tree has a similar appearance compared to *L. malabaricus*. *L. erythropterus* sequences appear in three mixed clusters (Fig. 1 A) and are grouped together with *L. malabaricus* sequences, *L. gibbus* sequences or *L. johnii* sequences.

A similar ambiguous situation was found in BOLD, where genetically identical taxa are indexed in a regimented BIN system, in which each BIN reflects a taxonomic operational unit and thus corresponds closely to a particular species. All BINs containing Malabar blood snapper, two-spot red snapper or crimson snapper specimens were composed in Table 3. In BOLD, *L. malabaricus* is grouped together in one BIN with *L. erythropterus*, in another BIN with *L. erythropterus* and *L. lutjanus*, and in yet another BIN with *L. erythropterus* and *L. gibbus*. The crimson snapper is found in three separate BINs with mixed species annotations (together with *L. malabaricus*; with *L. malabaricus* and *L. lutjanus*; with *L. gibbus* and *L. erythropterus*) and an additional single *L. erythropterus*-specimen forms a fourth BIN. *L. bohar* splits into three BINs, with one BIN containing 34 *L. bohar* specimens with only one other unidentified *Lutjanus* specimen and two BINs with only one *L. bohar* specimen among other species (*L. fulvus*; *L. stellatus* and *L. rivulatus*).

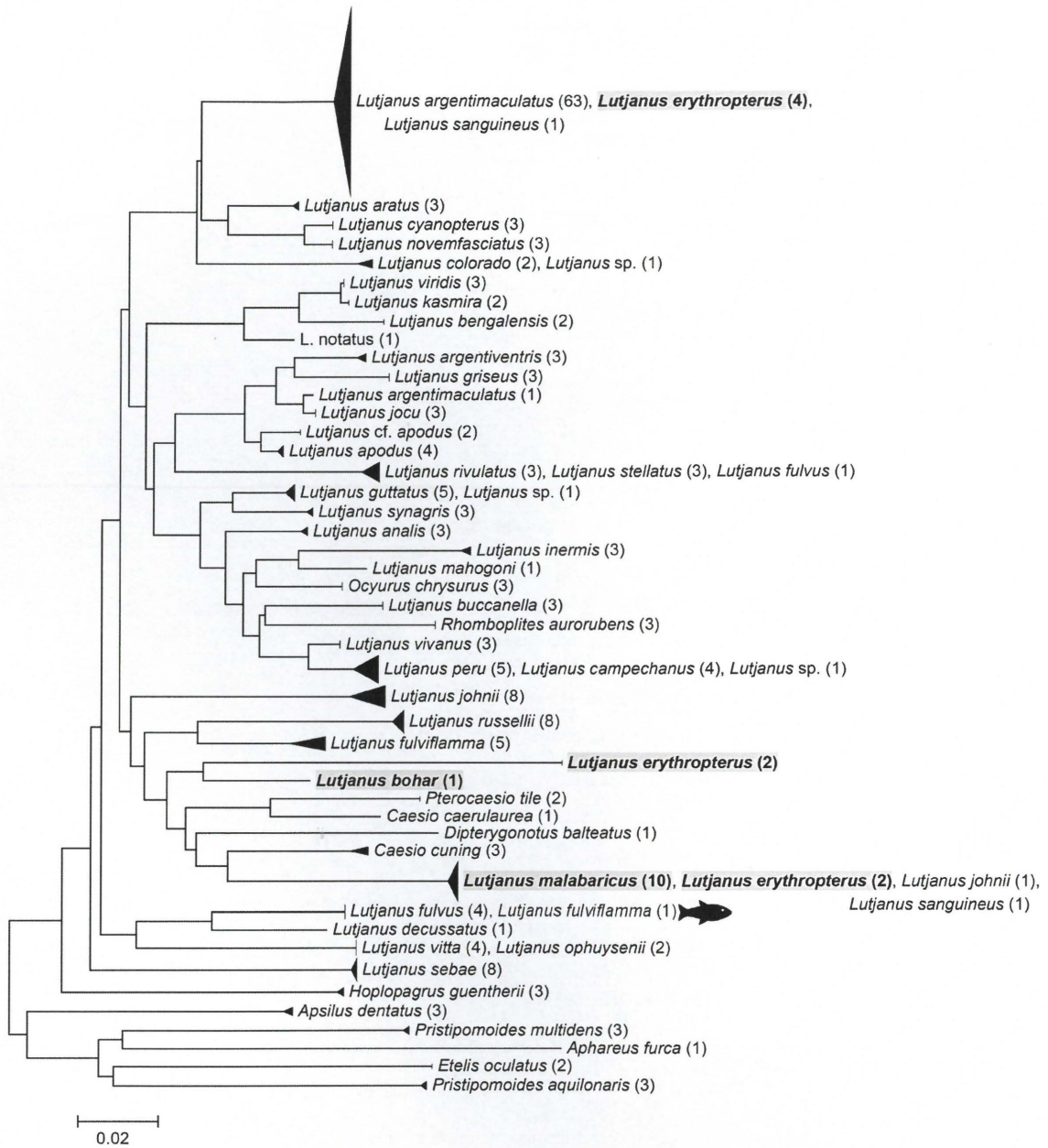

Fig. 2. NJ tree for a 357 bp *cytb* segment of 228 snapper sequences from GenBank. The genetic distances were calculated in MEGA7 (Kumar et al., 2016) using the K2P method (Kimura, 1980) and are in the units of the number of base substitutions per site. The Malabar blood snapper, the crimson snapper and the two-spot red snapper sequences are highlighted in grey and the placement of the 'Red Snapper' market sample is indicated with the symbol of a small fish.

3.2. Cytb sequences

With 'cytb' as a search term, 784 sequences for Lutjanidae entries were found in GenBank. The final alignment comprised 228 sequences and spanned 357 nucleotide positions. No stop codons were found. The NJ tree is shown in Fig. 2. From the 46 clusters, 6 clusters (13%) represented more than one species. The ten available *L. malabaricus* sequences from GenBank were all retained in the tree and formed a single mixed cluster with *L. erythropterus*, *L. johnii* and *L. sanguineus* sequences, whereas crimson snapper sequences were also present in two additional branches. The single two-spot red snapper sequence displayed a single branch in the tree.

3.3. 16S rDNA sequences

In GenBank, 189 sequences were found for Lutjanidae species with the search term '16S'. Again, it cannot be excluded that some sequences were missed because of varying annotations of the gene. Nevertheless, a search within all available sequences for the three snapper species in question did not reveal any additional sequences that had been missed with the search strategy. Short sequences were discarded and stretches of 'Ns' were deleted from two sequences, as they introduced large gaps into the alignment. The final alignment comprised 154 nucleotide sequences and 381 positions. Six (12.8%) of the 47 identified clusters were mixed. The only two Malabar blood snapper 16S rDNA sequences formed a mixed cluster with three of the four available crimson snapper sequences (see Fig. 3). In fact, the two *L. malabaricus* sequences

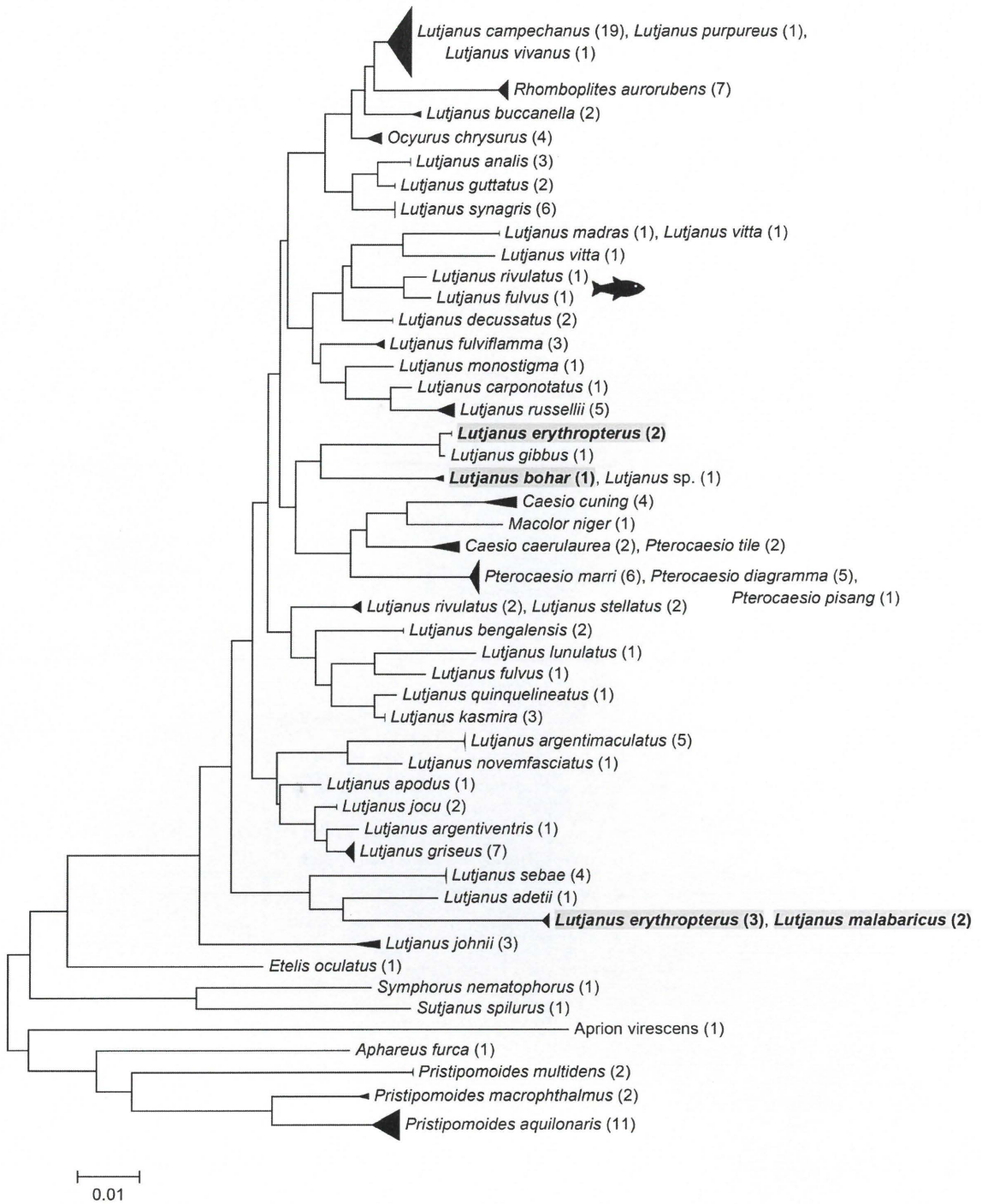

Fig. 3. 16S rDNA NJ tree calculated from the alignment of snapper sequences from GenBank comprising 154 nucleotide sequences and 381 positions. The genetic distances were computed using the K2P method (Kimura, 1980) and are in the units of the number of base substitutions per site. The NJ tree was generated with MEGA7 (Kumar et al., 2016). The Malabar blood snapper, the crimson snapper and the two-spot red snapper sequences are highlighted in grey and the placement of the 'Red Snapper' market sample is indicated with the symbol of a small fish.

represented a single specimen, because one of the sequences was the reviewed reference sequence for the complete mitochondrial genome, curated by the NCBI staff, and was derived from the other available sequence. The same applies to two of the five *L. erythropterus* sequences that did not cluster with the *L. malabaricus* sequences. The only occurring two-spot red snapper 16S rDNA sequence represented a separate branch with one other sequence that was only labelled with the genus *Lutjanus* (sp.).

3.4. 12S rDNA sequences

The search among Lutjanidae sequences with the term '12S' revealed 176 sequences in GenBank. As different *L. malabaricus* sequences were present which did not overlap, two separate alignments were prepared. The first alignment represented 127 sequences and comprised 179 nucleotide positions, whereas the second alignment had only 62 sequences but 403 positions. The corresponding NJ trees are displayed

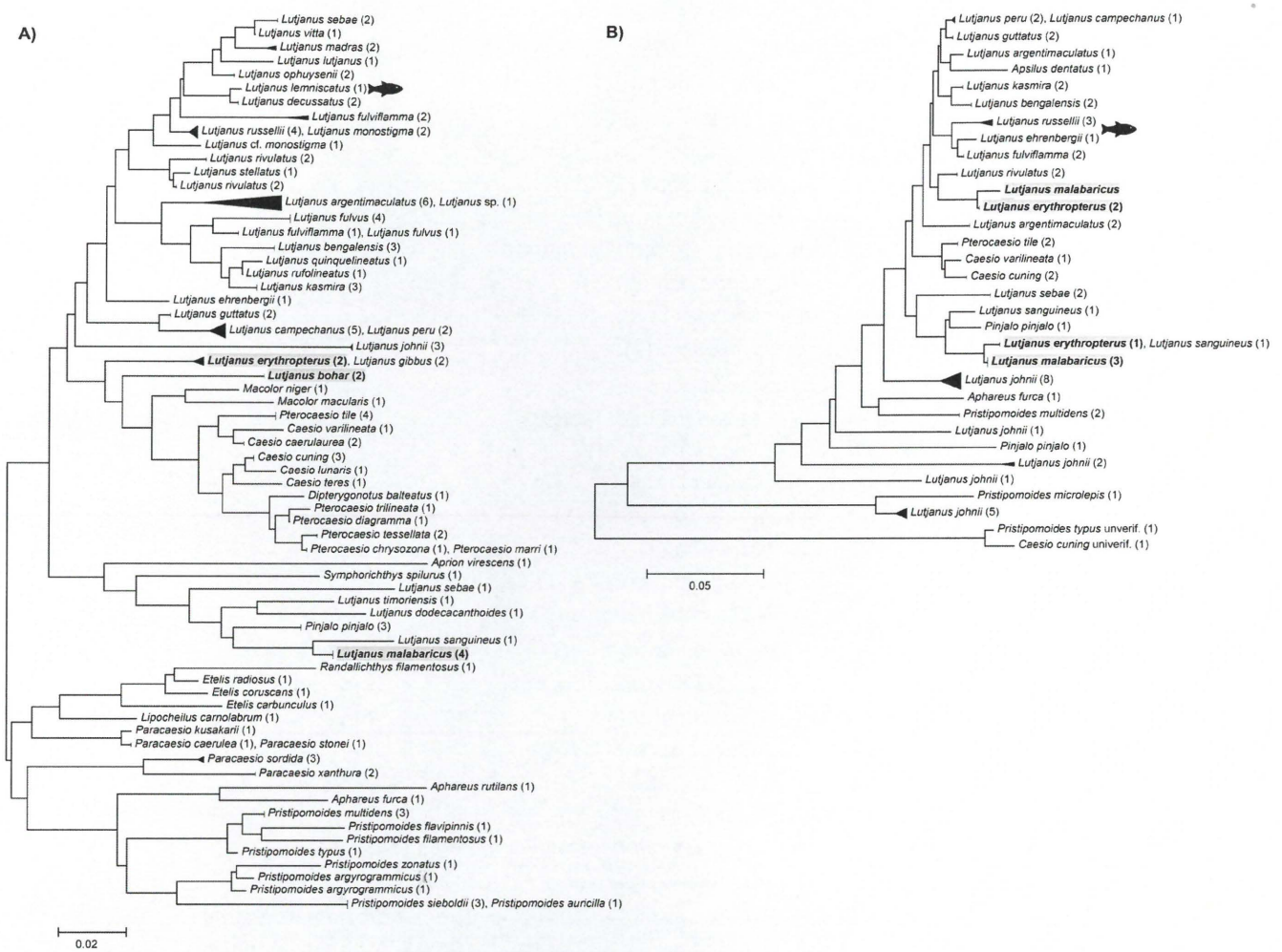

Fig. 4. 12S rDNA NJ trees of snapper sequences from GenBank. The first tree (A) was calculated from an alignment with 179 sequences from a 127 bp segment and the second tree (B) was calculated for 62 sequences and a 403-nucleotide segment. The NJ trees were computed in MEGA7 (Kumar et al., 2016) using K2P distances (Kimura, 1980), that are in the units of the number of base substitutions per site. The Malabar blood snapper, the crimson snapper and the two-spot red snapper sequences are highlighted in grey and the placement of the 'Red Snapper' market sample is indicated with the symbol of a small fish.

in Fig. 4. Six (9.1%) out of 66 clusters and two (6.3%) out of 32 clusters were mixed in the two trees, respectively. In the first tree (179 nt), there are four Malabar blood snapper sequences in a cluster without sequences from any other species. In the second tree (403 nt), the four *L. malabaricus* sequences are split up into two clusters both of which are very similar to the neighbouring clusters with *L. erythropterus* sequences. The two crimson snapper sequences (in fact only one specimen, see above) from the first alignment are grouped with *L. gibbus* sequences. In the second tree, the crimson snapper sequences are in close proximity to the Malabar blood snapper sequences. The two two-spot red snapper sequences in the first NJ tree form a separate branch without other species.

3.5. *rag1* sequences

For the nuclear gene encoding the recombination-activating gene 1 (*rag1* or RAG1), 35 sequences were downloaded from GenBank with the search term 'rag1'. One sequence (JN106041.1) could not be matched to the other sequences and on closer inspection this sequence indeed represented a rag2 sequence. The reason why this sequence was also downloaded with the import filter is that the term rag2 appeared in a line of text of this particular entry which read 'Cloning and expression analysis of recombination activating genes (RAG1/2) in red snapper (*Lutjanus sanguineus*)'. The remaining sequences could be aligned. One sequence was quite long (3944 nucleotides). The majority of the sequences were about 1400 nucleotides in length, but two sequences comprised only 752 nucleotides. The alignment was truncated to 726 nucleotide positions and displayed 14 ambiguities. Two sequences exhibited a gap with 31 nucleotides each, which was quite unexpected in a protein coding gene and can only be properly explained by a sequencing artefact. Most species were only represented by one sequence, and only *L. malabaricus* and *L. sanguineus* were present in a mixed cluster in the NJ tree (see Fig. 5). However, the sequences were very similar to the one occurring *L. erythropterus* sequence, which differed only by one nucleotide and one additional ambiguous position.

3.6. *ITS sequences*

The internal transcribed spacer (ITS) region represents a common genetic marker for the identification of plant or fungi specimens at species level (Wang et al., 2015), but for snappers there were 29 ITS sequences in GenBank as well. Four sequences only spanned the ITS1 region, however only two of these four sequences could be aligned to the sequence collection. The other two sequences were discarded, as they were also quite short so that the final alignment comprised 23 sequences and 919 nucleotide positions. The NJ tree (see Fig. 6)

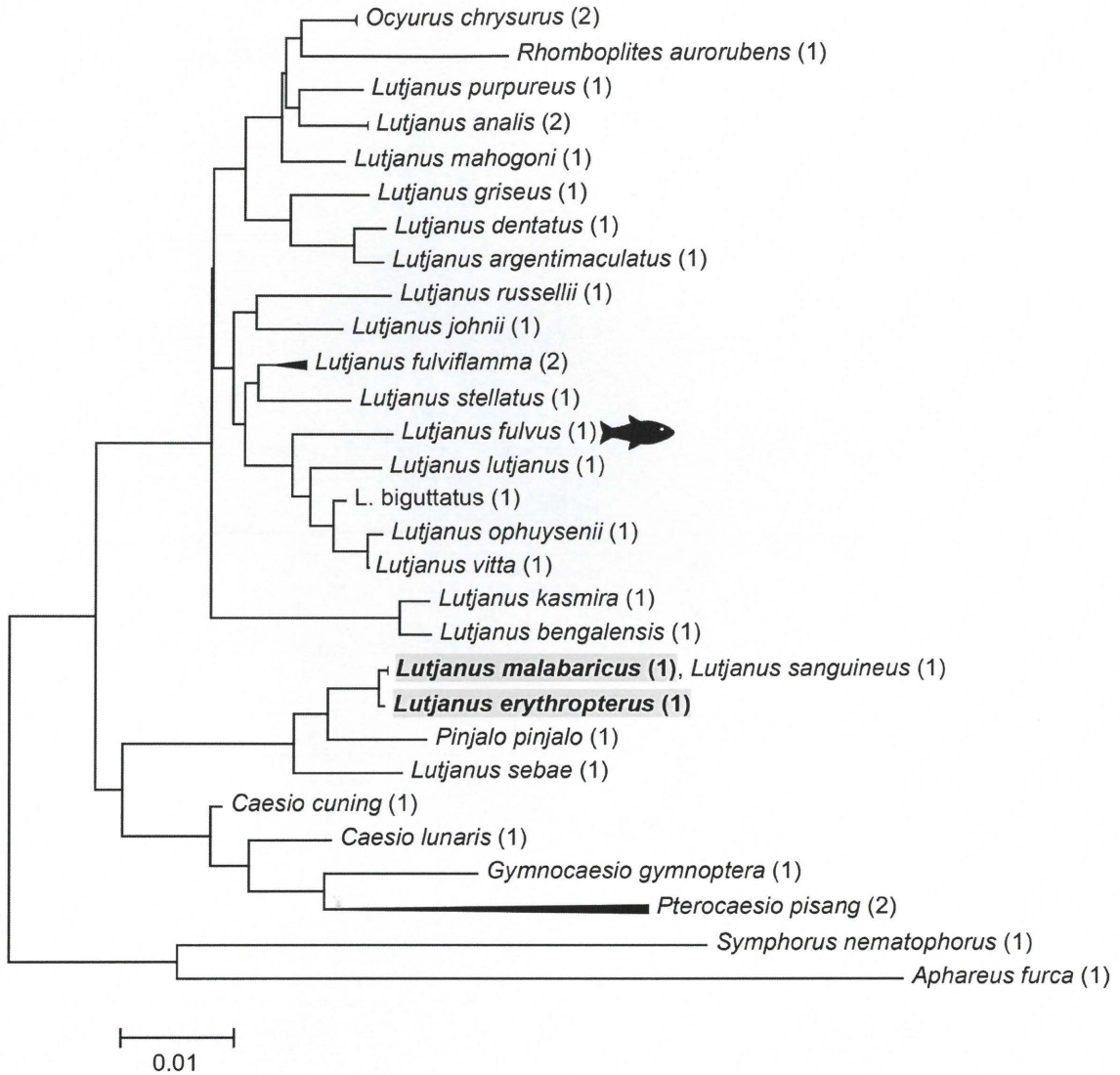

Fig. 5. *rag1* NJ tree of snapper sequences from GenBank. The alignment used for building the tree comprised 34 sequences and was truncated to 726 nucleotide positions. The genetic distances were computed in MEGA7 (Kumar et al., 2016) using the K2P method (Kimura, 1980) and are in the units of the number of base substitutions per site. The Malabar blood snapper and the crimson snapper sequences are highlighted in grey. There was no *rag1* sequence for the two-spot red snapper in GenBank yet. The placement of the 'Red Snapper' market sample is indicated with the symbol of a small fish.

Table 3
All BOLD BINs containing species annotations of *L. erythropterus*, *L. malabaricus*, *L. bohar* and *L. fulvus* as of September 10th, 2019. The number of specimens with the annotation of the particular species are given in parentheses. The fish symbol indicates the BIN in which the market sample was grouped.

BIN	Species
BOLD:AAA7594	*L. erythropterus* (22); *Lutjanus* sp. (3); *Lutjanus malabaricus* (2)
BOLD:AAA7595	*L. erythropterus* (40); *L. malabaricus* (30); *L. lutjanus* (5); *Lutjanus* sp. (4)
BOLD:AAB3276	*L. gibbus* (66); *L. malabaricus* (6); *L. erythropterus* (1)
BOLD:AAB4501	*L. bohar* (34); *Lutjanus* sp. MS-2015 (1)
BOLD:AAB4502	*L. fulvus* (45); *Lutjanus* sp. (1); *L. bohar* (1)
BOLD:AAB4503	*L. stellatus* (4); *L. rivulatus* (2); *L. bohar* (1)
BOLD:AAB7014	*L. lemniscatus* (18); *L. fulvus* (13); *Pristipomoides multidens* (1)
BOLD:AAB7015	*L. fulvus* (5)
BOLD:ACS0817	*L. erythropterus* (1)
BOLD:AAB7015	*L. fulvus* (5)

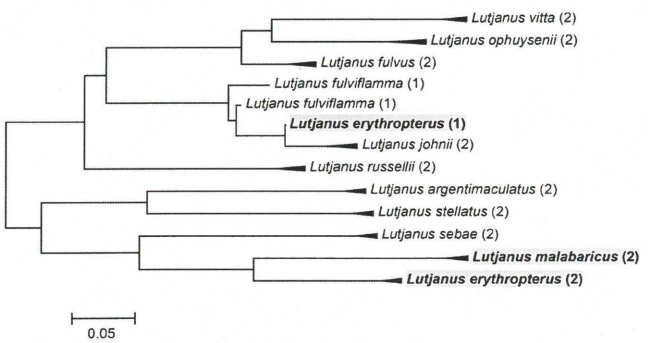

Fig. 6. NJ tree with ITS GenBank sequences from snappers. The tree was built in MEGA7 (Kumar et al., 2016) using K2P distances (Kimura, 1980) with 23 sequences for a 919 bp-segment. The distances are in the units of the number of base substitutions per site. ITS sequences for the two-spot red snapper were not present in GenBank. The Malabar blood snapper and the crimson snapper sequences are highlighted in grey. The attempts to determine the ITS sequence of the 'Red Snapper' market sample were unsuccessful.

exhibited 12 species, most of them represented by two sequences. The two Malabar blood snapper sequences were in agreement with one another, but of the three crimson snapper sequences one sequence differed substantially from the other two sequences. A two-spot red snapper ITS sequence was not present in GenBank.

3.7. Authentication of a red snapper market sample

The COI, *cytb* and 16S rDNA sequence raw data generated from the market sample were of very good quality and consensus sequences could be determined for the complete amplicons. The first 170 nucleotides of the forward strand of the 12S rDNA fragment could not be determined, but otherwise both strands were of good quality and the complete sequence of the amplicon could be assembled. After sequencing the *rag1* amplicon, a 1196 bp sequence could be assembled from the forward and the reverse strand, which covered the complete alignment of the database *rag1* sequences. The identified *rag1* sequence contained two ambiguous bases, which were determined in the forward strands of two independent sequencing reactions and showed electropherogram peaks with similar heights for the two bases A and G. Because of the reproducibility of this ambiguity and the similarities of the peak heights, it is quite likely that these positions represent heterozygous sites, which are not unlikely in nuclear genes. Several attempts to produce an ITS1 amplicon with the expected length and a reasonable DNA sequence were unsuccessful.

None of the determined sequences yielded a hit with more than or equal to 98% identity to any sequence with the annotation of *L. malabaricus* (see Table 4) in the BLAST search. When initially analysing the BLAST result of *cytb*, the sample seemed likely to be a blacktail snapper (*L. fulvus*), as it exhibited 99.28% to 99.75% identity to five *L. fulvus* entries. An additional *L. fulviflamma* hit with 99.75% identity seemed to be falsely annotated because eight further *L. fulviflamma* entries exhibited less than 93% identity to the sample's *cytb* sequence. This is also reflected by the *cytb* NJ tree in Fig. 2. However, the result for the database search with the sample's COI sequence was less clear: again, six hits of *L. fulvus* sequences with 99.38% to 100% identity were produced, but 14 more *L. fulvus* sequences showed less than 90% identity, which was also true for the FDA reference sequence for *L. fulvus* and can be seen in Fig. 1 B and D. Another conceivable species assignment on the basis of COI and 12S rDNA sequences could also be made to *L. lemniscatus* (yellowstreaked snapper), for which unfortunately neither *cytb*, 16S rDNA, *rag1* nor ITS sequences have been published so far. The assignment to *L. lemniscatus* is also supported by BOLD (see Table 3), although the relevant BOLD BIN comprises *L. lemniscatus* sequences as well as *L. fulvus* sequences. The 16S rDNA sequence as well as the *rag1* sequence do not add further details for clarification (see Figs. 3 and 5). On the whole, a valid species assignment for the respective market specimen sold under the scientific name '*Lutjanus malabaricus*' seems to be impossible.

4. Discussion

The marker with by far the most sequences for snapper species in GenBank was COI. This is most likely due to the Barcode of Life initiative. Nevertheless, *cytb* has also been a common marker for the identification of fish samples at species level for many years (see for example Quinteiro et al., 1998; Rehbein, Köppel, & Hankeln, 2012; Sotelo et al., 2001). Together, COI and *cytb* represent a valuable tool for species identification of fish samples and a technical specification has been published just recently on the use of both of them for the authentication of fish products (CEN, 2019). 16S rDNA and 12S rDNA, both also mitochondrial markers, have been analysed significantly less. Nevertheless, 16S rDNA is frequently used for phylogenetic studies, as it can quite well resolve higher phylogenetic relationships such as those of subfamilies or families (Miller & Cribb, 2007). Nuclear genes are rarely used for species assignments, as they are usually more conserved compared to mitochondrial markers. However, some nuclear regions are highly variable like the internal transcribed spacers between the small-subunit RNA and the large-subunit rRNA genes. Indeed, Wang et al. (2015) claim that the ITS region is a good candidate for a universal DNA barcode in eukaryotes. However, few studies on ITS sequences of fish have been published so far (see for example Chow et al., 2006; Guo, Su, Zhang, Ding, & Wang, 2009; Pérez, Vieites, & Presa, 2005). The attempts to determine the ITS1 sequence of the market sample in this study were unsuccessful and the reasons for this are not resolved. The primers might not match the ITS region of the specimen. Since the specimen could not be identified at species level, it cannot be known for sure whether a sequence for the species in question has already been published. Consequently, this possibility cannot be excluded. Another reason might be that the ITS PCR might require very precise PCR conditions, which had not been applied in this study. It could also be that the primers used here bind more efficiently to other genome regions of the species concerned. In this case the amplification of ITS1 may have been impeded. In any case, it would be most useful to know which primers and PCR conditions were used when generating the ITS sequences of snapper species available in GenBank.

The best genetic markers only lead to valid species assignments if the reference sequences are based on specimens with authentic species annotations. In the case of the Malabar blood snapper or the crimson snapper, hardly any of the investigated genetic markers are unambiguous nor are the BOLD BINs. The sequences are either found in distant sequence clusters in the NJ trees or form mixed clusters with the sequences from other species or both. Only the ITS region seems to give a clear picture for the Malabar blood snapper with the placement of two *L. malabaricus* sequences in one cluster well apart from the other species' clusters. Nevertheless, since these are only two sequences from one working group, it is most likely that other groups might define another species as the Malabar blood snapper. Strangely enough, one of the three crimson snapper ITS sequences differs significantly from the other two sequences, despite the fact that it was deposited by the same working group (albeit two years earlier than the other two sequences). Unfortunately, no publication which could explain this ambiguity was found in connection with these GenBank entries.

It is noticeable that the sequences for the Malabar blood snapper and the crimson snapper are often found together in mixed clusters of the NJ trees. This situation is also reflected in the BOLD BINs. Additionally, both of them are also mixed with sequences from other species, like *L. gibbus*, *L. johnii*, *L. lutjanus*, *L. sanguineus* or *L.*

Table 4
GenBank BLAST result for the 'Red Snapper' market sample. The 1000 BLAST hits were sorted by sequence identity and all species with ≥98% identity were recorded. The highest identities for the given species are indicated in parentheses.

Genetic marker	Species with hits ≥98% identity
COI	*L. fulvus* (100%); *L. lemniscatus* (100%); *Champsodon snyderi* (99.68%); *Pristipomoides multidens* (99.53%)
cytb	*L. fulviflamma* (99.75%); *L. fulvus* (99.75%)
16S rDNA	*L. fulvus* (99.09%); *L. decussatus* (98.61%); *L. madras* (98.18%)
12S rDNA	*L. lemniscatus* (99.77%); *L. decussatus* (99.10%)
rag1	*L. fulvus* (99.67%); *L. biguttatus* (99.21%); *L. vitta* (98.91%); *L. lutjanus* (98.91%); *L. ophuysenii* (98.83%); *L. fulviflamma* (98.66%); *L. johnii* (98.24%); *L. stellatus* (98.16%); *L. analis* (98.01%)

argentimaculatus. Obviously, these are all species which occur in the Indo-West Pacific or in the Western Indian Ocean (see https://www.fishbase.org). Interestingly, the respective DNA sequences excerpted from the *L. erythropterus* NCBI complete mitochondrial references sequence (NC_031331) are clustered with *L. gibbus* sequences in the COI tree, the 16S rDNA tree and one of the 12S rDNA trees but not with the majority of the other *L. erythropterus* sequences. The only *L. gibbus cytb* sequence from GenBank was not retained in the final *cytb* alignment because it did not cover the first part of it and is thus also not present in the *cytb* tree. But nevertheless, the *L. erythropterus* sequence from the complete mitochondrial genome sequence is 100% identical (albeit of course longer) compared to the *L. gibbus cytb* sequence. A misidentification of the specimen used for the determination of the complete mitochondrial reference sequence of the crimson snapper thus appears likely. The *L. malabaricus* reference sequence of the FDA and the NCBI reference sequence for the complete mitochondrial genome of the Malabar blood snapper are in accordance with one another. They also match the majority of the *L. malabaricus* sequences in the NJ trees of the investigated mitochondrial markers. It is therefore very likely that the complete mitochondrial reference sequence in GenBank (NC_012736) reflects the true Malabar blood snapper.

In GenBank, far fewer sequences are annotated as *L. bohar* compared to the other two investigated snapper species and the sequences are quite unambiguous: although the sequences for *L. bohar* split up into four distant clusters in the COI NJ tree and into three BINs in BOLD, the majority of the sequences are found in one tree cluster or in one BIN, respectively. Only single sequences are displayed in other clusters or BINs, likely reflecting misidentified specimens.

The sequence data are not only confusing for the Malabar blood snapper and the crimson snapper; a similar situation was found when checking the sequence data for the blacktail snapper (*L. fulvus*) as shown by the analysis of the market sample annotated as *L. malabaricus* in this study. The market specimen had clearly been mislabelled but could not be assigned to a particular species with certainty though the sequences clustered with *L. fulvus* sequences in some of the NJ trees. The closer inspection of the blacktail snapper sequences showed that they are split up into several clusters in the NJ trees as well as in three BOLD BINs.

The possible reasons for the confusing DNA sequence data for snapper species in GenBank and BOLD are manifold. First of all, mislabelling or insufficient labelling at the catch level has been reported. The Malabar blood snapper and the crimson snapper seem to be mislabelled regularly, especially because the juvenile and adult specimens vary in colouration (Bakar et al., 2018; Galal-Khallaf et al., 2019; Marko et al., 2004) and form mixed schools (Allen, 1985). As stated in the FAO species catalogue, the crimson snapper has been referred to as *Lutjanus malabaricus* (non Schneider) or *L. altifrontis* by many recent authors and the Malabar blood snapper is frequently misidentified as *L. sanguineus* (Allen, 1985). The speciation of several snappers might have occurred quite recently as it has been described for *L. campechanus* and *L. purpureus* (Pedraza-Marrón et al., 2019); two species which also show mixed sequence clusters in the NJ trees, especially in the COI tree (see dataset at Mendeley Data). Hajibabaei, Singer & Hickey state that species, which are difficult to resolve by conventional taxonomic methods, will also be a challenge for the barcoding approach (in Wang et al., 2010). There is also an overwhelming occurrence of cryptic diversity in taxa of the Indo-Pacific Ocean (Hubert et al., 2012), which might explain that DNA sequences with the annotation of a given species vary considerately and are placed at different locations in the NJ trees. Introgressive hybridisation and artificial breeding and hybridisation (Guo, Wang, Liu, Liu, & Liu, 2007) may also be a reason for the ambiguous placement of snapper species in the NJ trees. For instance, artificial breeding has been described for some high-value species, such as *L. sebae*, *L. argentimaculatus*, *L. malabaricus* and *L. erythropterus* (Guo et al., 2007). Since the mitochondrial DNA is maternally inherited, misidentified specimens may also represent hybrids, which have not been identified as such. In order to solve this, a combination of a mitochondrial marker and a nuclear marker should be used for the DNA-based barcoding of species, which are known to hybridise. Last but not least, single NCBI sequences differing substantially from the majority of sequences with the same species annotation as well as stand-alone BOLD BINs may indicate non-functional gene copies (nuclear-mitochondrial pseudogenes, NUMTs) (Phillips, Gillis, & Hanner, 2018). Nevertheless, this seems unlikely for the investigated sequences of the protein-coding genes COI, *cytb* and *rag1*, since stop codons, which are a typical sign for NUMTS, have not been identified in the sequences.

In order to allow investigative bodies to make valid statements about the authenticity of Malabar blood snapper, crimson snapper or blacktail snapper products, further research is necessary with regard to the genetic marker sequences of these species. However, this research is significantly hampered by the Nagoya Protocol on Access and Benefit Sharing (ABS) (see https://www.cbd.int/abs). This agreement – as important and useful as it undoubtedly is – does not allow scientists to generate and publish new DNA sequences from specimens of foreign countries without the explicit permission from the countries where the specimens originated. Yet, the procedures for obtaining authorization to use genetic material are often very time-consuming and need high administrative efforts. This hinders the work of scientists or even makes it impossible, as project deadlines must be respected. Further work on resolving the confused barcode data of particular snapper species can only be conducted by or in close collaboration with scientists from the countries of origin. Therefore, calls for international research funding should be initiated, in which experts from the different countries of origin of the snappers in question could work together on solving the problem. As a further parameter to improve the traceability and authenticity system in the global seafood trade, the CODEX standards (Codex Alimentarius international food standards) for fishery products, for instance the CODEX STAN 36-1981 ('CODEX Standard for quick frozen fin fish, uneviscerated and eviscerated') (Codex Alimentarius Commission, 1981) should be refined. A stricter labelling of the unique scientific name should be mandatory along the entire production chain and should not solely depend on the country where the fish product is sold. This change in standards could lead to a better understanding of the need for scientific names of fish species and stimulate more research on this issue.

5. Conclusion

This study reveals the ambiguity and inconsistency of nucleotide sequence data for some snapper species, in particular the Malabar blood snapper, the crimson snapper and the blacktail snapper in GenBank and BOLD, making valid DNA barcoding-based authentications of snapper products at species level difficult. The official food control laboratories and service laboratories should be aware that GenBank BLAST and BOLD results for snapper products should be treated with caution. The query results may appear quite clear, although data for a given species can be inconsistent due to the presence of multiple sequence entries with significant sequence variations. In this context, it is of utmost importance to carefully inspect the data available in the databases before the query results are assessed. Knowledge regarding the variation and concordance of the sequences for the species in question is essential. The GenBank BLAST results are often difficult to interpret, because they do not indicate whether all sequences that belong to the species represented by the best hits are unambiguous. Combining a BLAST search with a phylogenetic tree building approach therefore seems well suited to better judge the validity of the species assignments. Obviously, the data for the NJ tree building must be updated regularly, but for a first test of this approach the FASTA files from the dataset at Mendeley Data can be used. It is also recommended to analyse more than one marker. *Cytb* and COI seem to be the most suitable marker genes, because of the larger number of sequences in GenBank compared to the other genetic markers. This demonstration of ambiguous barcoding

data for snapper species can be seen as one example among others. Similar situations have been found when authenticating specimens from other fish families, such as parrotfishes, mullets, groupers etc., but also seafood samples, like crustaceans, molluscs or cephalopods (data not shown).

CRediT authorship contribution statement

Kristina Kappel: Conceptualization, Methodology, Investigation, Visualization, Writing - original draft. **Ute Schröder:** Writing - review & editing.

Declaration of competing interest

None.

Acknowledgements

We would like to thank Ulf Willeke for creating the computer scripts for downloading the GenBank sequences and renaming the sequences, and Phillip Mittelstedt for carrying out the laboratory work. Many thanks also to Heike Norman for revising the manuscript as a native speaker. Finally, we also thank the anonymous reviewers for their helpful comments.

References

Allen, G. R. (1985). *FAO species catalogue. v. 6: Snappers of the world. An annotated and illustrated catalogue of Lutjanid species known to date.* Rome: FAO. http://www.fao.org/docrep/009/ac481e/ac481e00.htm.

Allen, G. R., & Talbot, F. H. (1985). Review of the snappers of the genus *Lutjanus* (Pisces: Lutjanidae) from the Indo-Pacific, with description of a new species. *Indo-Pacific Fishes, 11*, 1–87.

Bakar, A. A., Adamson, E. A. S., Juliana, L. H., Nor Mohd, S. A., Wei-Jen, C., Man, A., et al. (2018). DNA barcoding of Malaysian commercial snapper reveals an unrecognized species of the yellow-lined *Lutjanus* (Pisces:Lutjanidae). *PloS One, 13*(9), e0202945.

Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Res. 42*(D1), D32–D37.

CEN (2019). *DIN SPEC 10703:2019-06 Foodstuffs - DNA barcoding of fish and fish products using defined mitochondrial cytochrome b and cytochrome c oxidase I gene segments.* Berlin: Beuth Verlag.

Chow, S., Nakagawa, T., Suzuki, N., Takeyama, H., & Matsunaga, T. (2006). Phylogenetic relationships among *Thunnus* species inferred from rDNA ITS1 sequence. *Journal of Fish Biology, 68*(A), 24–35.

Codex Alimentarius Commission (1981). Codex stan 36-1981 standard for quick frozen finfish, uneviscerated and eviscerated. Rome: Codex Alimentarius international food standards. Available online http://www.fao.org/fao-who-codexalimentarius/codex-texts/list-standards/en/.

Deeds, J. R., Handy, S. M., Fry, J. F., Granade, H., Williams, J. T., Powers, M., et al. (2014). Protocol for building a reference standard sequence library for DNA-based seafood identification. *Journal of AOAC International, 97*(6), 1626–1633.

Edwards, A., Zammit, A., & Farrell, H. (2019). Four recent ciguatera fish poisoning incidents in New South Wales, Australia linked to imported fish. *Communicable Diseases Intelligence, 43* Epub 1/02/2019.

Friedemann, M. (2019). Ciguatera fish poisoning outbreaks from 2012 to 2017 in Germany caused by snappers from India, Indonesia, and Vietnam. *J. Consumer Protect. Food Safe. 14*(1), 71–80.

Friedman, M. A., Fernandez, M., Backer, L. C., Dickey, R. W., Bernstein, J., Schrank, K., et al. (2017). An updated review of ciguatera fish poisoning: Clinical, epidemiological, environmental, and public health management. *Marine Drugs, 15*(3), 72.

Friedman, M. A., Fleming, L. E., Fernandez, M., Bienfang, P., Schrank, K., Dickey, R., et al. (2008). Ciguatera fish poisoning: Treatment, prevention and management. *Marine Drugs, 6*(3), 456–479.

Galal-Khallaf, A., Osman, A. G. M., El-Ganainy, A., Farrag, M. M., Mohammed-AbdAllah, E., Moustafa, M. A., et al. (2019). Mitochondrial genetic markers for authentication of major red sea grouper species (perciformes: Serranidae) in Egypt: A tool for enhancing fisheries management and species conservation. *Gene, 689*, 235–245.

Guo, M., Su, Y., Zhang, Z., Ding, S., & Wang, J. (2009). Differentiation of *Epinephelus moara* from *E. bruneus* by improved nest-tetra-primer-specific PCR. *Progress in Natural Science, 19*, 1221–1226.

Guo, Y. S., Wang, Z. D., Liu, C. W., Liu, L., & Liu, Y. (2007). Phylogenetic relationships of South China sea snappers (genus *Lutjanus*; family Lutjanidae) based on mitochondrial DNA sequences. *Marine Biotechnology, 9*(6), 682–688.

Hubert, N., Meyer, C. P., Bruggemann, H. J., Guérin, F., Komeno, R. J., Espiau, B., et al. (2012). Cryptic diversity in Indo-Pacific coral-reef fishes revealed by DNA-barcoding provides new support to the centre-of-overlap hypothesis. *PloS One, 7*(3), e28987.

Iwobi, A. N., Huber, I., Hauner, G., Miller, A., & Busch, U. (2011). Biochip technology for the detection of animal species in meat products. *Food Anal. Methods, 4*(3), 389–398.

Kappel, K., & Schröder, U. (2020). *Sequence alignments and neighbour joining trees of public DNA sequences with the annotation of Lutjanidae species.* Mendeley Data.

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

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution, 33*(7), 1870–1874.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics, 23*(21), 2947–2948.

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics, 30*(22), 3276–3278.

Laurent, D., Yeeting, B., Labrosse, P., & Gaudechoux, J.-P. (2005). *Ciguatera: Field reference guide. Un guide pratique. New caledonia: Secretariat of the pacific community.* Institute of Research for Developmenthttps://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers13-01/010036961.pdf.

Li, X., Shen, X., Chen, X., Xiang, D., Murphy, R. W., & Shen, Y. (2018). Detection of potential problematic *cytb* gene sequences of fishes in GenBank. *Frontiers in Genetics, 9*(30).

Marko, P. B., Lee, S. C., Rice, A. M., Gramling, J. M., Fitzhenry, T. M., McAlister, J. S., et al. (2004). Mislabelling of a depleted reef fish. *Nature, 430*, 309–310.

Miller, T. L., & Cribb, T. H. (2007). Phylogenetic relationships of some common Indo-Pacific snappers (Perciformes: Lutjanidae) based on mitochondrial DNA sequences, with comments on the taxonomic position of the Caesioninae. *Molecular Phylogenetics and Evolution, 44*(1), 450–460.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., et al. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species. *R Soc open sci, 2*, 150088.

Nei, M., & Saitou, N. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4*(4), 406–425.

Palumbi, S., Martin, A., Romano, S., McMillan, W. O., Stice, L., & Grabowski, G. (2002). *The simple fool's guide to PCR. Version 2.0.* Department of Zoology and Kewalo Marine Laboratory. University of Hawaiihttp://palumbi.stanford.edu/SimpleFoolsMaster.pdf.

Pedraza-Marrón, C.d. R., Silva, R., Deeds, J., Belleghem, S. M. V., Mastretta-Yanes, A., Domínguez-Domínguez, O., et al. (2019). Genomics overrules mitochondrial DNA, siding with morphology on a controversial case of species delimitation. *Proceedings of the Royal Society B: Biological Sciences, 286*(1900), 20182924.

Pérez, M., Vieites, J. M., & Presa, P. (2005). ITS1-rDNA-Based methodology to identify world-wide Hake species of the genus *Merluccius. Journal of Agricultural and Food Chemistry, 53*(13), 5239–5247.

Phillips, J. D., Gillis, D. J., & Hanner, R. H. (2018). Incomplete estimates of genetic diversity within species: Implications for DNA barcoding. *Ecol. Evolution*, 1–15.

Quenouille, B., Bermingham, E., & Planes, S. (2004). Molecular systematics of the damselfishes (Teleostei: Pomacentridae): Bayesian phylogenetic analyses of mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution, 31*(1), 66–88.

Quinteiro, J., Sotelo, C., Rehbein, H., Pryde, S. E., Medina, I., Pérez-Martín, R., et al. (1998). Use of mtDNA direct polymerase chain reaction (PCR) sequencing and PCR-restriction fragment length polymorphism methodologies in species identification of canned tuna. *Journal of Agricultural and Food Chemistry, 46*(4), 1662–1669.

Ratnasingham, S., & Hebert, P. D. N. (2007). Bold: The barcode of Life data system. *Molecular Ecology Notes, 7*(3), 355–364. www.barcodinglife.org.

Rehbein, H., Köppel, R., & Hankeln, T. (2012). Leitfaden für die Lebensmittelüberwachung zur Identifizierung der Fischart durch DNA-Sequenzierung von PCR-Produkten. *J. für Verbraucherschutz und Lebensmittelsicherheit, 7*, 255–260.

Reilly, A. (2018). *Overview of food fraud in the fisheries sector.* Rome: FAO. http://www.fao.org/3/I8791EN/i8791en.pdf.

Sotelo, C. G., Calo-Mata, P., Chapela, M. J., Pérez-Martín, R. I., Rehbein, H., Hold, G. L., et al. (2001). Identification of flatfish (Pleuronectiforme) species using DNA-based techniques. *Journal of Agricultural and Food Chemistry, 49*(10), 4562–4569.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., et al. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology, 25*, 929–942.

Wang, Z. D., Guo, Y. S., Tan, W., Li, L., Tang, E. P., Liu, C. W., et al. (2010). DNA barcoding, phylogenetic relationships and speciation of snappers (genus *Lutjanus*). *Science China Life Sciences, 53*(8), 1025–1030.

Wang, X.-C., Liu, C., Huang, L., Bengtsson-Palme, J., Chen, H., Zhang, J.-H., et al. (2015). ITS1: A DNA barcode better than ITS2 in eukaryotes? *Mol. Ecol. Resour. 15*(3), 573–586.