

A peer-reviewed version of this preprint was published in PeerJ on 6 January 2015.

[View the peer-reviewed version](https://peerj.com/articles/714) (peerj.com/articles/714), which is the preferred citable publication unless you specifically need to cite this preprint.

Bénard-Capelle J, Guillonnet V, Nouvian C, Fournier N, Le Loët K, Dettai A. 2015. Fish mislabelling in France: substitution rates and retail types. PeerJ 2:e714 <https://doi.org/10.7717/peerj.714>

Fish mislabelling in France : substitution rates and retail types

Julien Bénard-Capelle¹, Victoire Guillon², Claire Nouvian², Nicolas Fournier³, Karine Le Loët⁴, Agnès Dettai^{5*}.

5 ¹. Institut National de la Santé et de la Recherche Médicale, Unité 1001 Robustness and Evolvability of Life, Université Paris 5, Cochin, Paris

². BLOOM, 27, rue du Faubourg Montmartre, Paris

³. Oceana Europe, European Office, Brussels

⁴. Terra Eco, 1 Allée Cassard, Nantes

10 ⁵. Muséum national d'Histoire naturelle, Département Systématique et Évolution, UMR 7205 MNHN-CNRS-UPMC-EPHE, Paris, France

*corresponding author : adettai@mnhn.fr

Abstract

PeerJ PrePrints
15 IvanoThe development of citizen science has brought together scientific expertise and volunteer involvement to answer both scientific and societal questions. In this study, a consortium of citizens, journalist, scientists and non governmental organisations reports the first measure of the market-wide rate of fish mislabelling in France.

20 We collected in fishmonger shops, supermarkets and restaurants and sequenced 390 samples of fish either in fillets or prepared meals, which is the largest dataset assembled to date in an European country. The sampling covers 55 different commercial names

25 The overall substitution rate is one of the lowest observed for comparable surveys with large sampling in Europe. Remarkably, we detected no case of species mislabelling among the frozen fillets or in industrially prepared meals. We also investigated most of the mislabelling cases detected directly from the sellers. A number of them admitted that the substitution took place at the end of the supply chain.

The rate of mislabelling does not differ between species (3.7 %, ci 2.2-6.4%), except for bluefin tuna. Despite a very small sample size (n=6), this species stands in sharp contrast with the low substitution rate observed for the other species (rate between 36 and 99%).

30 This study shows that even in countries where species substitution rate is low, citizen science can enhance the management of natural resources and provide important insights for regulation policies.

Introduction

Fish are among the last harvested wild natural resource on the planet (Jaquet & Pauly 2008), and aquaculture still represents only a small part of the total of fish consumed : just over 40% in the world and slightly more than 30% for France (Meunier et al. 2013). Efficient management of existing natural resources is particularly important , as the stock collapses of several fish species in the 80s and 90s have demonstrated (Jaquet & Pauly 2008). But management depends on traceability and labelling, which is difficult because aquatic food is the most widely traded type of food (Cochrane et al., 2009). The origin of a given product can be very distant and, with a wide array of species now available on the market, effective control is complex. In Europe, a number of policies have been set up to regulate seafood labelling and traceability (Regulations (EC) 104/2000, 2065/2001, 178/2002, 1224/2009). In the public and political world, there is a growing concern about sustainability of food resources. There has been a multiplication of guides for fish consumers both in France and in other countries, to advise the often badly-informed consumers (Wong & Hanner, 2008 ; Jaquet & Pauly, 2008 ; Le Loët, 2014). These guides typically make recommendations in the form of species lists, in some cases with additional parameters like origin, size, or capture season. Taste and texture preferences also guide consumer choices, with both cultural and personal factors determining the most popular species. Nutritional advices and warnings also largely rely on recommended or discouraged species lists. Mislabelled fish puts the customer at risk of purchasing products not corresponding to their ethical or taste criteria (Rasmussen and Morrissey, 2008 ; Jaquet & Pauly, 2008). Some species can represent health hazards because they can transmit the ciguatera disease or present other risks like oilfish or pufferfish (Jaquet & Pauly 2008). The species designation therefore represents an important information for consumers; yet, the reliability of correspondence between the listed name and the species actually purchased varies largely by country and taxonomic group (Wong & Hanner, 2008 ; Machado-Schiaffino et al., 2008 ; Smith et al., 2008 ; Barbuto et al., 2009 ; Lowenstein et al., 2009 ; Vinas & Tudela, 2009 ; Filonzi et al., 2010 ; Miller & Mariani, 2010 ; FSAI, 2011 ; Hanner et al., 2011 ; Huxley-Jones et al., 2011 ; Garcia-Vasquez et al., 2011 ; Cawthorne et al., 2012 ; Cline, 2012 ; Miller et al., 2012 ; Di Pinto et al., 2013 ; Griffith et al., 2013 ; Curatelli et al., 2014). In Europe, fish is second on the list of products that are the most at risk of food fraud, and Europol has observed a rise in the number of general food fraud cases (Committee on the Environment, Public Health and Food Safety, European Parliament, 2013).

Although species substitution of commercial fish has been already studied in many countries, up to

65 now no results were available for France. A market-wide study was therefore needed for the first assessment of substitution rates over the less recognizable products: fillets both sold by fishmongers and industrially packaged or deep-frozen, and dishes ready-made or in restaurants. As the customers and control agencies cannot recognize species from their morphology in these instances, we could expect a high rate of species substitution.

70 Many DNA-based methods have been proposed for identification of fish specimens where the morphology cannot be used (reviewed in Rasmussen and Morrissey, 2008). Among these, the comparison of mitochondrial DNA sequences to reference datasets has proved its usefulness multiple times. The most frequently used sequences are those of partial cytochrome b, partial 16S or 12S ribosomal DNA, and partial cytochrome oxidase I. It is crucial for identification that a reference sequence for the searched species is present, so the more complete the reference datasets for the comparison, the better (Ekrem et al., 2007). With the success of the Barcode of Life Project for fish (Ward et al., 2009), its reference marker, cytochrome oxidase I, has been increasingly used for identifications, and represents the majority of substitution studies in the last years. Datasets for the regulatory identification of substitutions have been established for COI, for instance in the Regulatory Fish Encyclopedia of the FDA (Yancy et al., 2008). The Barcode of Life Database (BOLD, Ratnasingham & Hebert, 2007) currently contains almost 150,000 COI barcode sequences for almost 14,000 actinopterygian species, making it the largest dataset for fish identification. 80 Cytochrome b, the second largest, has only around 82,000 sequences listed in the GenBank nucleotide database. Additional features of the Barcode of Life project, such as linking sequences to vouchered specimens and specimen data, increase reliability compared to the notoriously high error rates in GenBank (Harris, 2003 ; Rasmussen and Morrissey, 2008).

85 The present study, the first of its type for France, aims to evaluate the extent of mismatch between the market names and the species for commercial samples of a number of the most common commercial marine fish species in France, focusing on the less recognisable processed or filleted forms. Two collecting efforts had been started in parallel by the NGO Oceana, associated to the magazine Terra Eco, and the NGO Bloom, in collaboration with researchers from the INSERM and 90 MNHN. They are both analysed and presented in this article, resulting in the largest European dataset for such a study.

Material and methods

Sampling

95 Samples and corresponding data were collected in a semi-participative way across France between
March and December 2013. Small samples were collected in pre-numbered tubes and stored until
extraction in either 95% ethanol (Fishlabel dataset) or in silica gel (Terra Eco dataset). For the
Fishlabel dataset (hereafter FL samples), each sample was divided at sampling in two tubes with the
same sampling number. Commercial name, latin name when indicated, date, collector, location,
brand name, shop or restaurant names were collected, as well as photographs of packages and
100 samples. The data were collected either on paper forms sent along with the samples or uploaded
online using the smartphone application Epicollect (Aanensen et al, 2010). The completeness of
collection data was checked, and the samples that lacked crucial data (defined as collection site,
retail name, dish name including species name and collector) were excluded.

PeerJ PrePrints
105 Only fillets and transformed fish were sampled, as they are not readily identifiable and thus
potentially easier to substitute. They were collected from fishmonger's shops, restaurants, and
supermarkets (at the fishmonger's department, industrially prepared -canned or fresh meals- and
deep-frozen fillets).

In the TE protocol, we asked the collectors to sample only three categories of products : cod,
anglerfish and tuna. For the FL protocol, we focused on 10 species chosen among the most
110 consumed fish species in France (according to www.franceagrimer.fr, checked april 2013): Bar or
Loup, Lieu noir, Cabillaud, Merlu or Colin, Baudroie or Lotte, Merlan, Sole, Pangas, Raie, Thon
(See table 1 for correspondance with species names and english names). We excluded all salmon
species because the French salmon market is dominated by cheap *Salmo salar* from aquaculture,
which is both easy to identify and expected to be less substituted because of its price. However,
115 collectors sampled a larger number of species than the recommended ones.

DNA extraction and sequencing

Most of the FL samples were extracted using an epMotion 5070 (Eppendorf) and Tissue extraction
kits (Macherey Nagel) following the manufacturer's instructions. Some samples were extracted
following the protocol in [Winnepenninckx et al. \(1993\)](#).

120 The partial cytochrome oxidase I was amplified using primers FishF1-5'
CAACCAACCACAAAGACATTGGCAC 3', FishF2-

5' TCGACTAATCATAAAGATATCGGCAC 3' and FishR1-5' TAGACTTCTGG

GTGGCCAAAGAATCA3' (Ward et al., 2005), TelF1-5'

TCGACTAATCAYAAAGAYATYGGCAC3' and TelR1-5'

125 ACTTCTGGGTGNCCAAARAATCARAA3' (Dettai et al., 2011). Shorter amplifications were used for samples with denatured DNA with the following newly defined primers flanking variable areas and diagnostic sites in the *Thunnus* sequences (Lowenstein et al., 2009): COIF268-5'

GAAACTGACTYATTCCTYTAATGAT 3', COIF270-5' AACTGACTTATTCCYYTAATGATYGG 3', COIR450-5' GAAGTTAATTGCCCAAGAATTGA 3', and COIR445-5'

130 AAGTTAATTGCTCCAAGAATTGAWGA 3'. Samples were first tested with the primers for the longest fragment, then the intermediate size, and finally the ones for the shortest fragment. PCR was performed in a 20 ml volume with 5% of DMSO, 5 mg of bovine serum albumine, 300 mM of each dNTP, 0.3 mM of Taq DNA polymerase (Qiagen), 2.5 ml of the corresponding buffer and 1.7 pM of each of the two primers. After denaturation for 2 min at 94 °C, the PCR ran for 40–50 cycles of
135 (20s, 94 °C; 25s, 52 °C and 45s, 72 °C), with a terminal elongation of 3 min at 72 °C on Biorad thermocyclers. Purification and sequencing of the PCR products were performed commercially by GATC (<http://www.gatc-biotech.com/>) using the same primers. Most sequences were obtained in only one direction, but a number chosen at random were sequenced in both directions. Samples where molecular ID differed from the commercial ID were extracted from the second sample
140 tube using the protocol in Winnepenninckx et al. (1993), amplified and sequenced a second time, when possible with a different pair of primer. Sequences were checked manually against their chromatogram using Codoncode Aligner (CodonCode Corporation) and then exported and aligned in Bioedit (Hall, 1999).

The Terra Eco samples (hereafter TE samples) were extracted, amplified and sequenced by a
145 commercial company specialised in molecular identification (<http://www.spygen.fr/>) with the same primers tailed as in Ivanova et al. (2007).

Molecular identification

Three datasets were assembled from the three sizes of amplifications and sequences: Complete: 652 bp (FishF1 or TelF1 forward and FishR1 or TelR1 reverse primers), medium: 442 bp (F268-FishR1
150 or TelR1 reverse primers), short: 208 bp (F268-R450). Distance trees were built to cluster similar sequences. These similar sequences were grouped in the alignment files, and sequence identities were checked on the alignments. Each distinct sequence was used to BLAST-search the Barcode of Life database. The complete dataset was compared to the Species Level Barcode Records, the

155 medium and short datasets to the Full Length Record Barcode Database to avoid issues due to
insufficient overlap of the sequences with the reference dataset. Both tree-based identification and
sequence similarity identification were used for the BOLD comparison. For species with low
interspecific divergences (*Gadus* species, *Thunnus* species), sequences were compared to each
other, to sequences from the BOLD, and to sequences from the FDA reference dataset for Seafood
identification

160 (<http://www.fda.gov/Food/FoodScienceResearch/DNASeafoodIdentification/ucm238880.htm>).
Species-specific characteristic attributes and characteristic combinations were checked on the
alignments following Lowenstein et al. (2009).

165 Sample descriptions and sequences are available in the Barcode of Life Database in the FSCF
project (FCSF001-14 to FCSF291-14 for Fishlabel, FCSF292-14 to FCSF404-14 for Terra eco), and
in GenBank. Collector names, brands and precise collection data were anonymized. Photographs
are included for samples when they do not threaten the anonymity of the data.

Molecular identifications were compared to the recorded commercial identifications only at the end
of the process. The admissible species that can be sold under each commercial name were
determined by consulting a governmental reference website

170 (<http://www.economie.gouv.fr/dgccrf/Poissons>, last checked on 25/02/2014).

Mislabelling determination

A sample was declared mislabelled if the commercial name indicated on the menu, the price tag, or
the box is not in the list of the admissible species.

175 We did not retain the commercial names obtained orally from the waiting staff in the calculations of
the substitution rates. We have kept this information in the data files since for one species (bluefin
tuna) the information given by the waiter appeared to be highly biased, a point we discuss below.

Grouping of commercial names

180 The total number of commercial denominations retrieved from the completed forms was high (55
different commercial names), preventing statistical analysis of a large part of the dataset. The
samples were thus grouped into broader commercial categories. For instance « cabillaud » (cod)
was grouped with « cabillaud du pacifique » (pacific cod) and « morue » (a French nomenclature
for dry and salted cod, whether Pacific or Atlantic) under « cabillaud ». This and similar cases
reduced the number of categories to 30. We further decreased the number of categories by keeping
only those for which at least 10 samples were available. All the other samples were grouped under

185 the « other » category. However, after a preliminary analysis it appeared that the mislabelling signal
detected for the « thon » (tuna) category was mostly attributable to the samples sold for « thon
rouge » (bluefin tuna). To account for this fact, this category was then split into « bluefin tuna » and
« tuna », although only 5 samples fall under the « bluefin tuna » name. This procedure ensures that
190 most categories have a large enough sample size for statistical analysis while being representative
of the French seafood market.

Note that for reading convenience and international comparison, the french fish names have been
translated into their english equivalent when available and used throughout this study (table 1).
Some could not be translated, like « colin » referring to a broad category of whitefleshed species,
and were kept in their original form. Furthermore, as the French vernacular names relate to the local
195 naming traditions, they might not designate the same species as in countries using the English
equivalent. For instance “albacore” refers in French to *Thunnus albacares*, while in English it refers
to *Thunnus alalunga*.

Statistical analysis

The substitution status of the samples was analysed as a binary variable (0=not substituted,
200 1=substituted) using a generalized linear model with a binomial error distribution and a logit link
function. The type of protocol (FL or TE), shop type (supermarket, restaurant or fishmonger shop),
species category and type of product sold (fresh fillets, frozen fillets, industrial meal or restaurant
meal) were included as explanatory variables, with interactions.

After removal of the non-significant interactions and variables, Tukey Honest Significant
205 Differences were calculated from the final model.

The influence of the price was investigated in a separate analysis for a subset of 156 samples for
which the information on the price was available and could be expressed in €/kg. The substitutions
were modelled as above with the price, the retail conditions and the type of shop as independent
variables, with interactions.

210 All the confidence intervals ($\alpha=0.05$) were calculated using Wilson's method. The statistical
analysis was performed with R (R Core Team, 2013) and both the script used and the original data
file are available on the figshare repository (
<http://dx.doi.org/10.6084/m9.figshare.978485>).

A follow-up investigation was performed by KL for samples for which mislabelling was detected.
215 They consisted in interviews of the retailers (Le Loët, 2014) to understand where the substitution

came from and, when possible, what motivated it.

Results

Sequencing

We collected 291 samples using the FL protocol, out of which 276 could be sequenced : 172 specimens for the whole barcode region (651 pb), 97 for the medium size amplification (441pb), and 7 specimens were only obtained for the shortest amplification (204bp). In addition, 15 samples (5,16%) could not be amplified at all, a failure rate comparable to other studies of this type (Hanner et al. 2011, Cawthorne et al. 2012). These included 9 ready-made dishes, one canned tuna sample (the only one present in the sampling), two smoked filets, and 3 restaurant dishes, all sources that are expected to show some DNA degradation. All 114 TE samples provided sequences: 45 samples for the longest amplification and 69 for the shortest. The intermediate size PCR was not used in this case.

Molecular identification

Identifications using the different approaches gave congruent results, although some approaches gave a better resolution than others in groups including very closely related species. Sequences in almost all species had a similarity between 99,19% (FLID089) and 100% (231 of the FL samples and 106 of the TE samples) with sequences present in BOLD. The single exception was « bar » samples (European Sea Bass, *Dicentrarchus labrax*). *Dicentrarchus labrax* is represented by three divergent clusters in BOLD. Part of our samples is almost identical to samples from the UK and Spain, the rest to samples from Turkey and Portugal. These two groups of sequences diverge by 2,5-3% from each other. However, one of these groups of sequences is identical (or with one base divergence) with the high reliability FDA208 *Dicentrarchus labrax* sequence in the FDA's Reference Standard Sequence Library for Seafood Identification, and the reliability of the identification for the second group is also supported by a number of reference sequences from independent datasets.

Almost all species included in the study were represented in the BOLD by barcode clusters that are single, cohesive, and non overlapping with other species clusters (Wong & Hanner, 2008), a prerequisite for good identification. Tree based identifications using the BOLD generated K2P trees placed all samples (except again the *D. labrax* samples, and tuna samples) reliably within large clusters of samples with the same identification as the sequences they were most similar to (grade A identification according to Costa et al. 2012). *D. labrax* samples had a grade C identification (Costa et al. 2012) if considering only the sequences in the BOLD. The sebaste (rockfish) sample was also

250 embedded within a large cluster of very closely related sequences, but this cluster contained a mix of sequences of *Sebastes viviparus*, *S. fasciatus*, *S. mentella*, *S. marinus* and unidentified *Sebastes* samples, so only genus identification was possible in this case. The same problem had already been encountered by Wong & Hanner (2008) and Hanner et al. (2011) on these same species. Direct sequence comparisons, looking for shared sites, were also congruent with the two previous approaches.

255 Most groups had relatively high interspecific divergences in the BOLD even between the more closely related species, making identification straightforward. However tuna species presented much less divergence, and pose problems with similarity or clustering based methods (Lowenstein et al., 2009 ; Vinas & Tuleda, 2009). Some species had clear shared sites in the sequences corresponding to the sites identified by Lowenstein et al. (2009), and also formed distinct clusters, as was the case with *Thunnus thynnus* (“thon rouge”), *Thunnus alalunga*, and our samples of
260 *Thunnus obesus*, so placement and identification posed no problem. However *Thunnus albacares* samples formed a group with more intraspecific variability, and the BOLD dataset also presented sequences identified as other *Thunnus* species clustering inside. The other sequences from these *Thunnus* species (*T. obesus* for instance) had their own distinct clusters in the BOLD, and misidentification of the samples that cluster within *T. albacares* is the probable explanation. This
265 makes distinction between these two species less reliable using this dataset.

Two pairs of mixed samples could be identified (FL0084 and FL0085, FL1263 and FL1266). The pairs originate from the same collection event and collector, and the identifications are exactly switched. The most probable explanation is exchange by the collector.

270 The molecular identifications can therefore be considered reliable. The summary of the commercial names and species determination is presented in table 1.

Species substitution

275 Nineteen samples for which important information were lacking or for which doubts on the quality of the collected data remained were removed, and 371 samples were kept for analysis. Among them, we found 14 cases of species substitution representing a rate of 3.7 % (ci 2.2-6.4%, table 2). As expected, most of these products were substituted for species with a lower market value. Five substitution cases were observed for bluefin tuna, although this species is represented by only 6 samples, i.e. the substitution rate is 83 % with a confidence interval of 36-99%.

The species representation was largely uneven with the top five species totalling 67 % of the samples and none of the remaining categories containing more than 18 samples (figure 1). The

280 samples were more evenly distributed among types of shops with 74 samples from fishmongers, 100 from restaurants and 197 from supermarkets.

The two datasets (Bloom and TE) are compatible as no effect of the protocol on the rate of substitution was detected in the full model. They were thus pooled. The variable « species » has an impact on the rate of mislabelling ($p < 0.001$, figure 2). Post-hoc test indicate that this is due to the
285 « bluefin tuna » category being significantly different from the three categories with the largest number of samples, i.e. « cod », « other » and « tuna » (respectively $p = 0.004$, $p = 0.006$, $p = 0.012$).

The different modes of retail also show a marginally significant differences among them ($p = 0.085$), as species substitution was found only for products sold as fresh fillets or as restaurants meals (figure 3).

290 No effect of the price on the probability of species substitution was observed (data not shown).

Reliability of participative collection

The whole TE sampling, and part of the FL sampling were done by volunteers. The sampling effort is very broadly distributed with the top 3 collectors contributing to 36 % of the sampling efforts (with respectively 72, 49 and 15 samples), while 75 % of the collectors contributed one or two
295 samples each (figure 4). Reliability of the collectors is always an issue, but the very low substitution number excludes a high error rate in collectors. All the substitutions cases were checked carefully and sequenced twice, so they can indeed be considered genuine.

Discussion

France ranks 7th among European countries for fish and seafood consumption with 33.7 kg/capita/year and represents the first market of Europe (after the Russian Federation) with more than 2 millions tons sold (data extracted from FOASTAT, 2009). The quality of the seafood supply is regularly controlled by governmental agencies but their conclusions are not made available to the public. This study is the first evaluation of the rate of fish species substitution in this country. We did not focus on species suspected of higher substitution rates based on the experience from other countries but rather on the comparison of different modes of retail.

We observed an overall rate of substitution of 3.7 % (ci 2.2-6.4%), which is low compared to the rates reported for other countries (table 3). However, the sample acquisition method here is not standardized, and it differs across studies. Ours was designed to sample a large array of fish products, and not only fresh fillets or restaurants meals, known to be prone to species substitution : we also sampled industrial products like deep-frozen fillets or ready-made meals. In previous studies, this type of products had been shown to present either very low substitution rate (below 1.5% for fish fingers, Huxley-Jones et al., 2011) or much higher high rates (above 30%, Garcia-Vasquez et al., 2011 ; Di Pinto et al., 2013).

Both the species easily available and the ones preferred by the consumers differ from country to country, so comparisons across studies and countries are limited. Despite these limitations, cod is the species represented over most studies by the largest number of samples and is the most adapted for comparisons. Our cod sampling is similar in size to the sampling of several cod-centered publications (Miller and Mariani, 2010 ; Miller et al., 2012 ; Di Pinto et al., 2013). In comparison, our substitution rate is one order of magnitude lower than in Italy (Di Pinto et al., 2013), or Ireland (Miller and Mariani, 2010 ; Miller et al., 2012, table 3), but close to UK (7.4% in Miller et al., 2012). This low rate for France is very encouraging. The explanation of the difference with the neighbouring countries needs to be investigated, however, and might even provide some answer for the amelioration of substitution rates elsewhere.

While the substitution rate is low, there is a systematic pattern in the substitutions : a species is replaced by one with a lower commercial value, a pattern also observed in other countries and hinting at economic reasons for the substitution. We did not observe substitutions of species claimed to be sustainable by species that are not, like in the UK for the Pacific cod (Miller et al., 2012). There was a very low proportion of *Gadus macrocephalus* in our samples, although it is an accepted

species under this name and preferable to the largely depleted *G. morhua*.

330 Since the number of samples per species shows a high variability (fig 1) and the substitution rate is low, we observed only a few cases of mislabelling per fish name category, preventing comparisons between them. However, we detected an effect of name category on the substitution rate. This effect was mostly due to bluefin tuna. Strikingly, we found this species to be highly mislabeled with 5 out of 6 samples being substituted (i.e. 83 %, ci : 36-99%), which stands in sharp contrast with the
335 overall low substitution rate. Moreover, for 16 samples collected in sushi restaurants the waiters replied upon enquiry that the tuna sold was bluefin tuna, which was never the case (data included in BOLD). This oral information was in contradiction with the commercial name indicated on the menu when it was photographed. Although we excluded these samples from our analysis, this shows an absence of care or knowledge in the usage of this commercial name. There are at least two
340 plausible explanations for the high mislabelling of this fish category. First, bluefin tuna is called « thon rouge » (« red tuna ») in French. This might confuse waiters and customers, as fresh tuna meat is reddish. Therefore, any uncooked tuna meat can appear as « red ».

Second, this species is considered on the French and other markets to have a high quality meat and would appear more attractive to the customers. Bluefin tuna catches have been subjected to large
345 debates and the species has been largely exposed in the media, which might in turn have increased the attention of the customers and then its prestige and price on the market. It is probably the most lucrative fisheries in the world, driven by a strong demand from Japanese market (80% of the global catches). This commercial importance led to severe overfishing during the 1990-2000s, with estimates of declines in stocks by 72% in the Eastern Atlantic, and by 82% in the Western Atlantic
350 (ICCAT, 2009). International concern over the species survival culminated in 2009 with the proposal to protect Bluefin tuna under the UN Convention on International Trade in Endangered Species (CITES, 2008), which was eventually rejected. Since then, the implementation of strengthened management measures resulted in reductions in catches and fishing mortality rates, indicating that the species may be slowly recovering (Fromentin, 2008 ; ICCAT 2013).

355 These two factors might have acted synergistically, the high demand of customers pushing the retailers to take advantage of the confusing French name of the species.

The probability of substitution is probably also influenced by the mode of distribution, although this trend is not statistically significant in our study. This might be due to the small number of substitutions observed (n=14), but several lines of evidence suggest that there might be a real
360 difference. First, we found no case of substitution in industrially processed food like prepared meals

(n=67) or deep-frozen fillets (n=33, figure 3). For a species heavily used by the industry, like the alaska pollock, we observed no case of substitution despite a decent sample size (n=33). Second, 10 out the 14 substitutions were investigated by interviews with 4 fish retailers, 2 restaurant owners, 4 supermarket executives and other actors along the supply chain. In 5 cases, the people responsible
365 for the last step before the fish reaches the consumer admitted the intentional substitution for increased profit or consumer expectation reasons (Le Loët, 2014), in good agreement with previous studies in other countries (Jaquet & Pauly, 2008).

The collaborative collecting approach proved reliable and efficient, as this is the largest dataset assembled for an European country. Citizen science has emerged in the last decades as a mean for
370 scientists to have access to large datasets extending the studies in space and time (Hochachka et al., 2012) or to have humans performing tasks that computers can not, as exemplified by the Galaxy Zoo (Clery, 2011) or FoldIt (Cooper et al., 2010) projects. However some authors have distinguished different types of collaborative work between scientists and citizen, depending on the involvement of citizens in the research tasks (Cooper et al., 2007). Our study provides an example
375 of a mixed type of research management. The study was initiated by non-scientists involved in controlling the economic use of natural marine resources who were then joined by professional scientists to ensure that the study will meet the stringent criteria of peer-reviewed science, a model referred to as « participatory action research » by Cooper et al. (2007). Finally, the volunteers were then recruited to enlarge the dataset, following a research model more common in citizen science.

380 So our study has mobilized two types of citizens : the initiators of the study, who actively participated to all the research tasks, and the collectors, who enabled the large scale of the study by collecting additional samples. Collectors were generally enthusiastic out of curiosity for the possible substitution cases they might have been the victim of. Citizen initiated science is made possible by two key factors in this type of studies. The first one is the developement of cheap, fast
385 and standard techniques such as DNA sequencing associated with universal libraries like BOLD. With these tools, high quality datasets can be produced for only a few thousands or even hundred euros, making them affordable and accessible for citizens grouped into non-governmental organizations. The second one is the facilitated access to scientific results and publications for interested citizens via the development of open journals which completely remove the cost of
390 reading and lower the cost of publishing. However, the analysis and interpretation of the results still needs qualified input, as the species identification is not straightforward yet, especially for very closely related species like tuna. Reliability of the reference sequences still remains a challenge, as repeatedly pointed out before (Wong & Hanner, 2008 ; Hanner et al., 2011), although there is

395 constant improvement of the reliability through curation and independent corroboration by addition
of sequences from projects run by different laboratories and researchers.

While this type of project generates high enthusiasm in the collectors, there are a few points of
interest for future studies that we can point out here. Clarity and precision in the protocols is
paramount. However, there can be a number of unforeseen problems with the samples. A repeated
400 problem for sample conservation is the size of the tissue sampled. Despite clear instructions, many
collectors were worried a small piece of muscle would not suffice and included a larger piece. The
conservation agent was not in sufficient amount to preserve the sample optimally, and some fresh
fillet samples presented amplification problems.

The pictures taken by the collectors, while appearing at first like a superfluous addition, were
incredibly useful to check and sometimes complete the informations on the samples. Photos of
405 packaging and menus are especially helpful ; taking a picture is a fast way to document a collecting
event, including all available information with a time stamp.

Campaign monitoring was made possible by the Epicollect application (Aanensen et al., 2010)
which allows collectors to fill a form from a smartphone and upload it on a database, along with a
picture and a GPS location. Although only a part of the data were collected by this means, they were
410 used to estimate the progress of the sampling which proved to be very useful.

While these results are encouraging, more research is needed with a higher number of samples per
species and provenance, with a more systematic sampling, focussing on species particularly subject
to substitution. Such a restricted taxonomic focus would make use of more precise, taxon-specific
tools (for instance Vinas & Tuleda, 2009 for tuna) and provide additional information on the quality
415 and precise origin of the samples. The present approach is also not adapted to the study of mixed
samples containing multiple species. The generalisation of next generation sequencing techniques is
very promising for manageable studies of this type of samples, as well as identifications at the
geographic sub-population level for species with such a structuration.

Despite limitations in a few taxa, DNA barcoding based on the COI sequence provided fast,
420 efficient and unambiguous identifications for most of our commercial fish samples, in line with
previous studies, even when only a short fragment was used (Hajibabaei et al., 2006 ; Meusnier et
al., 2008). Improved reliability of the reference dataset will make this tool even more useful and
straightforward to use in the following years, once the uncertain taxonomic situations and
identification problems of some samples are investigated. The BOLD hosted dataset gave a
425 resolution superior to the one in GenBank, and the tools available with the database permit an easier

evaluation of dataset quality and homogeneity.

430 The scientific names were indicated for only a low proportion of the samples at sale. In France like in other countries, legislation on labelling differs between restaurants, fresh sales and deep-frozen fish. For some groups, the vernacular name covers a large number of species (rays, tuna), including species with serious conservation concerns, so there is no way for knowledgeable consumers to choose according to sustainability criteria. We join Miller et al. (2012) and Jaquet et Pauly (2008) in their call for more precise and informative labelling up to the consumers, a requisite already present in the legislation but not enforced everywhere, and with a large number of exceptions allowed such as those for restaurants.

435 **Acknowledgements**

We thank all our collectors for their great and thorough work, and Baptiste Carton and Alexia Velasquez for help with the labwork and the sequence management. Laboratory access and assistance was provided by the "Service de Systematique Moleculaire" of the Muséum National d'Histoire Naturelle (CNRS UMS 2700). We thank Dusan Misevic for commenting on earlier
440 versions of this papern, and Gael Lancelot for language corrections.

References

- Aanensen DM, Huntley DM, Feil EJ, al-Own F, Spratt BG., 2009. EpiCollect: Linking Smartphones to Web Applications for Epidemiology, Ecology and Community Data Collection. *PLoS ONE* 4(9): e6968. doi:10.1371/journal.pone.0006968
- 445 Barbuto M, Galimberti A, Ferri E, Labra M, Malandra R, Galli P, Casiraghi M .2010. DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of “palombo” (Mustelus spp.). *Food Research International* 43: 376–38.
- Cawthorn D-M, Steinman H A, Witthuhn R C .2012. DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. *Food Research International* 46: 30–40.
- 450 CITES., 2008. Proposal to include Atlantic Bluefin Tuna (*Thunnus thynnus* (Linnaeus, 1758)) on Appendix I of CITES in accordance with Article II 1 of the Convention. <http://www.cites.org/sites/default/files/eng/cop/15/prop/E-15-Prop-19.pdf>
- Clery D. 2011 Galaxy Zoo volunteers share pain and glory of research. *Science* 33:173.
- 455 Cline, E. 2012. Marketplace substitution of Atlantic salmon for Pacific salmon in Washington State detected by DNA barcoding. *Food Research International* 45 388–393.
- Cochrane K; De Young C; Soto D; Bahri T (eds)., 2009. Physical and ecological impacts of climate change relevant to marine and inland capture fisheries and aquaculture. In : *Climate change implications for fisheries and aquaculture: overview of current scientific knowledge*. Cochrane K; De Young C; Soto D; Bahri T (eds). FAO Fisheries and Aquaculture Technical Paper. No. 530. Rome, FAO., 2009. 212p. http://www.fao.org/fileadmin/user_upload/newsroom/docs/FTP530.pdf
- 460 Committee on the Environment, Public Health and Food Safety, European Parliament, 2013. Draft Report on the food crisis, fraud in the food chain and the control thereof. De Lange, E. (Rapporteur). PR\1005774EN.doc. [http://www.europarl.europa.eu/sides/getDoc.do?](http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+COMPARL+PE-519.759+02+DOC+PDF+V0//EN&language=EN)
- 465 [pubRef=-//EP//NONSGML+COMPARL+PE-519.759+02+DOC+PDF+V0//EN&language=EN](http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+COMPARL+PE-519.759+02+DOC+PDF+V0//EN&language=EN)
- Consumer guides list under http://www.slowfood.com/slowfish/pagine/fra/pagina.lasso?-id_pg=94. Accessed 25/02/2014.
- Cooper CB, Dickinson J, Phillips T, and Bonney R., 2007. Citizen Science as a Tool for Conservation in Residential Ecosystems. *Ecology and Society* 12:11.
- 470 Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen L, Leaver-Fay A, Baker D, Popovi Z

>57,000 Foldit players. 2010. Predicting protein structures with a multiplayer online game. *Nature* 466:756.

Costa FO, Landi M, Martins R, Costa MH, Costa ME, Carneiro M, Alves MJ, Steinke D, Carvalho GR. 2012. A ranking system for reference libraries of DNA barcodes: application to marine fish species from Portugal. *PLoS One* 2012;7(4):e35858. doi: 10.1371/journal.pone.0035858.

Cutarelli A, Amoroso MG, De Roma A, Girardi S, Galiero G, Guarino A, Corrado F .2014. Italian market fish species identification and commercial frauds revealing by DNA sequencing. *Food Control* 37:46-50.

480 Dettai A, Lautredou A-C, Bonillo C, Goimbault E, Busson F, Causse R, Couloux A, Cruaud C, Duhamel G, Denys G, Hautecoeur M, Iglesias S, Koubbi P, Lecointre G, Moteki M, Pruvost P, Tercerie S, Ozouf-Costaz C. 2011. The actinopterygian diversity of the CEAMARC cruises: Barcoding and molecular taxonomy as a multi level tool for new findings. *Deep Sea Research II*. 58: 250-263 dx.doi.org/10.1016/j.dsr2.2010.05.021

485 Di Pinto A, Di Pinto P, Terio V, Bozzo G, Bonerba E, Ceci E, Tantillo G. 2013. DNA barcoding for detecting market substitution in salted cod fillets and battered cod chunks. *Food Chemistry* 141: 1757–1762.

Ekrem T, Willassen E, Stur E., 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetic and Evolution* 43, 530–542.

490 FAOSTAT., 2009. Fishery Statistical Collections. Consumption of Fish and Fishery Products. <http://faostat.fao.org/site/610/default.aspx#ancor>

Filonzi L, Chiesa S, Vaghi M, Nonnis Marzano F. 2010. Molecular barcoding reveals mislabelling of commercial fish products in Italy. *Food Research International* 43 :1383–1388.

Fromentin, J.M. (2008) Le thon rouge, une espèce surexploitée. Ifremer, Paris. http://wwz.ifremer.fr/content/download/41835/569837/file/13_01_28_DP_Thon_rouge.pdf

495 FSAI (Food Safety Authority of Ireland). 2011. Fish Labelling Survey. http://www.fsai.ie/resources_publications.html. Last accessed 28-03-2014

Garcia-Vazquez E, Perez J, Martinez JL, Pardinas AF, Lopez B, Karaiskou N, Casa MF, Machado-Schiaffino G, Triantafyllidis A. 2011. High Level of Mislabeling in Spanish and Greek Hake Markets Suggests the Fraudulent Introduction of African Species. *J. Agric. Food Chem.* 59: 475–480. DOI:10.1021/jf103754r

- Griffiths AM, Miller DD, Egan A, Fox J, Greenfield A, Mariani S . 2013. DNA barcoding unveils skate (Chondrichthyes: Rajidae) species diversity in ‘ray’ products sold across Ireland and the UK. *PeerJ* 1:e129; DOI 10.7717/peerj.129
- Hajibabaei M, Smith A, Janzen DH, Rodriguez JJ, Whitfield JB, Hebert PD., 2006. A minimalist
505 barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes* 6:959-964.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Research Symposium Series* 41: 95–98.
- Hanner RH, Becker S, Ivanova NV, Steinke D . 2011. FISH-BOL and seafood identification:
510 Geographically dispersed case studies reveal systemic market substitution across Canada. *Mitochondrial DNA*, 22(S1): 106–122
- Harris DJ., 2003. Can you bank on GenBank? *Trends in Ecology and Evolution* 18, 317–319.
- Hochachka WM, Fink D, Hutchinson RA, Sheldon D, Wong W and Kelling S. 2012. Data-intensive science applied to broad-scale citizen science. *Trends in Ecology and Evolution* 27:130
- 515 Huxley-Jones E, Shaw JLA, Fletcher C, Parnell J, Watts PC. 2012. Use of DNA Barcoding to Reveal Species Composition of Convenience Seafood. *Conservation Biology* 26(2): 367-371.
- ICCAT., 2009. Report of the 2008 Atlantic Bluefin Tuna Stock Assessment Session. Collective Volume of Scientific Papers ICCAT 64: 1-352.
http://www.iccat.int/Documents/Meetings/Docs/2008_BFT_STOCK_ASSESS_REP.pdf
- 520 ICCAT. 2013. Report of the 2012 Atlantic Bluefin Tuna Stock Assessment Session. Collective Volume of Scientific Papers ICCAT. In press.
http://www.iccat.int/Documents/CVSP/CV069_2013/colvol69.html
- Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN., 2007. Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes* 7:544-548.
- 525 Jacquet JL, & Pauly D., 2008. Trade secrets: renaming and mislabeling of seafood. *Marine Policy*, 32: 309–318.
- Le Loët K. 2014. Poissons : fraude à l’étiquette. *Terraeco* 55, march 2014: 42-52.
<http://www.terraeco.net/Poissons-Fraude-a-l-etiquette,53866.html>
- Lowenstein JH, Amato G, Kolokotronis S-O., 2009. The Real maccoyii: Identifying Tuna Sushi
530 with DNA Barcodes – Contrasting Characteristic Attributes and Genetic Distances. *PLoS ONE*

4(11): e7866. doi:10.1371/journal.pone.0007866

Machado-Schiaffino G, Martinez JL, Garcia-Vazquez E., 2008. Detection of Mislabeling in Hake Seafood Employing mtSNPs-Based Methodology with Identification of Eleven Hake Species of the Genus *Merluccius*. *J. Agric. Food Chem* 56, 5091–5095.

535 Meunier M, Daurès F, Girard S. 2013. « Etat des lieux des secteurs pêche et aquaculture et de la consommation des produits aquatiques. Approche nationale (France) et régionale (Bretagne) », Publications électroniques Amure, Série Rapports R-32-2013, 41p. Disponible : http://www.umr-amure.fr/electro_rap_amure/R_32_2013.pdf (Consulté le 01.03.2014*)

540 Meusnier I, Singer GA, Landry JF, Hickey DA, Hebert PD, Hajibabaei MA., 2008. Universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* 12(9):214. doi: 10.1186/1471-2164-9-214.

Miller DD, Jessel A, Mariani S. 2012. Seafood mislabelling: comparisons of two western European case studies assist in defining influencing factors, mechanisms and motives. *Fish and fisheries* 13: 345–358.

545 Miller DD, Mariani S. 2010. Smoke, mirrors, and mislabeled cod: poor transparency in the European seafood industry. *Frontiers in Ecology and the Environment* 8(10): 517-521.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Rasmussen RS, Morrissey MT., 2008. DNA-Based methods for the identification of commercial fish and seafood species. *Compr. Rev. Food Sci. Food Safety* 7, 280– 295.

550 Ratnasingham S, Hebert P., 2007. BOLD: The Barcode of Life Data System. (www.barcoding.life.org) *Mol Ecol Notes* 7:355–364. Last accessed 03-03-2014

Regulation (EC) <http://eur-lex.europa.eu/>. Last accessed march 2014.

Smith PJ, McVeagh SMM, Steinke D., 2008. DNA barcoding for the identification of smoked fish products. *Journal of Fish Biology* 72: 464–471. doi:10.1111/j.1095-8649.2007.01745.x

555 Vinas J, Tudela S., 2009. Validated methodology for genetic identification of tuna species (genus *Thunnus*). *PloSONE* 4:e7606.

Ward RD, Hanner R, Hebert PD., 2009. The campaign to DNA barcode all fishes. FISH-BOL. *Journal of Fisheries Biology* 74, 329–356.

560 Ward RD, Zemplak TS, Innes BH, Last PR, Hebert PD, 2005. DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London B Biological Science*

360:1847–1857.

Winnepenninckx B, Backeljau T, De Wachter R. 1993. Extraction of high molecular weight DNA from molluscs. *Trends in Genetics* 9 (1993), 407.

565 Wong EH-K, Hanner RH., 2008. DNA barcoding detects market substitution in North American seafood. *Food Research International* 41: 828–837

Yancy HF, Zemplak TS, Mason JA, Washington JD, Tenge BJ, Nguyen NT, Barnett JD, Savary WE, Hill WE, Moore MM, Fry FS, Randolph SC, Rogers PL and Hebert PDN., 2008. The Potential Use of DNA Barcodes in Regulatory Science: Applications of the Regulatory Fish Encyclopedia. *Journal of Food Protection* 71(1):210-7.

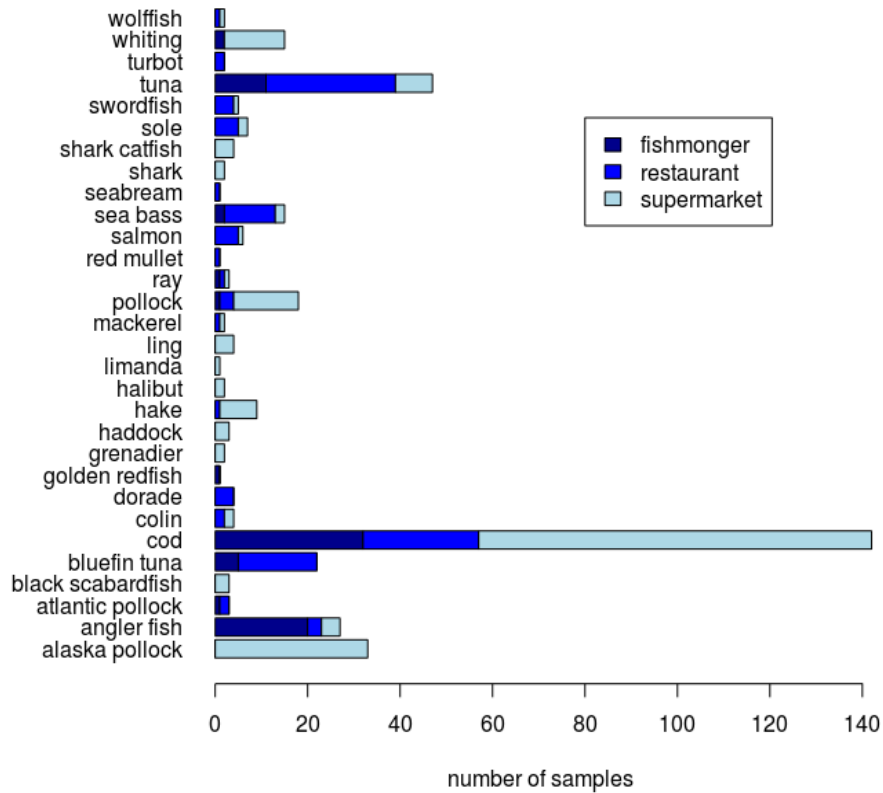


Figure 1. *Sampling by commercial name and shop types*

The sampling was very diverse, with 30 commercial names represented after grouping between names representing the same species of close ones. A few account for most of the samples (cod, tuna, alaska pollock and anglerfish) while a long tail of species include less than 4 samples. For the statistical analysis, only categories with more than 10 samples were kept and the rest were grouped together.

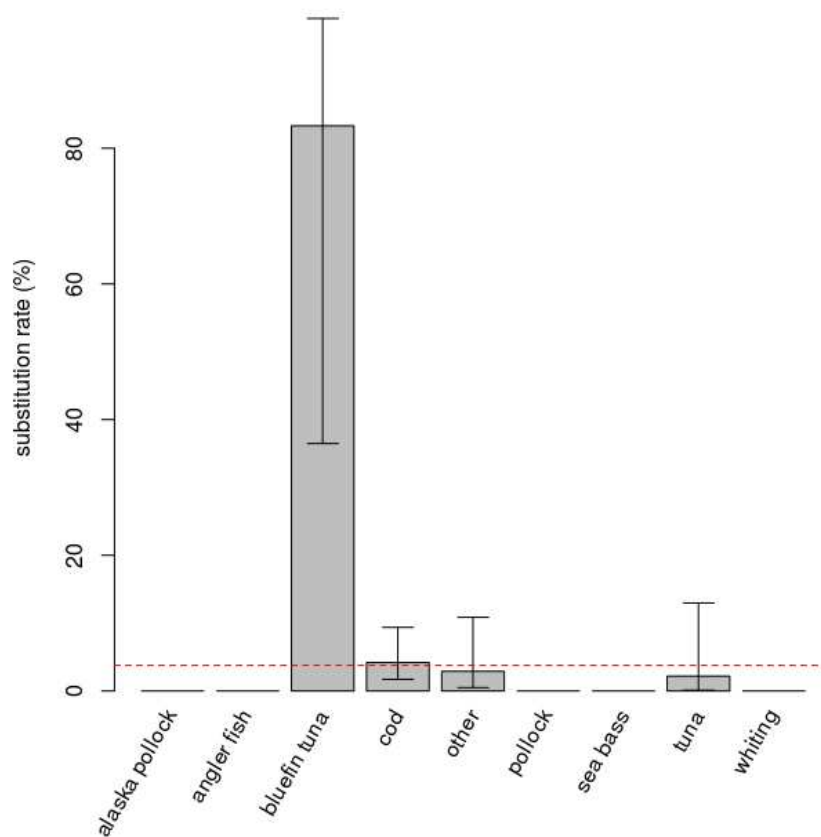


Figure 2. *Substitution rates for different commercial name category*

Substitutions are observed only in three of the species categories with more than 10 samples collected. They have comparable low substitution rates. Bluefin tuna displays an exceptionally high substitution rate, and was separated from other tuna species in the figure and in analyses, despite a very low number of samples ($n=6$). Error bars show the 95 % confidence interval. The red dashed line is the average substitution rate observed for the entire dataset.

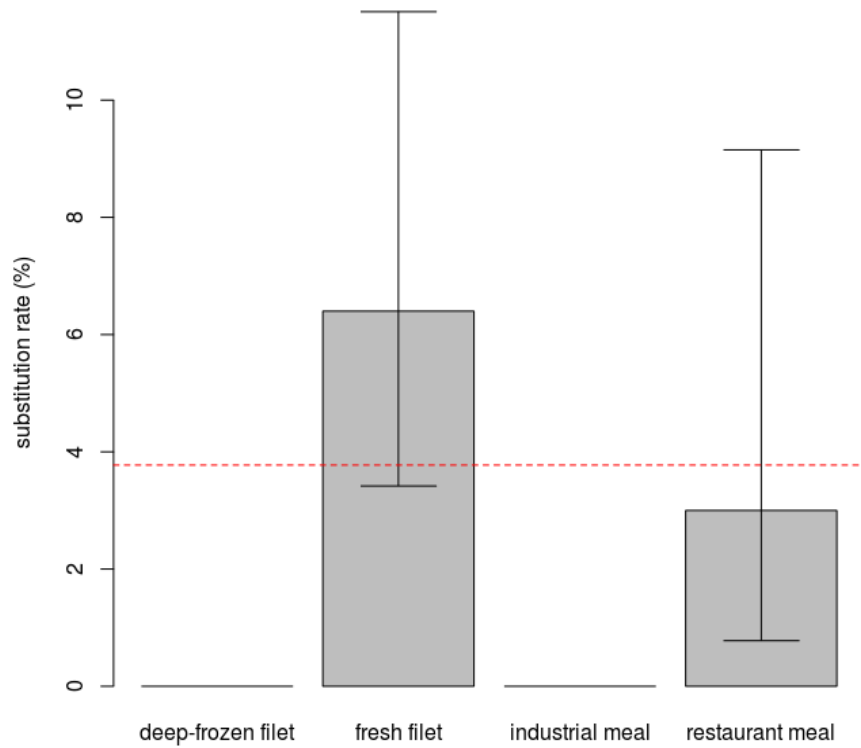


Figure 3. *Substitution rates for different retail modes*

585 Mislabelling was detected only for fresh filets and restaurants meals. Error bars show the 95 % confidence interval. The red dashed line is the average substitution rate observed for the entire dataset.

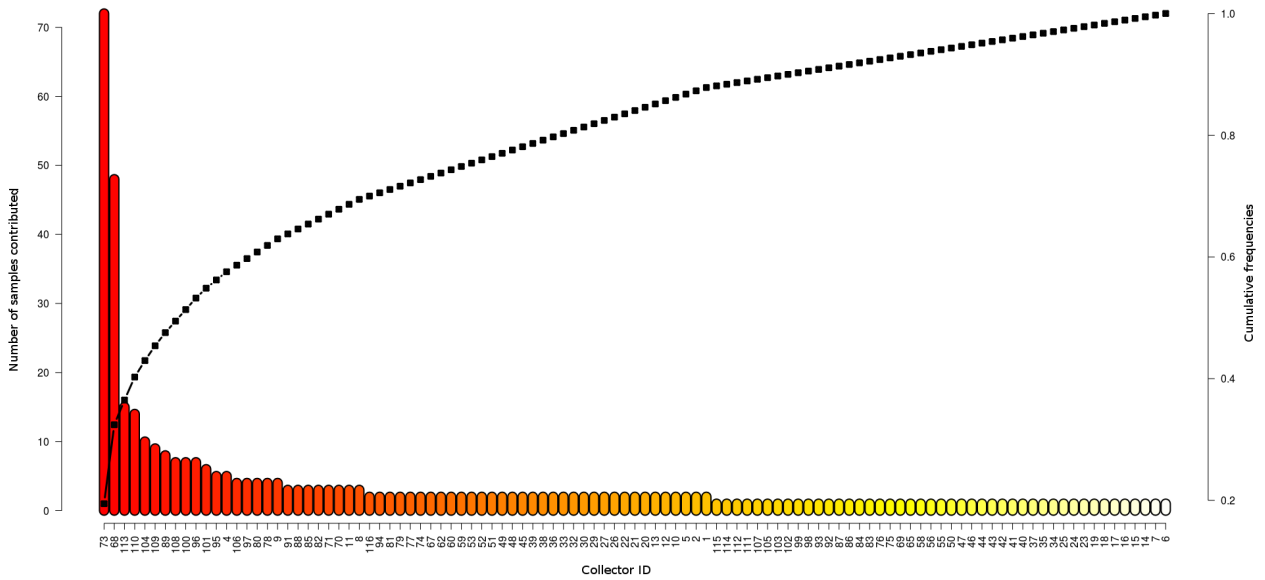


Fig 4. Pareto diagram of the number of samples per collector.

Bars represent the number of samples for each collector (identified by numbers on the x-axis).

590

Points represent the cumulative frequency of the samples. The sampling effort was unevenly distributed, with a few collectors providing most of the samples and most of the collectors sending 3 samples or less.

Tables

Commercial name	English name	Accepted species*	Fishmonger	Number of samples (substitution cases)			Ready-made dish	Total
				Supermarket fresh filet	Deep-frozen filets	Restaurants		
FL samples								
Thon*	Tuna	<i>Thunnus albacares</i> , others	5	1		8		14
Thon rouge	Atlantic bluefin tuna	<i>Thunnus thynnus</i>	1(1)			1(1)		2
Thon germon	Albacore tuna	<i>Thunnus alalunga</i>	1			1		1
Thon albacore	Yellowfin tuna	<i>Thunnus albacares</i>	2	4(1)	1	4	1	12
Thon listao	Skipjack tuna	<i>Katsuwonus pelamis</i>					1	1
Maquereau	Atlantic mackerel	<i>Scomber scombrus</i>				1	1	2
Cabillaud, morue	Cod (both Atlantic and Pacific cod)	<i>Gadus morhua</i> , <i>G. macrocephalus</i> , <i>G. ogac</i> , <i>Arctogadus glacialis</i> , <i>Boreogadus saida</i> , <i>Eleginus navaga</i> , <i>E. gracilis</i>	12	27(2)	8	19(1)	23	89
Colin d'Alaska, Lieu d'Alaska	Alaska pollock	<i>Gadus chalcogrammus (Theragra chalcogramma)</i>		2	4		27	33
Lieu jaune	Pollack	<i>Pollachius pollachius</i>	1			1		2
Lieu noir, lieu, colin (lieu)	Saithe	<i>Pollachius virens</i>	1	12	1	3	3	20
Aiglefin, églefine	Haddock	<i>Melanogrammus aeglefinus</i>		1			2	3
Merlan	Whiting	<i>Merlangius merlangus</i>	2	11			2	15
Julienne	Ling	<i>Molva molva</i>		3				3
Lingue	Blue ling	<i>Molva dypterygia</i>		1				1
Merlu blanc, merlu blanc du Cap, merlu (restaurant)	Hake	<i>Merluccius capensis</i> , <i>M. bilinearis</i> , <i>M. cadenati</i> , <i>M. gayi gayi</i> , <i>M. albidus</i> , <i>M. hubbsi</i> , <i>M. paradoxus</i> , <i>M. polylepis</i> , <i>M. productus</i> , <i>M. senegalensis</i>			2	1	5	8
Merlu, colin	European hake	<i>Merluccius merluccius</i>		1		2		3
Grenadier	Roundnose grenadier	<i>Coryphaenoides rupestris</i>		2				2
Lotte	Anglerfish	<i>Lophius americanus</i> , <i>L. piscatorius</i> , <i>L. budegassa</i>	1	1		4		6
Lotte du Cap	Devil anglerfish	<i>Lophius vomerinus</i>			1			1
Bar, Loup		<i>Dicentrarchus labrax</i>	2	2		11		15
Sole	Common sole	<i>Solea solea</i>		2(1)		5		7
Flétan	Atlantic halibut	<i>Hippoglossus hippoglossus</i> , <i>H. stenolepis</i>		2				2
Turbot	Turbot	<i>Scophthalmus maximus</i>				2		2
Limande du Nord	Yellowfin sole	<i>Limanda aspera</i>					1	1
Espadon	Swordfish	<i>Xiphias gladius</i>		1		4		5
Daurade royale	Gilthead seabream	<i>Sparus aurata</i>				4		4
Sébaste*	Rockfish	<i>Sebastes sp.</i>	1					1
Pagre (restaurant)	Red porgy	<i>Pagrus pagrus</i> , <i>Pagrus sp.</i>				1(1)		1
Loup de mer	Wolfish	<i>Anarhichas denticulatus</i> , <i>A. lupus</i> , <i>A. minor</i>		1		1		2
Sabre noir	Black scabbardfish	<i>Aphanopus carbo</i>		3				3
Saumon, saumon atlantique	Salmon	<i>Salmo salar</i>				4	1	5
Panga	Panga	<i>Pangasius hypophthalmus</i>		4				4
Raie	Ray	<i>Leucoraja naevus</i> , <i>Bathyraja brachyurops</i> + 25 species	1		1	1		3
Requin peau bleue	Blue shark	<i>Prionace glauca</i>		2				2
Rouget		<i>Mullus surmuletus</i>				1		1
							Total	276
TE samples								
			Fishmonger	Supermarket fresh filet	Deep-frozen filets	Restaurants	Ready-made dish	
Thon	Tuna		2			15		17
Thon albacore	Yellowfin tuna	<i>Thunnus albacares</i>		1				1
Thon germon, thon blanc	Albacore tuna	<i>Thunnus alalunga</i>	1					1
Thon rouge	Atlantic bluefin tuna	<i>Thunnus thynnus</i>	4(3)			16		20
Lotte	Anglerfish	<i>Lophius americanus</i> , <i>L. piscatorius</i> , <i>L. budegassa</i>	19	2				21
Cabillaud, morue	Cod (both Atlantic and Pacific cod)	<i>Gadus morhua</i> , <i>G. macrocephalus</i> , <i>G. ogac</i> , <i>Arctogadus glacialis</i> , <i>Boreogadus saida</i> , <i>Eleginus navaga</i> , <i>E. gracilis</i>	20(3)	12	15	6		53
Saumon	Salmon	<i>Salmo salar</i>				1		1
							Total	114

595 **Table 1.** Summary of the sampled species

The samples are presented per protocol (FL or TE) and commercial name. When several species are sold under one name, the species present in the dataset according to molecular ID are highlighted in bold.

600 The number in brackets are the number of substitutions observed after sample sorting. Table assembled using <http://www.economie.gouv.fr/dgccrf/Poissons>. * no qualifier is needed for restaurants the commercial name to be correct.

Sample	Commercial name (latin name if indicated)	English name	Dataset length	Similarity with BOLD sequences	Species identification	Origin	Zipcode
Tuna							
TE14	Thon rouge	Bluefin tuna	L	100.00%	<i>Thunnus albacares</i>	Fishmonger filet	75
TE32	Thon rouge (Thunnus thynnus)	Bluefin tuna	S	100.00%	<i>Thunnus obesus</i>	Fishmonger filet	75
TE109	Thon rouge	Bluefin tuna	L	100.00%	<i>Thunnus albacares</i>	Fishmonger filet	75
FL0183	Thon rouge	Bluefin tuna	L	100.00%	<i>Thunnus albacares</i>	Restaurant	50
FLID1031	Thon rouge	Bluefin tuna	L	100.00%	<i>Thunnus albacares</i>	Fishmonger filet	75
FLID116	Thon albacore	Yellowfin tuna	M	100.00%	<i>Thunnus obesus</i>	Supermarket fresh filet	75
Cod							
TE63	Cabillaud	Cod	S	100.00%	<i>Melanogrammus aeglefinus</i>	Fishmonger filet	75
TE112	Cabillaud (Gadus morhua)	Cod	L	99.69%	<i>Melanogrammus aeglefinus</i>	Fishmonger filet	75
TE190	Cabillaud	Cod	L	99.85%	<i>Melanogrammus aeglefinus</i>	Fishmonger filet	75
FL0196	Cabillaud	Cod	M	100.00%	<i>Pollachius virens</i>	Restaurant	75
FL0572	Cabillaud	Cod	M	100.00%	<i>Melanogrammus aeglefinus</i>	Supermarket fresh filet	77
FL0963	Cabillaud	Cod	L	100.00%	<i>Melanogrammus aeglefinus</i>	Supermarket fresh filet	76
Sole							
FLID089	Sole	Sole	L	99.19%	<i>Cynoglossus senegalensis</i>	Supermarket fresh filet	77
Red Porgy/seabream							
FL0007	Pagre	Red porgy	M	100.00%	<i>Gadus morhua</i>	Restaurant	75

L=Long primer, M = Medium-size primer, S=Short primer

Table 2. *Substitutions observed*

The “thon rouge” (bluefin tuna) category account for 5 out of the 14 substitutions observed in our sampling (n=371), although it contains only 6 samples.

Investigated country	Substitution rate	Nb of sequences for commercial samples	Taxonomic focus, if any	Origin of samples	Type of samples	Marker	Reference
EU							
Ireland	19.00%	111	diverse	F, S, R	FI, Fr, P, Rd	COI	FSAI 2011
Ireland	25.00%	156	cod	F, S, F&C	FI, Fr, Rd	COI	Miller & Mariani 2010
Ireland	28.20%	131	cod	F, S, F&C	FI, Fr, P, Rd	COI	Miller et al. 2012
Ireland/UK	na	98	Rajidae	F, S, F&C	FI, Rd	COI	Griffith et al. 2013
UK	7.40%	95	cod	F, S, F&C	FI, Fr, P, Rd	COI	Miller et al. 2012
UK	<1.5%	142	diverse	S	P	COI	Huxley-Jones et al. 2011
Italy	32.00%	69	diverse	F, S	FI, Fr	COI & Cytochrome b	Filonzi et al. 2010
Italy	77.80%	59	Mustelus sp.	F, S	FI	COI	Barbuto et al. 2009
Italy	56.36%	110	cod	S	FI, P	COI	Di Pinto et al. 2013
Italy	20.00%	18	diverse	Port authority	P	COI & Cytochrome b	Curatelli et al. 2014
Spain	>20%	40	Hake	S	Fr, P	Mt Control region SNPs	Machado-Schiaffino et al. 2008
Spain & Greece	>30%	279 (93*3)	Hake	S	Fr	5S rDNA, CytB RFLP	Garcia-Vasquez et al. 2011
France	3.7% (ci 2.2-6.4)	390	diverse	F, S, R	FI, Fr, P, Rd	COI	present study
Non-EU							
Japan	8.00%	26	Tuna	F, R	FI, Rd	COI, Mt Control region, ITS 1	Vinas & Tudela 2009
South Africa	21.00%	248	diverse	S, F	FI, Fr, P	COI	Cawthorne et al. 2012
Canada	41.20%	236	diverse	F, S, R	FI, Fr, Rd	COI	Hanner et al. 2011
US	32.35%	68	Tuna	R	Rd	COI	Lowenstein et al. 2009
US	11.00%	99	Salmon	R, S	FI	COI	Cline 2012
US & Canada	25.00%	90	diverse	F, R	FI, Rd	COI	Wong & Hanner 2008

F=Fishmongers, S=supermarkets, F&C=Fish and chips, R=restaurants
Fr=frozen, P=prepared dish (includes fishfingers and battered), FI=Fillet, Rd=Restaurant dish

Table 3. Comparison of substitution rates observed in similar studies

All these studies use molecular identification to estimate the rate of species substitution. Although their methodologies and scopes are not directly comparable, only a few of them reported rates below 10%.