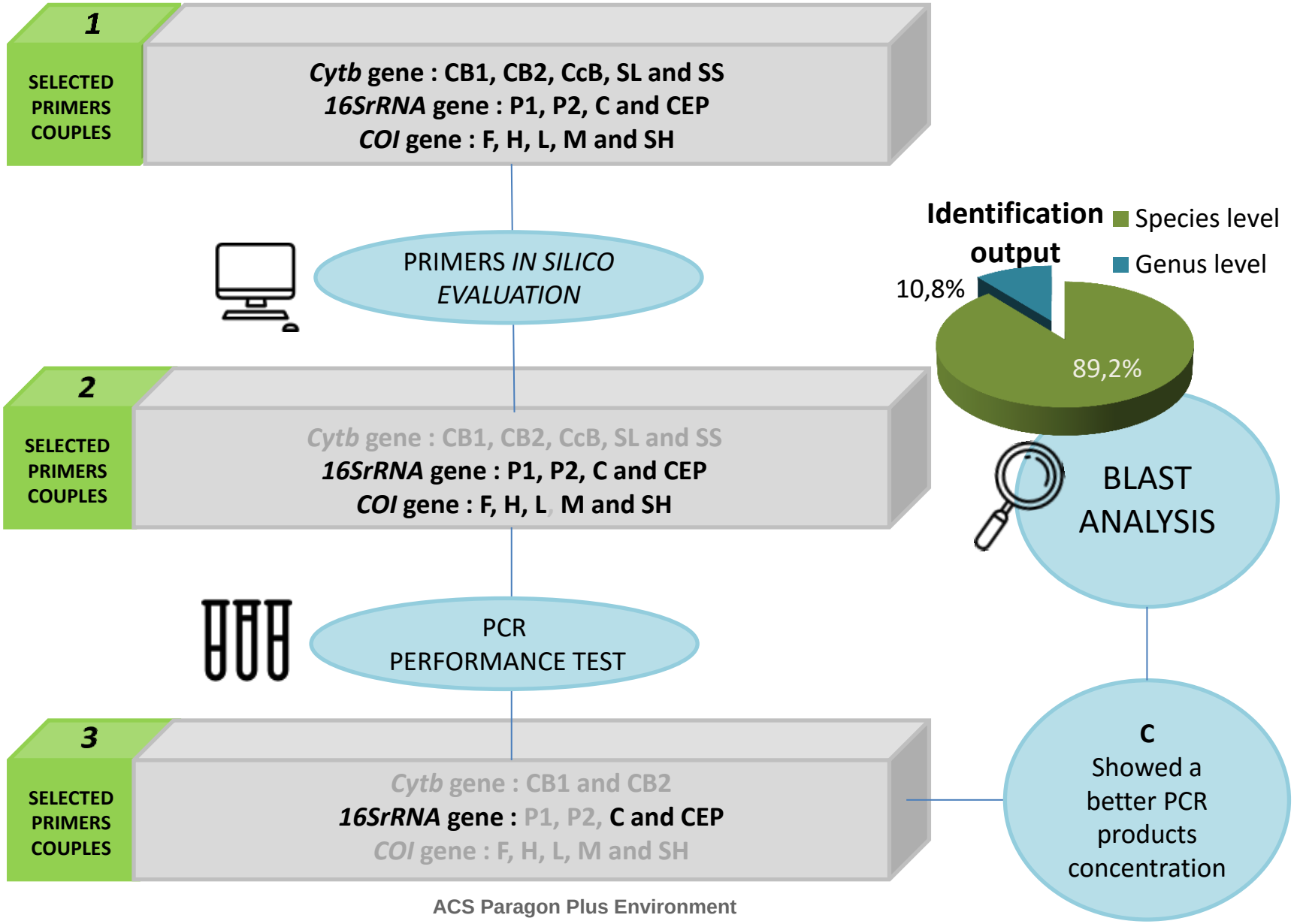
Seafood identification in multispecies products: assessment of 16srRNA, cytb and COI universal primer efficiency as preliminary analytical step for setting up metabarcoding Next Generation Sequencing (NGS) techniques

Journal:	*Journal of Agricultural and Food Chemistry*
Manuscript ID	jf-2016-058026.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Giusti, Alice; Universita degli studi di Pisa Dipartimento di Scienze Veterinarie Tinacci, Lara; Universita degli studi di Pisa Dipartimento di Scienze Veterinarie Sotelo, Carmen; Instituto de Investigaciones Marinas, Bioquímica de Alimentos Marchetti, Martina; Universita degli studi di Pisa Dipartimento di Scienze Veterinarie Guidi, Alessandra; Universita degli studi di Pisa Dipartimento di Scienze Veterinarie Zheng, Wenjie ; Tianjin Entry-Exit Inspection and Quarantine Bureau of the People's Republic of China Armani, Andrea; University of Pisa, Department of Veterinary Sceinces


1
SELECTED PRIMERS COUPLES
Cytb gene : CB1, CB2, CcB, SL and SS
16SrRNA gene : P1, P2, C and CEP
COI gene : F, H, L, M and SH
PRIMERS IN SILICO EVALUATION
2
SELECTED PRIMERS COUPLES
Cytb gene : CB1, CB2, CcB, SL and SS
16SrRNA gene : P1, P2, C and CEP
COI gene : F, H, L, M and SH
PCR PERFORMANCE TEST
3
SELECTED PRIMERS COUPLES
Cytb gene : CB1 and CB2
16SrRNA gene : P1, P2, C and CEP
COI gene : F, H, L, M and SH
Identification output
Species level
Genus level
10,8%
89,2%
BLAST ANALYSIS
C
Showed a better PCR products concentration

COI gene cytb gene

16SrRNA gene

COI gene

16SrRNA gene

16SrRNA gene

BEST CHOICE
FOR **METABARCODING**
NGS ANALYSIS
IN MULTISPECIES SEAFOOD PRODUCTS

Seafood identification in multispecies products: assessment of 16srRNA, cytb and COI universal primers' efficiency as a preliminary analytical step for setting up metabarcoding Next Generation Sequencing (NGS) techniques

Alice Giusti[1#], Lara Tinacci[1#], Carmen G. Sotelo[2], Martina Marchetti[1], Alessandra Guidi[1], Wenjie Zheng[3], Andrea Armani[1*]

[1] FishLab, Department of Veterinary Sciences, University of Pisa, Via delle Piagge 2, 56124, Pisa (Italy)

[2] Instituto de Investigaciones Marinas (IIM-CSIC), Eduardo Cabello 6, 36208 Vigo (Spain)

[3] Tianjin Entry-Exit Inspection and Quarantine Bureau of the People's Republic of China, Jingmen Road 158, Free trade Zone, Tianjin Port, 300461, Tianjin (China)

# These authors have equally contributed to this work

*corresponding author:

Postal address: *FishLab, (http://fishlab.vet.unipi.it/it/home/). Department of Veterinary Sciences, University of Pisa, Via delle Piagge 2, 56124, Pisa (Italy).*

Tel: +390502210207

E-mail address: andrea.armani@unipi.it (A. Armani)

Abstract

Few studies applying NGS have been conducted in the food inspection field, particularly on multispecies seafood products. A preliminary study screening the performance and the potential application in NGS analysis of 14 "universal primers" amplifying *16SrRNA, cytb* and *COI* genes in fish and cephalopods was performed. Species used in surimi preparation were chosen as target. An *in silico* analysis was conducted to test primers' coverage capacity, by assessing mismatches (number and position) with the target sequences. The 9 pairs showing the best coverage capacity were tested in PCR on DNA samples of 53 collected species to assess their amplification performance (amplification rate and amplicon concentration). The results confirm that primers designed for the *16SrRNA* gene amplification are the most suitable for NGS analysis also for the identification of multispecies seafood products. In particular, the primer pair of Chapela et al. (2002) ~~resulted~~ is the best candidate.

Keywords: metabarcoding, Next Generation Sequencing, multispecies seafood products, universal primers, fish, cephalopods.

Introduction

DNA-based methods are nowadays routinely applied in seafood species identification at laboratory level and in the last decades they have supported the transparency of seafood products trade and the compliance with regulations concerning IUU (Illegal Unreported Unregulated) fishing and labelling [1,2]. These methods, which mostly rely on PCR amplification, can be exploited for the analysis of an extremely wide range of seafood, from fresh to processed, mainly thanks to the relative thermal-stability of DNA[3]. Among the PCR-based methods, *Forensically Informative Nucleotide Sequencing* (FINS) and DNA Barcoding, both based on DNA sequencing, are the most frequently applied [4–6]. FINS generally relies on target regions of mitochondrial genes, such as 16S ribosomal RNA (*16SrRNA*), cytochrome b (*cytb*) and cytochrome oxidase subunit I (*COI*)[1], whereas, for the standard DNA Barcoding, the Consortium for the Barcode of Life (CBOL) has adopted a ~650 bp *COI* gene fragment [7].

DNA sequence-based identification generally uses the refined Sanger method, which is still the "gold standard" [8]. However, since Sanger method has been designed to produce a single sequence, generally from a single amplicon, it has been proved useful and reliable for the identification of products composed of individual species [9–14]. Therefore, even though applicable for the detection of species in mixed sources [2-15] it does not represent the elective method for this kind of products. ~~while it results poorly effective for the detection of species in mixed sources~~[14]. The development of innovative metabarcoding techniques, utilizing primers with broad binding affinity combined with Next Generation Sequencing (NGS), could allow identification of multiple ~~(NGS) could allow to identify multiple~~ species in a mixed sample[16,17]. NGS technologies, by massively parallel and clonal sequencing, have increased the ability to gain sequence information even from a single molecule within a complex or degraded DNA source [6,18,19]. NGS is becoming a standard approach in a large number of studies in many different fields, including sequencing of large genomes [20,21] and metagenomics studies [22–25]. Despite the benefits that this approach may provide to the species identification in the food

inspection field, only a few studies with this purpose have been conducted [19,26–28]. In particular, to the best of our knowledge, only one study applying NGS to seafood products has been reported to date [29]. This may be due to the lack of preliminary studies, necessary to practically approach the technique in the best way.

The selection of suitable universal primers with a wide fish species coverage (also called universality), represents a fundamental preliminary step for metabarcoding NGS analysis. In fact, the species detection could be affected by the variability of the primers' binding efficiency across taxa. On the contrary, the selection of universal primers would ideally allow to identify all the fish species contained in a mix, thus reducing the risk of false negatives [30]. To date, a wide variety of so-called universal primers, able to amplify fragments of different length from mitochondrial genes, have been proposed. Among them, those targeting the *16SrRNA* are often not degenerated due to the high degree of conservation of this gene [31]. The possibility to easily and concurrently amplify DNA fragments from a wide range of organisms has implied that the universal primers targeting *16SrRNA* have been employed in NGS studies, both in prokaryotes and eukaryotes genome analysis [27,32–34]. However, universal primers cannot always assure DNA amplification of all the species, due to the presence of mutations which cause mismatches in the primer sequences [35]. Moreover, in the case of a hypothetical NGS analysis of a DNA mixture, failed amplification of particular species could be masked by the recovery of amplicons from another one present in the sample, making protocol optimization difficult [36]. To overcome this issue, a detailed preliminary assessment for the selection of suitable primers is required before applying an NGS analysis for the identification of multispecies seafood products. The step preceding NGS amplification and sequencing requires a preliminary template preparation, in which fragments of DNA molecules are fused with adapters containing universal priming sites in order to convert the source nucleic acid material into standard libraries (composed by adaptors, primers and target DNA fragment) suitable for loading onto a sequencing instrument [18]. It is evident that robust library preparation producing a

representative, non-biased source of nucleic acid material from the genome under investigation is of crucial importance [37]. In this context, a meticulous choice of the DNA fragment that will be construed by the NGS machine, and as a direct consequence a proper primers selection, is undoubtedly required. In this study, 14 different pairs of primers (5 for the *cytb,* 4 for the *16SrRNA* and 5 for the *COI* genes) targeting fragments of different lengths, which have been reported in studies in the literature for the amplification of DNA from fish and cephalopod species, were tested on several species commonly used in the production of a commonly traded type of multispecies seafood such as surimi (**Table 1SM**), and compared to each other. The goal of this study was to supply a complete analysis of the universal primers targeting the three most employed genes for seafood species identification, also in order to provide practical backup for the setting up of subsequent NGS analysis targeting multi species products.

2. Materials and methods

Selection of target species and samples collection

A literature investigation was initially performed in order to identify fish and cephalopod species commonly used for surimi preparation and/or effectively identified in surimi-based products during forensic analysis. All these species (89 species, of which 84 fish and 5 cephalopods) are listed in **Table 1SM**. The most part of the species analysed in this study were collected according to this list, with the exception of other 5 fish species which, however, belonged to the same families/genera of the list: *Sardinella aurita* (family: Clupeidae; order: Clupeiformes), *Gadus morhua* (family: Gadidae; order: Gadiformes), *Dissostichus eleginoides* (family: Nototenidae; order: Perciformes), *Helicolenus barathri* (family: Sebastidae; order: Perciformes) and *Chelidonichthys lucernus* (family: Trilidae; order: Perciformes), which were collected to overcome the lack of some species belonging to **Table 1SM** or, in the specific case of the *G. morhua,* to enlarge the number of specimens belonging to the Gadus genus. ~~Thus~~Overall, ~~44 fish species~~ 49 fish species (144 specimens) and 4 cephalopod species (10 specimens) ~~(130 specimens) and 4 cephalopod species (10 specimens)~~

were collected out of those reported in **Table 1SM**, ~~jointly with other 5 fish species (14 specimens) belonging to families related to those collected, for a total of 49 fish species (144 specimens) and 4 cephalopod species (10 specimens).~~ All the collected species, ~~reported in **Table 1**,~~ were obtained in form of fresh, ethanol-preserved or dried tissue and were kindly provided by research institutes or directly collected in this study.

DNA extraction and evaluation

Ethanol-preserved, dried or lyophilized tissue samples were washed/rehydrated in a NaH_2PO_4 buffer (pH 8) for 15 min at room temperature on a digital Vortex-Genie® (Scientific industries, Inc. NY, 11716 USA). Total DNA extraction was performed from at least 100 mg of tissue following the protocol proposed by Armani et al.[38]. The amount and the purity of DNA was determined with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) by measuring the absorbance at 260 nm and the ratios A260/280 and A260/230. The DNA samples were provisionally stored at -20°C pending subsequent analysis.

Universal primers analysis

Primers selection. Initially, 14 pairs of universal primers reported in literature for the amplification of fish and cephalopod species were selected: 5 targeting the *cytb* gene (CB1, CB2, Ccb, SL and SS), 4 targeting the *16SrRNA* gene (P1, P2, C and CEP) and 5 targeting the *COI* gene (F, M, H, L and SH). The primers were conveniently divided in two groups on the basis of the amplicon length they produced: (i) LAL (Long Amplicon Length), which included pairs of primers for the amplification of a fragment longer than 500 bp (without adaptors); (ii) SAL (Short Amplicon Length), including primer pairs capable of amplifying a fragment shorter than 500 bp (**Table 2**).

Primers in silico evaluation. An *in silico* analysis of primers characteristics was performed in order to infer their amplification performance following Armani *et al*. (2016) [39]. For this purpose, all the available *cytb*, *16SrRNA* and *COI* sequences (complete and partial) of each fish and

cephalopod species reported in **Table 1SM** and **Table 1**, for a total of 94 species (89 fish and 5 cephalopods), were retrieved from GenBank and, in the case of the *COI* gene, also from BOLD. For each species, all the retrieved sequences belonging to each one of the three selected genes were aligned with the software Clustal W in BioEdit version 7.0.9 [40] and one representative sequence (complete when possible) of each haplotype per gene was chosen. Then, these sequences were aligned with the 14 primer pairs in order to evaluate two aspects: firstly, their *stricto sensu* coverage capacity through the direct count of mismatches between the primers and their respective matching region. In particular, the primers were divided in three distinct groups: (1) primers that presented no mismatches (perfectly complementary to the respective sequences); (2) primers that show 1 or 2 mismatches; (3) primers with 3 or more mismatches with the respective sequences; then, particular attention was given to the position of mismatches at the annealing regions, focusing especially on primers that present mismatches within the first four bases near the 3' end. ~~Then, these sequences were aligned with the 14 primer pairs in order to evaluate their coverage capacity on the basis of the number of mismatches between the primers and the respective matching region (also expressed in % value). In particular, the primers were divided in three distinct groups: (1) primers that presented no mismatches (perfectly complementary to the respective sequences); (2) primers that show 1 or 2 mismatches; (3) primers with 3 or more mismatches with the respective sequences. Moreover, particular attention was given to the position of mismatches at the annealing regions, focusing especially on primers that present mismatches within the first four bases near the 3' end.~~ On the basis of the preliminary *in silico* evaluation, 9 out of the 14 pairs of primers were selected. In particular, all the 5 primer pairs targeting the *cytb* gene were discarded. The workflow illustrating the whole process of primers evaluation and the output of each intermediate step is summarized in **Figure 1**.

Assessment of primers amplification performance. All the DNA samples extracted from fish and cephalopod specimens (**Table 1**) were amplified with the 4 selected *16SrRNA* primer pairs (P1, P2, C and CEP) and the 5 selected *COI* primers pairs (F, L, M, H and SH) (**Table 2**) on the peqSTAR

96 Universal Gradient thermocycler (Euroclone, Milan, Italy) according to the PCR protocols and programs reported in **Table 2SM**. Thus, five microliters of each PCR product were checked by electrophoresis on a 2% agarose gel, and the presence of expected amplicons was assessed by a comparison with the standard marker SharpMass TM50-DNA (Euroclone, Life Sciences Division, PV, Italia). The amplification results were analysed to calculate the amplification rate (expected bands obtained/n° of DNA samples amplified) and the amplicon concentration (bands intensity) for each pair of primers. As regards the amplicon concentration, 10 ng/µl was used as threshold for PCR products possibility to be sequenced [39].

Amplicon BLAST analysis. The amplicons obtained with the primer pairs which performed better in terms of amplification rate and amplicon concentration, retrieved from the sequence analysed in *section Primers in silico evaluation* (one representative sequence of each haplotype per gene chosen), were used to run a BLAST analysis on GenBank, in order to evaluate the diagnostic power, in term of specie specific identification, of each amplicon~~the effective discriminatory capacity of each amplicon~~. Due to the fact that the primers pair performing better was one of those amplifying the *16SrRNA* gene ~~(see *section 3.3.1*)~~, a top match with a sequence similarity of at least ~~99~~ 100% was used to designate potential species identification [41]. BLAST results are reported in **Table 7SM**.

3. Results and discussion

3.1 Target species selection and samples collection

Surimi represents a typical multispecies seafood product and, currently, 89 species (fish and cephalopods) are reported to be widely used for its production (**Table 1SM**). Such an elevated number of exploitable species essentially represents the reason why surimi-based products were selected as the starting point for the present analysis. The higher is the number of species included in the study, the more accurate the assessment of the

primers universality results. In details, the 84 fish species belong to 11 orders and 27 families, whereaswhile the 5 cephalopod species belong to 2 orders and 2 families (**Table 1SM**). The most part of the species analysed in this study were subsequently collected according to this list, with the exception of other 5 fish species which, however, belonged to the same families/genera of the list: *Sardinella aurita* (family: Clupeidae; order: Clupeiformes), *Gadus morhua* (family: Gadidae; order: Gadiformes), *Dissostichus eleginoides* (family: Nototenidae; order: Perciformes), *Helicolenus barathri* (family: Sebastidae; order: Perciformes) and *Chelidonichthys lucernus* (family: Trilidae; order: Perciformes), which were collected to overcome the lack of some species belonging to **Table 1SM** or, in the specific case of the *G. morhua*, to enlarge the number of specimens belonging to the Gadus genus.

3.2 Universal primers analysis

Primers selection. The available NGS studies inherent to species detection in mixed food source ustilized *16SrRNA* as the election molecular marker [19,26–28]. This gene has been shown to be a good marker also to differentiate fish species and it has been used in comparative intergeneric and interspecific studies in several fish families [41–43]. However, the *cytb* and *COI* genes, due to their comparable high interspecific es variation and low intraspecific variation, are nowadays the most widely used genetic markers for fish species identification, as reported in a large number of studies applied to food inspection [9–11,44]. A wide variety of universal primers is now available for the amplification of the three genes reported above. For this reason, the goal of this study was to provide an as much as possible complete analysis of 14 pairs of universal primers targeting these three mitochondrial genes, also in order to establish if the *16SrRNA* can be effectively considered the best one or if the other genetic markers present some advantages. All the primers pairs were selected due to the fact that they have been successfully tested of several fish and/or cephalopod species.

Primers in silico evaluation. The primers evaluation parameters considered in our analysis, directly interpretable by a visual check of the **Tables 3.1SM**, **3.2SM, 4.1SM**, **4.2SM**, **5.1SM** and **5.2SM**, were summarized in **Table 3**. ~~Basically, we considered appropriate to focalize the analysis on two features: firstly, the *stricto sensu* primers coverage capacity, evaluated through the direct count of mismatches between the primers and their respective matching region; secondarily, the observation of the mismatches position. In fact, it is known that the presence of mismatches within the first three bases near the 3' end affects PCR more dramatically than those located internally or at 5' end[30] and this aspect has shown to be actually more reliable in the amplification output prediction respect to the simple mismatches count[37]. About that, on the basis of our experience, we decide to consider as a negative prediction the presence of mismatches (one or more) in the first four bp starting from the 3'end[37]~~. In details, regarding the *16SrRNA* primers, ~~all~~ the pairs P1, P2 and CEP proved to be well performant in almost all the fish species, due to the low number of mismatches with all the sequences analysed (for the forward as for the reverse primer) and to the fact that those mismatches were in most cases located in regions distant from the 3' end (**Table 3**, **3.1SM** and **3.2SM**). In fact, it is known that the presence of mismatches within the first three bases near the 3' end affects PCR more dramatically than those located internally or at 5' end [31~~30~~] and this aspect has shown to be ~~actually~~ more reliable in the amplification output prediction ~~respect~~than ~~to~~ the simple mismatches count [39~~37~~]. ~~About that~~Thus, on the basis of our experience, we decided to consider as a negative prediction the presence of mismatches (one or more) in the first four bp starting from the 3'end [39]. For the cephalopod species, the P1 pair appeared to perform better than the P2, where the forward primer presented instead several mismatches at the 3' end (**Table 3**, **3.1SM** and **3.2SM**). The primer pair C seemed to perform better in cephalopod species with respect to fish, where the number of mismatches was in many cases higher than 2 and their position appeared critical especially on the forward primer, where all the sequences presented mismatches on the first four bp near the 3'end (**Table 3** and **3.2M**). As for the *COI* primers, substantial differences could be observed within the pairs. In particular, as

concerns fish species analysis, H, M and SH showed better outcomes in terms of number and position of mismatches respect to F and L primers set (**Table 3**, **4.1SM** and **4.2SM**). However, regarding the pair M, whereas the forward primer matched ~~enough properly~~, the reverse one presented mismatches in problematic positions ~~especially~~ for several fish species (**Table 3** and **4.1SM**). ~~although, especially in the case of H and M pairs, they could not be considered absolutely flawless, at least from a theoretical point of view (**Table 3**). Regarding the pair M, in particular, whereaswhile the forward primer seemed to match enough properly, the reverse one presented mismatches in problematic positions especially for several fish species (**Table 3** and **4.1SM**).~~ As regards cephalopod species, H and particularly SH primers did not performed well, considering the number and the position of the mismatches, whereas the M pair ~~seemed to~~ performed better (**Table 3**, **4.1SM** and **4.2SM**).

The pair L did~~seemed to~~ not perform properly on fish species, especially due to the forward primer, that showed a high number of mismatches, in many cases located near the 3' end, in almost all the species analysed (**Table 3** and **4.2SM**). It ~~seemed instead to~~ performed instead better ~~for~~ on cephalopods (**Table 3** and **4.2SM**). Similarly, the pair F presented a high number of mismatches (>3) with almost all the fish species, but it ~~seemed to~~ performed better on~~in~~ cephalopods (**Table 3** and **4.1SM**).

Regarding the *cytb* gene, CB1 and CB2 primer pairs analysis allows to hypothesize that they would not perform well on a large range of the fish species selected due to the fact that the number of mismatches was rather high. Furthermore in a great part of the sequences analysed they were located near the 3' end (**Table 3** and **5.1SM**).It was not possible to show their coverage capacity on cephalopods, since the primers did not match any of the analysed species. In the same way, the Ccb, SL and SS sets, presented a high number of mismatches positioned near the 3'end of the matching region on an extremely wide part of fish and cephalopod species analysed (**Table 3** and **Table 5.2SM**). Finally, all the *cytb* primers were discarded and they were not tested in the subsequent PCR amplification step. In fact, the pairs CB1 and CB2 could not amplify any cephalopod

187 species, while the pairs Ccb, SL and SL presented too many mismatches with the target species of the study. In all the species analysed, the average
188 number of mismatches was 7.2 (33% of the total primer length) in the forward primer and 7 (32% of the total primer length) in the reverse primers
189 for the pair Ccb, whereas for the pair SL and SS the forward primer (which is the same) presented 7.8 mismatches on average (36% of the total
190 primer length) on all the analysed species. In order to confirm the good results of those primers that performed well in this preliminary analysis and
191 to assess the amplification outputs of those primers for which interpretation resulted ambiguous, we decided to selected the pairs M, F, H, SH and L
192 (*COI*) and the pairs P1, P2 and CEP (*16SrRNA*) for the subsequent amplification step, jointly with the pair C (*16SrRNA*) that, even if seemed to not
193 perform as the other selected, it showed an average mismatches number of 5.2 (22% of the total primer length) in the forward primer and of 2.5
194 (11% of the total primer length) in the reverse one, which is substantially better than the results showed by the discarded *cytb* primers. The primers
195 pair C, F and L were included in this group also due to the fact that their sufficiently good performance for cephalopod species could be
196 undoubtedly exploited in analysis concerning surimi-based products or other complex matrix that contain cephalopods.

197 *Primers amplification performance assessment*. The amplification rate and the PCR products concentration were assessed after PCR
198 amplification with all the primer pairs mentioned above and the results are reported in **Table 4** and **Table 6SM**.

199 In details:

200 *(i) COI primer pairs:* The primers pair H amplified the DNA from all the fish species (100% amplification rate), yielding PCR products with an
201 average concentration of 20 ng/µl (± 5.59 ng/µl). On the contrary, these primers did not amplify DNA from any cephalopod species, confirming that
202 the H pair, specifically designed on fish DNA sequences, is not suitable for cephalopod species identification. Differently, the primer pair M
203 amplified the DNA from all the cephalopod species (100%amplification rate) with a concentration of 25 ng/µl in all the tested species, whereas for

the fish the amplification rate was 71.4%, with an average concentration of 9.6 ng/µl (±7.82 ng/µl). Moreover, 7 amplified species showed a concentration of 5 ng/µl, which is lower than that required for sequencing. The primer pair F performed well in all cephalopod species amplification, yielding PCR products with a concentration of 25 ng/µl, whereas for the fish the amplification rate was 34.7% and the average concentration 6.8 ng/µl (±9.82 ng/µl). The primer pair SH amplified the DNA from 95.9% of the fish species with an average concentration of 17.3 ng/µl (±6.54 ng/µl). Only in 2 species, PCR products showed an average concentration lower than what required for sequencing. This pair also amplified one of the four species of cephalopods included in the study, but the average concentration was low (5 ng/µl). Finally, the pair L did not amplify any fish and cephalopod species despite it had performed well on cephalopods in the *in silico* analysis. Due to these constraints affecting the *COI* primer pairs they were not considered as the optimum choice for NGS analysis.

(ii) 16SrRNA primer pairs: The primer pair P1 amplified the DNA from all fish and cephalopod species (100% amplification rate) giving an average PCR product concentration of 20.6 ng/µl (±5.16 ng/µl) for fish and of 20 ng/µl (±4 ng/µl) for cephalopods. Also the primer pair P2 amplified the DNA of all fish and cephalopod species, but whereas in the case of the fish an average DNA concentration of 17.6 ng/µl (±3.96 ng/µl) was obtained, for the cephalopods the average concentration was only 5 ng/µl. Also, the primer pair CEP amplified the DNA from all fish and cephalopod species (100% amplification rate) with an average PCR product concentration of 19.6 ng/µl (±1.99 ng/µl) for fish and of 20 ng/µl for all cephalopod species. Unexpectedly, also the primer pair C performed well amplifying the DNA from almost all fish species (97.8% amplification rate), with the only exception of *Dissostichus eleginoides*, and from all cephalopod species (100% amplification rate), with a slight predilection for cephalopods from the PCR products concentration point of view. In fact, the average concentration of PCR products was 20.9 ng/µl (±5.23 ng/µl) for fish and 30 ng/µl for cephalopod species. These results, despite the difference observed between the four primer pairs, substantially confirmed that the *16SrRNA*

gene is effectively widely conserved not only between species, but also between different classes. As known, the *16SrRNA* sequences show the lowest mean genetic p-distances at the taxonomic level, from species to order, in a large range of taxa, including fish, while higher values have been observed for *COI* and *cytb*[30]. Moreover, we could assert that, in a hypothetical use in NGS studies, the best pair of primers were the CEP and the C ones. In fact, in addition to showing excellent amplification rates and high products concentration both for fish and cephalopods, they fully meet the current NGS platforms requirements which, as mentioned above, work better with shorter amplicons. Among these two pairs, in particular, the C one seemed to perform better as regards the products concentration in both fish and cephalopod species and thus it proved to be the best one among all the pairs analyzed. In fact, despite the fact that *D. eleginoides* was not amplified, these primers could be easily used in NGS studies for surimi species detection due to the fact that this species has rarely been reported in this seafood product.

BLAST analysis. Performing a BLAST analysis with the fragment comprised between the C primer pair, in case of cephalopods, always resulted in an identity value of 100% with only one species. In addition, the nearest identity values obtained for other species were always lower than 98%. Therefore, all the amplicons allowed a species-specific identification, confirming the ability of this fragment in discriminating cephalopods species. A similar result was obtained for 58,.1% of the amplicons retrieved from the fish species investigated. Among the remaining amplicons, 74.1% showed an identity of 100% with only one species. However, in this case, specie specific identification was not unambiguously achieved due to an identity value of 99% with other species belonging to the same genus. Therefore, 78.2% of the amplicons only allowed a genus-level identification. In particular, ambiguity among species belonging to the same genus were highlighted during the analysis of the amplicons belonging to the species *T. chalcogramma, G. ogac, M. hubbsi*, ~~and~~ *M. australis*, *T. japonicus* and *N. japonicus* that showed overlapping identity values with *T. finmarchica, G. macrocephalus, M. merluccius*/*M. productus*, ~~and~~ *M. poutassou, T. declivis* and *N. virgatus,* respectively. In the particular case of

the *T. chalcogramma*, however, a taxonomical study proposed by Ursvik et al.[45] asserted that they could represent a single species. More problematic identification were encountered with the amplicons of the species belonging to the genus *Oreochromis* spp. and *P. medius*, since they presented an identity value of 99-100% also with species belonging to other genera. Overall, the fragment amplified using C primer pair demonstrated its ability in discriminating between different value species, and particularly it allowed to detect the presence of less valuable freshwater species mislabelled as species commonly caught in open sea. Moreover, these primers have shown their capacity in effectively discriminating between fish and cephalopods and this feature could also be exploited in studies aimed to detect the presence of potential allergenic species in complex seafood matrix.

Selection of the best primers pairs

All the analytical phases for primers evaluation developed in this study (schematized in **Figure 1**) have lead to the final choice of the best primer pair among all those analysed. The results of the primers performance test were schematically reported in **Table 5**. As already highlighted, the primers selection was based on those features that would be essential in an NGS study. Thus, jointly with the amplification rate and the products concentration, in this final selection step we also considered the amplicon length

The size of the target DNA fragments in the final library is a key parameter for NGS library construction[46]. In fact, each available platform disposes of a defined own read length, which unavoidably affects the primers selection. The read length of Roche 454, which was initially 100-150 bp in 2005, has nowadays reached 700 bp; the SOLiD system length read raised from 35 bp before 2007 to 85 bp in 2010; Illumina and Ion Torrent PGM ~~GA/HiSeq~~ maximum reads length ~~system read length~~ is nowadays ~~up to 600~~2x150 bp and 400 bp, respectively. Considering the great impact

and the success of NGS technique in the scientific world, it is absolutely appropriate to consider the possibility to target a longer amplicon, certainly more informative, ~~that such features will further improve~~ in the ~~next years~~future. This is the reason why both LAL and SAL universal primers were analysed in this study (**Table 2**). ~~In fact, even if to date all the platforms substantially required relatively short target template, it is not excluded the eventuality to utilize a longer one, certainly more informative, in the future. Therfore, w~~We decided to prioritize those primers that were able to amplify a fragment <500 bp, even if, on the basis of the obtained results, also the pair P1 could obviously be utilized in NGS studies in the case of future progresses ~~and enrichments of this new sequencing technique~~. In this case anyway, the primer pair C resulted to perform best, even if further studies aimed at its efficiency improvement could be required.

~~The available NGS studies inherent to species detection in mixed food source utilized *16SrRNA* as election molecular marker[17,25-27]. This gene has been shown to be a good marker also to differentiate fish species and it has been used in comparative intergeneric and interspecific studies in several fish families[39,44,45]. However, the *cytb* and *COI* genes, due to their comparable high interspecies variation and low intraspecific variation, are nowadays the most widely used genetic markers for fish species identification, as reported in a large number of studies applied to food inspection[8-10,46]. To date, the introduction of DNA barcoding has determined a growth in the use of the *COI* with respect to the *cytb* gene as a genetic marker for species identification and for biodiversity analysis. A wide variety of universal primers is now available for the amplification of the three genes reported above. For this reason, the goal of this study was to provide an as much as possible complete analysis of 14 pairs of universal primer targeting these three mitochondrial genes, also in order to establish if the *16SrRNA* can be effectively considered the best one or if the other genetic markers present some advantages. All the primers pairs, reported in **Table 2**, were selected due to the fact that they have been successfully tested of several fish and/or cephaloped species.~~

The analysis conducted in this study, taking in consideration and evaluating a broad number of universal primers sets able to give amplicons from all the three most utilized mitochondrial genes (*16SrRNA*, *cytb* and *COI*) in fish and cephalopods species detection, could represent a first step in the future advancement of this technique. The results of this study confirmed that the primers designed on the *16SrRNA* gene effectively showed a higher universality respect to those designed on *COI* or *cytb* genes and, therefore, they could be more suitable for the NGS analysis finalized to seafood species detection.

Metabarcoding NGS techniques could effectively become a turning point in the food inspection field, overcoming the limits of the standard analytical methods in the detection of multispecies matrices that are now spreading on the market in different shapes (fish sticks, fish cakes, fish balls, hamburgers). The current lack of a solid background of studies, especially in the seafood inspection field, highlights the necessity to further deepen, improve and consolidate this research field. In the case of seafood products, the extremely wide number of species potentially detectable on the market let easily perceive that one of the most substantial hindrance is represented by the research of suitable universal primers able to hypothetically discriminate any species in a seafood complex matrix. This would promote fair trade in the seafood industry, by preventing illegal substitution in the supply chain and contrast IUU fishing and overexploitation that frequently affect species most requested from the fishery. Finally, metabarcoding NGS techniques would increase consumers' protection level and prevent health frauds, when cephalopod species, which are known to be allergenic, are used.

Acknowledgments:

The authors wish to thank all of the people who kindly provided reference tissues of fish species utilized in this study and the anonymous reviewers for their valuable comments and suggestions.

Funding Institution

The present work was funded by the University of Pisa

Supporting Information description:

Table 1SM. Fish and cephalopod species reported in literature as utilized for surimi-based product preparation; in grey boxes: species effectively recovered in surimi-based products during forensic analysis; NR: not reported

Table 2SM. Amplification protocols and programs of each couple of primers selected in this study: [(a)(b)] 5Prime, Gaithersburg, USA; [(c)]dNTPmix, Euroclone S.p.A-Life Sciences Division; [(d)]PerfectTaq DNA Polymerase (5Prime, Gaithersburg, USA); [(e)]purified BSA 100×, New England Biolabs; [(f)]Water Mol. Bio. Grade, DNase−RNase and Protease free, 5Prime, Gaithersburg, USA; FV: up to the final volume achievement

Table 3.1SM. Alignment between the P1 and P2 couples of primers used in this study and the available 16SrRNA gene sequences of the fish and cephalopod species used in this work. Mismatches have been highlighted in grey. O: Order; F: Family; S: Species; NAS: Not available sequences; NUS: Not usable sequences

304 **Table 3.2SM.** Alignment between the C and CEP couples of primers used in this study and the available 16SrRNA gene sequences of the fish and
305 cephalopod species used in this work. Mismatches have been highlighted in grey. O: Order; F: Family; S: Species; NAS: Not available sequences;
306 NUS: Not usable sequences
307 **Table 4.1SM.** Alignment between the M and F couples of primers used in this study and the available COI gene sequences of the fish and
308 cephalopod species used in this work. Mismatches have been highlighted in grey. O: Order; F: Family; S: Species; NAS: Not available sequences;
309 NUS: Not usable sequences
310 **Table 4.2SM.** Alignment between the H, SH and L couples of primers used in this study and the available COI gene sequences of the fish and
311 cephalopod species used in this work. Mismatches have been highlighted in grey. O: Order; F: Family; S: Species; NAS: Not available sequences;
312 NUS: Not usable sequences
313 **Table 5.1SM.** Alignment between the CB1 and CB2 couples of primers used in this study and the available cytb gene sequences of the fish species
314 used in this work. Mismatches have been highlighted in grey. O: Order; F: Family; S: Species; NAS: Not available sequences; NUS: Not usable
315 sequences
316 **Table 5.2SM.** Alignment between the Ccb, SL and SS couples of primers used in this study and the available cytb gene sequences of the fish and
317 cephalopod species used in this work. Mismatches have been highlighted in grey. O: Order; F: Family; S: Species; NAS: Not available sequences;
318 NUS: Not usable sequences

Table 6SM. PCR products concentration obtained from reference species DNA amplification. The estimation were done by comparison with the molecular marker SHARPMASS 50® (EuroClone SPA Figino al Pero, Milan). In grey boxes were highlighted the unamplified DNA (0) or the PCR products whose concentrations were below the limit required for sequencing

Table 7SM. BLAST analysis of the C pairs fragment on the available sequences analysed; In the second column obtained identity value of 100% were reported; in the third column, values immediately below than 100% were reported; H: haplotype.

References:

(1) Armani, A.; Castigliego, L.; & Guidi, A. Fish frauds: the DNA challenge. *Anim. Sci. Rev.* **2012**.
(2) Helyar, S. J.; Lloyd, H. A. D.; De Bruyn, M.; Leake, J.; Bennett, N.; Carvalho, G. R. Fish product mislabelling: Failings of traceability in the production chain and implications for Illegal, Unreported and Unregulated (IUU) fishing. *PLoS One* **2014**, *9* (6), 1–7.
(3) Teletchea, F. Molecular identification methods of fish species: Reassessment and possible applications. *Rev. Fish Biol. Fish.* **2009**, *19* (3), 265–293.
(4) Hebert, P. D. N.; Cywinska, A.; Ball, S. L.; Dewaard, J. R. Biological identifications through DNA barcodes. **2003**.
(5) Bartlett, S. E.; Davidson, W. S. FINS (forensically informative nucleotide sequencing): a procedure for identifying the animal origin of biological specimens. *Biotechniques* **1992**, *12* (3), 408–411.
(6) Rasmussen Hellberg, R. S.; Morrissey, M. T. Advances in DNA-based techniques for the detection of seafood species substitution on the commercial market. *J. Lab. Autom.* **2011**, *16* (4), 308–321.
(7) Hebert, P.; Ratnasingham, S.; DeWaard, J. R. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. London* **2003**, *270* (August 2003), 96–99.
(8) Hutchison, C. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res.* **2007**.
(9) Xiong, X.; Guardone, L.; Giusti, A.; Castigliego, L.; Gianfaldoni, D.; Guidi, A.; Armani, A. DNA barcoding reveals chaotic labeling and misrepresentation of cod (Xue) products sold on the Chinese market. *Food Control* **2016**, *60*, 519–532.
(10) Armani, A.; Guardone, L.; Castigliego, L.; D'Amico, P.; Messina, A.; Malandra, R.; Gianfaldoni, D.; Guidi, A. DNA and Mini-DNA barcoding for the identification of Porgies species (family Sparidae) of commercial interest on the international market. *Food Control* **2015**, *50*, 589–596.
(11) Armani, A.; Guardone, L.; La Castellana, R.; Gianfaldoni, D.; Guidi, A.; Castigliego, L. DNA barcoding reveals commercial and health issues in ethnic seafood sold on the Italian market. *Food Control* **2015**, *55*, 206–214.
(12) Di Pinto, A.; Di Pinto, P.; Terio, V.; Bozzo, G.; Bonerba, E.; Ceci, E.; Tantillo, G. *DNA barcoding for detecting market substitution in salted*

cod fillets and battered cod chunks; 2013; Vol. 141.
(13) Cawthorn, D. M.; Steinman, H. A.; Witthuhn, R. C. DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. *Food Res. Int.* **2012**, *46* (1), 30–40.
(14) Barbuto, M.; Galimberti, A.; Ferri, E.; Labra, M.; Malandra, R.; Galli, P.; Casiraghi, M. DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of "palombo" (Mustelus spp.). *Food Res. Int.* **2010**, *43* (1), 376–381.
(15) Ripp, F.; Krombholz, C. F.; Liu, Y.; Weber, M.; Schäfer, A.; Schmidt, B.; Köppel, R.; Hankeln, T. All-Food-Seq (AFS): a quantifiable screen for species in biological samples by deep DNA sequencing. *BMC Genomics* **2014**, *15* (1), 639.
(16) Galal-Khallaf, A.; Ardura, A.; Borrell, Y. J.; Garcia-Vazquez, E. Towards more sustainable surimi? PCR-cloning approach for DNA barcoding reveals the use of species of low trophic level and aquaculture in Asian surimi. *Food Control* **2016**, *61* (January 2016), 62–69.
(17) Littlefair, J.; Clare, E. Barcoding the food chain: from Sanger to high-throughput sequencing 1. *Genome* **2016**.
(18) Metzker, M. L. Sequencing technologies - the next generation. *Nat. Rev. Genet.* **2010**, *11* (1), 31–46.
(19) Tillmar, A. O.; Dell'Amico, B.; Welander, J.; Holmlund, G. A universal method for species identification of mammals utilizing next generation sequencing for the analysis of DNA mixtures. *PLoS One* **2013**, *8* (12), 1–9.
(20) Green, R.; Krause, J.; Ptak, S.; Briggs, A.; Ronan, M. Analysis of one million base pairs of Neanderthal DNA. *Nature* **2006**.
(21) Miller, W.; Drautz, D.; Ratan, A.; Pusey, B.; Qi, J.; Lesk, A. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **2008**.
(22) Brandon-Mong, G.; Gan, H.; Sing, K. DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bull. Entomol. Res.* **2015**.
(23) Hancock□Hanser, B.; Frey, A.; Leslie, M. Targeted multiplex next□generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Mol. Ecol.* **2013**.
(24) Yu, D.; Ji, Y.; Emerson, B.; Wang, X.; Ye, C. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol.* **2012**.
(25) Hajibabaei, M.; Shokralla, S.; Zhou, X.; Singer, G. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One* **2011**.
(26) Bertolini, F.; Ghionda, M. C.; D'Alessandro, E.; Geraci, C.; Chiofalo, V.; Fontanesi, L. A next generation semiconductor based sequencing approach for the identification of meat species in DNA mixtures. *PLoS One* **2015**, *10* (4), 1–16.
(27) Coghlan, M.; Haile, J.; Houston, J.; Murray, D. Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. *PLoS Genet.* **2012**.
(28) Muñoz-Colmenero, M.; Martínez, J.; Roca, A. NGS tools for traceability in candies as high processed food products: Ion Torrent PGM versus conventional PCR-cloning. *Food Chem.* **2017**.
(29) Park, J.; Lee, S.; An, C.; Kang, J.; Kim, J.; Chai, J. Comparative study between Next Generation Sequencing Technique and identification of microarray for Species Identification within blended food products. *Biochip J.* **2012**.
(30) Vences, M.; Thomas, M.; van der Meijden, A.; Chiari, Y.; Vieites, D. R. Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. *Front. Zool.* **2005**, *2*, 5.
(31) Kochzius, M.; Seidel, C.; Antoniou, A.; Botla, S. Identifying fishes through DNA barcodes and microarrays. *PLoS One* **2010**.

(32) Shokralla, S.; Spall, J.; Gibson, J. Next□generation sequencing technologies for environmental DNA research. *Molecular* **2012**.
(33) Fouts, D.; Szpakowski, S.; Purushe, J.; Torralba, M. Next generation sequencing to define prokaryotic and fungal diversity in the bovine rumen. *PLoS One* **2012**.
(34) Claesson, M.; Wang, Q.; O'sullivan, O. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic acids* **2010**.
(35) Kwok, S.; Chang, S.; Sninsky, J. Design and use of mismatched and degenerate primers. *PCR Prim. A Lab.* **1995**.
(36) Deagle, B. E.; Jarman, S. N.; Coissac, E.; Pompanon, F.; Taberlet, P. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biol. Lett.* **2014**, *10* (9).
(37) Dijk, E. van; Jaszczyszyn, Y.; Thermes, C. Library preparation methods for next-generation sequencing: tone down the bias. *Exp. Cell Res.* **2014**.
(38) Armani, A.; Castigliego, L.; Tinacci, L.; Gandini, G.; Gianfaldoni, D.; Guidi, A. A rapid PCR-RFLP method for the identification of Lophius species. *Eur. Food Res. Technol.* **2012**, *235* (2), 253–263.
(39) Armani, A.; Giusti, A.; Guardone, L.; Castigliego, L.; Gianfaldoni, D.; Guidi, A. Universal Primers Used for Species Identification of Foodstuff of Animal Origin: Effects of Oligonucleotide Tails on PCR Amplification and Sequencing Performance. *Food Anal. Methods* **2016**, *9* (5), 1199–1209.
(40) Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**.
(41) Armani, A.; Tinacci, L.; Xiong, X.; Castigliego, L.; Gianfaldoni, D.; Guidi, A. Fish species identification in canned pet food by BLAST and Forensically Informative Nucleotide Sequencing (FINS) analysis of short fragments of the mitochondrial 16s ribosomal RNA gene (16S rRNA). *Food Control* **2015**, *50*, 821–830.
(42) Cawthorn, D.-M.; Steinman, H. A.; Witthuhn, R. C. Evaluation of the 16S and 12S rRNA genes as universal markers for the identification of commercial fish species in South Africa. *Gene* **2012**, *491*, 40–48.
(43) Finizio, A. Di; Guerriero, G.; Russo, G. L.; Ciarcia, G. Identification of gadoid species (Pisces, Gadidae) by sequencing\nand PCR–RFLP analysis of mitochondrial 12S and 16S rRNA\ngene fragments. *Eur Food Res Technol* **2007**, *225*, 337–344.
(44) Pepe, T.; Trotta, M.; Di Marco, I.; Anastasio, A.; Bautista, J. M.; Cortesi, M. L. Fish Species Identification in Surimi-Based Products. *J. Agric. Food Chem.* **2007**, *55* (9), 3681–3685.
(45) Ursvik, A.; Breines, R.; Christiansen, J. S.; Fevolden, S.-E.; Coucheron, D. H.; Johansen, S. D. A mitogenomic approach to the taxonomy of pollocks: Theragra chalcogramma and T. finnmarchica represent one single species. *BMC Evol. Biol.* **2007**, *7*, 86.
(46) Head, S. R.; Kiyomi Komori, H.; LaMere, S. A.; Whisenant, T.; Van Nieuwerburgh, F.; Salomon, D. R.; Ordoukhanian, P. Library construction for next-generation sequencing: Overviews and challenges. *Biotechniques* **2014**, *56* (2), 61–77.
(47) Sevilla, R. G.; DIEZ, A.; NORÉN, M.; MOUCHEL, O.; JÉRÔME, M.; VERREZ-BAGNIS, V.; VAN PELT, H.; FAVRE-KREY, L.; KREY, G.; CONSORTIUM, T. F.; et al. Primers and polymerase chain reaction conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome b and nuclear rhodopsin genes. *Mol. Ecol. Notes* **2007**, *7* (5), 730–734.
(48) Verma, S.; Singh, L. Novel universal primers establish identity of an enormous number of animal species for forensic application. *Mol. Ecol.*

Notes **2003**.
(49) Chapela, M. J.; Sotelo, C. G.; Calo-Mata, P.; Pérez-Martín, R. I.; Rehbein, H.; Hold, G. L.; Quinteiro, J.; Rey-Méndez, M.; Rosa, C.; Santos, A. T. Identification of Cephalopod Species (Ommastrephidae and Loliginidae) in Seafood Products by Forensically Informative Nucleotide Sequencing (FINS). *J. FOOD Sci.* **2002**, *67* (5).
(50) Santaclara, F. J.; Espiñeira, M.; Vieites, J. M. Genetic Identification of Squids (Families Ommastrephidae and Loliginidae) by PCR–RFLP and FINS Methodologies. *J. Agric. Food Chem.* **2007**, *55* (24), 9913–9920.
(51) Palumbi, S. R. Nucleic acids II: the polymerase chain reaction. *Mol. Syst.* **1996**, *2* (1), 205–247.
(52) Armani, A.; Castigliego, L.; Tinacci, L.; Gianfaldoni, D. Multiplex conventional and real-time PCR for fish species identification of Bianchetto (juvenile form of Sardina pilchardus), Rossetto (Aphia minuta), and Icefish in. *Food Chem.* **2012**.
(53) Folmer, 0; Black, M.; Hoeh, W.; Lutz, R.; Vrijenhoek+, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **1994**, *3* (5), 294–299.
(54) Mikkelsen, P.; Bieler, R.; Kappner, I. Phylogeny of Veneroidea (Mollusca: Bivalvia) based on morphology and molecules. *Zool. J. Linn. Soc.* **2006**, *148*, 439–521.
(55) 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. A single-laboratory validated method for the generation of DNA barcodes for the identification of fish for regulatory compliance. *J. AOAC Int.* **2011**, *94* (1), 201–210.
(56) Leray, M.; Yang, J.; Meyer, C. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral. *Front. Zool.* **2013**, *10* (1), 1.

Figure caption:

Figure 1: workflow illustration of the whole process of primers evaluation and output of each intermediate step

Tables:

Table 1. Fish and cephalopods species utilized for surimi preparation collected in this study; **in grey boxes:** species not reported in Table 1SM selected and collected in this study.

Order	Family	Species	Research Institute
FISHES			
Clupeiformes	Clupeidae	*Clupea harengus*	Fishlab – Depatment of Veterinary Science – University of Pisa

		Clupea pallasii	Maslenikov – University of Washington Fish Collection School os Aquatic and Fishery Sciences and Burke Museum of Natural History and Culture
		Sardina pilchardus	Fishlab – Depatment of Veterinary Science – University of Pisa
Cypriniformes	Cyprinidae	*Ctenopharyngodon idella*	Korwin-Kossakowski – The Stanislaw Sakowicz Inland Fisheries Institute in Olsztyn, Pond Fishery Department in Zabieniec, Poland
		Cyprinus carpio	Stratev – Department of Food Hygiene and Control, Veterinary Legislation and Management
		Hypophthalmichthys molitrix	Korwin-Kossakowski – The Stanislaw Sakowicz Inland Fisheries Institute in Olsztyn, Pond Fishery Department in Zabieniec, Poland
		Hypophthalmichthys nobilis	Korwin-Kossakowski – The Stanislaw Sakowicz Inland Fisheries Institute in Olsztyn, Pond Fishery Department in Zabieniec, Poland
Gadiformes	Gadidae	*Gadus ogac*	CSIC – IIM, Instituto de Investigaciones Marinas – Vigo (Spain)
		Melanogrammus aeglefinus	Fishlab – Depatment of Veterinary Science – University of Pisa
		Micromesistius poutassou	Fishlab – Depatment of Veterinary Science – University of Pisa
		Theragra chalcogramma	NOAA Fisheries, Alaska Fisheries Science Center
	Merluccidae	*Macruronus magellanicus*	CSIC – IIM, Instituto de Investigaciones Marinas – Vigo (Spain)
		Merluccius australis	CSIC – IIM, Instituto de Investigaciones Marinas – Vigo (Spain)
		Merluccius capensis	CSIC – IIM, Instituto de Investigaciones Marinas – Vigo (Spain)
		Merluccius gayi	IZSTO - Istituto Zooprofilattico Sperimentale del Piemonte, Liguriae Valle d'Aosta - Torino
		Merluccius hubbsi	IZSTO - Istituto Zooprofilattico Sperimentale del Piemonte, Liguriae Valle d'Aosta - Torino
		Merluccius merluccius	CSIC – IIM, Instituto de Investigaciones Marinas – Vigo (Spain)
		Merluccius productus	Maslenikov – University of Washington Fish Collection School os Aquatic and Fishery Sciences and Burke Museum of Natural History and Culture
Perciformes	Carangidae	*Trachurus japonicus*	Fishlab – Depatment of Veterinary Science – University of Pisa
		Trachurus picturatus	Fishlab – Depatment of Veterinary Science – University of Pisa

	Cichlidae	*Oreochromis aureus*	Penman Institute of Aquacolture - University of Stirling - Scotland
		Oreochromis mossambicus	Penman Institute of Aquacolture - University of Stirling - Scotland
		Oreochromis niloticus	Fishlab – Depatment of Veterinary Science – University of Pisa
	Mullidae	*Parupeneus indicus*	Max Rubner Institut – Department of Safety and Quality of Milk and Fish Products – Hamburg - Germany
		Pseudupeneus prayensis	Fishlab – Depatment of Veterinary Science – University of Pisa
		Upeneus tragula	Fishlab – Depatment of Veterinary Science – University of Pisa
	Nemipteridae	*Nemipterus furcosus*	Fishlab – Depatment of Veterinary Science – University of Pisa
		Nemipterus japonicus	Department of Biology – Chinese University of Hong Kong
		Nemipterus virgatus	Fishlab – Depatment of Veterinary Science – University of Pisa
	Priacanthidae	*Priacanthus macracanthus*	Department of Biology – Chinese University of Hong Kong
	Trichiuridae	*Trichiurus lepturus*	Fishlab – Depatment of Veterinary Science – University of Pisa
	Scianeidae	*Larimichthys crocea*	IZSTO - Istituto Zooprofilattico Sperimentale del Piemonte, Liguriae Valle d'Aosta - Torino
		Larimichthys polyactis	IZSTO - Istituto Zooprofilattico Sperimentale del Piemonte, Liguriae Valle d'Aosta - Torino
	Sparidae	*Evynnis cardinalis*	Biodiversity Research Center Academia Sinica – Nankang, Taipei, Taiwan
		Evynnis tumifrons	Biodiversity Research Center Academia Sinica – Nankang, Taipei, Taiwan
	Sphyraenidae	*Sphyraena sphyraena*	IZSTO - Istituto Zooprofilattico Sperimentale del Piemonte, Liguriae Valle d'Aosta - Torino
Pleuronectiformes	Pleuronectidae	*Atheresthes evermanni*	Maslenikov – University of Washington Fish Collection School os Aquatic and Fishery Sciences and Burke Museum of Natural History and Culture
		Atheresthes stomias	Biodiversity Institute and Natural History Museum – University of Kansas
			Fishlab – Depatment of Veterinary Science – University of Pisa

Salmoniformes	Salmonidae	*Oncorhynchus gorbuscha*	Maslenikov – University of Washington Fish Collection School os Aquatic and Fishery Sciences and Burke Museum of Natural History and Culture
		Oncorhynchus keta	Maslenikov – University of Washington Fish Collection School os Aquatic and Fishery Sciences and Burke Museum of Natural History and Culture
		Oncorhynchus mykiss	Fishlab – Depatment of Veterinary Science – University of Pisa
Scorpaeniformes	Anoplopomatidae	*Anoplopoma fimbria*	Maslenikov – University of Washington Fish Collection School os Aquatic and Fishery Sciences and Burke Museum of Natural History and Culture
	Hexagrammidae	*Pleurogrammus monopterygius*	Maslenikov – University of Washington Fish Collection School os Aquatic and Fishery Sciences and Burke Museum of Natural History and Culture
	Sebastidae	*Helicolenus barathri*	CSIC – IIM, Instituto de Investigaciones Marinas – Vigo (Spain)
	Triglidae	*Chelidonichthys lucernus*	Fishlab – Depatment of Veterinary Science – University of Pisa
Siluriformes	Pangasiidae	*Pangasianodon hypophthalmus*	CSIC – IIM, Instituto de Investigaciones Marinas – Vigo (Spain)
CEPHALOPODS			
Myopsida	Loliginidae	*Doryteuthis gahi*	Fishlab – Depatment of Veterinary Science – University of Pisa
		Doryteuthis pealeii	Fishlab – Depatment of Veterinary Science – University of Pisa
Teuthida	Ommastrephidae	*Dosidicus gigas*	Fishlab – Depatment of Veterinary Science – University of Pisa
		Ommastrephes bartramii	Fishlab – Depatment of Veterinary Science – University of Pisa

439

440 **Table 2.** List of the primers considered in this study; LAL: Long Amplicon Length; SAL: Short Amplicon Length; NR: Not reported due to the
441 impossibility of matching the primer on the target sequence; [(a)] Evaluated with mitochondrial complete genome sequence of *G. chalcogrammus*
442 (GenBank accession code: AB182308); [(b)] Evaluated with mitochondrial complete genome sequence of *D. gigas* (GenBank accession code:
443 NC_009734)

Code	Group	Target gene	Original primer name	Primer sequences (5'-3')	Amplicon length without primers (base pairs)	Melting Temperature (°C)	References	Target

CB1	LAL	*cytb*	GluFish-F	AACCACCGTTGTTATTCAACTACAA	806[(a)] NR[(b)]	58.1	47	Teleost fishes
			CytBI-4R	AGGAAGTATCATTCGGGCTTAATATG		60.1		
CB2	SAL		mcb398	TACCATGAGGACAAATATCATTCTG	421[(a)] NR[(b)]	58.1	48	Mammals, birds, reptiles, fishes
			mcb869	CCTCCTAGTTTGTTAGGGATTGATCG		63.2		
Ccb	SAL		Cytb1F	CAGCTATTCCATATGTTGGTGA	297[(a)(b)]	56.5	49	Loliginidae and Ommastrephidae species
			Cytb1R	GGTTACTAAAGGATTAGCTGGA		56.5		
SL	LAL		CEF-H	TTATGGKTGRGTRYTDCGTTAT	605[(a)(b)]	55.6	50	Loliginidae, Ommastrephidae, Sepiidae and Octopodidae species
			CEF-L	TACHCCYCCWARTTTWYTAGGAAT		57.6		
SS	SAL		CEF-H	TTATGGKTGRGTRYTDCGTTAT	160[(a)(b)]	55.6		
			H15149AD	GCICCTCARAATGAYATTTGTCCTCA		62.4		
P1	LAL	*16SrRNA*	16sar-L	CGCCTGTTTATCAAAAACAT	571[(a)] 511[(b)]	51.1	51	All animal species
			16sbr-H	CCGGTCTGAACTCAGATCACGT		62.1		
P2	SAL		FOR16Spc	TGCCCGTGCAGAAGCGG	295[(a)] 339[(b)]	60.0	52	Fish species belonging to Clupeidae, Engraulidae, Salangidae, Scombridae
			REV16Spc	CAACATCGAGGTCGTAAACCC		59.8		
C			16sf-var	CAAATTACGCTGTTATCCCTATGG	209[(a)] 148[(b)]	59.3	49	Cephalopod species belonging to

			16sr-var	GACGAGAAGACCCTAATGAGCTTT		61.0		Ommastrephidae and Loliginidae
CEP			16FORF-CEP3	GAGAAGACCCTDTKGAGCTT	206[(a)] 145[(b)]	57.0	Modified from-[41]	Cephalopod species
			16REVF-CEP	GCTGTTATCCCTAKGGTAAC		56.3		
F	LAL	*COI*	LCO1490	GGTCAACAAATCATAAAGATATTGG	658[(a)]	56.4	[53]	Metazoan invertebrates
			HC02198	TAAACTTCAGGGTGACCAAAAAATCA		58.5		
M			COIF-ALT	ACAAATCAYAARGAYATYGG	658[(a)] 657[(b)]	51.1	[54]	Mollusca
			COIR-ALT	TTCAGGRTGNCCRAARAAYCA		56.9		
H			FISHCOILBC	CTCAACYAATCAYAAAGATATYGGCAC	655[(a)(b)]	61.2	[55]	Fishes, mammals and birds
			FISHCOIHBC	ACTTCYGGGTGRCCRAARAATCA		60.6		
L	SAL		mlCOIintF	GGWACWGGWTGAACWGTWTAYCCYCC	313[(a)(b)]	64.8	[56]	Fishes
			jgHCO2198	TAIACYTCIGGRTGICCRAARAAYCA		63.2		
SH			FISHCOILBC	CTCAACYAATCAYAAAGATATYGGCAC	139[(a)(b)]	61.2	[55]	Fishes
			Revshort1	GGYATNACTATRAAGAAAATTATTAC		54.5	[10]	

Table 3. Primers *in silico* evaluation summary. GMP**:** Good Mismatches Position, which indicate the percentage of sequences, among those that presented mismatches, that do not have any mismatches on the first four bp near the 3'end. NR: Not reported due to the impossibility of matching the primer on the target sequence.

Gene	Primers set	Primers name	Analysed fish sequences number	Mismatches number			GMP	Analysed cephalopod sequences number	Mismatches number			GMP
				0	1-2	≥3			0	1-2	≥3	
16SrRNA	P1	16sar-L	56	1.8%	98.2%	0%	94.5%	4	100%	0%	0%	-
		16sbr-H	50	100%	0%	0%	-	4	100%	0%	0%	-
	P2	FOR16Spc	78	14.1%	83.3%	2.6%	100%	5	0%	100%	0%	100%
		REV16Spc	75	100%	0%	0%	-	5	100%	0%	0%	-
	C	16sf-var	78	0%	0%	100%	0%	5	0%	100%	0%	0%
		16sr-var	80	0%	52.5%	47.5%	83.8%	5	0%	100%	0%	0%

	CEP	16FORF-CEP3	77	100%	0%	0%	-	5	100%	0%	0%	-
		16REVF-CEP	77	83.1%	16.9%	0%	100%	5	100%	0%	0%	-
Cytb	CB1	GluFish-F	45	51.1%	26.7%	22.2%	40.9%	-	NR	NR	NR	-
		CytBI-4R	61	3.3%	77%	19.7%	66.1%	-	NR	NR	NR	-
	CB2	mcb398	72	0%	27.8%	72.2%	65.3%	-	NR	NR	NR	-
		mcb869	68	0%	8.8%	91.2%	41.2%	-	NR	NR	NR	-
	Ccb	Cytb1F	63	0%	0%	100%	7.9%	5	0%	20%	80%	20%
		Cytb1R	67	0%	0%	100%	3%	5	0%	20%	80%	20%
	SL	CEF-H	67	0%	0%	100%	0%	3	66.7%	33.3%	0%	100%
		CEF-L	67	0%	20.9%	79.1%	19.4%	3	0%	66.7%	33.3%	66.7%
	SS	H15149AD	68	16.2%	73.5%	10.3%	100%	6	0%	16.7%	83.3%	100%
COI	H	FISHCOILBC	41	7.3%	75.6%	17%	86.8%	2	0%	50%	50%	0%
		FISHCOIHBC	48	20.8%	70.8%	8.3%	76.3%	3	0%	100%	0%	66.7%
	M	COIF-ALT	43	2.2%	95.6%	2.2%	100%	3	33.3%	66.7%	0%	100%
		COIR-ALT	48	25%	75%	0%	5.6%	4	0%	100%	0%	0%
	F	LCO1490	43	2.3%	0%	97.7%	59.5%	3	0%	0%	100%	100%
		HC02198	49	2%	10.2%	87.8%	35.4%	3	0%	0%	100%	33.3%
	SH	Revshort1	111	47.7%	49.5%	2.7%	100%	6	16.7%	0%	83.3%	100%
	L	mlCOIintF	129	10.9%	58.9%	30.2%	67%	5	50%	50%	0%	50%
		jgHCO2198	48	95.8%	2.1%	2.1%	0%	3	100%	0%	0%	-

Table 4. Amplification rate (expected bands obtained/n° of DNA samples amplified) and average PCR products concentration of the different couples of primers on all the fish and cephalopod species tested in this study; SD: Standard deviation.

Gene	Primer couple	Amplification rate		Average amplicon concentration (±SD) (ng/µl)	
		Fish species	Cephalopod species	Fish species	Cephalopod species
16SrRNA	P1	100%	100%	20.6 (±5.16)	20 (±4)
	P2	100%	100%	17.6 (±3.96)	5 (±0)
	C	97.8%	100%	20.9 (±5.23)	30 (±0)
	CEP	100%	100%	19.6 (±1.99)	20 (±0)
COI	F	34.7%	100%	6.8 (±9.82)	25 (±0)
	L	0%	0%	0 (-)	0 (-)
	H	100%	0%	20 (±5.59)	0 (-)
	M	71.4%	100%	9.6(±7.82)	25 (±0)
	SH	95.9%	25%	17.3 (±6.54)	1.25 (±2.5)

452 **Table 5**. Schematically primers performance output evaluation. According to Table 4, for the amplification rate, the sign (+) was assigned to
453 average percentages >90%, while percentages <90% were indicated as (-). For the amplicon concentration, the sign (+) was assigned to average
454 concentrations >15 ng/ µl , while the sign (-) was assigned to average concentration <15 ng/ µl. For the fragment length, according to Table 2, LAL
455 were indicated with the sign (-) and SAL with the sign (+). The performance evaluation was indicated with the sign (+) if all the respective column
456 reported the sign (+).

Gene	Primers couple	Fish species amplificability	Fish species amplicon concentration	Cephalopod species amplificability	Cephalopod species amplicon concentration	Fragment lenght	Performance evaluation
16SrRNA	P1	+	+	+	+	-	-
	P2	+	+	+	-	+	-
	C	+	+	+	+	+	+
	CEP	+	+	+	+	+	+
COI	H	+	+	-	-	-	-
	L	-	-	-	-	-	-
	M	-	-	+	+	-	-
	F	-	-	+	+	-	-
	SH	+	+/-	+	-	+	-

457