Elsevier Editorial System(tm) for Food Control

Manuscript Draft

Manuscript Number: FOODCONT-D-16-02836R1

Title: Is raw better? A multiple DNA barcoding approach (full and mini) based on mitochondrial and nuclear markers reveals low rates of misdescription in sushi products sold on the Italian market.

Article Type: Research Paper

Keywords: Sushi products, DNA barcoding, COI gene, 16S rRNA gene, PEPCK gene, misdescription,

Corresponding Author: Dr. Andrea Armani,

Corresponding Author's Institution: University of Pisa

First Author: Andrea Armani

Order of Authors: Andrea Armani; Lara Tinacci; Raniero Lorenzetti; Alessio Benvenuti; Francesca Susini; Laura Gasperetti; Enrica Ricci; Marcella Guarducci; Alessandra Guidi

Abstract: New dietary habits have favored an ever growing popularity of Eastern country cooking style and in particular of sushi. Even though the Reg. (EU) 1379/2013 does not apply to restaurants and caterers, the Reg. (EU) 1169/2011 establishes that all the information they provided to the final consumer have to meet the transparency requirements as regards the description of the ingredients used for the preparation of food. The present study aimed at performing a molecular based survey to identify the seafood species used in the sushi preparations at the retail level. A total of 185 samples were collected from sushi venues and supermarkets and DNA barcoding, followed by a pairwise divergence and Neighbour Joining clustering analysis, was applied in order to verify the information declared at purchase. A low misdescription rate (3.4%) was found, showing a proper training of Food Business Operators working in catering activities for what concerns labeling and consumer information. In addition, the common practice of proposing standardized menus involving recurrent species could have further limited the risk of misdescription occurrence.

Is raw better? A multiple DNA barcoding approach (full and mini) based on mitochondrial and nuclear markers reveals low rates of misdescription in sushi products sold on the Italian market.

Armani A.[1°*], Tinacci L.[1°], Lorenzetti R.[2], Benvenuti A.[1], Susini F.[3], Gasperetti L.[3], Ricci E.[3], Guarducci M.[3], Guidi A.[1]

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

[2] *Experimental Zooprophylactic Institute of Lazio and Tuscany, Rome District, via Appia Nuova, Rome*

[3] *Experimental Zooprophylactic Institute of Lazio and Tuscany, Pisa District, S.S. dell'Abetone e del Brennero 4, 56123 Pisa, (Italy)*

°These authors contributed equally to this work.

*corresponding author:

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

Tel: +390502210207

Fax: +390502210213

Email: andrea.armani@unipi.it

Abstract

New dietary habits have favored an ever growing popularity of Eastern country cooking style and in particular of sushi. Even though the Reg. (EU) 1379/2013 does not apply to restaurants and caterers, the Reg. (EU) 1169/2011 establishes that all the information they provided to the final consumer have to meet the transparency requirements as regards the description of the ingredients used for the preparation of food. The present study aimed at performing a molecular based survey to identify the seafood species used in the sushi preparations at the retail level. A total of 185 samples were collected from sushi venues and supermarkets and DNA barcoding, followed by a pairwise divergence and Neighbour Joining clustering analysis, was applied in order to verify the information declared at purchase. A low misdescription rate (3.4%) was found, showing a proper training of Food Business Operators working in catering activities for what concerns labeling and consumer information. In addition, the common practice of proposing standardized menus involving recurrent species could have further limited the occurrence of misdescriptions.

Keywords

Sushi products, DNA barcoding, *COI* gene, *16S rRNA* gene, PEPCK gene, misdescription,

1. Introduction

Ethnic foods are increasingly present in Western dietary habits as a result of last decades' trade globalization, innovation in conservation technologies, improvement of transportation networking and increasing migrations phenomena that contribute to the crossing and overlapping of different food spheres (Stano, 2015). Sushi, literally 'seasoned rice', is a typical Japanese bite-sized food prepared with acidified rice and various toppings and fillings based on seafood, seaweed and vegetables. This product has become increasingly popular among European consumers thanks to its nutritional properties and refined presentation (Mouritsen, 2009), perfectly matching the consumer appeal for a "high culinary aesthetic". Subsequently, the promotion of prepackaged ready to eat sushi, also available in supermarkets, contributed to the product accessibility (Hsin-I Feng, 2012). Finally, the increasing business around sushi has been boosted following the diffusion firstly in the United States and later in Europe of low cost sushi bars and take-away venues, generally owned by people of not-Japanese origin (Chietzka, 2005; Hsin-I Feng, 2012). In Italy, for example, sushi bars and wok-sushi are mainly managed by Chinese restaurateurs (Mudu, 2007). In these venues strong limits in hygiene management procedures and products traceability have been found, leading to lack of conformities in health and commercial requirements (Guidi *et al.*, 2010; Armani et al., 2015a).

A great variety of seafood is currently used for sushi making, including: tuna, salmon, swordfish, yellowtail, white muscle fishes (e.g. sea bass, gilt head seabream), preserved fish (e.g. smoked eel, smocked mackerel), octopus, squid, shrimps or prawns, scallop and flying fish roes (Mouritsen, 2009). Sushi based specialties are generally presented to the final consumer with the phonetic translation of the original Japanese script using Latin alphabet (Stano, 2015) together with a brief description of the ingredients. In particular, the name of the seafood category (e.g. "shrimp" "tuna") and not the specific official denomination is usually reported (authors' note).

The EU fishery and aquaculture products' market is regulated by Reg. (EU) 1379/2013, which introduced specific requirements for a common organization of the market and establishes traceability and labeling rules both for caught and farmed seafood, integrating the mandatory

provisions of Reg. (EU) 1169/2011 on food labeling. The Regulation applies to pre-packed and non-prepacked products sold along the supply chain and at retail level even including restaurants and caterers (Chapter I, Art. 5, g). However, restaurant owners are not required to detail all label information on their menus during administration to the final consumer, except for specific indications concerning allergens and mandatory information on specific products (raw fish and cooked crustaceans) that fully fall within the scope of Reg. (EU) 1379/2013 (Art. 35). Restaurateurs are nonetheless required to maintain and make available all the information for the authorities or the consumers at their request (D'Amico *et al.*, 2016). Therefore, all information provided to the consumer have to meet the transparency requirements defined by the EU Reg. (EU) 1169/2011 as regards the general description of the product.

Recent studies pointed out restaurants and catering activities as a potential weak link of the traceability system with respect to seafood mislabeling (Warner *et al.*, 2013; Bérnard-Capelle *et al.*, 2015; Vandamme *et al.*, 2016). Mislabeling and misrepresentation of seafood have major consequences for both consumers and producers in terms of human health risk and economic losses. Moreover, they can affect the conservation status of overfished or endangered species. Finally, they can foster illegal, unreported and unregulated (IUU) fishing (Mariani *et al.*, 2015; Helyar *et al.*, 2014; Jacquet and Pauly, 2008).

Molecular biology methods based on sequencing, particularly the DNA barcoding approach, have proven as effective tools in fish species identification. Mitochondrial DNA genes, have emerged as near-universal markers for this purpose (Armani *et al.*, 2016; Clark, 2015). At present the *COI* gene is the most targeted and exploited mitochondrial marker, thanks also to the ever improving international molecular identification system FISH-BOL (www.fishbol.org) and to the continuous updating of the reference sequences databases (Ward, 2012; Hanner *et al.*, 2011). However, nuclear genes can also represent alternatives target for species discrimination and phosphoenolpyruvate carboxykinase (*PEPCK*) and sodium–potassium ATPase a-subunit (*NaK*) have been successfully applied in phylogenetic studies within Penaeoidea providing a useful

instrument for the classification of these species (Ma *et al.*, 2009; Tsang *et al.*, 2008). Finally, the application of *COI* Mini DNA Barcoding protocols (Armani *et al.*, 2015a; Armani *et al.*, 2015b) or the selection of alternative genes, such as the *16S rRNA* gene (Armani *et al.*, 2016), represent a useful approach.

Barcoding techniques applied from 2009 to these days to investigate seafood labeling at the retail level (restaurant, grocery stores, take away venues) (Table 1), pointed out divergent results on the species substitution rate on the European and US market. From these studies a lower rate of frauds of the European market with respect to the US is evident. Given the few studies available on sushi products at the European level, conducted particularly in the UK, France and Belgium (Vandamme *et al.*, 2016; Bérnard-Capelle *et al.*, 2015; http://eu.oceana.org/sites/default/files/421/oceana_factsheet_seafood_fraud_brussels_eng.pdf) and the lack of similar studies in Italy, a two years-survey was carried out on sushi products, in particular *nigiri* (fish topped rice ball), *hosomaki* and *uramaki* (fish filled rice roll), directly purchased from restaurants, self-services, take-away bars and grocery markets located in four different provinces in Tuscany (Pisa, Florence, Leghorn and Lucca). The aim was to verify the authenticity of the products and assess the level of misdescription by a multiple DNA barcoding approach (full and mini) based on mitochondrial and nuclear markers. Results will also provide data on the rate of species substitution at the end point of the seafood chain.

2. Material and methods

2.1 Sample collection and storage

Sushi products were directly purchased from 23 sushi restaurants and supermarkets. Each product was composed of a variable number (3-8) of different types of pieces (*nigiri*, *hosomaki*, *uramaki)*, which correspond to the samples singularly analyzed in this work (Figure 1). The sampling was conducted in two sampling cycles, the first from April to October 2014 (88 samples) and the second from March to September 2015 (97 samples) for a total of 185 samples (Table 1SM). All the products were stored at -20°C until the DNA analysis.

131 For all the take-away samples purchased from restaurants, information on seafood used in the
132 preparation was directly collected from the menus available online or in the venues or from details
133 orally reported by the caterers; for pre-packaged products purchased at supermarkets label
134 information was transcribed.

135 ***2.2 Molecular analysis***

136 *2.2.1 DNA extraction, quantification and evaluation of DNA fragmentation.* Total DNA
137 extraction for each sample was performed in double according to the salting out procedure proposed
138 by Armani *et al.*, 2014 starting from 150 mg of tissues in the case of fishes, cephalopods and
139 crustacean samples and from 250 mg of tissue in the case of eggs samples. Final DNA
140 concentration and purity were evaluated with Nanodrop ND-1000 spectrophotometer (NanoDrop
141 Technologies, Wilmington, DE, US) by two subsequent measurements of the absorbance value at
142 260nm and calculation of A260/A280 and of A260/230 ratio according to manufacturer's
143 indications (http://www.nanodrop.com/Library/T009-NanoDrop%201000-&-NanoDrop%208000-
144 Nucleic-Acid-Purity-Ratios.pdf). A260/A280 $\geq$ 2.0 and A260/A230 $\geq$ 2.0 ratios were considered as
145 top values of nucleic acid purity and ratio values A260/230 >1.80 as the minimum value for
146 satisfactory nucleic acid purity.

147 One thousand ng of the total DNA extracted from each sample were run on 1% agarose gel
148 (GellyPhorLE®, Euroclone, Pero, MI) previously stained with GelRed™ Nucleid Acid Gel Stain
149 (Biotium, Hayward, CA, USA) and visualized under UV light. The degree of DNA fragmentation
150 was assessed by comparison to the standard marker SharpMass™50- DNA ladder and
151 SharpMass™1-DNA ladder (Euroclone, Wetherby, UK). According to the UV visible pattern the
152 samples were classified as 1) low or not fragmented (>1000 bp or complete); 2) medium
153 fragmented (500-1000 bp) and 3) highly fragmented/ Not Evaluable (<500bp or NE).

154 *2.2.2 DNA Amplification and Sequencing.* Cytochrome C oxidase subunit I (*COI*) gene was
155 selected as the "first choice target gene" for fish and mollusks and crustacean identification**.** The
156 PCRs were set according to the DNA fragmentation level highlighted by the previous

electrophoresis: for all those DNA samples showing a medium-low fragmentation a 655-658bp Full Length Barcode (FLB) was firstly amplified for the three sample categories (fish, mollusks, crustacean) by the application of three distinct primers couple proposed by Handy *et al.*, (2011), Mikkelsen *et al.*, (2006) and Folmer *et al.*, (1994) respectively. For the amplification of total DNA extracted from crustacean tissues a second target gene, the nuclear Phosphoenolpyruvate carboxykinase (*PEPCK*), was selected and the primers couple PEPCK2-FOR/PEPCK3-REV (Tsang *et al.*, 2008) for the amplification of a 598 bp fragment of the gene was applied in case of failure of the first amplification protocol.

In case of highly fragmented DNA or failure of the FLB amplification a Mini-DNA Barcoding (MDB) protocol proposed by Armani *et al*., 2015b was selected for the amplification of a 139 bp *COI* fragment. Finally, for two highly fragmented DNA samples belonging to scallop tissue that have failed both long fragment and MDB *COI* amplification, the primer couple proposed by Armani *et al.*, 2015c was used for the amplification of a fragment of around 118 bp of the *16SrRNA* gene. All the PCR reactions were set in a final volume of 20 µl containing 2 µl of a 10 X PCR buffer (5 Prime, Gaithersburg, USA), 100 ng of total DNA, 100 mM of each dNTP (Euroclone Spa, Milano), 250 nM of each primer, 25 ng/mL of BSA (New England BioLabs, Inc. USA), 1.25U PerfectTaq DNA Polymerase (5 Prime, Gaithersburg, USA), and DNase free sterile water (5 Prime, Gaithersburg, USA). Primers information and amplification programs applied in the study are summarized in Tables 2 and 2SM.

All the PCR products (5 µL) were checked on a 1.8% agarose gel (GellyPhorLE, Euroclone, Milano) stained with GelRed™ Nucleid Acid Gel Stain (Biotium, Hayward, CA, USA) and the presence of fragments of the expected length was assessed by comparison with the standard molecular marker SharpMass™50-DNA. PCR products were purified with EuroSAP PCR Enzymatic Clean-up kit (EuroClone Spa, Milano) according to the manufacturer instructions and finally stored at -80°C up to the further sequencing. DNA sequencing was carried out by Biotechnology Office belonging to Lazio and Tuscany's Experimental Zooprophylactic Institute,

183 Rome district (Rome, Italy) to obtain a sequence in forward and reverse direction for each PCR
184 product.
185 *2.2.3 Sequence analysis, comparison with databases and phylogenetic analysis*. The forward and
186 reverse direction sequences were manually checked and edited with Bioedit 7.0 software (Hall,
187 1999) and then aligned by the use of the software Clustal W. The final sequences were queried by
188 Basic Local Analysis Search Tool (BLAST) and Identification System (ID's) (Ratnasingham &
189 Hebert, 2007) against the reference sequences available on GenBank (http://www.ncbi.nlm.nih.gov)
190 and BOLD (http://www.boldsystems.org/) databases. The highest similarities percentages obtained
191 within the first 100 top match records by BLAST and ID's query were registered (Table 1SM). As
192 regards the *COI* barcode the specimen identification at species level was considered achieved when
193 the identity rate showed less than 2% difference with the reference sequences (Barbuto *et al.*, 2010).
194 In case of *PEPCK* and genes the identity score of 100% was used as as cut-off for the species
195 identification (Armani *et al.*, 2015c). Moreover, in the case of *16SrRNA* gene, given the limited
196 length of the fragment, the lowest expect value (E-value) (0.0) was taken into consideration to
197 evaluate the significance of the identity score obtained. The identity values were then verified by
198 pairwise divergence and Neighbour Joining clustering analysis. To this purpose *COI*, PEPCK and
199 *16S rRNA* gene reference sequences were collected from BOLD and GenBank. In particular, *COI*
200 gene reference sequences were collected for some species belonging to Scombridae, Moronidae,
201 Sparidae, Osmeridae, Exocoetidae, Octopodidae, Loliginidae and Sepiidae families and
202 Istiophoriformes Anguilliformes and Beloniformes orders. PEPCK gene reference sequences were
203 collected for the Penaeidae family and *16S rRNA* gene reference sequences for the Pectinidae
204 family. Up to five vouchered sequences per species were collected (for details on sequences see
205 Table 7SM). These sequences, together with those obtained in this study from commercial samples,
206 were used to produce distinct sequence clusters and alignment datasets. A separate phylogenetic
207 analysis was performed for each categories for which the comparison with databases did not
208 allowed unequivocal species identification. As regard fish products one dataset was produced

grouping together sequences obtained from seabass, seabream, salmon, swordfish, salmon roe, lumpfish roe and flying fish roe, one grouping the sequences obtaining from tuna and one those obtained from eel. Three dedicated dataset were also obtained for cephalopods shrimps and scallops, respectively. In addition, for each sequence cluster one or two datasets were produced on the basis of the sequence length obtained from the commercial samples (long or short fragment) (see Table 4SM). For each sequence cluster a pairwise distance matrices within and between species by the use of p-distance model with 1000 non-parametric bootstrap replicates were produced. Finally, unrooted Neighbour Joining (NJ) dendrograms with 1000 bootstrap re-samplings (Saitou & Nei, 1987) were prepared in order to visualize the clustering pattern. All the analysis were computed on MEGA 7.0 software package. Totally 10 trees were produced (see Table 4SM).

2.3 Comparison of the molecular results with purchasing information

The molecular results obtained were compared to the information collected at purchase points. The samples were declared mislabeled when the commercial name provided by the official Italian list (Min. Decree MIPAAF, 2008), corresponding to the species molecularly identified, did not matched with the name of the seafood category declared for that product.

Results and discussion

3.1 Sample collection

The sampling strategy plays a key role for the independence of the data analyzed (Vandamme *et al.*, 2016). Therefore, the sampling plan was designed on four separate provinces, in order to reduce repeated sampling on a single sushi venue. In addition, the collection was done over an extensive period of time so as to guarantee the sampling of independent product lots and suppliers for the single restaurateur. Regarding the sampling size, the final number collected in the study (N=185) is consistent with that analyzed both in Europe and US in the latest 5 year studies on the same subject (Table 1).

Sushi products prepared with 5 different categories of seafood were collected. Specifically, 78% of the samples contained fresh fish (71.3%) or processed fish (7%) whereas, the remaining 23%

were mollusks, crustaceans or fish eggs (7.5%; 9.7% and 4.5% respectively). Samples details are reported in Table 1SM. More than 50% of fish samples consisted of fresh tuna. These kind of products were highly represented since they are widely used in the preparation of all three sushi types (*nigiri*, *hosomaki*, *uramaki*) and they are commonly known and frequently purchased by consumers (Girard & Mariojouls, 2008; Verbeke *et al*., 2007). The "white fish" category was the second most represented group in sample numbers (16.2%).

3.2 Molecular analysis

3.2.1 DNA quality and fragmentation. The spectrophotometric analysis confirmed medium high yield and quality (A260/A280 and A260/A230 ratio >2.0) (Table 3) for almost all of the total DNA samples analyzed with the exception of the samples belonging to mollusks bivalves or fish eggs and highly processed fish (canned tuna and smoked/cooked eel) samples for which medium low yield (average value of 0,09 µg/mg, 0.03 µg/mg and 0.06 µg/mg, respectively) and low A260/A230 ratio (<1.3, <1.2, <1.4 respectively) were registered. In the case of mollusks and fish eggs the lower DNA recovery and the decrease of the quality were conceivably due to the presence of organic compounds (mucopolysaccharides, phenolic compounds and lipids) not effectively removed before the final DNA precipitation (Pereira *et al.*, 2011; Winnepenninckx *et al*., 1993). The low DNA yield and low spectrophotometric quality (and the high absorbance at 230nm) obtained for processed samples was plausibly related both to organic compounds residues (free fatty acids) and high content of free nucleic acids resulting from canning processing and DNA subsequent degradation (Armani *et al.*, 2015c).

Total DNA electrophoresis highlighted that 32% of DNA samples were highly fragmented (≤500 bp or NE), 30% medium fragmented (500-1000bp) and 38% low or not fragmented (>1000 bp) (Table 3, Table 3SM). Thermal treatments, ingredients mixing and storage conditions are the most important factors that can induce DNA degradation and, particularly, pH and temperature variations are known as major chemical-physical factors for depurination, hydrolysis, oxidation and subsequent DNA fragmentation (Teletchea, 2009; Armani *et al.*, 2015c). In this study the DNA

261 degradation could be induced by the use of a vinegar solution during sushi preparation. In fact, rice
262 acidification to 4.5 pH values is required for the inhibition of microbial growth (Lee, C. J., &
263 Heacock, 2016). Moreover, DNA fragmentation may have been induced by compulsory freezing
264 introduced in order to prevent and control parasitological risk related to presence of anisakid larvae
265 in seafood administered raw to the consumer (Regulation (EC) No 853/2004), followed by incorrect
266 thawing procedures and/or prolonged storage of the product (Armani *et al.*, 2015c; Rodriguez-
267 Ezpeleta *et al.*, 2013).

268 *3.2.2 DNA amplification and sequencing.* The *COI* gene has been selected as the first choice
269 considering that it has already been validated and selected as the preferred diagnostic marker for
270 molecular species identification (Dawnay *et al.*, 2007; Handy *et al.*, 2011). The two primer pairs
271 proposed by Handy *et al.*, (2011) and Mikkelsen *et al.*, (2006) were selected since they had already
272 been successfully applied for fish (Armani *et al.*, 2015b) and cephalopods (Yancy *et al.*, 2008)
273 DNA amplification. A single exception in the amplification setting was represented by the DNA
274 samples of *Dicentrarchus labrax*, for which, according to our lab experience, no PCR products
275 could be obtained with Handy *et al.*, (2011) primers. In this specific case, the amplification was
276 obtained by the application of a newly developed forward primer, FORUNICOI (Table 2) coupled
277 with the primer reverse used for the amplification of Full and Mini Barcodes. Concerning
278 crustacean DNA samples, a first amplification trial was set using a primer pair designed by Folmer
279 *et al.*, (1994) for the amplification of a 655 bp *COI* fragment. However, despite the successful
280 application of the primer pair in identification studies on decapoda (Rajkumar *et al.*, 2015; Bilgin *et*
281 *al.*, 2014), no PCR products or PCR products unsuitable for sequencing (low concentrated) were
282 obtained, even with repeated trials. In fact, successful PCR amplification of the *COI* FDB has
283 proven to be a challenge particularly in decapods (Costa et al., 2007; da Silva et al., 2011).
284 Moreover, the possible presence of nuclear mitochondrial pseudogenes has been highlighted for this
285 gene (Song et al., 2008; Gislason et al., 2013). Therefore, an alternative target was chosen for this
286 category. The nuclear PEPCK gene was selected as it had been previously applied in phylogeny and

molecular characterization studies of the superfamily Penaeoidea (Tsang *et al.*, 2008; Ma *et al.*, 2009). It proved to be performant in our preliminary amplification trials and it was successfully applied for the analysis of all the crustacean DNA samples.

On the basis of total DNA electrophoresis results (section 3.2.1) the MDB PCR protocol, already verified for the analysis of samples belonging to distant seafood taxa (Armani et al., 2015a) was applied for the amplification of evidently fragmented total DNA samples (< 500bp) (Table 3) or as alternative PCR protocol in absence of suitable long fragment PCR products. Finally, for the obtaining of a PCR product even from highly degraded mollusks bivalves DNA samples, the primers pair proposed by Armani *et al.*, (2015c) for the amplification a short fragment of *16S rRNA* gene was applied. In fact, also Marin *et al.*, (2015) selected the *16S rRNA* gene for the identification and species discrimination in Pectinidae.

Despite the high and medium DNA degradation, the application of the multiple DNA Barcoding approach based (mitochondrial and nuclear markers of different lengths) allowed us to recover at least one PCR product for each analyzed samples, giving an overall amplification success of 100%.

A total of 185 PCR products were collected and purified for further sequencing analysis. In particular, 165 *COI* gene PCR products of which 114 FDB and 51 MDB fragments; 18 PEPCK gene PCR products; 2 *16S rRNA* short fragment PCR products were sequenced.

Sequencing failures were registered only for seabass (n=2), octopus (n=3) and fish roe (n=2) samples (Table 1SM and Table 4). Therefore, interpretable sequences were obtained for 96.2% (178/185) of the PCR products, corresponding to a sequencing success of 99% for fish, 78% for mollusks, 100% for crustacean and 75% for PCR products belonging to fish roe total DNA (Table 4).

3.2.3 Sequence analysis and molecular identification. The final length of the sequences ranged from 69.3 to 100% of the expected amplicon. In particular, *COI* FDB sequences ranged from 454 to 655 bp, *COI* MDB sequences rangend from 120 to 139 bp, PEPCK sequences ranged from 440-538 bp and*16S rRNA* gene sequences measured 118 pb (Table 1SM).

3.2.4 Comparison with databases and phylogenetic distance analysis

By the combination of BLAST and BOLD ID's results all samples were identified at least at genus level (Table 1SM), with the exception of 6 MDB sequences belonging to tuna samples (see below). A species-specific level attribution was reached for 55.6% (103/185) of the samples analyzed (Table 4). In particular, all sequences derived from the commercial samples belonging to salmon (n=57), gilt-head seabream (n=13), sea bass (14), swordfish (n=1), salmon roe (n=2), lumpfish roe (n=2), flying fish roe (n=2), octopus (n=7), squid (n=1) and cuttlefish (n=1) allowed species specific identification with a top match identity value of 99-100% with a single species. The samples SUS 12.4 (flying fish roe) was the only one for which a specific identification was not achieved (neither with phylogenetical analysis): it was identified at genus level as *Cheilopogon* sp. (Fig 2SM). FDB and MDB showed the same discrimination power in the case of the aforesaid species (Table 1SM). The distance NJ tree method was applied to these samples to confirm the BLAST results and verify the species allocation. Specifically, 4 trees were obtained (see Table 4SM). All dendrograms showed distinct family clades and well-defined subclusters with Bootstrap Values (BV) always higher than 75% at both genus and species level (Fig. 1SM, 2SM, 3SM, 4SM). All the commercial samples were allocated within subclusters corresponding to the species previously highlighted by the comparison analysis confirming the identification at species level with the exception of the sample SUS 12.4, for which a MDB was obtained. On the contrary, MDBs obtained from cephalopods allowed to identify the commercial samples at species level as *O. vulgaris* (BV values of 97%).

About the 47 tuna products (44 fresh and 3 canned; see Section 3.1): 41 (87.2%) were confirmed belonging to the genus *Thunnus* sp. by the comparison of the FDB with database. The remaining 6 MDB (13%) (SUS2.3; SUS3.2; SUS25.2; SUS25.4; SUS26.3; SUS32.2), 3 of which originated from canned samples, could not be assigned to any genus since a top match of 100% was obtained both for *Thunnus* sp. and for the species *Katsuwonus pelamis* (skipjack tuna), due to the reduction of the discriminating power due to the decreasing length of the analyzed fragment (Armani *et al.*,

2015c). This aspect was confirmed by the p-distance analysis within the Scombridae family that highlighted a substantial decrease of the overall mean divergence estimated by the use of the MDB (0.035, SE 0.007) with respect to the FDB (0.059, SE 0.004). The same patterns were obtained within the *Thunnus* sp. In particular, the mean divergence values confirmed a very low interspecific distance estimate of 0.010 (SE 0.002) for the FDB fragment and 0.006 (SE 0.004) for the MDB barcodes. In particular, by the use of the MDB barcode, a mean decrease of the distance estimates values between *Thunnus* sp. and *Katsuwonus pelamis* from 0.095 to 0.013 was observed (Fig 5SM). The inferred NJ tree obtained by the analysis of FDB fragments for the tuna products showed six separate clusters strongly supported by BV (from 78 to 100%) collecting *Scomber scombrus*, *Sarda sarda*, *Auxis thazard*, *A.rochei*, *Katsuwonus pelamis* and *Thunnus* sp. Within *Thunnus* sp. two significant sub clusters were found, the first grouping *Thunnus orientalis* and *Thunnus alalunga* and the second one the remaining species belonging to the *Thunnus* genus. Although all the commercial samples were allocated within species-specific clusters they were not identified at species level (BV always lower than 75%, FIG 5SM). The NJ tree obtained by the application of MDB fragments showed three main clusters collecting *Scomber scombrus*, *Sarda sarda* separately and a unique third cluster for the genus *Thunnus* sp.; *Auxis* sp. and *Katsuwonus pelamis* (Fig. 6SM). Therefore, the analysis only allowed to identify these samples at family level.

Despite a higher discriminatory ability within the genus *Thunnus* sp. has been highlighted for other molecular markers such as cytochrome b (*cytb*) and nuclear First Internal Transcribed Spacer for rDNA (ITS-1) (Santaclara *et al*., 2015), the *COI* gene was enough informative for the purpose of this study (see section 3.3). Overall, the verification of the information furnished during the selling of the products sold as tuna was achieved for 87% of the samples. They were in fact all identified as belonging to the genus *Thunnus*.

Five FDB and 5 MDB were produced from the 10 products labeled as smoked eel. By the use of the BLAST analysis 4 products were directly assigned to a species. In particular, 1 was identified as *Anguilla japonica* (n=1, SUS32.1) by using FDB and 3 as *A. anguilla* (n=3, SUS7.1; SUS12.3; SUS

18.4) by MDB. One product was identified as *Conger* sp. (SUS21.4) by using MDB and 5 as *Anguilla* sp., since more than one species were included within the 98-100% range settled as cut-off for species identification. In particular, 4 out of the latter 5 products (SUS5.3, SUS8.2, SUS8.4, SUS13.4) identified as *Anguilla* sp. highlighted a top match identity with the species *Anguilla bicolor pacifica*. Therefore, a variable level of discriminatory ability was associated to FDB and MDB in the comparison with the databases. The results were all confirmed by the NJ analysis that was conducted separately for FDB and MDB barcodes (Fig. 7SM and 8SM). The FDB NJ tree showed well defined species clusters, all of them strongly supported by BV of 80 to 100%. SUS 32.1 was unambiguously identified as *A. japonica* and SUS5.3, SUS8.2, SUS8.4, SUS13.4 were confirmed to be *A. bicolor pacifica* (Fig. 7SM). The NJ analysis confirmed MDBs to be less efficient in species identification. About *Conger* sp. 4 distinct species clusters, supported by BV from 85 to 99% were highlighted for *C. japonicus*, *C. cinereus*, *C. triporiceps* and *C. orbignianus*; a fifth cluster collected the remaining species considered in the analysis (*C. conger*, *C. myriaster*, *C. oceanicus* and *C. wilsoni*) that, as already verified from pairwise divergence matrix, presented a reduced number of diagnostic sites within the MDB fragment (Fig. 8SM). As for *Anguilla* sp. two clusters were shown: the first collecting *A. rostrata* and *A. anguilla*, themselves separated into two distinct sub-clusters, the second grouping all the remaining species (*A. nebulosa*, *A. bengalensis*, *A. marmorata*, *A. luzonensis*, *A. japonica*, *A. malgumora*, *A. bicolor* and *A. bicolor pacifica*). Three out of 5 commercial samples (SUS7.1; SUS12.3; SUS18.4) were unambiguously verified as *A. anguilla*. Conversely SUS18.4 and SUS21.4, previously allocated to *Anguilla* sp. and *Conger* sp. weren't assigned to any species due to the lack of diagnostic sites in the MDB fragment.

By BLAST analysis, all the 18 PEPCK sequences obtained from shrimp products were attributed to *Litopenaeus* sp. since no identity values higher than 99% were recorded with any referenced sequence in the database. On this regard, it is likely that the discriminatory ability of this gene was further limited by the low number of reference sequences. In fact, no available reference sequences were deposited for 3 species (*L. occidentalis; L. schmitti; L. stilirostris*) out of the 5

ascertained species of *Litopenaeus* sp. (http://www.sealifebase.org/Nomenclature/ScientificNameSearchList.php?crit1_fieldname=SYNONYMS.SynGenus&crit1_fieldtype=CHAR&crit1_operator=EQUAL&crit1_value=Litopenaeus&crit2_fieldname=SYNONYMS.SynSpecies&crit2_fieldtype=CHAR&crit2_operator=contains&crit2_value=&group=summary&backstep=-2&sortby=validname). Fifteen genera of this family were chosen and used for the distance analysis within the genera that had shown the highest identity values with BLAST analysis (Table 1SM). Six main clusters collecting one or more genera were obtained, although none of them was supported by BP values higher than 75%. Also in this case, the reference sequences of the two species of the genus *Litopenaeus* sp. were grouped in a subcluster collecting together also species belonging to the genus *Fenneropenaeus* sp.; *Penaeus* sp. and *Heteropenaeus*. However, considering that our sequences showed low identity values (93 to 95%) with reference sequences of other genera of the Penaeidae family, the samples were considered identified as *Litopenaeus* sp.. Therefore, also in this case, the molecular target was informative enough to verify the information collected at purchasing (all the samples were in fact sold as shrimp).

According to the BLAST analysis the pectin samples SUS 8.1 and SUS39.4 were confirmed as *Pecten* sp. The species wasn't univocally identified due to the presence of more than one species with a top identity value of 98 to 100%. The NJ on Pectinidae was specifically set on 10 species belonging to the family within the genera *Argopecten*, *Pecten*, and *Nodipecten*. Firstly, a significantly lower mean overall divergence rate within *Pecten* sp. (0.014 SE 0.003) than within *Argopecten* sp. (0.318 SE 0.012) was observed. The NJ analysis produced 5 separate clusters, all of them well supported by high BV (75% to 99%), 4 of which represented by mono-species clades for *Argopecten irradians*, *A. purpuratus*, *A. ventricosus* and *Nodipecten subnodosus* and the fifth cluster collecting all the species selected for the *Pecten* genus. Thus, although 3 subsequent subclusters were revealed within *Pecten* sp., the low BV and the lack of mono-species clustering

did not allow the final allocation of the two samples to a unique species (Fig. 10 SM). Also in this case, the identification provided by the analyzed marker allowed to verify the product information.

3.3 Products misdescription

In this study, considering that the Regulation (EU) 1379/2013 (Art 35-Cape IV) does not apply at catering level and the EU Reg. (EU) 1169/2011 only requires a general description of the product, the name of seafood category (e.g. "shrimp" "tuna") and not the specific official denomination was verified in the light of the molecular results.

A noncompliance was highlighted for the sample SUS9.2 labeled as "white fish". In fact, this general description of the product represents itself a sort of "misdescription" since the term "white fish", doesn't comply to the Reg. (EU) 1169/2011 requirements in term of "*sufficient descriptive clarity towards consumers for the discrimination of the product with which it could be confused*" (Art. 2, paragraph 2p). This name was in fact not adopted by the Italian official list for any species and, consequently, it was impossible to verify the results of the molecular analysis.

While the *COI* FDB was enough informative to allow the verification of the name of the seafood category in almost all tuna products the MDBs, obtained from 6 tuna samples, failed this goal. Therefore, the identification at family level (as Scombridae) was not enough for assessing the correctness of these products. In fact, the fish species marketed as tuna in UK, *Tonno* in Italy, *thon* in France, a*tun* in Spain and *thunfisch* in Germany, are only those belonging to *Thunnus* spp. (Xiong et al., 2016). On the contrary, other genus belonging to the Scombridae family are usually referred as skipjack tuna (*K. pelamis), Tonnetto* in Italy, Atlantic bonito (*S. sarda*), *Palamita* in Italy etc. Therefore, the 6 tuna samples identified at family level were not included in the calculation of the misdescription rate that was therefore assessed on 177 out the 185 samples analyzed.

Overall, the samples were generally found to match with the information provided by the restaurateurs with an overall compliance rate of 96.6% (171/177). Thus, a misdescription rate of only 3.4% (6/177) was found. In details (Table 1SM), misdescriptions with species substitution

were highlighted for: 2 seabass samples (SUS 11.2; SUS 20.4) that were identified as *Sparus aurata* (Gilt-head sea bream); 1 smoked eel samples (SUS 21.4) that was were identified as *Conger* sp.; 2 samples (SUS1.2, SUS 17.7) sold as lumpfish roes and 1 sample (SUS 13.3) sold as flying fish roe that were found to be *Mallotus villosus* (Capelin). All the aforesaid cases of misdescription can be attributed to errors during the preparation of the product because of the lack of training of the operators (Guidi *et al.*, 2010) or, especially as regards preserved products of Asian importation, to well-known deficiencies in the identification and labeling system of fish products in the exporting country (Armani *et al.*, 2015a). Therefore, these substitutions could be considered involuntary.

Our study confirms the allegation made by Van Damme *et al.*, (2016) with respect to a general lower levels of mislabelling in restaurants across UK. However, the low misdescription rate highlighted in this study could also be referred to the fact that only the name of the seafood category, and not the commercial denomination, was verified. In fact, while the commercial denomination usually refers to one or few species, the seafood category comprises many species. Comparison with previous studies conducted at the same point of the distributing chain (Table 1) cannot be performed due to a different approached used in assessing products' mislabeling. This highlights that the different approach of the EU countries on regards of seafood labelling (Xiong *et al*., 2016) could hamper the comparison of mislabeling/misdescription rates across Europe. Only the adoption of a standardized nomenclature for seafood products, following the approach one species-one name, could allow to compare data coming from different studies. Finally, it is of the utmost importance that the mandatory information established by the Regulation (EU) 1379/2013 (Art 35-Cape IV) will be requested along all stages of the fishery logistic chain.

4. Conclusions

The present study represents the first survey conducted in Italy on sushi venues for verifying the authenticity of information given by caterers on fish and seafood commonly used for sushi products making. In our opinion, more than to a proper training of Food Business Operators working at the catering level, the low mislabeling rate found in this study could be ascribed to the standardization

of the products sold in ethnic restaurants. In fact, the preparation of few kind of recipes always relying on the same (inexpensive) species of fish could limit the risk of mislabeling occurrence. The selection of a "universally-validated" barcode, such as *COI*, usually represent the best choice when approaching species identification in seafood products. However, limitation due to amplification failure or, at a less extent, lack of discrimination ability can affect the DNA barcoding efficiency. Therefore, alternative molecular target could be selected to overcome technical limits in obtaining reliable barcodes. In this case, a preliminary assessment of the availability of the reference sequences for the seafood category under investigation (group by group) could improve the change to reach species identification.

Funding: The study was supported by the Italian Ministry of Health (Projects code: Ricerca Corrente RC0052011 "Valutazione della qualità igienico-sanitaria e commerciale dei prodotti ittici destinati alla ristorazione scolastiche, ospedaliere ed etniche") and by the Tuscany Region (Project title: "Analisi delle problematiche igienico-sanitarie e commerciali legate alla vendita e somministrazione dei prodotti alimentari etnici all'interno della Regione Toscana").

Figure captions

Figure 1. Sushi products collected in the study. 1) Salmon hosomaki, 2a/b Salmon and Tuna nigiri, 3) Uramaki filled with both vegetable and white fish.

Fig 1SM. Distance tree inferred using the Neighbor-Joining method for the analysis of FDB obtained from the following commercial products: salmon (n=34), gilt-head seabream(n=13), sea bass(n=12), white fish (n=1) swordfish (n=1), lumpfish roe (n=1). The distance analysis was computed using the p-distance involving 188 nucleotide sequences (n=61 from commercial products and 127reference sequences). Bootstrap values (BP) >75% obtained from 1000 replicates are shown below the branches.

Fig 2SM. Distance tree inferred using the Neighbor-Joining method for the analysis of MDB obtained from the following commercial products: salmon (n=23), gilt-head seabream(n=1), sea bass(n=2), salmon roe (n=1) lumpfish roe (n=2), Flying fish roe(n=1). The distance analysis was computed using the p-distance involving 158 nucleotide sequences (n=31 from commercial products and 127 reference sequences). Bootstrap values (BP) >75% obtained from 1000 replicates are shown below the branches.

Fig 3SM. Distance tree inferred using the Neighbor-Joining method for the analysis of FDB obtained from the following commercial products: octopus (n=5), squid (n=1), cuttlefish(n=1). The distance analysis was computed using the p-distance involving 154 nucleotide sequences (n=7 from commercial products and 147 reference sequences). Bootstrap values (BP) >75% obtained from 1000 replicates are shown below the branches.

Fig 4SM. Distance tree inferred using the Neighbor-Joining method for the analysis of MDB obtained n=2 octopus products. The distance analysis was computed using the p-distance involving 51 nucleotide sequences (n=2 from commercial products and 49 reference sequences). Bootstrap values (BP) >75% obtained from 1000 replicates are shown below the branches.

Fig 5SM. Distance tree inferred using the Neighbor-Joining method for the analysis of FDB from Tuna (n=41) products. The distance analysis was computed using the p-distance involving 105 nucleotide sequences (n=41 from commercial products and 64 reference sequences). Bootstrap values (BP) >75% obtained from 1000 replicates are shown below the branches.

Fig 6SM. Distance tree inferred using the Neighbor-Joining method for the analysis of MDB obtained from Tuna (n=6) products. The distance analysis was computed using the p-distance involving 72 nucleotide sequences (n=6 from commercial products and 64 reference sequences). Bootstrap values (BP) >75% obtained from 1000 replicates are shown below the branches.

Fig 7SM. Distance tree inferred using the Neighbor-Joining method for the analysis of FDB obtained from eel (n=5) products. The distance analysis was computed using the p-distance

involving 81 nucleotide sequences (n=5 from commercial products and 76 reference sequences). Bootstrap values (BP) >75% obtained from 1000 replicates are shown below the branches.

Fig 8SM. Distance tree inferred using the Neighbor-Joining method for the analysis of MDB obtained from eel (n=5) products.The distance analysis was computed using the p-distance involving 81 nucleotide sequences (n=5 from commercial products and 76 reference sequences). Bootstrap values (BP) >75% obtained from 1000 replicates are shown below the branches.

Fig 9SM. Distance tree inferred using the Neighbor-Joining method for the analysis of PEPCK fragments obtained from shrimp (n=18) products. The distance analysis was computed using the p-distance involving 76 nucleotide sequences (n=18 from commercial products and 58 reference sequences). Bootstrap values (BP) >75% obtained from 1000 replicates are shown below the branches.

Fig 10SM. Distance tree inferred using the Neighbor-Joining method for the analysis of short 16S rRNA fragments obtained from scallop (n=2) products. The distance analysis was computed using the p-distance involving 37 nucleotide sequences (n=2 from commercial products and 35 reference sequences). Bootstrap values (BP) >75% obtained from 1000 replicates are shown below the branches.

References

1. Armani, A., Guardone, L., La Castellana, R., Gianfaldoni, D., Guidi, A., & Castigliego, L. (2015a). DNA barcoding reveals commercial and health issues in ethnic seafood sold on the Italian market. *Food Control*, 55, 206-214.
2. Armani, A., Guardone, L., Castigliego, L., D'Amico, P., Messina, A., Malandra, R., Guidi, A. (2015b). DNA and Mini-DNA barcoding for the identification of Porgies species (family Sparidae) of commercial interest on the international market. *Food Control*, 50, 589-596.

3. Armani, A., Tinacci, L., Xiong, X., Castigliego, L., Gianfaldoni, D., & Guidi, A. (2015c). 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*, 50, 821-830.
4. Armani, A., Giusti, A., Guardone, L., Castigliego, L., Gianfaldoni, D., & Guidi, A. (2016). Universal Primers Used for Species Identification of Foodstuff of Animal Origin: Effects of Oligonucleotide Tails on PCR Amplification and Sequencing Performance. *Food Analytical Methods*, 9(5), 1199-1209.
5. 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(1), 376-381.
6. Bérnard-Capelle, J., Guillonneau, V., Nouvian, C., Fournier, N., Le Loët, K., & Dettai, A. (2015). Fish mislabelling in France: substitution rates and retail types. *Peer J*, 2, e714.
7. Bilgin, R., Utkan, M.A., Kalkan, E., Karhan, S.U., & Bekbölet, M. (2014). DNA barcoding of twelve shrimp species (Crustacea: Decapoda) from Turkish seas reveals cryptic diversity. Mediterranean Marine Science, 16(1), 36-45.
8. Clark (2015). The current status of DNA barcoding technology for species identification in fish value chains. *Food policy*,54, 85-94
9. Costa, F. O., DeWaard, J. R., Boutillier, J., Ratnasingham, S., Dooh, R. T., Hajibabaei, M., & Hebert, P. D. (2007). Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences*, *64*(2), 272-295.
10. Cwiertka, K. J. (2005). From ethnic to hip: circuits of Japanese cuisine in Europe. *Food & Foodways*, 13(4), 241-272.
11. da Silva, J. M., Creer, S., Dos Santos, A., Costa, A. C., Cunha, M. R., Costa, F. O., & Carvalho, G. R. (2011). Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca). *PLoS one*, *6*(5), e19449.
12. D'Amico, P., Armani, A., Gianfaldoni, D., & Guidi, A. (2016). New provisions for the labelling of fishery and aquaculture products: Difficulties in the implementation of Regulation (EU) n. 1379/2013. Marine Policy, 71, 147-156.
13. Dawnay, N., Ogden, R., McEwing, R., Carvalho, G.R., & Thorpe, R. S. (2007). Validation of the barcoding gene COI for use in forensic genetic species identification. *Forensic Science International*, 173(1), 1-6.

14. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3,294-299
15. Girard, S., & Mariojouls, C. (2008). What is the demand for farmed fish on the European markets?. http://hdl.handle.net/1957/36945
16. Gíslason, Ó. S., Svavarsson, J., Halldórsson, H. P., & Pálsson, S. (2013). Nuclear mitochondrial DNA (numt) in the Atlantic rock crab Cancer irroratus Say, 1817 (Decapoda, Cancridae). *Crustaceana*, *86*(5), 537-552.
17. Guidi, A., Armani, A., Castigliego, L., Li, X. N., Fanzone, F., Fusco, S., Facibeni, E., & Gianfaldoni, D. (2010). Labeling of ethnic food in the Prato Chinese community.*Veterinary Research communications*, 34(1), 163-166.
18. Hall, T. A. (1999). *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*. Nucleic Acids Symposium Series, 41, 95e98.
19. Handy, S. M., Deeds, J. R., Ivanova, N. V., Hebert, P. D., 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 AOAC International*, 94(1), 201-210.
20. Hanner, R., Desalle, R., Ward, R. D., & Kolokotronis, S. O. (2011). The fish Barcode of Life (FISH-BOL) special issue. *Mitochondrial DNA*, 22(1), 1-2.
21. Helyar, S. J., Lloyd, H. A. D., de Bruyn, M., Leake, J., Bennett, N., & Carvalho, G. R. (2014). Fish product mislabelling: failings of traceability in the production chain and implications for illegal, unreported and unregulated (IUU) fishing. *PloS One*, 9(6), e98691.
22. Hsin-I Feng, C. (2012). The tale of sushi: history and regulations. *Comprehensive Reviews in Food Science and Food Safety*, 11(2), 205-220.
23. Jacquet, J. L., & Pauly, D. (2008). Trade secrets: renaming and mislabeling of seafood. *Marine Policy*, 32(3), 309-318.
24. Khaksar, R., Carlson, T., Schaffner, D. W., Ghorashi, M., Best, D., Jandhyala, S., ... & Amini, S. (2015). Unmasking seafood mislabeling in US markets: DNA barcoding as a unique technology for food authentication and quality control. *Food Control*, *56*, 71-76.
25. Lee, C. J., & Heacock, H (2016). Safety and pH Measurements of Sushi Rice in Japanese Restaurants in Burnaby BC, Canada. http://www.beta.ncceh.ca/sites/default/files/BCIT-Lee-2014.pdf.

611 26. Ma, K. Y., Chan, T. Y., & Chu, K. H. (2009). *Phylogeny of penaeoid shrimps (Decapoda:*
612 *Penaeoidea) inferred from nuclear protein-coding genes. Molecular Phylogenetics and*
613 *Evolution*, 53(1), 45-55
614 27. Mariani S, Griffiths AM, Velasco A, Kappel K, Jérôme M, Perez-Martin RI, Schröder U,
615 Verrez-Bagnis V, Silva H, Vandamme SG, Boufana B. (2015). Low mislabeling rates
616 indicate marked improvements in European seafood market operations. *Frontiers in Ecology*
617 *and the Environment* 13(10):536-540
618 28. Marin, A., Fujimoto, T., & Arai, K. (2015). The variable 5′ end of the 16S rRNA gene as a
619 novel barcoding tool for scallops (Bivalvia, Pectinidae). *Fisheries science*, 81(1), 73-81.
620 29. Mikkelsen P.M., Bieler R., Kappner I., Rawlings T.A. (2006). Phylogeny of Veneroidea
621 (Mollusca: Bivalvia) based on morphology and molecules. Zool. J. Linnean Soc., 148 (3),
622 439-521.
623 30. Ministerial Decree of the Italian Ministry of Agriculture,Food and Forestry (MIPAAF) of 31
624 January 2008 .Assignment of the Italian trade name to fish species of commercial interest
625 Amendments to the list of the Ministerial Decree of January 14, 2005.
626 31. Mouritsen O.J (2009) SUSHI Food for the eye, the body & the soul. Science+Business
627 Media B.V, USA, 978-1-4419-0618-2 (Online) springer.com
628 32. Mudu, P. (2007). The people's food: the ingredients of "ethnic" hierarchies and the
629 development of Chinese restaurants in Rome. *Geo Journal*, 68(2-3), 195-210.
630 33. Pereira, J. C., Chaves, R., Bastos, E., Leitão, A., & Guedes-Pinto, H. (2011). *An efficient*
631 *method for genomic DNA extraction from different molluscs species.* International journal of
632 molecular sciences, 12(11), 8086-8095
633 34. Ratnasingham, S., Hebert, P.D.N. (2007). Bold: the barcode of life data system
634 (http://www.barcodinglife.org). *Molecular Ecology Notes*, 7(3):355-364
635 35. Rajkumar, G., Bhavan, P. S., Udayasuriyan, R., & Vadivalagan, C. (2015). Molecular
636 identification of shrimp species, Penaeus semisulcatus, Metapenaeus dobsoni, Metapenaeus
637 brevicornis, Fenneropenaeus indicus, Parapenaeopsis stylifera and Solenocera crassicornis
638 inhabiting in the coromandel coast (Tamil Nadu, India) using MT-COI gene. *Internationl*
639 *Journal of Fisheries and Aquatic Studies*, 2, 96-106.
640 36. Regulation (EC) No 853/2004 of the European parliament and of the Council of 29 April
641 2004 laying down specific hygiene rules for on the hygiene of foodstuffs. Official Journal of
642 European Union L139:55
643 37. Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25
644 October 2011 on the provision of food information to consumers, amending Regulations

(EC) No 1924/2006 and (EC) No 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC and Commission Regulation (EC) No 608/2004. Official Journal of the European Union L 304.

38. Regulation (EU) No 1379/2013 of the European Parliament and of the Council of 11 December 2013 on the common organisation 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. Official Journal of the European Union L 354.
39. 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 Investigaciones Marinas*, 20, 77-87.
40. Santaclara, F. J., Velasco, A., Pérez-Martín, R. I., Quinteiro, J., Rey-Méndez, M., Pardo, M. A., Jimenez, E., & Sotelo, C. G. (2015). Development of a multiplex PCR–ELISA method for the genetic authentication of *Thunnus* species and *Katsuwonus pelamis* in food products. *Food chemistry*, *180*, 9-16
41. Song, H., Buhay, J. E., Whiting, M. F., & Crandall, K. A. (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences*, *105*(36), 13486-13491.
42. Stano. S. (2015). Eating the others. Translation of the culinary codes. Cambridge Scholars Publishing, Lady Stephenson Library, Newcastle upon time, NE6 2PA, UK
43. Teletchea, F. (2009). Molecular identification methods of fish species: reassessment and possible applications. Reviews in Fish Biology and Fisheries , 19, 265-293
44. Tsang, L. M., Ma, K. Y., Ahyong, S. T., Chan, T. Y., & Chu, K. H. (2008). Phylogeny of Decapoda using two nuclear protein-coding genes: origin and evolution of the Reptantia. *Molecular Phylogenetics and Evolution*, 48(1), 359-368.
45. Vandamme, S. G., Griffiths, A. M., Taylor, S. A., Di Muri, C., Hankard, E. A., Towne, J. A., Watson, M., & Mariani, S. (2016). Sushi barcoding in the UK: another kettle of fish. PeerJ, 4, e1891.
46. Verbeke, W., Vermeir, I., & Brunsø, K. (2007). Consumer evaluation of fish quality as basis for fish market segmentation. Food Quality and Preference, 18(4), 651-661.

47. Ward, R. D. (2012). FISH-BOL, a case study for DNA barcodes. DNA Barcodes: Methods and Protocols, 423-439.
48. Warner K, Timme W, Lowell B, Hirshfield M. (2013). Oceana study reveals seafood fraud nationwide. Available at http:// oceana.org/ reports/ oceana-study-reveals-seafood- fraud-nationwide
49. Winnepenninckx, B., Backeljau, T. R. D. W., & Dewachter, R. (1993). Extraction of high-molecular-weight DNA from mollusks. *Technical tips*, TIG 9 (12), 407.
50. Yancy, H. F., Fry, F. S., Randolph, S. C., Deeds, J., Ivanova, N. V., Grainger, C. M., ... & Ormos, A. (2008). A protocol for validation of DNA-barcoding for the species identification of fish for FDA regulatory compliance. *FDA Laboratory Information Bulletin. LIB*, 4420.
51. Xiong, X., D'Amico, P., Guardone, L., Castigliego, L., Guidi, A., Gianfaldoni, D., & Armani, A. (2016). The uncertainty of seafood labeling in China: A case study on Cod, Salmon and Tuna. *Marine Policy*, *68*, 123-135.

***Highlights (for review)**

- A two years-survey was carried out on sushi products sold in Italy at retail level
- 185 sushi samples were collected and subdued to DNA barcoding analysis
- A low misdescription rate (3.3%) was found
- Results reflect a proper training of FBO towards labelling

Figure

Click here to download high resolution image

Table

Table 1: Overview on the most recent studies about fish mislabeling at retail level (restaurants, supermarkets) with particular emphasis on sampling carried out on sushi products. * http://eu.oceana.org/sites/default/files/421/oceana_factsheet_seafood_fraud_brussels_eng.pdf

Reference	Retail point	Country	Samples number	Fish/Seafood category	Overall mislabeling rate (%)
Vandamme *et al.*, 2016	Sushi restaurants	UK	115	Tuna, Eel, Yellowtail, Mackerel, Swordfish, Kingfish, Seabass, Seabream, Black cod, Barramundi, Snapper, flying fish	10,4
Bérnard-Capelle *et al.*, 2015	Fishmongers, Supermarkets, Restaurants	France	371 (**16 sushi products**)	30 different categories (Tuna, cod, hake, plate fishes several species, Pangasius, Anglerfish, Seabass)	3,7
Oceana report, 2015*	Canteen, Restaurant, Suhi venue	Belgium	280 (21 sushi products)	Tuna, cod, hake, sole	38 (sushi venue 54)
Kakhsar et al., 2015	Supermarkets, Sushi venues	US	216	Tuna, Salmon, Pacific salmon, Steelhead, Catfish Halibut, Sole, Black mussels, Little neck clam, Cuttlefish, Squid	12,8
Warner *et al.*, 2013	Restaurants, grocery stores sushi venues		1200 (**197 sushi products**)	Tuna. Snapper, Yellotail, Seabass, Grouper, Halibut, Sole	33 (**Sushi venues 74**)
Lowestein *et al.*, 2009	Sushi restaurants		68	Tuna	32,3
Wong and Hanner, 2008	Supermarket, restaurants		91	Snapper, Tilapia, Monkfish, Mackerel, Seabass, salmon, Kinfish, Plate fish (several species)	25
Faisal *et al.*, 2012	Sushi restaurants	Malaysia (Penang Island)	7	Tuna, Salmon, Butterfish,	0

Table

Table 2: Universal primers for the amplification of total DNA belonging to the sushi products. FORUNICOI1 was obtained by modifying FISHCOILBC_ts: highlighted in bold the nucleotide substitutions.PL: primer length; AL: amplicon length. [a:] primer tailed with Steffens 1993 oligonucleotides highlighted in grey, in bracket the primer length without the tail. * Amplicon length refers to the fragment generated using FISHCOILBC_ts as primer forward ; ** Amplicon length calculated on the sequences FR849595 of *Sardina pilchardus*

Primer code	TARGET GENE	REFERENCE	Primer sequence (5'-3')	PL(bp)	AL with and without primers
FISHCOILBC_ts[a]	COI	Handy *et al.*, 2011	**CACGACGTTGTAAAACGA**CTCAACYAATCAYAAAGATATYGGCAC	45(27)	705/655
FISHCOIHBC_ts[a]			**GGATAACAATTTCACACAGG**ACTTCYGGGTGRCCRAARAATCA	43(23)	
FORUNICOI1		This study	**CACGACGTTGTAAAACGA**CTCAAC**W**AATCA**T**AAAGA**Y**AT**T**GG**H**AC	45(27)	
COIFALT		Mikkelsen *et al.*, 2006	ACAAATCAYAARGAYATYGG	20	698/650
COIRALT			TTCAGGRTGNCCRAARAAYCA	21	
LCO1490		Folmer *et al.*, 1994	GGTCAACAAATCATAAAGATATTGG	25	710/659
HCO2198			TAAACTTCAGGGTGACCAAAAAATCA	26	
REVshort1[a]		Armani *et al.*, 2015b	**GGATAACAATTTCACACAGG**GGYATNACTATRAAGAAAATTATTAC	46(26)	*192/139
PEPCK for2	PEPCK	Tsang *et al.*, 2008	GCAAGACCAACCTGGCCATGATGAC	25	644/598
PEPCK rev3			CGGGYCTCCATGCTSAGCCARTG	23	
FOR16S-2	16S rRNA	Armani *et al.*, 2015c	CTTMGGTTGGGGCGACC	17	≈**152 /117
REV16S-2			CTGTTATCCCTAGGGTAACT	20	

Table

Table 3: Total DNA quality and fragmentation of the samples. The average yield and the mean absorbance values were calculated by grouping the samples for seafood categories. Highlighted in grey the categories with the lowest quality and DNA yield.

PRODUCTS CATEGORY	Samples n.	DNA QUANTIFICATION AND QUALITY									DNA FRAGMENTATION (samples n.)		
		Yield (µg/mg)			A260/A280			A260/A230			< 500bp or NE	500-1000pb	>1000
		Med.	Max	Min	Med.	Max	Min	Med.	Max	Min			
Raw Fish	132	0,53	0,97	0,21	2,08	2,17	1,94	2,07	2,36	1,82	29	43	60
Processed Fish	13	0,06	0,13	0,01	2,04	2,31	1,89	1,32	1,53	0,98	8	5	0
Crustacean	18	0,33	0,47	0,19	2,08	2,14	1,98	2,03	2,29	1,79	3	7	9
Cephalopods	12	0,28	0,40	0,21	2,01	2,08	1,92	1,99	2,22	1,78	9	1	2
Bivalves	2	0,09	0,11	0,06	2,01	2,04	1,98	1,22	1,28	1,15	2	0	0
Fish eggs	8	0,03	0,07	0,01	1,94	2,04	1,89	1,14	1,35	0,84	8	0	0
TOTAL	**185**										**60(32%)**	**55(30%)**	**71(38%)**

Table

Table 4: Results of sequencing and post sequencing analysis

Menu/label Description	Samples number	Sequences number	Sequencing rate (%)	Species specific identification	Successfull species specific identification (%)
FISHES					
Seabream	13	13	100	13	100
Seabass	16	14	87	14	100
"White fish"	1	1	100	1	100
Salmon	57	57	100	57	100
Tuna	44	44	100	0	0
Swordfish	1	1	100	1	100
Canned tuna	3	3	100	0	0
Eel	10	10	100	4	60
Category total	**145**	143	**99**	**90**	**63.4**
MOLLUSCS					
Octopus	10	7	70	7	100
Squid	1	1	100	1	100
Cuttlefish	1	1	100	1	100
Scallop	2	2	100	0	0
Category total	**14**	11	**78**	**9**	**82**
CRUSTACEAN					
Shrimp	16	16	100	0	00
Whiteleg shrimp	2	2	100	0	0
Category total	**18**	18	**100**	**0**	**0**
FISH ROE					
Lumpfish roe	3	2	67	2	100
Flying fish roe	3	2	67	1	50
Salmon roe	2	2	100	2	100
Category total	**8**	**6**	**75**	**5**	**83.3**
Total	**185**	**178**	**96**	**103**	**55.6**

e-component

Click here to download e-component: Table 1SM revised 08-03-17.docx

e-component

Click here to download e-component: Table 2SM rev.docx

***Manuscript**

Click here to download e-component: Table 3SM.docx

e-component

Click here to download e-component: Table 4SM.docx

e-component

Click here to download e-component: Table 5SM.docx

e-component

Click here to download e-component: Table 6SM.docx

e-component

Click here to download e-component: Fig 1SM.jpg

e-component

Click here to download e-component: Fig 2SM.jpg

e-component

Click here to download e-component: Fig 3SM.jpg

e-component

Click here to download e-component: Fig 4SM.jpg

e-component

Click here to download e-component: Fig 5SM.jpg

e-component

Click here to download e-component: Fig 6SM.jpg

e-component

Click here to download e-component: Fig 7SM.jpg

e-component

Click here to download e-component: Fig 8SM.jpg

e-component

Click here to download e-component: Fig 9SM.jpg

e-component

Click here to download e-component: Fig 10SM.JPG