Elsevier Editorial System(tm) for Food Control Manuscript Draft

Manuscript Number: FOODCONT-D-14-00722R1

Title: DNA and Mini-DNA Barcoding for the identification of Porgies species (family Sparidae) of commercial interest on the international market.

Article Type: Research Article

Keywords: DNA Barcoding, Mini-DNA Barcoding, Sparidae, COI gene, mislabeling, seafood identification.

Corresponding Author: Dr. Andrea Armani,

Corresponding Author's Institution:

First Author: Andrea Armani, Dr.

Order of Authors: Andrea Armani, Dr.; Lisa Guardone, Dr.; Lorenzo Castigliego; Priscilla D'Amico; Antonino Messina, Dr.; Renato Malandra; Daniela Gianfaldoni, Prof.; Alessandra Guidi, Prof.

Abstract: The morphological similarity among Sparidae species, which are characterized by a different market price, represents a serious problem for their trade and for stock management, since it encourages fraud for substitution. The most accredited morphological method for their identification is based on the dental-plate, but this approach is not simple and cannot be used for prepared products. When molecular methods are used the DNA degradation induced by cooking is the main drawback. In this work, we collected 314 reference tissues belonging to 75 Sparidae species and we produced a dataset of full (FDB) and mini-barcode (MDB) reference sequences starting from DNA extracted from fresh and ethanol-preserved tissues using universal primes. Moreover, some fresh samples were cooked. The FDB was successfully amplified in 91% (fresh), 50% (cooked) and 81% (ethanolpreserved) samples, while the amplification rates of the MDB were considerably higher in case of cooked (100%) and ethanol-preserved (94%) samples. The same primers were used for the amplification of the DNA obtained from 58 market samples (MS). All the DNA barcodes were compared with BOLD and GenBank using IDs and BLAST analysis. FDB was able to provide unambiguous specieslevel identifications for 53 (78%) and 44 (64.7%) reference samples analyzed on BOLD and GenBank, respectively. Mini-DNA barcode (MDB) showed a lower discriminating power with 32 (45.7%) and 29 (41.4%) sequences unambiguously matched to a species on BOLD and GenBank. However, the MDB allowed to identify all the reference sequences as belonging to the Sparidae family. FDB and MDB showed a similar performance in analyzing the MS, allowing to highlight 21 (38%) mislabeled MS. Our study, while confirming the FDB as a reliable tool for fish authentication, proposes the MDB as a promising tool to recover molecular information in case of cooked products.

Dear Editor,

We would like to submit the following manuscript for possible publication.

"*DNA and Mini-DNA Barcoding for the identification of Porgies species (family Sparidae) of commercial interest on the international market*"

Among the globally marketed fish, the species belonging to the family Sparidae are excellent foodfishes of high economic value. This family includes about 115 species divided in 33 genera and nowadays 85 species of Sparidae are commercialized worldwide.

The morphological similarity among Sparidae species, which are characterized by a different market price, represents a serious problem for their trade and for stock management. The specialized dentition is the most used criterion for their identification but, the marked similarities, which represent a problem even in the presence of whole specimens, make it almost impossible to distinguish the prepared or processed products during the inspection.

The DNA-based techniques are a useful tool to overcome the problems related to morphological identification and DNA barcoding has been successfully used to enforce traceability regulations in the seafood chain. Despite excellent performances when applied to fresh products, DNA barcoding has shown some weaknesses in case of processed products. For this reason, and considering that targeting a shorter region would increase the likelihood of successful amplification from degraded DNA, in this study, together with the full-barcode, the ability of a mini-DNA barcode was also assessed to produce a correct identification of Sparidae species.

In this work, we collected 314 reference tissues belonging to 75 Sparidae species and we produced a dataset of full and mini-barcode reference sequences using universal primes. The same primers were used for the amplification of the DNA obtained from 58 market samples (MS). All the DNA barcodes were compared with BOLD and GenBank using IDs and BLAST analysis. Full-DNA barcode was able to provide unambiguous species-level identifications for an higher percentage of samples than the mini-barcode on both databases. However, the mini-barcode allowed to identify all the reference sequences as belonging to the Sparidae family. Both barcodes showed a similar performance in analyzing the MS highlighting 21 mislabeled MS.

Our study, while confirming the full-DNA barcoding as a reliable tool for fish authentication, shows that the mini-barcode is a valid approach to recover molecular information from processed samples, allowing to assess the authenticity of imported products preventing commercial fraud, but also to enforce fishery control.

Best regards,

Andrea Armani

Dear Editor,

we revised the manuscript as suggested by the Reviewer and here below you can find our answers, comments and rebuttals.

Best Regards

Andrea Armani

Reviewers' comments:

The manuscript from Armani *et al*. is interesting since they have used both the full and minibarcode methodology to analyze a commercially important fish family, the Sparidae or Porgies. In addition, they have also developed a reference dataset of COI sequences for 75 Sparidae fish species using universal primers.

However, the manuscript gets confused when the authors choose to test a range of problems that may affect amplification by PCR and species identification by BOLD and Genbank.

Moreover, I could find at least four aims in the manuscript:

- (1) Development of a COI Barcode dataset for Sparidae,
- (2) Testing the full and mini-barcodes;
- (3) Market mislabeling and

(4) Factors affecting PCR amplification (i.e. ethanol preservation and cooking) when using full and mini barcodes.

All these aims together make the manuscript very hard to read. I suggest splitting the manuscript into at least four distinct sections and results presented separately. I would recommend looking at the market samples' analysis and data set development forensically instead of discussing other technical related problems.

We appreciated the Reviewer's suggestions and decided to follow his advice to make easier the reading of the manuscript.

The chapter 3 (Results and discussions) have been reorganized in four different sections.

Moreover, some "too much specific parts" have been removed or summarized.

Specifically:

Section 3.1 (Sample collection) line 206-211 (original manuscript) has been moved in section 2.1 (Sample collection: reference and market samples) line 122-125 (revised manuscript)

Line 222-224 (original manuscript) have been removed

Line 311-324 (original manuscript) have been summarized line 283-288 (revised manuscript)

Line 459-462 (original manuscript) have been removed

A new sentence line 501-505 (revised manuscript) and a new table (Table 3) have been added in section 3.4.3 (Mislabeled products: what and why?)

Table 6SM has been changed in table 2 (and thus inserted in the text) to facilitate the comprehension of the results.

Other minor corrections have been made but not highlighted

After reading the manuscript organized this way and observing it was much more readable, we decided not to split the section Results and Discussion. In fact, this would have implied some repetitions, in order to reintroduce topics of discussion every time.

Abstract

The morphological similarity among Sparidae species, which are characterized by a different market price, represents a serious problem for their trade and for stock management, since it encourages fraud for substitution. The most accredited morphological method for their identification is based on the dental-plate, but this approach is not simple and cannot be used for prepared products. When molecular methods are used the DNA degradation induced by cooking is the main drawback. In this work, we collected 314 reference tissues belonging to 75 Sparidae species and we produced a dataset of full (FDB) and mini-barcode (MDB) reference sequences starting from DNA extracted from fresh and ethanol-preserved tissues using universal primes. Moreover, some fresh samples were cooked. The FDB was successfully amplified in 91% (fresh), 50% (cooked) and 81% (ethanol-preserved) samples, while the amplification rates of the MDB were considerably higher in case of cooked (100%) and ethanol-preserved (94%) samples. The same primers were used for the amplification of the DNA obtained from 58 market samples (MS). All the DNA barcodes were compared with BOLD and GenBank using IDs and BLAST analysis. FDB was able to provide unambiguous species-level identifications for 53 (78%) and 44 (64.7%) reference samples analyzed on BOLD and GenBank, respectively. Mini-DNA barcode (MDB) showed a lower discriminating power with 32 (45.7%) and 29 (41.4%) sequences unambiguously matched to a species on BOLD and GenBank. However, the MDB allowed to identify all the reference sequences as belonging to the Sparidae family. FDB and MDB showed a similar performance in analyzing the MS, allowing to highlight 21 (38%) mislabeled MS. Our study, while confirming the FDB as a reliable tool for fish authentication, proposes the MDB as a promising tool to recover molecular information in case of cooked products.

Keywords: DNA Barcoding, Mini-DNA Barcoding, Sparidae, *COI* gene, mislabeling, seafood identification.

1. Introduction

Trade globalization is one of the main challenges for the identification of fishery products. In fact, due to the depletion of the stocks of the most requested fish on the market, alternative and underutilized species are now exploited. As a consequence, the number of products commercialized over the world is widely increased, especially in the western Countries. In Italy, the number of official denominations for seafood species has augmented from around two hundred to more than nine hundred in about ten years.

The international authorities, due to an increased attention on nutritional, ecological and safety concerns related to seafood, have issued a traceability legislation in the fishery sector. The European Union has adopted a very stringent approach: seafood must be labeled with the commercial and the scientific name, the production method, the catch area (EU Reg. No. 104/2001 64 and 404/201) and, from the $1st$ January 2015, the category of fishing gear (EU Reg. No. 1379/2013). A global seafood traceability network requires the harmonization of regulatory and commercial practices across the whole fishing sector. However, some developing Countries still have difficulties to conform to the rules of the international trade chain (Environmental Justice Foundation 2012; Armani, D'Amico, Castigliego, Sheng & Gianfaldoni, 2012a; Cawthorn, Steinman &Witthuhn, 2012; Clarke, 2009). Moreover, considering that a single commercial name can be used at the international level for different species, unscrupulous traders could take profit from this confusion by selling illegal products. Recent surveys showed that frauds are becoming widespread and seafood mislabelling has reached alarming levels (Armani, Tinacci, Giusti, Castigliego & Gianfaldoni, 2013; Carvalho, Neto, Brasil & Oliveira, 2011;Wong & Hanner, 2008). Among the globally marketed fish, the species belonging to the family Sparidae (Porgies) are

excellent food-fishes of high economic value (Antonucci, Costa, Aguzzi & Cataudella, 2009).

This family includes about 115 species divided in 33 genera (Nelson, 2006) although, according to Fishbase, the species are 133 and the genera 35 78 (http://www.fishbase.org/Nomenclature/FamilySearchList.php?). On the basis of the official lists

consulted (Table 1SM), 85 species of Sparidae are commercialized worldwide with different commercial designations, and other unexploited species could attract the interest of the market in the future.

Porgies are very similar to each other and their morphological identification can only be performed by skilled operators. The specialized dentition, on the basis of which the Sparidae family 84 has been grouped in six subfamilies, is the most used criterion for their identification (Smith & Smith 1986; Akazaki, 1962). These marked similarities, which represent a problem even in the presence of whole specimens, make it almost impossible to distinguish the prepared or processed products during the inspection.

The DNA-based techniques are a useful tool to overcome the problems related to the morphological identification (Armani, Castigliego & Guidi, 2012c) and the DNA barcoding, based on the analysis of the first part of the cytochrome c-oxidase I (*COI*) gene sequence, is the most promising approach (Hebert, Ratnasingham, & de Waard, 2003). In fact, this DNA region usually shows a greater interspecific than intraspecific variation (Hajiababei, Singer, Hebert & Hickey, 2007; Hebert *et al.,* 2003) allowing discrimination among species. Consequently, many researchers have investigated the use of DNA barcoding to enforce traceability regulations and to fight illegal fishing and frauds (Handy, Deeds, Ivanova, Hebert & Hanner, 2011; Ward, Hanner, & Hebert, 2009; Yancy, Zemlak, Mason, Washington & Tenge, 2008). Even though this method has been successfully used for the identification of fresh seafood products (Di Pinto, Di Pinto, Terio, Bozzo & Bonerba, 2013; Cawthorn *et al*., 2012; Barbuto, Galimberti, Ferri, Labra & Malandra, 2010; Wong & Hanner, 2008), it has shown some weaknesses in the case of processed products, due to the DNA fragmentation induced by heating (Cawthorn *et al.* 2012; Wong & Hanner, 2008). At the same time, the DNA degradation induced by prolonged storage in ethanol, which can occur in museum reference samples (Hajibabaei, de Waard, Ivanova, Ratnasingham & Dooh, 2005), could affect the amplification of the full *COI* barcode region, limiting the construction of sequence datasets, necessary for seafood "molecular inspection". These considerations and the possibility that fish substitutions could occur not only at the market level but also during catering activities, has prompted us to assess, together with the full-DNA barcode (FDB) fragment, also the capability of a mini-DNA barcode (MDB) in identifying the Sparidae species of commercial interest for the international market.

In this work, we collected 75 species of Sparidae, from fresh and ethanol-preserved reference tissues, and we produced a dataset of full-length *COI* barcode reference sequences by using universal primers. Then, by aligning these sequences and those retrieved from databases, we developed a new reverse primer to amplify a mini-DNA *COI* barcoding region of ~ 190bp. The FDB and MDB obtained from the reference samples and from 58 market samples were compared to BOLD and GenBank databases. Lastly, a phylogenetic analysis using the Neighbor-Joining (NJ) method was performed. The information on the label of the market samples were evaluated in the light of the molecular results.

2. Materials and Methods

2.1 Sample collection: reference and market samples

Eighty whole fresh fish were collected and morphologically identified by the Official veterinarian of the wholesale market of Milan. Two hundred thirty four ethanol-preserved reference tissues were kindly provided by Research Institutes. Overall, we collected 75 species, distributed across 26 genera, out of the 133 included in the Sparidae family (Table 2SM), and 72 out of the 85 species of commercial interest included in the official lists consulted (Table 1SM). The mean number of the collected specimens per species was 4.2 (range 1-11). Fifty-eight market samples (MS) were collected from retail markets, large-scale distribution and restaurants (Table 3SM). Each fish/tissue was labeled with an internal code and stored at -20°C.

2.2 Preparation of processed samples

Thirty-four whole fresh fish were used for the preparation of processed samples according to standard recipes. Part of them were baked as whole in an oven preheated at 180°C for a variable

time (25-40 min) depending on the size. The rest were filleted and cooked in a frying pan for 10-15 min.

Fresh muscle tissue samples were collected before and after cooking and used for DNA extraction.

2.3 DNA extraction and evaluation of DNA fragmentation by gel electrophoresis

The ethanol-preserved reference samples were re-hydrated in 100 mM TRIS-base (pH 7.8) for 30 min at Room Temperature (RT) on a thermoshaker. Total DNA extraction was performed starting from at least 20 mg of tissue as described by Armani, Castigliego, Tinacci, Gandini & Gianfaldoni, (2012b). DNA from fresh and cooked samples was extracted as described by Armani, Tinacci, Xiong, Titarenko & Guidi (2014). The DNA quality and quantity was determined with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, US).

One thousand nanograms of total DNA were electrophoresed on 1% agarose gel GellyPhorLE (Euroclone, Wetherby, UK), stained with GelRed™ Nucleid Acid Gel Stain (Biotium, Hayward, CA, USA) and visualized via UV transillumination. DNA fragment size was estimated by comparison with the marker SharpMass™50-DNA ladder (Euroclone, Wetherby, UK).

2.4 Amplification and sequencing of the full-COI barcode (FDB)

Some universal primers for the FDB region (Table 4SM) were aligned with the *COI* complete sequences of the Sparidae species available in GenBank. Those proposed by Handy *et al.* (2011) were selected. The reverse primer (SPACOIREV) was slightly modified and tailed as proposed by Steffens, Sutter, & Roemer (1993) (Table 4SM).

A 655bp fragment of the *COI* gene was firstly amplified from the DNA extracted from fresh reference specimens with the following PCR protocol: 20 µl reaction volume containing 2 µl of a 10x buffer (5Prime, Gaithersburg, USA), 100 µM of each dNTP (Euroclone, Pavia, Italy), 300 nM of forward primers, 400 nM of reverse primer, 25 ng/µL of BSA (New England BIOLABS® Inc. Ipswich, MA, USA), 1.25 U PerfectTaq DNA Polymerase (5Prime, USA), 100 ng of DNA and DNase free water (5Prime, USA) with the following cycling program: denaturation at 94 °C for 3 min; 45 cycles at 94°C for 30s, 53°C for 30s, 72°C for 35s; final extension at 72°C for 10 min. Five µL of PCR products were checked by electrophoresis on a 1.8% agarose gel and the presence of expected amplicons was assessed by a comparison with the standard marker SharpMass™50-DNA ladder. Amplicons were purified and sequenced by High-Throughput Genomics Center (Washington, USA). The same PCR protocol was used for the amplification of cooked, ethanol-preserved and market DNA samples. The ethanol-preserved and the market DNA samples that gave the expected amplicon were sequenced.

2.5 Full-DNA barcode (FDB) sequence analysis and comparison with databases

The obtained sequences were analyzed using Clustal W in MEGA version 6 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). Fine adjustments were manually made after visual inspection. Before the upload on the database, all the sequences were used to run a BLAST analysis on GenBank and analyzed using the Identification System (IDs) on BOLD (Species Level Barcode Records) (Ratnasingham & Hebert, 2007) to assess the concordance between the morphological and the molecular analysis (Ratnasingham & Hebert, 2013). A top match with a sequence similarity of at least 98% was used to designate potential species identification (Barbuto *et al*., 2010). Then, all the reference sequences were deposited on BOLD and GenBank (Table 5SM). Moreover, the sequences deposited on BOLD were used to produce a Barcode Index Number discordance report (BINdr). The mean genetic distances were calculated within species, genus and family using the Kimura 2-parameter model (Kimura, 1980) using the Distance Summary tool on BOLD.

The 55 *COI* sequences from MS, not originating from expert-identified specimens, were not submitted to the databases and were only used to assess the discriminatory ability of the barcoding region (Table 3SM).

2.6 Reverse primer design for the amplification of a mini barcoding region of the COI gene

Five hundred and sixty two reference sequences belonging to 73 Sparidae species available on GenBank and BOLD were downloaded and aligned with those produced in this study using Clustal 181 W in MEGA. Once a potential region was found spanning from the $140th$ and the 190th bp, all the

sequences were examined for the presence of polymorphisms. The projected reverse primer (REVshort1) (Table 4SM) was tailed (Steffens *et al*., 1993).

2.7 Amplification and sequencing of the mini-barcode (MDB)

The DNA of the reference samples was used to test the performance of the primer pair FISHCOILBC_ts/REVshort1 for the amplification of a ~190bp DNA region (139bp without 187 primers). The PCR was made in 20 µl reaction volume, containing 2 µl of a 10x buffer (5Prime, USA), 100 µM of each dNTP, 300 nM of primers, 25 ng/µL of BSA, 1.25 U PerfectTaq DNA Polymerase, 100 ng of DNA and DNase free water. The cycling program was the following: denaturation at 94 °C for 3 min; 45 cycles at 94°C for 25s, 51°C for 30s, 72°C for 10s; final extension at 72°C for 5 min. This protocol was also applied to samples for which the amplification of the 655bp *COI* barcoding region failed. All the PCR products were sequenced as reported in section 2.4.

2.8 Mini-DNA barcode (MDB) sequence analysis and comparison with databases

The obtained MDB were checked as reported in section 2.5 and those obtained from the reference samples were deposited in the European Bioinformatics Institute (EBI) (Table 5SM) due to the fact that BOLD and GenBank do not allow the submission of sequences shorter than 200bp. All the sequences were compared to the databases as reported in section 2.5. The mean genetic distances were calculated using the Kimura 2-p model in MEGA.

The sequences obtained from the MS were only used to assess labeling non conformities.

2.9 Phylogenetic analysis.

Two datasets were used to produce NJ dendrograms in MEGA computing the distance using the Kimura 2-parameter model with 2000 bootstrap re-samplings (Saitou & Nei, 1987).

In case of the FDB 460 reference sequences of 546bp (219 from this study and 241 from databases) and 52 sequences from MS were used while for the MDB 478 reference sequences of 138bp (254 from this study and 224 from databases) and 55 sequences from MS were used.

3. Results and Discussion

3.1 Development of a COI Barcode dataset for Sparidae

3.1.1 Full DNA barcode (FDB): primers a*mplification performances and DNA amplificability.*

Since the different origin and preservation of tissue samples may affect the primers amplification performances, we calculated the specificity and the rate of successful amplifications on the number of the species collected rather than on the totality of the samples analyzed. The primers selected in this study presented a specificity of 100% for the target region. Overall, the rate of successful amplifications was 95%, and rose to 100% for fresh samples.

The overall DNA amplificability was 85%. The DNA of the fresh specimens was successfully amplified in 91% of the cases; the rate drastically decreased to 50% after cooking. The DNA amplificability of ethanol-preserved tissue was 81%.

3.1.2 Full DNA barcode (FDB) sequence analysis. Sequencing yielded 225 *COI* FDB with an average length of 650bp (520-655), without stop codons, insertions or deletions. We obtained at least one FDB for 68 species (91%), with an average of 3.3 (range 1-8) per species.

The sequences belonging to the species *Acanthopagrus palmaris*, *A. sivicolus*, *Calamus arctifrons*, *C. proridens*, *Dentex angolensis*, *D. canariensis*, *D. gibbosus*, *D. maroccanus*, *Diplodus noct*, and *Pagrus africanus* were obtained in this study for the first time.

As expected, the congeneric divergence was found to be higher than the conspecific divergence, with mean pairwise genetic distances of 0.43%, 9.16%, and 16.18% for conspecific, congeneric and confamilial, respectively. These values were very similar to those obtained by Keskin & Atar, (2013) and Ward *et al*., 2009.

228 3.1.3 Mini DNA barcode (MDB): primer design for the amplification of a 139 bp region. DNA barcoding should be effective in recovering "molecular information" even from processed products, whose DNA is expected to be severely degraded. In this study, considering that DNA extracted from different kind of samples did not yield the expected amplicon, we designed a new reverse 232 primer (REVshort1) for the amplification of $a \sim 190$ bp MDB.

As well as for the FDB, the specificity was 100% and the overall rate of successful amplification was 93%. The DNA of the 3 species that were not amplified had been preserved in formalin or in ethanol for a long time.

The DNA amplificability was 95%, 100% and 94% for fresh, cooked and ethanol-preserved tissues. In case of cooked and ethanol-preserved samples the rates were considerably higher than with the FDB (91% for fresh, 50% for cooked, and 81% for ethanol-preserved samples) and we obtained molecular data also for *D. cervinus* and *P. africanus*.

3.1.5 Mini DNA barcode (MDB) sequence analysis. Thirty four MDB with an average length of 135bp (60-139bp) were produced and registered. No insertions, deletions or stop codons were found, indicating that nuclear DNA sequences (NUMTs) were not amplified (Zhang & Hewitt, 1996).

3.2 Testing the full (FDB) and mini-barcodes (MDB)

3.2.1. BOLD: full barcode (FDB) IDs results and BINdr. The BOLD System includes a tool for the characterization of unknown specimens, the Identification System (IDs) resource, that delivers a species identification if the query sequence shows a divergence less than 1% to a reference sequence. When less than 1% divergence is found with two or more taxa all possible species assignments are shown (Ratnasingham & Hebert, 2007). On the other hand, the BIN module assigns new *COI* sequences longer than 500bp to an existing or a new BIN, clustering them into OTUs independently from their previous taxonomic assignment. This analysis allows to confirm the concordance between barcode sequence clusters and species designations.

The IDs results and the BINdr are summarized in Table 5SM and 2, respectively. A maximum species identity in the range of 98–100% was obtained for 220 sequences (98%). For *C. arctifrons, D. canariensis* and *D. gibbosus*, the absence of reference sequences in the database resulted in "no match". The identification approach based on IDs results was coherent with the morphological 257 approach for 39 species out of 68 (57.4%), according to an identity value \geq 98%. Usually, when a sequence matches with more than one species, the highest value is obtained for the species inferred from the morphological identification (Table 5SM). A previous work suggested that a threshold value of 2% was effective in distinguish different species (Hebert *et al*., 2003). In this work this threshold did not allow to identify the remaining 29 species (42.6%). However, among these "non-identifiable" species, 9 (13.2%) were not identified due to the lack of reference sequences (Table 5SM).

We found that inconsistencies, such as indecision among species, were confirmed in most of the cases by the BINdr (Table 2). Among the 259 sequences that obtained a BIN, 37 were discordant at 266 the genus level and 56 at the species level.

Considering the high number of "ambiguous" results we further investigate the issues highlighted by the IDs analysis and the BINdr, with the aim to interpret and possibly solve them.

In most of the cases, only a few sequences were responsible for the discordance at the genus level. In particular, among the most interesting cases, the sequences of *Boops boops*, *Pagellus acarne*, *Pagellus erythrinus*, and *Pagrus pagrus*, for which all the discordances were related to sequences of *O. melanura* produced in an unpublished work. The probable misidentification of these sequences was already supposed by Keskin & Atar, (2013). However, a 7% mean genetic distance between our sequences and those of Keskin & Atar, (2013) highlights a remarkable intraspecific variation within the specimens of *O. melanura.* The reliability of our morphological identification is supported by the fact that our sequences show a mean identity value of 99.7% with other private sequences available on BOLD. Interesting to note that, while our specimens were collected in the Western part of the Mediterranean Sea, those analyzed by Keskin & Atar, (2013) came from the Eastern Mediterranean. Similar values of intraspecific divergence have been reported for the most diverse fish groups, and often attributed to cryptic species (April, Mayden, Hanner & Bernatchez, 2011; Ward, Holmes & Yearsley, 2008).

Also in the case of the sequences of *Evynnis cardinalis*, *V. acromegalus*, and *Rhabdosargus haffara*, which showed misidentification with *E. tumifrons*, *P. acarne*, and *S. aurata* respectively, the discordance might come from mislabeled specimens. In particular, in case of *R. haffara*, considering the different geographical origin of the two species, it could be possible that the specimen identified as *S. aurata* was a misidentified specimen of *R. haffara* migrated through the Suez Canal (Golani, 1992).

All the discrepancies at the genus level are reported in Table 5SM and 2. These findings could be due to the fact that the barcodes are not filtered as they enter BOLD, even if they show deep sequence divergence from existing records (Ratnasingham & Hebert, 2007). This eventuality could distort the outcomes of studies relying on database comparison.

Regarding the discrepancies at the species level, different issues were found. For instance, the species belonging to the genus *Acanthopagrus* are very similar from both a genetic and a morphological point of view (Hsu, Guillén Madrid, Burridge, Cheng & Gwo, 2011). There have been many re-descriptions within this genus and currently 15 species and 2 subspecies are recognized (Hsu *et al*., 2011). The impossibility encountered in this work to distinguish *A. pacificus* from *A. berda* could be due to a misidentification of specimens or to an identification based on previous classification, considering that *A. pacificus*, very similar in overall appearance to *A. berda,* has been recently re-described as a new species (Iwatsuki, Kume, & Yoshino, 2010). The barcodes were not even able to distinguish among *A. schlegelii*, *A. schlegelii schlegelii*, and *A. sivicolus*, which are closely related species belonging to the "black seabream complex" (Hsu *et al*., 2011).

Moreover, the occurrence of hybrid-like individuals among the *Acanthopagrus* species makes the study of this group even more difficult (Hsu *et al*., 2011). In fact, by using a mitochondrial gene, only the matrilineal lineage is examined (Carvalho *et al*., 2011; Costa, Landi, Martins, Costa & Costa, 2012). In this case, supplemental analyses on nuclear genes would be advisable.

When two or more species of the same genus cluster together, misidentification among them could have occurred (Costa *et al*., 2012).

The reason why the DNA barcode has not been capable to distinguish among *Pagrus major* and *P. auratus* could be related to the fact that they might be two subspecies, as suggested by Tabata & Taniguchi (2000). As well, the system was neither able to distinguish the *D. sargus* subspecies due to the close phylogenetic relationship of the genus *Diplodus*, which includes 13 species and 11 subspecies (Summerer, Hanel & Sturmbauer, 2001).

However, the DNA barcoding approach is always capable to distinguish this genus from the other belonging to the family Sparidae.

On the basis of this elaboration process, 53 additional sequences (belonging to 14 species) were considered resolvable and therefore the IDs could discriminate 53 species out of 68 (78%), strongly increasing the ability of the FDB in discriminating among Porgies species. Summarizing, the system was not able to identify 15 species due to the lack of reference sequences (n=9) or due to close phylogenetic relationship among species (n=6) (Table 1).

3.2.2. Full barcoding (FDB) BLAST analysis on GenBank: A maximum species identity in the range of 98–100% were obtained in GenBank for 208 sequences (92.4%) belonging to 37 species out of 68 (54.4%).

The impossible identification of the remaining 31 species was related to the absence of reference sequences or to the presence of problematic sequences (Table 5SM). In particular, identity values lower than 98% were obtained for *A. pacificus*, *C. arctifrons*, *C. leucosteus*, *C. proridens*, *D. canariensis*, *D. gibbosus*, *D. spariformis*, *V. acromegalus*, *O. melanura* and *A. spinifer* (Table 5SM).

As for BOLD, when a sequence matched with more than one species, the highest identity value was attained for the species inferred from the morphological identification (Table 5SM).

In the case of *D. puntazzo* and *P. aeneum*, the ambiguous identification was due to sequences of *D. labrax* and *P. sordida* (Moronidae and Lutjanidae family), while in the case of *D. holbrookii*, *D. vulgaris*, *E. cardinalis*, *P. bellottii*, *P. auratus*, *P. major*, *P. pagrus*, and *S. cantharus* the identification problems were the same observed on BOLD (section 3.5.1). However, for all of them, with the exception of *E. cardinalis*, the system was able to correctly identify the sequences at the genus level.

Summarizing, the BLAST analysis could clearly discriminate 44 species out of 68 (65%), increasing the ability of the FDB in discriminating among Porgies species (Table 1), while it was not able to identify 24 species (35.3%), due to absence of reference sequences (n=17) or due to close phylogenetic relationships (n=7).

3.2.3 Full DNA barcoding (FDB): comparison between BOLD and GebBank.. Even though the DNA barcoding is a useful tool for the species identification, many cases of ambiguous results due to species misidentification, wrong labeling or mistakes during sequences submission have been reported (Barbuto *et al*., 2010; Carvalho *et al*., 2010). These types of mistakes are readily detected when specimens from different orders or families cluster together, but must be carefully considered and analyzed when species belonging to the same genus are involved.

We observed that the discriminatory ability of the FDB was strictly related to the availability of correctly identified reference sequences. In fact, after the correction of the ambiguous results, BOLD was able to identify 53 species (78%) while GenBank only 44 (64%). The higher resolution of BOLD compared to GenBank agrees with the results obtained by Wong *et al*. (2008) and Cawthorn *et al*. (2012), who analyzed different groups of fish. In our study, this could be due to the fact that on BOLD only 9 reference sequences were missed, while on GenBank the lacking sequences were almost twice.

Our results are similar to those obtained by Barbuto *et al*. 2010, who, using the DNA barcoding approach for the identification of *Palombo,* recognized at the species level 34 out of 45 (75.6%) samples. In fact, in case of *Mustelus* spp., the high genetic correlations and morphological similarities made difficult their recognition by the IDs system, as in the case of the species belonging to the genus *Acanthopagrus* and *Diplodus.* On the contrary, in other studies the FDB allowed to unequivocally identify a higher percentage of samples (Cawthorn *et al*., 2012; Keskin & Atar, 2013). On the basis of this data, it seems that the DNA barcoding approach is more precise when applied to species belonging to different genus and families.

Interestingly to note that on BOLD the number of problematic sequences that could lead to misinterpretation and need thorough analysis were higher (n= 73) that on GenBank (n=59), making this latter database more suitable for "non-skilled" users. A systematic revision (elaboration process) of the "raw data" obtained by the IDs system should be performed to resolve "ambiguity" produced by unreliable sequences. Therefore, considering that published sequences are susceptible to occasional inaccuracies, a more stringent process of confirmation and validation is desirable.

3.2.4 Phylogenetic analysis of the full-barcode (FDB). The NJ phylogenetic analysis of the FDB allowed to solve the most part of issues highlighted with the DNA barcoding analysis. In particular, the most part of the species and subspecies formed discrete clusters (Fig. 1SM), with bootstrap values > 70%, showing the presence of unique and diagnostic polymorphism. However, a few species still could not be distinguished, such as: *D. maroccanus* from *D. angolensis*, *P. auratus* from *P. major*, *A. sivicolus* from *A. schlegelliii*, *D. cervinus* from *D. cervinus hottentotus*, *S. chrysops* from *S. caprinus*.

3.2.5 139bp mini DNA barcodes (MDB) sequence analysis and comparison with databases. Hajibabaei *et al*., (2005) have tested "*in silico*" the possibility to use MDB of 218bp and 109bp for the identification of fishes, observing that they generally provided sequence variability comparable to that of FDB at both intraspecific and intrageneric levels.

Meusnier, Singer, Landry, Hickey & Hebert, (2008) found that, even though the FDB performed slightly better (97% species resolution), 250bp MDB gave only slightly lower rates (95%), while with 100bp MDB resolution decreased to 90%.

The MDB sequences were compared with BOLD and GenBank databases. The BINdr could not be performed due to the limit of the system in processing sequences shorter than 500bp.

Only 251 MDB were used on BOLD because sequences shorter than 80bp cannot be processed by the IDs. All the analyzed sequences retrieved a max identity value from 98 to 100% allowing to unequivocally identify 28 species (40%). Of the remaining species, 10 (14.3%) were not identified due to the absence of reference sequences, and 32 (45.7%) where not identifiable or showed

ambiguous results. After an interpretation process, the number of correctly identified species rose to 32 (45.7%) (Table 1). Furthermore, the MDB allowed identifying at the genus level 50% of the remaining not identifiable 28 species

Two hundred fifty five sequences were analyzed by BLAST analysis on GenBank and a max identity value ranging from 98 to 100% was obtained for 243 sequences (95.2%). Sequences from *C. arctifrons, D. macrophthalmus*, *D. spariformis*, *O. melanura*, *R. haffara*, and *V. acromegalus* gave lower identity values (95-97%). MDB allowed to unequivocally identifying 26 species (37.1%). For the remaining species, 18 (25.7%) were not identified due to the absence of reference sequences, 26 (37.1%) showed ambiguous results or were not identifiable to the species level. Once that this issues have been resolved the number of correctly identified species rose to 29 (41.4%). However, the 139 mini-barcode allowed to identify at the genus level 13 (56%) of the unidentifiable 23 species (Table 1).

The analysis of the MDB highlighted a similar discriminatory power on both databases, with a comparable number of species correctly identified (32 and 29, respectively) (Table 1). Even though the discriminatory power was lower than the FDB the MDB allowed to identify 60% and 65% of the species correctly identified analyzing the FDB on BOLD and GenBank, respectively. The higher discriminatory power associated to GenBank could be explained considering that, in this database, also shorter sequences are used by the identification engine.

Finally, the MDB allowed to unambiguously identify all the reference sequences as belonging to the Sparidae family. This is a further advantage when Porgies species are replaced with species belonging to different group of fish.

3.2.6 Phylogenetic analysis of the mini-barcode (MDB). The NJ phylogenetic analysis obtained with the MDB (Fig. 2SM), despite the average lower bootstrap values at species and subspecies level, were able to correctly cluster most of the reference sequences with the exception of: *D. maroccanus*, *D. angolensis*, *D. canariensis*, *P. auratus*, *P. major*, *E. cardinalis*, *P. edita*, *S. emarginatum*, *S. cantharus*, C*. nodosus*, *C. calamus*, *D. sargus*, *D. noct*, *D. holbrookii*, *D.*

argenteus, *A. sivicolus*, *A. schlegelliii*, *D. cervinus, D. cervinus hottentotus*, *S. chrysops* and *S. caprinus*.

3.3 Factors affecting PCR amplification when using full (FDB) and mini barcodes (MDB)

The DNA electrophoresis clearly showed that the cooked samples had a more degraded DNA with respect to the fresh ones (data not shown) and that the degradation was extremely variable among the samples. In some cases, the degradation patterns revealed a scarce presence of fragments longer than 300bp. In particular, the level of degradation was higher in fish of smaller dimensions. No marked differences were observed between cooking processes.

In case of ethanol preserved specimens the degradation patterns were variable, with a smear in the range of 100 to 1000bp, not always comparable between samples belonging to the same batch (Institution).

Considering that other DNA samples of the same species were amplified with the same primers, the amplification failure of the DNA extracted from fresh samples cannot be explained with an improper primers annealing, but it might be more likely caused by DNA degradation. In fact, in 427 some cases, the DNA obtained from fresh tissues after 5 days of storing at 4^oC can be fully degraded (Rodriguez-Ezpeleta, Mendibil, Álvarez & Cotano, 2013).

The reduced amplificability of the DNA extracted from the cooked products agrees with the observed degradation patterns. Thermal treatments, ingredients and storage conditions are among the most important factors that can induce DNA degradation (Armani *et al*., 2013; Armani, Castigliego, Tinacci, Gianfaldoni & Guidi, 2012d; Rodriguez-Ezpeleta *et al*., 2013). In fact, even though the cooking procedure used in this study was not comparable to that caused by canning 434 processes, the amplificability was strongly affected. Similar problems were reported by Wong & Hanner, (2008) and Cawthorn *et al*., (2012), who were not able to produce the FDB from smoked, pickled and canned products, confirming that DNA degradation is the main obstacle to the application of the "classical DNA barcoding" approach.

The lower rate of DNA amplificability of ethanol-preserved could be due to the preservation of samples in formalin or in ethanol for a long time. Many evidences suggest that formaldehyde induces DNA degradation (Diaz-Cano & Brady, 1997), whereas alcoholic reagents yield superior results in terms of DNA amplificability (Srinivasan, Sedmak & Jewell, 2002). Therefore it is generally difficult to recover the FDB from museum specimens (Hajibabaei *et al*., 2005). Nevertheless, even short-term conservation can affect DNA integrity. Rodriguez-Ezpeleta *et al*., (2013) found that fish muscle stored in ethanol for 120 days showed a lower DNA integrity than those stored for only 30 days. In accordance, we found that samples that were soaked in ethanol just before the shipping showed a higher rate of DNA amplificability than those preserved for a longer time.

In the light of the aforesaid issues, it would be advisable to collect many samples per species in order to obtain at least 3 reference barcodes.

3.4 Mislabeling of commercial samples

Fifty eight samples (43 from market and 15 from restaurant) have been collected throughout Italy. The 55 DNA FDB (average length 653bp) and 58 MDB (average length 139bp) (55 extrapolated from the FDB) obtained have been compared to the databases and used for the phylogenetic analysis.

3.4.1 Full-DNA barcodes (FDB) comparison with BOLD and GenBank. A maximum species identity in the range of 98–100% was obtained in BOLD for 54 sequences (98%) and in GenBank for 47 sequences (85%). On the basis of the identity value obtained and considering the correction factors already discussed (section 3.2) for the reference sequences, 45 samples (83%) and 38 samples (81%) were unambiguously identified at the species level on BOLD and GenBank, respectively. Only considering a top match of 100% the number of MS identify at the species level rises to 50 (91%) on BOLD and to 42 (89%) on GenBank (Table 3SM). Even though, on both databases 100% of the remaining MS not identified at the species level were identified at the genus level, this did not allow to verify the traceability information on the remaining samples.

Overall, the analysis performed on both databases matched and allowed to highlight 21 mislabeled samples (38%). In particular, we found 7 (33%) mislabeled restaurants products and 14 (67%) mislabeled samples from retail food and large-scale markets distribution.

3.4.2 Mini DNA barcodes (MDB)comparison with BOLD and GenBank. A maximum species identities in the range of 98–100% were obtained in BOLD for 58 sequences (100%) and in GenBank for 57 sequences (98.2%). On the basis of the identity value obtained, and considering the correction factors already discussed (section 3.2), 37 samples (64%) and 42 samples (74%) were unambiguously identified at the species level on BOLD and GenBank, respectively. Only considering a top match of 100% the number of MS identified to species level rises to 47 (81%) on BOLD and to 51 (89%) on GenBank (Table 3SM). The MDB confirmed the mislabeling already detected by the barcode. No additional mislabeling was found for the three MS for which only the short fragment was amplified.

In summary, we found that FDB and MDB applied to MS were characterized by a similar discriminatory power on GenBank (89% vs 89%) while on BOLD a discrepancy was observed (91% vs 81%). Interestingly, all the MS were correctly identified with the NJ analysis using the FDB (Fig. 1SM), while using the MDB 5 MS could not be unequivocally assigned to a species (Fig. 2SM).

3.4.3 Mislabeled products: what and why?

This study confirmed that, as reported by Cawthorn *et al*., (2012) and Stiles, Lahr, Lahey, Shaftel & Bethel, (2011) more than one third of the commercialized fish is mislabeled.

On the contrary, our data are quite different from most of the studies reporting that the mislabeling rate is usually higher in processed products (Carvalho *et al*., 2011; Cawthorn *et al*., 2012). In this work, 71% of the mislabeled samples were sold as whole fish while the rest were fillets. This could be explained taking into consideration the high morphological similarity among Porgies.

Some of the mislabeling, such as *S. salpa* sold as S. *auratus*, *Diplodus* spp. sold as *O. melanura*, and *Spicara maena* sold as *S. salpa*, could be voluntary and aimed at charging higher prices on low commercial value species.

Other cases were due to the improper use of commercial denomination, such as the utilization of a generic name for the whole genus rather than the specific commercial name stated in the Italian list: Seabream (Pagello) instead of Red Pandora (Pagello fragolino) for *P. erythrinus*, Seabream (Sarago) instead of Sharp snout seabream (Sarago pizzuto) for *D. puntazzo*, Dentex (Dentice) instead of Canary dentex (Dentice atlantico) for *D. canariensis*.

In some European countries, such as Italy, many different commercial names have been issued for the different species of Sparidae, while in the UK, all the species of the family Sparidae except *Boops boops* (Bogue), *Diplodus sargus* (White sea bream) and *Pagrus auratus* (Golden seabream) can be referred to as Porgy. The ratio among the total number of commercial denominations and the total number of Porgies species considered in the official lists of seafood products analyzed in this study reflects the different national approaches for the management of seafood products. In particular, the percentage of family coverage varies from more than 79% (Australia, Canada and Italy) to 2% for UK (Table 3). This discrepancy is probably due to different culinary traditions and to a different attention paid to the preservation of the local products (D'Amico, Armani, Castigliego, Sheng & Gianfaldoni, 2014). In this light, trade names associated to single species, which often include geographical adjectives, can clearly differentiate national products from the imported ones.

Unfortunately, the different approaches adopted from different countries can enormously complicate the fair commerce of seafood species.

Conclusion

In this study, the DNA barcoding was confirmed as a reliable approach for supporting the traceability in the seafood chain and ensure the correct information of consumers, in agreement with what reported by the EU Reg. No. 1379/2013.

The analysis of MS sequences and their comparison with our dataset of reference sequences, supported by the comparison performed on BOLD and GenBank, allowed to highlight commercial frauds in the trade of Porgies' species.

Moreover, considering that targeting a shorter region would increase the likelihood of successful amplification from degraded DNA, for the first time a mini DNA barcoding approach was proposed for the identification of seafood species. In fact, considering that it is not possible to establish *a priori* the degradation level of a DNA sample, the utilization of a MDB represents a valid, and sometimes the only, approach to recover molecular information from an unknown sample.

Finally, our work highlighted that both BOLD and Genbank still lack of reference sequences and host different kind of problematic sequences. For these reasons, it would be beneficial to use both the databases, supported by a NJ analysis, and to perform a careful and aware analysis and elaboration of the raw data in order to solve ambiguous results that could create misidentification.

Acknowledgments

The Authors wish to sincerely thank all the people who kindly provided reference tissues of Sparidae: Andy Bentley of Biodiversity Institute, University of Kansas (Lawrence KS, USA); Dr. Barbara A. Brown and Lowell Flanders of Department of Ichthyology, American Museum of Natural History (New York NY, USA); Christopher Caine and Peter Caine of Fishweights.net (Jeffreys Bay, South Africa); Dave Catania of California Academy of Sciences (San Francisco CA, USA); Dr. Prosanta Chakrabarty and Caleb McMahan of Louisiana State University, Museum of Natural Science (Baton Rouge LA, USA); Dr. Satoru Chiba of Center for Molecular Biodiversity Research, National Museum of Nature and Science (Ibaraki, Japan); Dr. Achille De Sanctis of Scoglitti Fish Market (Ragusa, Italy); Dr. Jonathan Deeds of US FDA Center for Food Safety and Applied Nutrition (College Park MD, USA); Dr. Arfang Diamanka of Laboratoire de Parasitologie Générale, Faculté des Sciences et Techniques, Université Cheikh Anta Diop de Dakar (Dakar, Sénégal); Dr. Paolo Galli of Biotechnology and Bioscience Department, University of Milan Bicocca (Milano, Italy); Gabriela M. Hogue of the North Carolina Museum of Natural Sciences

(Raleigh NC, USA); Erling Holm and Margaret Zur of Department of Natural History, Royal Ontario Museum (Toronto, Canada); Dr. Seishi Kimura of Fisheries Research Laboratory, Mie University (Mie, Japan); Mark Lisher of South African Institute for Aquatic Biodiversity (Grahamstown, South Africa); Terry A. Lott of Florida Museum of Natural History, University of Florida (Gainesville FL, USA); Dr. Ione Madinabeitia of Department of Environmental Dynamics and Management, Hiroshima University (Hiroshima, Japan); Dr. Stefano Mariani of the School of Environment & Life Sciences, University of Salford (Salford, UK); Mark McGrouther, Ichthyology, Australian Museum (Sydney, Australia); Dr. Gui M. Menezes of Departamento de Oceanografia e Pescas, Universidade dos Açores (Horta, Portugal); Dr. Natacha Mesquita of Museu Nacional de História Natural e da Ciência (Lisbon, Portugal); Dr. Gavin Partridge of Australian Centre for Applied Aquaculture Research Challenger Institute of Technology (Fremantle, Australia); Dr. Mark Sabaj Pérez of Department of Ichthyology, The Academy of Natural Sciences of Philadelphia (PA, USA); Dr. Matt E. Roberts of Mississippi Museum of Natural Science (Jackson MS, USA); Dr. Hiroshi Senou of Kanagawa Prefectural Museum of Natural History (Kanagawa, Japan); Dr. Shao Kwang-Tsao of Biodiversity Research Center, Academia Sinica (Taipei, Taiwan); Karen Shearer and Rod Asher of Cawthron Institute (Nelson, New Zealand); Dr. Nir Stern of Steinhardt Museum of Natural History and National Research Center at Tel Aviv University (Tel Aviv, Israel); Dr. Jose Tavera of University of California, Davis (CA, USA); Dr. Alfred W. Thomson and Molly Phillips of Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute (St. Petersburg FL, USA); Dr. Nicholas Tuckey of The New Zealand Institute for Plant & Food Research (Nelson, New Zealand); Dr. H.J. Walker of Marine Vertebrate Collection Scripps Institution of Oceanography, University of California (La Jolla CA, USA) and Dr. Gregory J. Watkins-Colwell of Yale Peabody Museum of Natural History (New Haven CT, USA)

References

1. Akazaki, M. (1962). *Studies on the spariform fishes. Anatomy, phylogeny, ecology, and taxonomy*. Misaki Marine Biology Institute, Kyoto University, Spec. Rep. 1. Osaka: Kosugi Ltd.

2. Antonucci, F., Costa, C., Aguzzi, J., & Cataudella, S. (2009). Ecomorphology of Morpho-Functional Relationships in the Family of Sparidae: A Quantitative Statistic Approach. *Journal of Morphology*, 270, 843–855.

3. April, J., Mayden, R. L., Hanner, R., & Bernatchez, L. (2011). Genetic calibration of species diversity among North America´s freshwater fishes. *PNAS*, 108(26), 10602–10607.

4. Armani, A., D'Amico, P., Castigliego, L., Sheng, G., Gianfaldoni, D., & Guidi, A. (2012)a. Mislabeling of an "unlabelable" seafood sold on the European market: the jellyfish. *Food Control*, 26(2), 247-251.

5. Armani, A., Castigliego, L., Tinacci, L., Gandini, G., Gianfaldoni, D., & Guidi, A. (2012)b. A rapid PCR–RFLP method for the identification of Lophius species. *European Food Research and Technology*, 235(2), 253-263.

6. Armani, A., Castigliego, L., & Guidi, A. (2012)c. Fish fraud: The DNA challenge. *CAB Animal Science Reviews*, 7, 227-239.

7. Armani, A., Castigliego, L., Tinacci, L., Gianfaldoni, D., & Guidi, A. (2012)d. Multiplex conventional and real-time PCR for fish species identification of Bianchetto (juvenile form of *Sardina pilchardus*), Rossetto (*Aphia minuta*), and Icefish in fresh, marinated and cooked products. *Food Chemistry*, 133(1), 184-192.

8. Armani, A., Tinacci, L., Giusti, A., Castigliego, L., Gianfaldoni, D., & Guidi, A. (2013). What is inside the jar? Forensically informative nucleotide sequencing (FINS) of a short mitochondrial COI gene fragment reveals a high percentage of mislabeling in jellyfish food products. *Food Research International*, 54(2), 1383-1393.

9. Armani, A., Tinacci, L., Xiong, X., Titarenko, E., Guidi, A., & Castigliego, L. (2014). Development of a simple and cost-effective bead-milling method for DNA extraction from fish muscles. *Food Analytical Methods*, 7, 946–955.

10. Barbuto, M., Galimberti, A., Ferri, E., Labra, M., Malandra, R., Galli, P., & Casiraghi, M. (2010). DNA barcoding reveals fraudulent substitutions in shark seafood products: the Italian case of ''palombo" (Mustelus spp.). *Food Research International*, 43, 376–381.

11. Carvalho, D. C., Neto, D. A., Brasil, B. S., & Oliveira, D. A. (2011). DNA barcoding unveils a high rate of mislabeling in a commercial freshwater catfish from Brazil. *Mitochondrial DNA,* 22(S1), 97-105.

- 12. Cawthorn, D. M., Steinman, H. A., & Witthuhn, R. C. (2012). DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. *Food Research International*, 46, 30–40.
- 13. Clarke, S., (2009). Understanding China's fish trade and traceability. TRAFFIC, East Asia, Available at www.traffic.org/fisheries-reports/traffic_pub_fisheries9.pdf Accessed 22.03.13
- 14. Costa, F. O., Landi, M., Martins, R., Costa, M. H., Costa, M. E., Carneiro, M., Alves, M. J.,
- Steinke, D., & Carvalho, G. R. (2012). A ranking system for reference libraries of DNA barcodes:
- application to marine fish species from Portugal. *PloSone*, 7(4), e35858.
- 15. D'Amico, P., Armani, A., Castigliego, L., Sheng, G., Gianfaldoni, D., & Guidi, A. (2014).
- Seafood traceability issues in Chinese food business activities in the light of the European provisions. *Food Control*, 35, 7-13.
- 16. Díaz-Cano, S. J., &Brady, S. P. (1997). DNA extraction from formalin-fixed, paraffin-embedded tissues: Protein digestion as a limiting step for retrieval of high-quality DNA. *Diagnostic Molecular Pathology*, 6(6), 342-346.
- 17. Di Pinto, A., Di Pinto, P., Terio, V., Bozzo, G., Bonerba, E., Ceci, E., & Tantillo, G. (2013). DNA barcoding for detecting market substitution in salted cod fillets and battered cod chunks. Food Chemistry, 141, 1757–1762.
- 18. Enviromental Justice Foundation (2012). Pirate fishing exposed: the fight against illegal fishing in West Africa and the EU. Avaiable at: http://ejfoundation.org/sites/default/files/public/Pirate%20Fishing%20Exposed.pdf
- 19. European Commission (EC) Council Regulation No. 104/2000 of 17 December 1999. (21st January 2000). On the common organization of the markets in fishery and aquaculture products. Official Journal of the European Union, L 17/22.
- 20. European Union (EU) Commission Implementing Regulation No. 404/2011 of 8 April 2011.
- (30th April2011). Laying down detailed rules for the implementation of Council Regulation (EC)
- No 1224/2009 establishing a Community control system for ensuring compliance with the rules of
- the Common Fisheries Policy. Official Journal of the European Union, L 112/1.
- 21. European Union (EU) Regulation No 1379/2013 of the European Parliament and of the Council of 11 December 2013 on the common organization of the markets in fishery and aquaculture products, amending Council Regulations (EC) No 1184/2006 and (EC) No 1224/2009 and repealing Council Regulation (EC) No 104/2000
- 22. Golani, D. (1992). *Rhabdosargus haffara* (Forsskål, 1775) and *Sphyraena flavicauda* Rüppell, 1833—new Red Sea immigrants in the Mediterranean. *Journal of Fish biology*, 40(1), 139-140.
- 23. Hajibabaei, M., de Waard, J. R., Ivanova, N. V., Ratnasingham, S., Dooh, R. T., Kirk, S. L.,
- Mackie, P. M., & Hebert, P. D. N. (2005). Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360, 1959-1967.
- 24. Hajibabaei, M., Singer, G. A. C., Hebert, P. D. N., & Hickey, D. A. (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in*
- *Genetics*, 23, 167–172.
- 25. Handy, S. M., Deeds, J. R., Ivanova, N. V., Hebert, P. D. N., Hanner, R. H., Ormos, A.,
- Weigt, L. A., Moore, M. M., &Yancy, H. F. (2011). A single-laboratory validated method for the
- generation of DNA barcodes for the identification of fish for regulatory compliance. *Journal of*
- *Association of Official Analytical Chemists International*, 94, 201-210.
- 26. Hebert, P. D. N., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B*, 270, 596–599.
- 27. Hsu, T. H., Guillén Madrid, A. G., Burridge, C. P., Cheng, H. Y., & Gwo, J. C. (2011). Resolution of the *Acanthopagrus* black seabream complex based on mitochondrial and amplified fragment‐length polymorphism analyses. *Journal of fish biology*, 79(5), 1182-1192.
- 28. Iwatsuki, Y., Kume, M., & Yoshino, T. (2010). A new species, *Acanthopagrus pacificus* from the Western Pacific (Pisces, Sparidae). *Bulletin of the National Science Museum*, 36, 115-130.
- 29. Keskin, E., & Atar, H. H. (2013). DNA barcoding commercially important fish species of Turkey. *Molecular Ecology Resources*, 13(5), 788-797.
- 30. Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16, 111–120.
- 31. Meusnier, I., Singer, G. A., Landry, J. F., Hickey, D. A., Hebert, P. D., & Hajibabaei, M.
- (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC genomics*, 9(1), 214.
- 32. Nelson, J. S., (2006). *Fishes of the world*. (6th ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.
- 33. Ratnasingham, S., & Hebert, P. D. (2007) BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*, 7, 355–364.
- 34. Ratnasingham, S., & Hebert, P. D. (2013). A DNA-based registry for all animal species: The Barcode Index Number (BIN) System. *PLoSOne*, 8(7), e66213.
- 35. Rodriguez-Ezpeleta, N., Mendibil, I., Álvarez, P., &Cotano, U. (2013). Effect of fish sampling and tissue storage conditions in DNA quality: considerations for genomic studies. *Revista*
- *de Investigación Marina, AZTI-Tecnalia*, 20(6), 77-87
- 36. Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4, 406–425.
- 37. Smith, J. L.B. & Smith M. M. (1986). *Sparidae*. In Smith, M. M., & Heemstra P. C. (Eds.), Smiths' sea fishes. Berlin: Springer-Verlag.
- 38. Srinivasan, M., Sedmak, D., & Jewell, S. (2002). Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *The American Journal of Pathology*, 161(6), 1961-1971.
- 39. Steffens, D. L., Sutter, S. L., & Roemer, S. C. (1993). An alternate universal forward primer for improved automated DNA sequencing of M13. *Biotechniques*, 15, 580-582.
- 40. Stiles, M. L., Lahr, H., Lahey, W., Shaftel, E., Bethel, D., Falls, J., & Hirshfield, M. S.,
- (2011). Oceana, Bait and switch: how seafood fraud hurts our Oceans, our wallets and our health. http://oceana.org/sites/default/files/reports/Bait_and_Switch_report_2011.pdf
- 41. Summerer, M., Hanel, R., & Sturmbauer, C. (2001). Mitochondrial phylogeny and biogeographic affinities of sea breams of the genus *Diplodus* (Sparidae). *Journal of Fish Biology*, 59(6), 1638-1652.
- 42. Tabata, K., & Taniguchi, N. (2000). Differences between *Pagrus major* and *Pagrus auratus* through mainly mtDNA control region analysis. *Fisheries Science*, 66(1), 9-18.
- 43. Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725-2729.
- 44. Ward, R. D., Holmes, B. H., & Yearsley, G. K. (2008). DNA barcoding reveals a likely second species of Asian sea bass (barramundi) (*Lates calcarifer*). *Journal of Fish Biology*, 72, 458– 463.
- 45. Ward, R. D., Hanner, R., & Hebert, P. D. N. (2009). The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, 74, 329-356.
- 46. Wong, E.H.K., & Hanner, R.H. (2008). DNA barcoding detects market substitution in North American seafood. *Food Research International*, 41, 828–837.
- 47. Yancy, H. F., Zemlak, T. S., Mason, J. A., Washington, J. D., Tenge, B. J., Nguyen, N. L.,
- Barnett, J. D., Savary, W. E., Hill, W. E., Moore, M. M., Fry, F. S., Randolph, S. C., Rogers, P. L.,
- & Hebert, P. D. (2008). Potential use of DNA barcodes in regulatory science: applications of the
- Regulatory Fish Encyclopedia. *Journal of Food Protection*, 71, 210-217.
- 48. Zhang, D.X., Hewitt G.M. (1996). Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution*, 6, 247-251.
-
- Similarities among Sparidae species complicate morphological identification
- DNA barcoding has proven to be a useful tool for seafood products inspection
- Full and mini-DNA barcodes have been compared for the identification of Sparidae
- Full-barcode shows higher discriminatory ability but a lower amplification rate
- Analysis of marketed samples confirmed widespread mislabeling in the seafood chain

Table 1. Summary of the results of the IDs analysis on BOLD and of the BLAST analysis on GenBank using the full and the mini DNA barcodes (655bp and 139bp, respectively), before and after the elaboration of the results. * Include the sequences that were not identified due to the presence of sequences belonging to misidentified specimens in the databases or to close relationship between species.

Table 2: BIN discordance report.

Table 3. Percentage of coverage of the commercial denominations for the Sparidae family in different Countries.

Supplementary material for online publication only [Click here to download Additional Files: Tab.1_SM.doc](http://ees.elsevier.com/foodcont/download.aspx?id=338768&guid=6fd62274-5cc5-4b2c-b289-e2ab3420b304&scheme=1)

				(Pagrus spp.)		Seabream		
Pagrus major	Pagro del Giappone			Brasse, Meer, Dorade (Pagrus spp.)	Porgy, Sea Bream	Silver Seabream, Japanese Seabream, Genuine Porgy		Japanese seabream
Pagrus pagrus	Pagro	Pargo		Sack-Brasse	Porgy	Seabream, Red Porgy, Porgy		Red porgy
Polysteganus coeruleopunctatus			Denté à points bleu					Blueskin seabream
Pterogymnus laniarius			Panga de l'Atlantique S- E Spare panga		Porgy			Panga seabream
Rhabdosargus globiceps			Sargue de l'Atlangique $S.-E.$	Stumpfnase, Weiße				White stumpnose
Rhabdosargus sarba	Sarago dorato		Sarguedorée				Tarwhine	Goldlined seabream
Sarpa salpa	Salpa	Salema	Saupe	Goldstriemen				Salema
Sparidentex hasta							SobaityBream	Sobaity seabream
Sparus aurata	Orata	Dorada		Gold-Brasse		Gilthead Bream	Bream	Gilthead seabream
Spondyliosoma cantharus	Tanuta	Chopa	Griset, Doradegrise	Meer-Brasse Streifen-Brasse, Dorade				Black seabream
Stenotomus caprinus					Porgy	Shiner, Seabream, Porgy, Longspined Porgy		Longspine porgy
Stenotomus chrysops					Porgy, Scup	Scup, Porgy		Scup

Table 1 SM. Official Trade Names of the species of commercial interest belonging to the Sparidae family according to the lists of Italy (Ministerial Decree of the Italian Minister of Agriculture, Food and Forestry (MIPAAF) of 27th March 2002 and subsequent integrations), Spain (Resolución de 22 Marzo 2011 de la Secretaría General del Mar), France [\(http://www.economie.gouv.fr/dgccrf/Consommation/Etiquetage-des-produits/Produits-de-la-](http://www.economie.gouv.fr/dgccrf/Consommation/Etiquetage-des-produits/Produits-de-la-mer-et-d-eau-douce/Listes-des-denominations-commerciales)

mer-et-d-eau-douce/Listes-des-denominations-commerciales), Germany

(http://www.fischinfo.de/pdf/HANDELSBEZEICHNUNGEN %28DEUTSCH%29.pdf), United Kingdom (Food Standard Agency of United Kingdom), USA (US Food and Drug Administration (USFDA), Regulatory Fish Encyclopedia (RFE), 2012), Canada (Canadian Food Inspection Agency, CFIA Fish List, 2012), Australia (Australia Government, Seafood Services Australia Ltd Fishery Research Development Corporation). Moreover, the FAO English names are reported (^aAquatic Sciences and Fisheries Information System (ASFIS) [http://www.fao.org/fishery/collection/asfis/en\)](http://www.fao.org/fishery/collection/asfis/en). ^aFor all species of the family Sparidae except *Boops boops* the legal name is Sea bream or Porgy; ^bTrade denomination assigned to the species;

NR = Not Reported.

.

Table 2 SM. Reference samples collected in the study, with the indications of the Institutions, the geographical origin and the number of full and mini barcode obtained. ^aSpecies not considered in the International Official Trade lists; ^bDNA samples only used for testing the amplification performance of primers; NS: Not Sequenced.

Table 3SM. Results of the IDs analysis (BOLD) and of the BLAST analysis (GenBank) of market samples (MS), with the information reported on the label. Mislabeled samples are highlighted with a grey background. ^a Sequences not available on both databases; ^b Sequences not available in Genbank; MI: Max Identity.

Supplementary material for online publication only [Click here to download Additional Files: Tab.4_SM.doc](http://ees.elsevier.com/foodcont/download.aspx?id=338765&guid=d71751df-fe55-414e-9f12-393581f392ae&scheme=1)

Table 4SM. Universal primers for the amplification of the *COI* gene from fish (Armani et al, 2012c with modification). * The length refers to the amplicon generated using the forward FISHCOILBC_ts

Table 5SM: The results of the IDs analysis on BOLD and of the BLAST analysis on GenBank for the full and for the mini DNA barcode. The BOLD codes and the NCBI access number are reported when available (no code is assigned in BOLD to sequences <200bp). The species not reported in **bold type** have been considered originating from incorrectly identified or mislabeled specimens. When two or more values of MI are reported they are referred to a range (if separated by a -) or to different MI retrieved (if separated by a semicolon ;). ^a No sequences were available for this species in consulted databases; ^bNo sequences were available for this species in Genbank database;*Species not considered in the International Official Trade lists; *D. sargus* subspecies: *D. sargus ascensionis, D. sargus capensis, D. sargus helenae; D. sargus kotschyi, D. sargus lineatus, D. sargus sargus.*

Supplementary material for online publication only [Click here to download high resolution image](http://ees.elsevier.com/foodcont/download.aspx?id=338769&guid=8cfeca22-6c06-4db9-bca4-86550c9465a2&scheme=1)

Supplementary material for online publication only [Click here to download high resolution image](http://ees.elsevier.com/foodcont/download.aspx?id=338763&guid=fd962de9-1192-431d-bd34-ecc10f3b96f8&scheme=1)

