

**A high-quality genetic reference database for European commercial fishes reveals substitution fraud of processed Atlantic cod (*Gadus morhua*) and common sole (*Solea solea*) at different steps in the Belgian supply chain**

Authors: Dumas Deconinck<sup>1,2\*</sup>, Filip A.M. Volckaert<sup>2</sup>, Kris Hostens<sup>1</sup>, Remigiusz Panicz<sup>3</sup>, Piotr Eljasik<sup>3</sup>, Miguel Faria<sup>4</sup>, Carolina Sousa Monteiro<sup>4</sup>, Johan Robbens<sup>1</sup>, Sofie Derycke<sup>1,5</sup>

<sup>1</sup> Flanders Research Institute for Agriculture, Fisheries and Food, ILVO, Aquatic Environment and Quality, Ankerstraat 1, B-8400 Oostende, Belgium

<sup>2</sup> Laboratory of Biodiversity and Evolutionary Genomics, KU Leuven, Ch. Deberiotstraat 32, B-3000 Leuven, Belgium

<sup>3</sup> Department of Meat Science, Faculty of Food Science and Fisheries, West Pomeranian University of Technology Szczecin, Kazimierza Królewicza 4 71-550 Szczecin, Poland

<sup>4</sup> LAQV-REQUIMTE, Faculty of Pharmacy, University of Porto, R. Jorge de Viterbo Ferreira 228, 4050-313 Porto, Portugal

<sup>5</sup> Department of Biology, Ghent University, Krijgslaan 281, Building S8, 9000 Ghent, Belgium

\* Corresponding author:

Sofie Derycke, Email: [sofie.derycke@ilvo.vlaanderen.be](mailto:sofie.derycke@ilvo.vlaanderen.be); phone: +32 59 56 98 89; ORCID: 0000-0003-3763-6187

Other authors:

Dumas Deconinck, Email: [dumas.deconinck@ilvo.vlaanderen.be](mailto:dumas.deconinck@ilvo.vlaanderen.be); phone: +32 59 56 98 56; ORCID: 0000-0002-2854-7708

Filip Volckaert, Email: [filip.volckaert@kuleuven.be](mailto:filip.volckaert@kuleuven.be); phone: +32 16 32 39 72; ORCID: 0000-0003-4342-4295

Kris Hostens, Email: [kris.hostens@ilvo.vlaanderen.be](mailto:kris.hostens@ilvo.vlaanderen.be); phone: +32 59 56 98 48; ORCID: 0000-0002-8476-1362

Remigiusz Panicz, Email: [rpanicz@zut.edu.pl](mailto:rpanicz@zut.edu.pl); ORCID: 0000-0003-4858-2233

Piotr Eljasik, Email: [peljasik@zut.edu.pl](mailto:peljasik@zut.edu.pl); +48 605 268 292, ORCID: 0000-0001-7220-5484

Miguel Faria, Email: [mfaria@ff.up.pt](mailto:mfaria@ff.up.pt); phone: +351 220 428 639; ORCID: 0000-0001-9366-0443

Carolina Sousa Monteiro, Email: [carolinasousamonteiro@gmail.com](mailto:carolinasousamonteiro@gmail.com); ORCID: 0000-0003-4168-5862

Johan Robbens, Email: [johan.robbens@ilvo.vlaanderen.be](mailto:johan.robbens@ilvo.vlaanderen.be); phone: +32 59 56 98 50

Journal: SEAFOODTOMORROW Special Issue (SI) in the Food and Chemical Toxicology (FCT)

## Abstract

Seafood is an important component of the human diet. With depleting fish stocks and increasing prices, seafood is prone to fraudulent substitution. DNA barcoding has illustrated fraudulent substitution of fishes in retail and restaurants. Whether substitution also occurs in other steps of the supply chain remains largely unknown. DNA barcoding relies on public reference databases for species identification, but these can contain incorrect identifications. The creation of a high quality genetic reference database for 42 European commercially important fishes was initiated containing 145 *Cytochrome c oxidase subunit I (COI)* and 152 *Cytochrome b (cytB)* sequences. This database was used to identify substitution rates of Atlantic cod (*Gadus morhua*) and common sole (*Solea solea*) along the fish supply chain in Belgium using DNA barcoding. Three out of 132 cod samples were substituted, in catering (6 %), import (5%) and fishmongers (3%). Seven out of the 41 processed sole samples were substituted, in wholesale (100%), food services (50%), retailers (20%) and catering (8%). Results show that substitution of *G. morhua* and *S. solea* is not restricted to restaurants, but occurs in other parts of the supply chain, warranting for more stringent controls along the supply chain to increase transparency and trust among consumers.

## Graphical abstract (separate file, optional)

### Running title

Substitution fraud of Atlantic cod and common sole along the Belgian food supply chain

### Key words

DNA barcoding, *COI* and *cytB*, EU fish, substitution fraud, Belgian supply chain, Atlantic cod, common sole

### Abbreviations

BoLD            Barcode of Life Database

COI	Cytochrome oxidase c subunit I
cytB	Cytochrome b
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EUMOFA	European Marketing Observatory for Fisheries and Aquaculture Products
FAM	Carboxyfluorescein
FAO	Food and Agriculture Organization of the United Nations
IUU	Illegal, unreported and unregulated
PCR	Polymerase Chain Reaction
qPCR	quantitative real time PCR
SQL	Structured Query Language
TAE	Tris-acetate-EDTA
VIC	2'-chloro-7'-phenyl-1,4-dichloro-6-carboxy-fluorescein
VLAM	Vlaams centrum voor agro- & visserijmarketing

## 1. Introduction

Seafood has a worldwide key role in the human diet as a source of popular and healthy food. After a steady increase since the 1950s, worldwide global fish production peaked at 171 million tonnes in 2016, of which 151 million tonnes were used for human consumption (FAO, 2018). As a consequence, fish stocks are under pressure globally (Bryndum-Buchholz et al., 2019; Worm et al., 2006), which is aggravated by climate change (Cheung et al., 2013; Smalås et al., 2019). Pressure on supplies has led to fraudulent practices, such as illegal, unreported and unregulated (IUU) fisheries (Schmidt, 2005) and substitution of higher value species with cheaper alternatives (Fox et al., 2018; “Young’s Seafood,” n.d.). Apart from economic gain, substitution of seafood can also be driven by (1) economically valuable species becoming a limited resource with increased demand (FAO, 2018; Rehbein, 2008), (2) a wide range of visually similar fish being traded (Piñeiro et al., 2001), (3) consumers seeking easy access to low-cost foods (Fox et al., 2018), and (4) attempting to conceal IUU fishing practices (Fox et al., 2018; Vandamme et al., 2016).

Substitution can have economic, ethical and health consequences (Fox et al., 2018; Spink and Moyer, 2011), the latter through toxins and allergens (Kan et al., 1999; Noguchi and Arakawa, 2008; Weinsier, 2016) or through accumulation of harmful substances (Hobbs et al., 2019; Spink and Moyer, 2011). Furthermore, substitution of fish seems to be present mainly at the end point of the supply chain, more specifically in restaurants, canteens and food services, as a consequence of less stringent controls and of processing, which makes the fish less recognisable (Christiansen et al., 2018; Kappel, 2016; Shehata et al., 2017). Little is known about the prevalence of substitution in other parts of the supply chain (Manning and Soon, 2014). A knock-on-effect can be expected, as every subsequent part of the supply chain adds to the original deception (Gordoa et al., 2017). Correct information is critical to help consumers make informed choices (Bénard-Capelle et al., 2015; Meikle and McDonalds, 2013; Mottola et al., 2013) and to increase transparency and safety in the seafood industry.

Substitution might be more prevalent in highly processed seafood products, because morphological characteristics are no longer visible. The rise of molecular techniques which allow distinguishing species without morphologically deterministic traits now enables investigating substitution in highly processed products (Ardura et al., 2016; Bénard-Capelle et al., 2015; Devloo-Delva et al., 2016). DNA barcoding is a popular technique to quickly discover mislabelling (Bénard-Capelle et al., 2015; Christiansen et al., 2018; Hebert et al., 2003; Rees et al., 2014), relying on the amplification of marker genes, like *Cytochrome oxidase c subunit 1 (COI)* and *Cytochrome b (cytB)*. Both regions in the mitochondrial genome allow for the identification of species upon comparing sequences to a reference database (Ward et al., 2005). The *COI* gene is an excellent barcode marker as the universal primers are robust, enabling recovery of its 5' end from representatives of most animal phyla (Folmer et al., 1994; Hebert et al., 2003). In addition, the *cytB* gene is used frequently for fish identifications, as its phylogenetic performance is comparable to that of *COI*, but its greater length increases species level identifications for some groups (Kochzius et al., 2010; Sevilla et al., 2007; Zardoya and Meyer, 1996).

Unfortunately, public databases still contain erroneous sequences that are linked to incorrect species names (Li et al., 2018; Mioduchowska et al., 2018), which complicates the correct assignment of species names to DNA sequences when used as a control or comparison tool. To cope with this, high quality databases have been created, such as the Barcode of Life Data Systems (BoLD) ("Bold Systems v4," n.d.) and Fishtrace ("FishTrace," n.d.). BoLD serves as a reference database for *COI* sequences, containing species name, voucher data, collection record, identifier, used primers and trace files. Fishtrace is a genetic catalogue for *cytB* and *rhodopsin* sequences of commercially important fish species (Ratnasingham and Hebert, 2007; Sevilla et al., 2007). A high quality reference database containing DNA sequences for both most commonly used barcode markers, obtained from morphologically identified and digitally vouchered specimens for the most important fish species traded on the European market is needed to further increase the accuracy of DNA barcoding of fishes in Europe.

As there are differences in substitution rates between species and countries (Bénard-Capelle et al., 2015), this study focused on substitution fraud for two fish species in the Belgian food supply chain. Atlantic cod (*Gadus morhua*) was selected due to its popularity (“Vis - VLAM,” n.d.) and its high selling price in Belgium. It is a highly imported fish species in Belgium, in spite of declining stocks in the southern part of the Atlantic (Cook, 2018). Common sole (*Solea solea*) is one of the most commonly landed fish species by Belgian fishermen. Sole is an expensive product and considered a delicacy in Belgium (“Vis - VLAM,” n.d.).

In this study, three specific objectives were addressed: (1) to initiate the creation of a reliable reference database for both *COI* and *Cytb* sequences of commercially important species in Europe, to ensure correct identification through DNA barcoding; (2) to obtain insight into the current situation of seafood and fish trade in the Belgian supply chain and (3) to evaluate the occurrence of substitution fraud for Atlantic cod *G. morhua* and common sole *S. solea* along the different steps of the Belgian food supply chain.

## **2. Materials and methods**

### **2.1. The SEAFOOD<sup>TOMORROW</sup> reference database**

#### **2.1.1. Creating a digital platform**

Existing databases either contain the *COI* gene (Bold) or the *cytB* gene (Fishtrace) or do not allow to upload images of specimens (GenBank). Therefore, an online platform was developed using an SQL database (Microsoft, Redmond, USA) to store all sequences and images and all metadata of the collected fish specimens of different European commercially important fish species. The database features a hierarchical storage system with digital images of morphologically identified seafood specimens, information on sampling protocols, lab protocols, storage location, next to DNA sequence data in fasta format for each vouchered specimen. The SEAFOOD<sup>TOMORROW</sup> database is hosted by smarterasp.net (BusinessICS Intl Limited, Monterey Park, CA, US) and is freely accessible after registration (<http://seafoodtomorrowdata.eu/>).

### 2.1.2. Collection and sequencing of voucher specimens

Fish specimens from different geographic regions (Baltic sea, Inland waters of North East Europe, Northeast Pacific, North Sea, Oceanic Northeast Atlantic waters) and from aquaculture were collected by researchers from different institutes (ILVO, ZUT and the University of Porto), vouchered and stored at -20°C. From every specimen, three fin clips and three muscle tissue samples were taken and stored in ethanol at -20°C. DNA was extracted from one of these tissues and also stored at -20°C. The *COI* gene, *cytB* gene and *Rhodopsin* gene were amplified using the different institutes' protocols. *Rhodopsin* was used only in cases where the two other markers did not suffice to discern species. Detailed information on the primer sets used by each institute is presented in supplementary Table S1. Details on the DNA extraction kit, Polymerase Chain Reaction (PCR) protocols and all available metadata for each specimen are listed in the SEAFOOD<sup>TOMORROW</sup> database.

### 2.1.3. Sequence data analysis

To ensure all sequences were assigned to the correct species name, phylogenetic trees for each marker gene were built in R (R Core Team, 2018) using the following packages: *seqRFLP* (Ding and Zhang, 2012), *msa* (Bodenhofer et al., 2015), *seqinr* (Charif and Lobry, 2007), *ape* (Paradis and Schliep, 2018), *Biostrings* (Pagès et al., 2018), *ggplot2* (Wickham, 2016) and *ggtree* (Yu et al., 2018). Sequences from the SEAFOOD<sup>TOMORROW</sup> database were converted to a fasta file after which they were aligned using ClustalW (Thompson et al., 1994). A matrix of pairwise distances from the aligned sequences was calculated based on sequence similarity. Inter- and intraspecific p-distances, calculated using the *adhoc* package (Sonet et al., 2013), were investigated to identify the presence of a barcoding gap. Thresholds for both *COI* and *cytB* genes were identified, below which samples were considered to belong to the same species. P-distance frequency graphs and error rates at different thresholds graphs were plotted using R (R Core Team, 2018) and *ggplot2*. Cumulative errors were calculated based on false positives and false negatives for a range of threshold values, to define the optimal threshold for



discriminating between two species (Wiemers and Fiedler, 2007). A false negative occurs when two sequences of the same species are identified as different species (threshold too low). A false positive occurs when two sequences of different species are identified as the same species (threshold too high).

## **2.2. Substitution of cod and sole in the Belgian food supply chain**

### **2.2.1. Scoping and sampling the Belgian fish supply chain**

The Belgian supply chain for fish was assessed through interviews with local stakeholders and scientists, and additional information from reports and literature (Blondeel et al., 2016; Fox et al., 2018; Verlé et al., 2016).

As identifying intact fish usually does not require molecular techniques, this study investigated only 'processed' fish products, i.e. where the whole specimen was no longer present or recognisable. A power analysis indicated that at least 30 samples per trader and per fish species were needed to ensure enough power to detect statistical differences in substitution rates between the different steps in the supply chain. For sole, this was problematic because most sole were traded directly from the auction to fishmongers and restaurants, or were sold only as whole specimens along the food supply chain.

All samples were anonymously purchased in Belgium. Import samples were analysed after being purchased by a company, which allowed samples to be traced back to the international exporter. The following metadata were recorded: collection date, name of the purchaser, species name and commercial name of the food product as mentioned by the trader, scientific name of the fish (if applicable), a photograph of the product (when possible), name and location of the trader, type of food product (fillet, meal or heavily processed), and brand (if applicable). After collection, the fish products were given a unique code and stored at -20°C.

### **2.2.2. Extraction, amplification and Sanger sequencing**

Samples from the substitution part of the study were not stored in ethanol. Total DNA was isolated from approximately 200 mg of food product using the NucleoSpin® Food kit (Macherey - Nagel GmbH

& Co. KG, Düren, Germany) following the manufacturer instructions. For amplification three different primer sets were used (supplementary Table S1). The first genetic marker was *COI*, for which primer set 1 was used. The *COI* marker gene was incapable of clearly distinguishing *Limanda aspera* from *Limanda limanda*, therefore the *cytB* marker gene was amplified with primer set 2. In case of degraded DNA, a shorter fragment (364 bp) of the *cytB* gene was amplified with primer set 3.

PCR reactions were performed in volumes of 40 µl containing 25 µl VWR Red Taq DNA Polymerase Master Mix (VWR International, Oud-Heverlee, Belgium). For each reaction 0.2 µM (4 µl) forward and reverse primers were added. The same volume of DNA was added to all reactions for the same genetic marker (2 µl for the *COI* gene and 2.5 µl for the *cytB* gene). PCR graded water was added until a total reaction volume of 40 µl was obtained. PCR cycling conditions for the *COI* fragment (primer set 1) consisted of 2 minutes at 94°C, followed by 35 cycles of 30 seconds at 94°C, 40 seconds at 52°C and 60 seconds at 72°C. Reactions were finished after a final extension step of 10 minutes at 72°C. For the *cytB* fragments (primer set 2 and 3), reactions were preheated for 5 minutes at 95°C, followed by 40 cycles of 30 s at 95°C, 30 s at 50°C and 60 s at 72°C. Reactions were finished after a final extension step for 10 minutes at 72°C. PCR products were loaded on a 1 % agarose gel, which was made of 100 ml 1 x TAE buffer (a buffer mixture of Tris base, acetic acid and EDTA), 1 g agarose and 10 µl GelRed® Nucleic Acid Gel Stain (Biotium Inc, Fremont, US). After successful gel electrophoresis, the PCR product was cleaned using Wizard® SV Gel and PCR Clean-Up System (Promega, Madison, USA), following the manufacturer's protocol. Samples were Sanger sequenced by Macrogen (Amsterdam, Netherlands) and GeneWiz (Bishop's Stortford, UK).

### **2.2.3. Sequence analysis**

For visual inspection of the sequencing chromatograms and creation of consensus sequences, Bionumerics v7.6.3 (Applied Maths, Sint-Martens-Latem, Belgium) was used. DNA sequences were then compared against the SEAFOOD<sup>TOMORROW</sup> database (SeafoodTomorrow, n.d.) with BLAST+ (Basic Local Alignment Search Tool) (Altschul et al., 1990) (NCBI, Bethesda, USA). From the list of BLAST hits,

the species with the highest query coverage, lowest E-value and highest identity were chosen. In case identity was below 99 %, samples were also blasted against the public library of GenBank (NCBI, Bethesda, USA) to ensure that this was not caused by absence of the species in the SEAFOOD<sup>TOMORROW</sup> database.

Samples were considered substituted when the detected species did not match the scientific name on the packaging or sales label. When the scientific name for a certain sample was not mentioned, the resulting scientific name was compared against the allowed commercial names for that species ([https://ec.europa.eu/fisheries/cfp/market/consumer-information/names\\_en](https://ec.europa.eu/fisheries/cfp/market/consumer-information/names_en)). The sampled specimen/species was considered substituted if the Dutch or French name did not match (“Gemeenschappelijke Marktordening | Departement Landbouw & Visserij,” n.d.).

#### **2.2.4. Quantitative PCR (qPCR) identification of Atlantic cod**

Alongside barcoding, a species specific assay with real-time PCR was used to identify Atlantic cod *G. morhua*. Briefly, the SureFood® Fish ID *Gadus morhua* IAAC (R&D version) kit (R-Biopharm, Darmstadt, Germany) was applied following the manufacturer’s protocol. The kit distinguishes two signals at different wavelengths. A signal in the VIC channel indicates the presence of fish (internal control), while a signal in the FAM channel indicates the presence of *G. morhua*. As such, when a sample contains *G. morhua* both channels should detect a signal. If it is a different species, only the VIC channel will indicate the presence of fish. A Light Cycler 480 I (Roche, Basel, Switzerland) was used to run the qPCR assay. When a product was suspected of substitution, the sample was sent for DNA sequencing, following the protocols as described in paragraph 2.2.2.

#### **2.2.5. Data visualisation**

Barplot graphs and maps were designed in R (R Core Team, 2018). The following packages were used: ggplot2 (Wickham, 2016), forcats (Wickham, 2019a), readxl (Wickham and Bryan, 2019), dplyr (Wickham et al., 2019), tibble (Müller and Wickham, 2019), stringr (Wickham, 2019b), ggspatial

(Dunnington, 2018), sf (Pebesma, 2018), rnatuarearth (South, 2017a), rnatuarearthdata (South, 2017b), pryr (Wickham, 2018) and ggpubr (Kassambara, 2019).

### 3. Results

#### 3.1. The SEAFOOD<sup>TOMORROW</sup> reference database

Currently, the SEAFOOD<sup>TOMORROW</sup> reference database holds information about 42 fish species of economic importance in Europe. In total 300 sequences were generated: 145 *COI*, 152 *cytB* and 3 *Rhodopsin* genes (Table 1).

Table 1 Overview of the collected species and generated sequences for the *COI*, *cytB* and *Rhodopsin (Rho)* genes that are currently (January 31, 2020) in the SEAFOOD<sup>TOMORROW</sup> database. Aqc: Aquaculture, BS: Baltic sea, IwNEE: Inland waters North East Europe, NEP: Northeast Pacific, NS: North Sea, ONA: Oceanic northeast Atlantic

Species	COI							Cytb							Rho	Total
	Aqc	BS	IwNEE	NEP	NS	ONA	Total COI	Aqc	BS	IwNEE	NEP	NS	ONA	Total Cytb	Aqc	
<i>Abramis brama</i>			3				3			3				3		6
<i>Chelidonichthys cuculus</i>												3		3		3
<i>Chelidonichthys lucerna</i>					1	1	2					1	1	2		4
<i>Chelon labrosus</i>			1			3	4			1			3	4		8
<i>Chelon ramada</i>			1			1	2			1			1	2		4
<i>Clupea harengus</i>		3			3		6		3			3		6		12
<i>Cyprinus carpio</i>			3				3			3				3		6
<i>Engraulis encrasicolus</i>						3	3						3	3		6
<i>Gadus macrocephalus</i>								1						1		1
<i>Gadus morhua</i>		3			2		5		3			3		6		11
<i>Limanda limanda</i>					3		3					4		4		7
<i>Melanogrammus aeglefinus</i>					1		1					1		1		2
<i>Merlangius merlangus</i>					3		3					3		3		6
<i>Merluccius merluccius</i>					1		1					1		1		2
<i>Micromesistius poutassou</i>						3	3						3	3		6
<i>Microstomus kitt</i>					1		1									1
<i>Oncorhynchus keta</i>				3			3				3			3		6
<i>Oncorhynchus kisutch</i>				3			3				3			3		6
<i>Oncorhynchus mykiss</i>	7						7	8						8		15
<i>Oncorhynchus nerka</i>				3			3				3			3		6
<i>Oncorhynchus tshawytscha</i>				3			3				3			3		6
<i>Oncorhynchus gorboscha</i>				3			3				3			3		6
<i>Pegusa lascaris</i>					1		1					1		1		2
<i>Platichthys flesus</i>		3			3	3	9		3			3	3	9		18
<i>Pleuronectes platessa</i>		3			4		7		3			4		7		14
<i>Pollachius pollachius</i>					1		1					1		1		2
<i>Pollachius virens</i>					1		1					1		1		2

<i>Salmo salar</i>	3					3	3						3	6		
<i>Salmo trutta</i>	3	3				6	3	3					6	12		
<i>Salvelinus alpinus</i>	1					1							1	2		
<i>Salvelinus fontinalis</i> x <i>Salvelinus alpinus</i>	6					6	6						6	3 15		
<i>Sardina pilchardus</i>			2	3		5					2	3	5	10		
<i>Sardinella aurita</i>				1	1							1	1	2		
<i>Scomber colias</i>				3	3							3	3	6		
<i>Scomber scombrus</i>	1		2	3		6		1			2	3	6	12		
<i>Solea senegalensis</i>				3	3							3	3	6		
<i>Solea solea</i>			3	2		5					5	2	7	12		
<i>Sparus aurata</i>				3	3							3	3	6		
<i>Sprattus sprattus</i>	4		3			7		4			3		7	14		
<i>Suaeda vera</i>														1		
<i>Trachurus trachurus</i>			4	3		7					4	3	7	14		
<i>Trigla lyra</i>				3	3							3	3	6		
<i>Trisopterus luscus</i>				4		4					4		4	8		
<b>Total</b>	<b>19</b>	<b>17</b>	<b>11</b>	<b>15</b>	<b>43</b>	<b>39</b>	<b>145</b>	<b>20</b>	<b>17</b>	<b>11</b>	<b>15</b>	<b>49</b>	<b>38</b>	<b>152</b>	<b>3</b>	<b>300</b>

Phylogenetic trees of the *cytB* and *COI* genes (supplementary material Figure S1 and S2 respectively) showed that conspecifics cluster together and that congeneric species were more closely related than non-congeneric species, illustrating that all sequences were linked to the correct species names. The frequency plots (Figure 1A and 1C) represent p-distances created by comparing all sequences present in the SEAFOOD<sup>TOMORROW</sup> database for *COI* and *cytB* genes respectively, and illustrate no overlap between intra- and interspecific distances. The optimal threshold is defined as the threshold at which species can be identified without any false negative or false positive result. The optimal threshold for the *COI* gene ranges from 2 to 4 %, the optimal threshold for the *cytB* gene ranges from 1.6 to 5.6 % (Figure 1B and 1D).

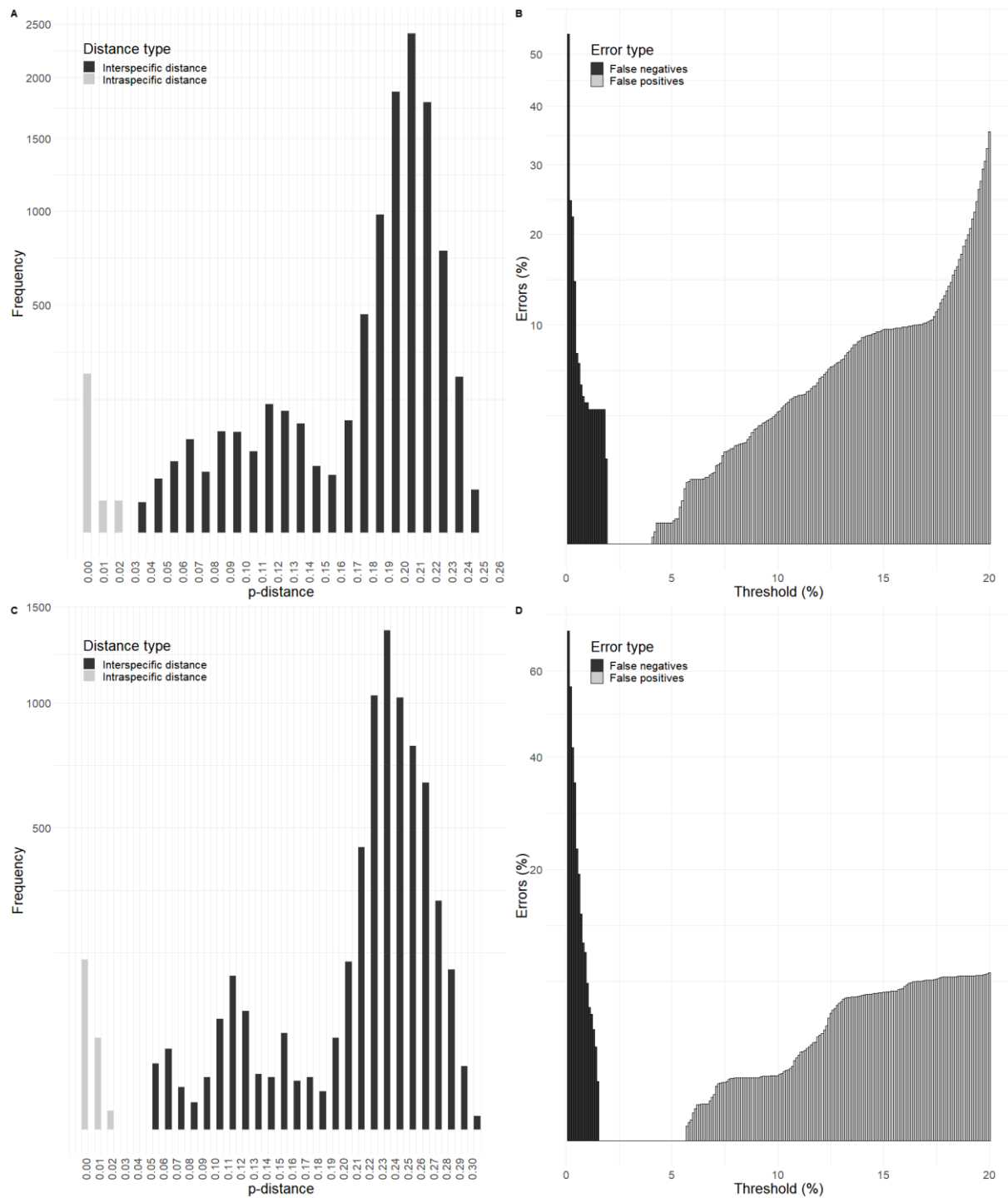


Figure 1 (A,C) Frequency distribution of intraspecific and interspecific genetic divergences in the COI and cytB genes for the 42 fish species in the Seafood<sup>Tomorrow</sup> database. (B,D) Cumulative error for the COI and cytB genes based on false positives and false negatives for thresholds ranging from 0.1 to 20 %. Optimal threshold value for COI between 2 and 4 %; for cytB between 1.6 and 5.6 %.

### 3.2. The Belgian fish supply chain

Eleven steps in the Belgian fish food supply chain were identified: (1) fishermen (the people who catch the fish); (2) the fish auction (where the catch is landed); (3) export (fish sold to countries outside of Belgium); (4) import (fish sold by countries outside of Belgium); (5) the fishermen's market (sells fish bought by fishermen directly to customers); (6) wholesale (bulk acquisition and selling of fish food products between traders); (7) processing, which includes both preparation (filleting, gutting and boning of a fish) and effective processing (the addition of ingredients, changes in temperature or breeding); (8) fishmongers (specialised in trading fish); (9) retailers (stores where a large variety of food products are sold, including fish); (10) catering (services where food tailored to needs is ordered individually, e.g. restaurants); and (11) food services (where food is prepared in bulk, e.g. canteens).

After fish are landed, the catch is transported to three potential locations: the fishermen's market, "De Vistrap", located in Oostende and Nieuwpoort, the auction hall and to exporters. The fish auction hall, in turn, sells fish to four potential customers: processors, wholesalers, fishmongers and companies outside Belgium (export). Wholesalers and processors also import fish from outside Belgium or export their products. Retailers and food services buy fish from wholesalers and processors. Fishmongers buy fish from the auction hall and from wholesalers. Catering buy fish from wholesalers, fishmongers and, in certain cases, directly from the fish auction. Additionally, wholesalers and processors sell products back and forth between one another and, often, companies will engage in all of these activities (preparation, processing and wholesale). The network is presented in Figure 2.

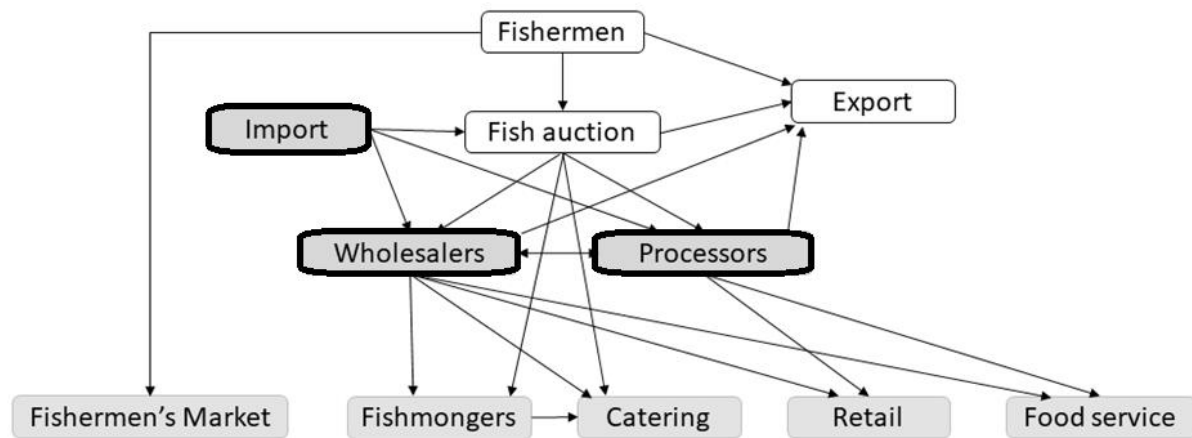


Figure 2 Network of 11 traders in the Belgian seafood supply chain and interactions between them. In grey without a black border are the end points of the supply chain that were sampled; in grey with a black border are points higher up in the supply chain that were sampled in this study. White traders were not sampled because they only provide whole cod and sole or the fish products are exported outside Belgium, which falls outside the scope of this study.

### 3.3. Substitution of cod and sole in the Belgian food supply chain

As this study focused on 'processed' fish products, we did not assess fishermen nor the fish auction. Also, 'Export' was not accessed as this study focused on the 'real' Belgian food supply chain, leaving eight steps of the Belgian food supply chain that were sampled for 'processed' cod and sole (Figure 2).

A total of 139 'processed' cod products were sampled: 2 from the fishermen's market, 25 from import, 1 from wholesalers, 18 from processors, 33 from fishmongers, 16 from catering, 35 from retail, and 9 from food service. A total of 46 'processed' sole products could be sampled: 3 from wholesalers, 1 from processors, 26 from catering, 14 from retail and 2 from food service. Samples for both cod and sole were collected mainly in the western (Flemish) part of Belgium (Figure 3A). Imported cod samples originated mainly from the United Kingdom, Iceland and Denmark, next to different East European countries and Russia (Figure 3B).



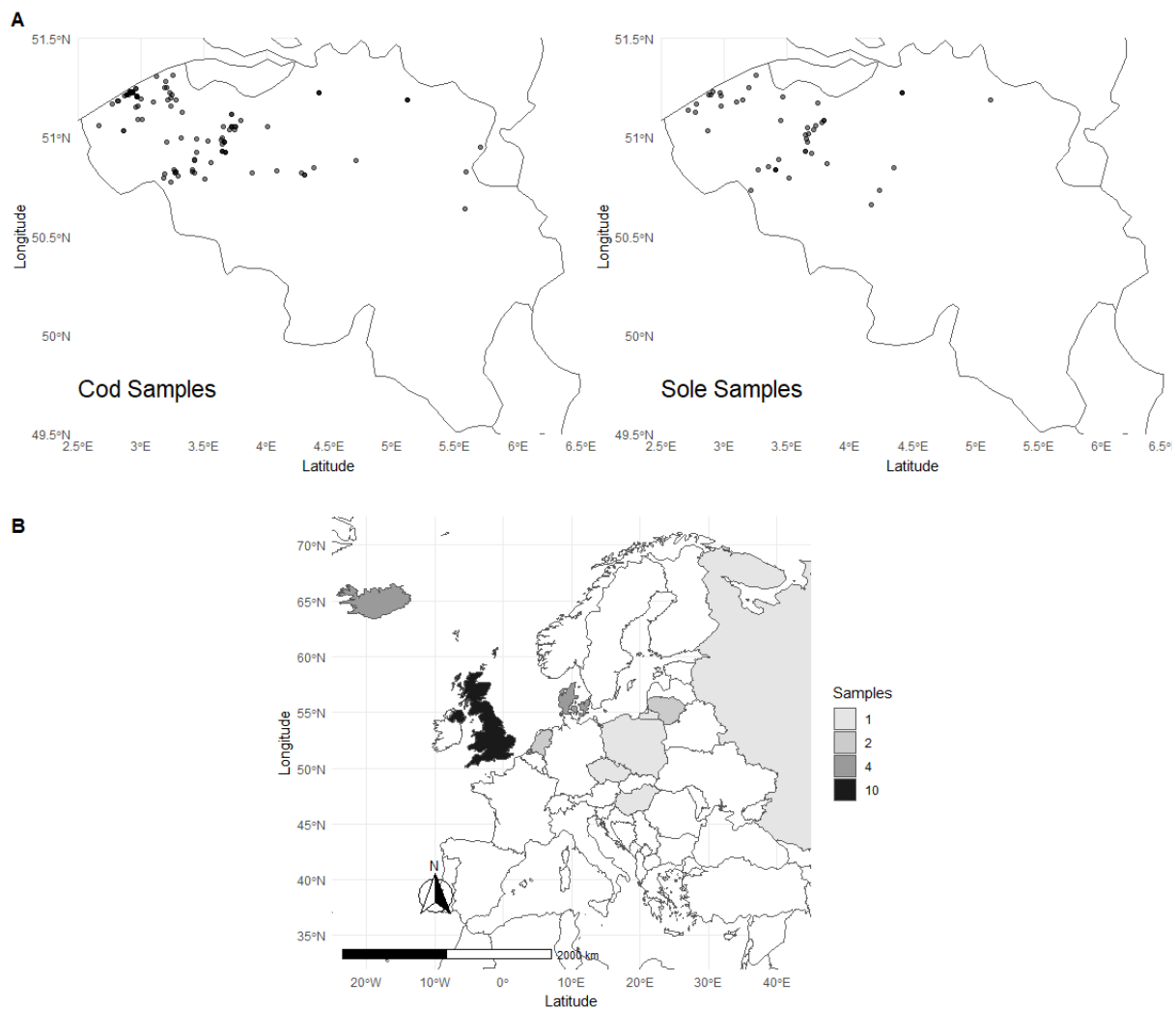


Figure 3 Maps representing the samples taken in this study. (A) Locations of samples collected for each species, dots are grey and get darker as they overlap (cod left, sole right). (B) All countries from which samples were imported. The shades of grey reflect the number of samples imported from each country.

Of the 185 samples collected, 182 were identified successfully. One cod sample from a fishmonger could not be amplified with either qPCR or normal PCR using any of the primer sets, as the DNA was found to be of poor quality. Two cod samples that could not be identified by qPCR were properly identified by DNA barcoding. One of the two cod samples that could not be identified with barcoding, was successfully identified using qPCR. One sole sample from a wholesaler could be amplified with PCR for both the *COI* and *cytB* gene, but sanger sequencing chromatograms were highly contaminated. A second sole sample was amplified with the *cytB* gene and sanger sequenced but could not be identified reliably, as the best match on GenBank was only 91 % with the species *Cynoglossus joyneri*. This clearly

indicated that this sample was not sole, and as such was retained in the further analyses. As a result, a total of 138 cod samples and 45 sole samples were used for further analysis. Table 2 shows the ratios of successful identifications for both DNA barcoding and qPCR for different food types. Some samples were barcoded with both *COI* and *cytB*, and when cod samples were suspected of being fraud, they were also barcoded to confirm the substitution. Additionally, qPCR was not used at the start of the study and not every cod was analysed with qPCR. No perceivable difference in the success rates of identification was found between the techniques (for cod samples only) or the grade of processing (both cod and sole samples, Table 2).

Table 2 Ratio of successful identifications for DNA barcoding and qPCR (the latter only for cod samples) for the different 'processed' cod and sole food products to the total number of samples analysed.

	FILET	PROCESSED	MEAL	TOTAL
QPCR	73/75	20/20	21/22	114/117
BARCODING	25/26	10/10	47/48	82/84

In total, 138 cod and 45 sole food products were identified molecularly. Certain food products from the same retailer were collected multiple times, sometimes in a different city. The outcome was identical for the same products of the same retailer but from different stores. To avoid inflation of the substitution rate, only one sample of these food products was retained. As a result, substitution rates were based on 132 cod samples and 41 sole samples.

For cod, one out of 16 catering samples (6 %) was a meal substituted with *Pollachius virens*; one of 21 import samples (5 %) was a fillet substituted with *Melanogrammus aeglefinnus*; one of 32 fishmonger samples (3 %) was a processed fish substituted with *Gadus chalcogrammus*. Substitution occurred in all three food types (Figure 4A and C, Figure 5A and C).

Substitution rates for sole were as follows: 2 out of 2 for wholesale (100 %), 1 out of 2 for food service (50 %), 0 out of 1 for processors, 2 out of 10 for retail (20 %) and 2 out of 26 for catering (8 %) (Figure

4B). Substitution of sole was found in two ‘processed’ food products: 2 out of 10 fillets (20 %) and 5 out of 31 meals (16 %) (Figure 4D). For the meal sample from a food service, *S. solea* was replaced with *L. aspera* (Figure 5B and Figure 5D). Two substituted samples from wholesale (both fillets) were *Cynoglossus sp.* and *S. senegalensis*, the latter sold as ‘farmed sole’. In five meal samples from retail sole was substituted with *L. aspera* (4 samples) and *Pangasianodon hypophthalmus* (1 sample). Additionally, one of the retail samples sold as sole returned both *L. aspera* and *Lepidopsetta polyxtra*. Also, the substituted meal sample from a food service contained *L. aspera*. Two sole catering samples were substituted with *L. aspera* and *Microstomus kitt*. Two retail samples were sold as “Tongrolletjes” a Belgian dish normally prepared with *S. solea*, but the label stated it to be lemon sole (*M. kitt*). These samples were also classified as substituted, as misleading the consumer is equally considered as fraud.

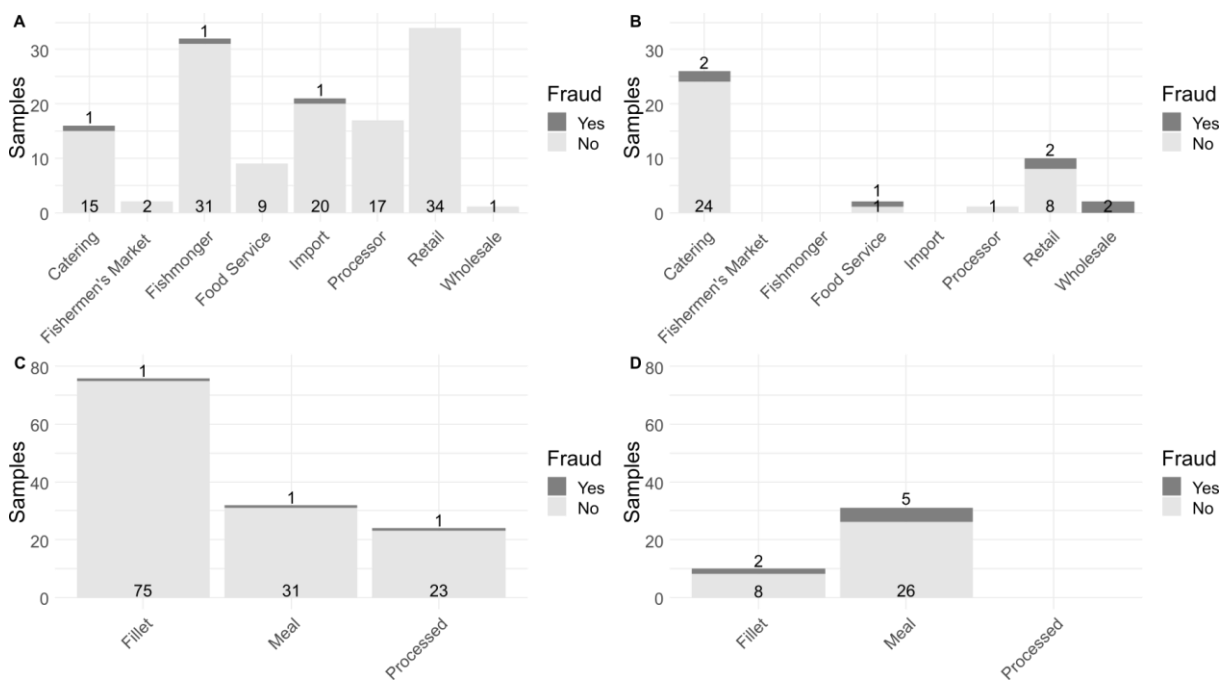


Figure 4 Substitution rates for ‘processed’ cod and sole in the Belgian food supply chain. Total number of unique samples gathered per seller and food type for cod (A, C) and for sole (B, D). In dark grey are the number of substituted samples and in light grey are the correctly labelled samples.

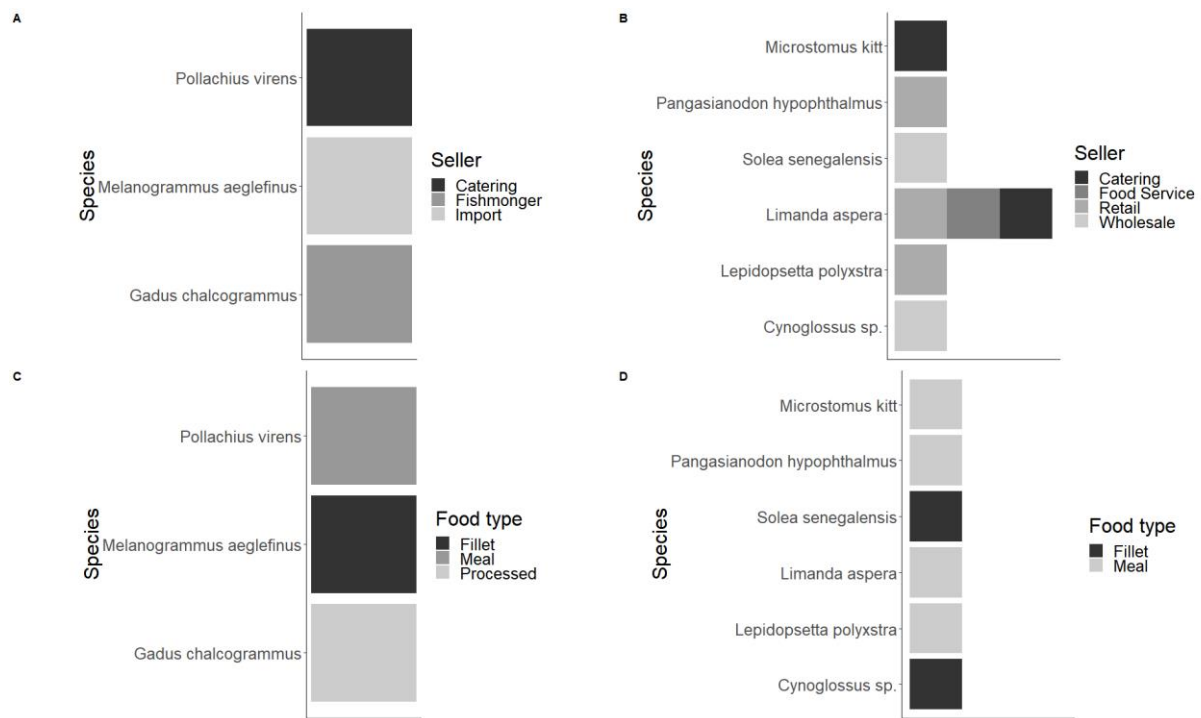


Figure 5 Substitution species, per seller and food type for cod (A, C) and for sole (B, D).

## 4. Discussion

### 4.1. The SEAFOOD<sup>TOMORROW</sup> reference database

Through DNA barcoding, the amplified marker gene sequence is compared with known sequences linked to species names registered in reference databases (Hebert et al., 2003). The importance of a reliable reference database for correct species identification cannot be overstated (Li et al., 2018; Mioduchowska et al., 2018). Therefore, we initiated the creation of a reference database for economically important European fish species and their prominent substitutes. All specimens were morphologically identified by experts and in-depth quality control of the obtained sequences was performed. For every collected specimen a picture is uploaded and the metadata can be traced back to the specimen itself, but also to the extracted tissues, the extracted DNA from the tissue, the laboratory protocols and the DNA sequences for both *COI* and *cytB* genes. The SEAFOOD<sup>TOMORROW</sup> database is different from the *COI* targeted BoLD database, as it focuses on commercially important European species and also includes *cytB* sequences, which allows better species resolution for a

number of closely related fishes than the *COI* gene (Kochzius et al., 2010). Compared to Fishtrace, the current database contains more information, as both *cytB* and *COI* sequences for commercially important European fish species are registered, and photographic material is included, which allows to investigate morphological identifications at any point.

Frequency distribution plots of intra- and interspecific distances showed no overlap for both *COI* and *cytB* genes. Based on this barcoding gap, correct identification of new samples is expected when sequence similarity is above 96% similarity for *COI* and above 95% for *cytB*. The barcoding gap is expected to narrow down when more species are added (Hubert et al., 2008; Kochzius et al., 2010; Pereira et al., 2013; Ward, 2009). Moreover, pitfalls related to species identification using the barcoding gap are well known (Meier et al., 2008), including the incompleteness of the database (Virgilio et al., 2012). One example that showcases this is the occurrence of *L. aspera* as a substitute in the current study, which is not a European fish, and as such was not included in the SEAFOOD<sup>TOMORROW</sup> database. However, this species is genetically highly similar to *L. limanda* (BLASTs reaching >98 % identity for *COI* and >96 % identity for *cytB*), as such, even though the identity appears to fall in the previously described thresholds, sequences below 99 % identity were additionally BLASTed against GenBank, here revealing that it was more likely to be *L. aspera*. Since similarity of the *cytB* sequences was very low when compared to the SEAFOOD<sup>TOMORROW</sup> reference database, the sequences of this species were compared against the public databases, and yielded similarities of 100% with *L. aspera* sequences from Genbank. The effect of the incompleteness of the reference database for species level identification is further illustrated for the sole sample that appeared most similar to species from the *Cynoglossus* genus. Not all species of the *Cynoglossus* genus are currently present in public databases so species level identification for this specimen was not possible regardless of the reference database used. This indicated that reliance solely on a threshold does not guarantee that the identification is correct and multiple genes should be checked in order to increase the reliability of the identification.

#### **4.2. The Belgian fish supply chain**

The Belgian fisheries supply chain is a complex, but important part of the Belgian economy. In 2019, a total of 13 754 tonnes (59.1 million euros) of fishery products were landed across all Belgian harbours, a total of 125 380 tonnes (740.2 million euros) of fishery products were imported, and 17 684 tonnes (51.8 million euros) of fishery products were exported (*EUMOFA*, 2019; Flanders department of Agriculture and Fisheries).

Atlantic cod has two entry points in the Belgian supply chain: Belgian fishermen and import. In 2019, 355 tonnes (1.1 million euros) of cod were landed across all Belgian harbours, 3 880 tonnes (43.5 million euros) of cod were imported, and 21 tonnes (136 thousand euros) of cod were exported (*EUMOFA*, 2019; Flanders department of Agriculture and Fisheries). Cod is Belgium's most popular fish consumed at home, being purchased by 45% of all Belgian families ("Vis - VLAM," n.d.). Cod caught by Belgian fishermen is mostly sold by the fish auction to local restaurants and fishmongers. Retailers rarely buy cod from Belgian fishermen, since its supply is unstable and cannot be guaranteed. Moreover, retailers often require an MSC label, which is not the case for Belgian caught cod, related to the detrimental nature of the beam trawl fishing technique (Depestele et al., 2016; "Sustainable Fishing | MSC | Marine Stewardship Council," n.d.). Wholesalers and processors thus import cod from other European countries, mostly from the United Kingdom, and mainly as filets. Filets purchased by wholesalers and processors are then sold to fishmongers, retailers, catering and food services. Processors can turn the cod into processed samples (often battered cod fillets) or premade meals (e.g. bouillabaisse), which are then sold to retailers and food services. In most scenarios, cod is already unrecognisable when it enters the Belgian supply chain.

Like cod, common sole also has two entry points in the Belgian supply chain: Belgian fishermen and import. In 2019, 2045 tonnes (25.9 million euros) of common sole were landed across all Belgian harbours, 2938 tonnes (16.7 million euros) of common sole were imported, and 2 tonnes (43 thousand euros) of common sole were exported (*EUMOFA*, 2019; Flanders department of Agriculture and Fisheries). In contrast to cod, sole comes primarily from the Belgian fishermen and fish auction. Sole is

rarely processed and mostly sold as complete fish to wholesalers, fishmongers and catering. Also, retailers mostly sell complete fish, traded from wholesalers. The few companies that do process sole, buy it from the fish auction and convert it to fillets or fish meals, which are sold to retailers and food services. The difference in supply chains for both cod and sole created the ideal scenario to thoroughly evaluate substitution in the entire fish food supply chain in Belgium.

#### **4.3. Substitution of cod and sole along the Belgian food supply chain**

Atlantic cod *G. morhua* was only substituted in three out of 132 (2%) samples along the Belgian supply chain. This is lower than the mislabelling rates found in the meta-analysis by Luque and Donlan (2019). No substitutions were found in retail, confirming a study across North Atlantic countries, which also did not find mislabelling of cod retailed in Belgium or other European countries (Bréchon et al., 2016). In contrast, a study in France found 11 % mislabelling of cod products collected from fishmongers, supermarkets and restaurants, while another study found 55 % mislabelling for retailers in Italy (Bénard-Capelle et al., 2015; Pinto et al., 2013). The 4 % substitution rate recorded for cod in food services and restaurants is lower than the 13 % described for food services and restaurants in Brussels, the capital of Belgium (Christiansen et al., 2018). In contrast, the current samples were gathered closer to the Belgian coast, where the consumer may have an increased knowledge about fish and seafood products, as such deterring substitution fraud.

All three substituted species for cod belonged to the Gadidae and are very similar to *G. morhua* in both taste and texture. *Polachius virens* and *Melanogrammus aeglefinus* are frequently landed at the Belgian fish auction ("Vis - VLAM," n.d.), while *Gadus chalcogrammus* is mainly imported as filleted or processed product ("Vis- & zeevruchtengids - voor professionele gebruikers | Zeevruchten gids," n.d.). Similar substitutions were found in other studies (Christiansen et al., 2018; Herrero et al., 2010; Luque and Donlan, 2019; Tomás et al., 2017). The processor that imported the *M. aeglefinus* fillet that was sold as *G. morhua* claimed to be unaware of this substitution fraud. Although the substitution rate for cod was low, the results show that substitution does not only occur at the end of the food supply chain.

The substitution of sole in the Belgian supply chain (17 %) was much higher than the substitution rate recorded for cod, and comparable to the 20 % average for sole noted in the meta-analysis of Luque and Donlan (2019). The substitution rate of 8 % found in catering is comparable to the 11 % described by Christiansen et al. (2018) for Belgium, but substantially lower than the 50 % detected in Germany (Kappel, 2016). The substitution in Belgian retail (20 % of the samples) is lower than the 30 % recorded in Spanish shops and supermarkets (Herrero et al., 2012) or the 50 % recorded for French supermarkets, although only two samples were identified in the latter study (Bénard-Capelle et al., 2015). Belgian retailers seem not to be responsible for the substitution fraud themselves, as meals and filleted soles sold in supermarkets are usually prepared and packaged at other steps in the Belgian food supply chain.

Other studies found that sole was often substituted by cheaper, morphologically similar species from the Soleidae family (Luque and Donlan, 2019), such as *Cynoglossus senegalensis*, *Solea senegalensis*, *Synaptura lusitanica* and *Pegusa lascaris*, but also with less similar species such as *Pangasius sp.*, *Lepidopsetta polyxstra* and *Limanda aspera* (Bénard-Capelle et al., 2015; Christiansen et al., 2018; Herrero et al., 2012). We mainly found morphologically similar substitutes, being *Cynoglossus sp.* and *S. senegalensis.*, when looking at ‘pure’ sole fillets. However, several other species (*L. aspera*, *P. hypophthalmus*, *L. polyxstra* and *M. kitt*) were used as substitutes when fillets were further processed, i.e. rolled up and covered in sauce. Also in Germany, *L. aspera* and *Cynoglossus* were common substitution alternatives for sole (Kappel, 2016; Rehbein, 2008). One of the Belgian processors confirmed that *L. aspera* is an increasingly popular alternative to sole. Sole is priced between 20 and 40 euros per kg, while *L. aspera* costs around 6 euros per kg, making it prone to economic fraud (“ISPC,” n.d., “VisOnline,” n.d.).

Other studies stated that substitution rates in restaurants may be higher due to less stringent controls (Christiansen et al., 2018; Fox et al., 2018; Hanner et al., 2011; Kappel, 2016; Khaksar et al., 2015; Shehata et al., 2017; Vandamme et al., 2016). This was not the case in our study, as substitution was



detected at almost every step in the Belgian food supply chain. The course of action against species substitution varies depending on the point of the supply chain it takes place in (Fox et al., 2018) and emphasizes the importance of extensive knowledge of substitution rates in different parts of the supply chain. The current study followed to a large extent the recommendations of the European Commission that were invoked for a coordinated control plan for fish species substitution across its member states (SANCO/12569/2014) in that major species of the Belgian consumer market were targeted in a variety of products (meals, fillets and other kinds of processed products) in different parts of the supply chain. However, our results indicate that the recommendation of collecting a total of 100 samples across different fishes and different steps in the supply chain may be insufficient to achieve adequate substitution rates, since substitution in cod was low but nevertheless present. We encourage future studies to collect a minimum and balanced number of samples in each step of the supply chain and for each species when the aim is to identify steps in the supply chain that are more prone to substitution practices. Such studies will greatly contribute in creating effective control and mitigation measures throughout the supply chain (Fox et al., 2018).

## 5. Conclusions

The SEAFOOD<sup>TOMORROW</sup> DNA barcoding reference database is void of misidentifications and provides a reliable basis for the identification of processed samples for commercially important European fish species, based on both *COI* and *cytB* gene markers. In addition, when the query sequence retrieves low similarity scores with any of the sequences in the database, this will point to a non-European fish species as substitute. To further identify such sequences, a comparison with public reference databases is needed. Sequences for commercially important export and import species may be added to the SEAFOOD<sup>TOMORROW</sup> database, provided they are vouchered and morphologically and genetically identified by experts.

The Belgian food supply chain appears as a complex web of potential interactions between traders. Still, most fish products have a more or less distinguished pathway, with cod mainly brought into the

supply chain via import from other European countries, and sole mainly originating from local Belgian fisheries and the fish auction.

Substitution fraud does occur for sole (17 %) and to a lesser extent for cod (2 %), two popular fish species in Belgium, at different steps along the Belgian food supply chain. Considering the huge economic value of these species, these substitution rates also impact the Belgian economy. More stringent control measures are needed to ensure more transparency for consumers, such that they can trust their purchases and the labelled information.

## **6. Acknowledgments**

We thank the scientists at ILVO who helped with sampling, Tom Putteman for help in the lab, and Bart Ampe for help with the initial power calculations and design for the substitution experiment. This research is part of the EU Horizon2020 project SEAFOOD<sup>TOMORROW</sup>. This project has received funding from the European Union's Horizon 2020 funding programme, Grant Agreement no. 773400 (SEAFOOD<sup>TOMORROW</sup>). This output reflects the views of the author(s), and the European Commission cannot be held responsible for any use that might be made of the information contained therein. M. A. Faria thanks "Fundação para a Ciência e Tecnologia" the researcher contract and financial support from the project UID/QUI/50006/2019.

## 7. References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool.

J. Mol. Biol. 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)

Ardura, A., Zaiko, A., Borrell, Y.J., Samuiloviene, A., Garcia-vazquez, E., Devloo-delva, F., Miralles, L.,

Ardura, A., Borrell, Y.J., Pejovic, I., Tsartsianidou, V., Garcia-vazquez, E., 2016. Novel tools for early detection of a global aquatic invasive , the zebra mussel *Dreissena polymorpha*. MPB 105, 23–29. <https://doi.org/10.1002/aqc.2655>

Bénard-Capelle, J., Guillonneau, 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>

Blondeel, L., Verlé, K., Vlaemyneck, G., Braeckman, R., Vermeersch, X., Vanderperren, E., Torreele, E.,

2016. VERWERKVIS - De Vlaamse visserijsector en de visverwerkende nijverheid status, noden en kansen. ILVO Meded. 209.

Bodenhofer, U., Bonatesta, E., Horejs-Kainrath, C., Hochreiter, S., 2015. msa: an R package for multiple sequence alignment. Bioinformatics 31, 3997–3999.

<https://doi.org/10.1093/bioinformatics/btv494>

Bold Systems v4 [WWW Document], n.d. URL <http://www.boldsystems.org/> (accessed 12.6.19).

Bréchon, A.L., Hanner, R., Mariani, S., 2016. A systematic analysis across North Atlantic countries unveils subtleties in cod product labelling. Mar. Policy 69, 124–133.

<https://doi.org/10.1016/j.marpol.2016.04.014>

Bryndum-Buchholz, A., Tittensor, D.P., Blanchard, J.L., Cheung, W.W.L., Coll, M., Galbraith, E.D.,

Jennings, S., Maury, O., Lotze, H.K., 2019. Twenty-first-century climate change impacts on marine animal biomass and ecosystem structure across ocean basins. Glob. Chang. Biol. 25, 459–472. <https://doi.org/10.1111/gcb.14512>

Charif, D., Lobry, J.R., 2007. Seqin{R} 1.0-2: a contributed package to the {R} project for statistical computing devoted to biological sequences retrieval and analysis., in: Bastolla, U., Porto, M., Roman, H.E., Vendruscolo, M. (Eds.), *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations, Biological and Medical Physics, Biomedical Engineering*. Springer Verlag, New York, pp. 207–232.

Cheung, W.W.L., Watson, R., Pauly, D., 2013. Signature of ocean warming in global fisheries catch. *Nature* 497, 365–368. <https://doi.org/10.1038/nature12156>

Christiansen, H., Fournier, N., Hellemans, B., Volckaert, F.A.M., 2018. Seafood substitution and mislabeling in Brussels' restaurants and canteens. *Food Control* 85, 66–75.  
<https://doi.org/10.1016/j.foodcont.2017.09.005>

Cook, R.M., 2018. Stock collapse or stock recovery? Contrasting perceptions of a depleted cod stock. *ICES J. Mar. Sci.* <https://doi.org/10.1093/icesjms/fsy190>

Depestele, J., Ivanović, A., Degrendele, K., Esmaeili, M., Polet, H., Roche, M., Summerbell, K., Teal, L.R., Vanellander, B., O'Neill, F.G., 2016. Measuring and assessing the physical impact of beam trawling. *ICES J. Mar. Sci.* 73, i15–i26. <https://doi.org/10.1093/icesjms/fsv056>

Devloo-Delva, F., Miralles, L., Ardura, A., Borrell, Y.J., Pejovic, I., Tsartsianidou, V., Garcia-vazquez, E., 2016. Detection and characterisation of the biopollutant *Xenostrobus securis* ( Lamarck 1819 ) Asturian population from DNA Barcoding and eBarcoding. *MPB* 105, 23–29.  
<https://doi.org/10.1016/j.marpolbul.2016.03.008>

Ding, Q., Zhang, J., 2012. seqRFLP: Simulation and visualization of restriction enzyme cutting pattern from DNA sequences.

Dunnington, D., 2018. ggspatial: Spatial Data Framework for ggplot2.

FAO, 2018. World Fisheries and Aquaculture Sofia Report.

FishTrace [WWW Document], n.d. URL <https://fishtrace.jrc.ec.europa.eu/> (accessed 12.6.19).

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 3, 294–299. <https://doi.org/10.1371/journal.pone.0013102>

Fox, M., Mitchell, M., Dean, M., Elliott, C., Campbell, K., 2018. The seafood supply chain from a fraudulent perspective. *Food Secur.* 10, 939–963. <https://doi.org/10.1007/s12571-018-0826-z>

Gemeenschappelijke Marktordening | Departement Landbouw & Visserij [WWW Document], n.d. URL <https://lv.vlaanderen.be/nl/visserij/visserijbeleid/gemeenschappelijk-visserijbeleid-gvb/illegale-visserij/gemeenschappelijke> (accessed 6.5.19).

Gordoa, A., Carreras, G., Sanz, N., Viñas, J., 2017. Tuna species substitution in the Spanish commercial chain: A knock-on effect. *PLoS One* 12, 1–15. <https://doi.org/10.1371/journal.pone.0170809>

Hanner, R., Becker, S., Ivanova, N. V., Steinke, D., 2011. FISH-BOL and seafood identification: Geographically dispersed case studies reveal systemic market substitution across Canada. *Mitochondrial DNA* 22, 106–122. <https://doi.org/10.3109/19401736.2011.588217>

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* 270, 313–321. <https://doi.org/10.1098/rspb.2002.2218>

Herrero, B., Lago, F.C., Vieites, J.M., Espiñeira, M., 2012. Real-time PCR method applied to seafood products for authentication of European sole (*Solea solea*) and differentiation of common substitute species. *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* 29, 12–18. <https://doi.org/10.1080/19440049.2011.623682>

Herrero, B., Madriñán, M., Vieites, J.M., Espiñeira, M., 2010. Authentication of atlantic cod (*Gadus morhua*) Using real time PCR. *J. Agric. Food Chem.* 58, 4794–4799. <https://doi.org/10.1021/jf904018h>

- Hobbs, C.A.D., Potts, R.W.A., Walsh, M.B., Usher, J., Griffiths, A.M., 2019. Using DNA barcoding to investigate patterns of species utilisation in UK shark products reveals threatened species on sale 1–10. <https://doi.org/10.1038/s41598-018-38270-3>
- Hubert, N., Hanner, R., Holm, E., Mandrak, N.E., Taylor, E., Burrige, M., Watkinson, D., Dumont, P., Curry, A., Bentzen, P., Zhang, J., April, J., Bernatchez, L., 2008. Identifying Canadian freshwater fishes through DNA barcodes. *PLoS One* 3. <https://doi.org/10.1371/journal.pone.0002490>
- ISPC [WWW Document], n.d. URL <https://www.ispc.be/nl> (accessed 12.6.19).
- Kan, K., Ushiyama, H., Shindo, T., Uehara, S., Yasuda, K., 1999. Outbreak of histamine poisoning due to ingestion of fish, “Abura-sokomutsu” (*Lepidocybium flavobrunneum*). *J. Food Hyg. Soc. Japan (Shokuhin Eiseigaku Zasshi)* 41, 3–24.
- Kappel, K., 2016. Substitution of high-priced fish with low-priced species : Adulteration of common sole in German restaurants 59, 478–486. <https://doi.org/10.1016/j.foodcont.2015.06.024>
- Kassambara, A., 2019. ggpubr: “ggplot2” Based Publication Ready Plots.
- Khaksar, R., Carlson, T., Schaffner, D.W., Ghorashi, M., Best, D., Jandhyala, S., Traverso, J., Amini, S., 2015. Unmasking seafood mislabeling in U.S. markets: DNA barcoding as a unique technology for food authentication and quality control. *Food Control* 56, 71–76. <https://doi.org/10.1016/j.foodcont.2015.03.007>
- Kochzius, M., Seidel, C., Antoniou, A., Botla, S.K., Campo, D., Cariani, A., Vazquez, E.G., Hauschild, J., Herve, C., Hjörleifsdottir, S., Hreggvidsson, G., Kappel, K., Landi, M., Magoulas, A., Marteinson, V., Nölte, M., Planes, S., Tinti, F., Turan, C., Venugopal, M.N., Weber, H., Blohm, D., 2010. Identifying fishes through DNA barcodes and microarrays. *PLoS One* 5, 1–15. <https://doi.org/10.1371/journal.pone.0012620>
- Li, X., Shen, X., Chen, X., Xiang, D., Murphy, R.W., Shen, Y., 2018. Detection of potential problematic Cytb gene sequences of fishes in GenBank. *Front. Genet.* 9, 1–5.

<https://doi.org/https://doi.org/10.3389/fgene.2018.00030>

Luque, G.M., Donlan, C.J., 2019. The characterization of seafood mislabeling: A global meta-analysis.

Biol. Conserv. 1–15. <https://doi.org/10.1016/j.biocon.2019.04.006>

Manning, L., Soon, J.M., 2014. Developing systems to control food adulteration. Food Policy 49, 23–

32. <https://doi.org/10.1016/j.foodpol.2014.06.005>

Meier, R., Zhang, G., Ali, F., 2008. The use of mean instead of smallest interspecific distances

exaggerates the size of the “barcoding gap” and leads to misidentification. Syst. Biol. 57, 809–

813. <https://doi.org/10.1080/10635150802406343>

Meikle, J., McDonalds, H., 2013. Cameron tells supermarkets: horsemeat burger scandal

unacceptable | World news | The Guardian [WWW Document]. URL

<https://www.theguardian.com/world/2013/jan/16/tesco-burgers-off-shelves-horsemeat>

(accessed 5.24.19).

Mioduchowska, M., Czyz, M.J., Gołdyn, B., Kur, J., Sell, J., 2018. Instances of erroneous DNA

barcoding of metazoan invertebrates: Are universal cox1 gene primers too “universal”? PLoS

One 13, 1–16. <https://doi.org/10.1371/journal.pone.0199609>

Mottola, A., Marchetti, P., Bottaro, M., Pinto, A.D.I., Cunha, H.A., Da Silva, V.M.F., Santos, T.E.C.,

Moreira, S.M., Do Carmo, N.A.S., Solé-Cava, A.M., DEMESURE, B., SODZI, N., PETIT, R.J.,

Shehata, H.R., Bourque, D., Steinke, D., Chen, S., Hanner, R., Keskin, E., Atar, H.H., Trivedi, S.,

Aloufi, A.A., Ansari, A.A., Ghosh, S.K., Saitou, N., Nei, M., Pereira, L.H.G., Hanner, R., Foresti, F.,

Oliveira, C., DEMESURE, B., SODZI, N., PETIT, R.J., Moore, J.C., Spink, J., Lipp, M., Uddin, M.,

Okazaki, E., Helyar, S.J., Lloyd, H.A.D., De Bruyn, M., Leake, J., Bennett, N., Carvalho, G.R.,

Horreo, J.L., Ardura, A., Pola, I.G., Martinez, J.L., Garcia-Vazquez, E., Paracchini, V., Petrillo, M.,

Lievens, A., Puertas Gallardo, A., Martinsohn, J.T., Hofherr, J., Maquet, A., Silva, A.P.B., Kagkli,

D.M., Querci, M., Patak, A., Angers-Loustau, A., 2013. A set of universal primers for

amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *J. Food Sci.* 4, 1–7. <https://doi.org/10.1111/j.1365-294X.1995.tb00201.x>

Müller, K., Wickham, H., 2019. *tibble: Simple Data Frames*.

Noguchi, T., Arakawa, O., 2008. Tetrodotoxin - Distribution and accumulation in aquatic organisms, and cases of human intoxication. *Mar. Drugs* 6, 220–242. <https://doi.org/10.3390/md20080011>

Pagès, H., Aboyoun, P., Gentleman, R., DebRoy, S., 2018. *Biostrings: Efficient manipulation of biological strings*.

Paradis, E., Schliep, K., 2018. *ape 5.0: an environment for modern phylogenetics and evolutionary analyses in {R}*. *Bioinformatics* 35, 526–528.

Pebesma, E., 2018. *Simple Features for R: Standardized Support for Spatial Vector Data*. R J.

Pereira, L.H.G., Hanner, R., Foresti, F., Oliveira, C., 2013. Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? *BMC Genet.* 14–20. <https://doi.org/doi:10.1186/1471-2156-14-20>

Piñeiro, C., Vázquez, J., Marina, A.I., Barros-Velázquez, J., Gallardo, J.M., 2001. Characterization and partial sequencing of species-specific sarcoplasmic polypeptides from commercial hake species by mass spectrometry following two-dimensional electrophoresis. *Electrophoresis* 22, 1545–1552. [https://doi.org/10.1002/1522-2683\(200105\)22:8<1545::AID-ELPS1545>3.0.CO;2-5](https://doi.org/10.1002/1522-2683(200105)22:8<1545::AID-ELPS1545>3.0.CO;2-5)

Pinto, A. Di, Pinto, P. Di, 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 Chem.* 141, 1757–1762. <https://doi.org/10.1016/j.foodchem.2013.05.093>

R Core Team, 2018. *R: A Language and Environment for Statistical Computing*.

Ratnasingham, S., Hebert, P.D.N., 2007. The Barcode of Life Data System. *Mol. Ecol. Notes* 7, 355–364. <https://doi.org/10.1111/j.1471-8286.2006.01678.x>



- Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R.M., Gough, K.C., 2014. The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. *J. Appl. Ecol.* 51, 1450–1459. <https://doi.org/10.1111/1365-2664.12306>
- Rehbein, H., 2008. New fish on the German market: Consumer protection against fraud by identification of species. *J. fur Verbraucherschutz und Leb.* 3, 49–53. <https://doi.org/10.1007/s00003-007-0301-9>
- Schmidt, C.C., 2005. Economic drivers of illegal, unreported and unregulated (IUU) fishing. *Int. J. Mar. Coast. Law* 20, 479–507. <https://doi.org/10.1163/157180805775098630>
- SeafoodTomorrow, n.d. Home Page - SEAFOODTOMORROW [WWW Document]. URL <http://seafoodtomorrowdata.eu/authentication> (accessed 5.24.19).
- Sevilla, R.G., Diez, A., Norén, M., Mouchel, O., Jérôme, M., Verrez-Bagnis, V., Van Pelt, H., Favre-Krey, L., Krey, G., Bautista, J.M., 2007. Primers and polymerase chain reaction conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome b and nuclear rhodopsin genes. *Mol. Ecol. Notes* 7, 730–734. <https://doi.org/10.1111/j.1471-8286.2007.01863.x>
- Shehata, H.R., Li, J., Chen, S., Redda, H., Cheng, S., Tabujara, N., Li, H., Warriner, K., Hanner, R., 2017. Droplet digital polymerase chain reaction (ddPCR) assays integrated with an internal control for quantification of bovine, porcine, chicken and turkey species in food and feed. *PLoS One* 12, 1–17. <https://doi.org/10.1371/journal.pone.0182872>
- Smalås, A., Strøm, J.F., Amundsen, P., Dieckmann, U., Primicerio, R., 2019. Climate warming is predicted to enhance the negative effects of harvesting on high-latitude lake fish. *J. Appl. Ecol.* 1–13. <https://doi.org/10.1111/1365-2664.13535>
- Sonet, G., Jordaens, K., Nagy, Z.T., Breman, F.C., Meyer, M. De, Backeljau, T., Virgilio, M., 2013. Adhoc: an R package to calculate ad hoc distance thresholds for DNA barcoding identification. *Zookeys* 365, 329–336.

South, A., 2017a. rnaturalearth: World Map Data from Natural Earth.

South, A., 2017b. rnaturalearthdata: World Vector Map Data from Natural Earth Used in  
“rnaturalearth.”

Spink, J., Moyer, D.C., 2011. Defining the public health threat of food fraud. *J. Food Sci.* 76.  
<https://doi.org/10.1111/j.1750-3841.2011.02417.x>

Sustainable Fishing | MSC | Marine Stewardship Council [WWW Document], n.d. URL  
<https://www.msc.org/home> (accessed 10.22.19).

The EU Fish Market 2016 Edition. EUMOFA, 2019. <https://doi.org/10.2771/168390>

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: Improving the sensitivity of progressive  
multiple sequence alignment through sequence weighting, position-specific gap penalties and  
weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.  
<https://doi.org/10.1093/nar/22.22.4673>

Tomás, C., Ferreira, I.M.P.L.V.O., Faria, M.A., 2017. Codfish authentication by a fast Short Amplicon  
High Resolution Melting Analysis (SA-HRMA) method. *Food Control* 71, 255–263.  
<https://doi.org/10.1016/j.foodcont.2016.07.004>

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.  
<https://doi.org/10.7717/peerj.1891>

Virgilio, M., Jordaens, K., Breman, F.C., Backeljau, T., de Meyer, M., 2012. Identifying insects with  
incomplete DNA barcode libraries, African fruit flies (Diptera: Tephritidae) as a test case. *PLoS*  
*One* 7. <https://doi.org/10.1371/journal.pone.0031581>

Vis- & zeevruchtengids - voor professionele gebruikers | Zeevruchten gids [WWW Document], n.d.  
URL <http://www.zeevruchtengids.org/nl> (accessed 11.14.19).

Vis - VLAM [WWW Document], n.d. URL <https://www.vlam.be/nl/feitenencijfers/vis> (accessed 6.5.19).

VisOnline [WWW Document], n.d. URL <https://www.visonline.be/home> (accessed 12.6.19).

Ward, R.D., 2009. DNA barcode divergence among species and genera of birds and fishes. *Mol. Ecol. Resour.* 9, 1077–1085. <https://doi.org/10.1111/j.1755-0998.2009.02541.x>

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., Hebert, P.D.N., 2005. DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 1847–1857.  
<https://doi.org/10.1098/rstb.2005.1716>

Weinsier, J., 2016. Restaurant busted for passing off cheaper fish for white tuna [WWW Document]. URL <https://www.local10.com/news/restaurant-reports/restaurant-busted-for-passing-off-cheaper-fish-for-white-tuna> (accessed 5.24.19).

Wickham, H., 2019a. *forcats: Tools for Working with Categorical Variables (Factors)*.

Wickham, H., 2019b. *stringr: Simple, Consistent Wrappers for Common String Operations*.

Wickham, H., 2018. *pryr: Tools for Computing on the Language*.

Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H., Bryan, J., 2019. *readxl: Read Excel Files*.

Wickham, H., François, R., Henry, L., Müller, K., 2019. *dplyr: A Grammar of Data Manipulation*.

Wiemers, M., Fiedler, K., 2007. Does the DNA barcoding gap exist? - A case study in blue butterflies (Lepidoptera: Lycaenidae). *Front. Zool.* 4, 1–16. <https://doi.org/10.1186/1742-9994-4-8>

Worm, B., Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, B.S., Jackson, J.B.C., Lotze, H.K., Micheli, F., Palumbi, S.R., Sala, E., Selkoe, K.A., Stachowicz, J.J., Watson, R., 2006. Impacts of biodiversity loss on ocean ecosystem services. *Science* (80-. ). 314, 787–790.

<https://doi.org/10.1126/science.1132294>

Young's Seafood [WWW Document], n.d. URL <https://youngsseafood.co.uk/> (accessed 5.24.19).

Yu, G., Lam, T.T.-Y., Zhu, H., Guan, Y., 2018. Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Mol. Biol. Evol.* 35, 3041–3043.

<https://doi.org/10.1093/molbev/msy194>

Zardoya, R., Meyer, A., 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Mol. Biol. Evol.* 13, 933–942.

<https://doi.org/10.1093/oxfordjournals.molbev.a025661>