Food Control 79 (2017) 297–308

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Novel nuclear barcode regions for the identification of flatfish species

Valentina Paracchini [a], Mauro Petrillo [a], Antoon Lievens [a], Antonio Puertas Gallardo [a], Jann Thorsten Martinsohn [a], Johann Hofherr [a], Alain Maquet [b], Ana Paula Barbosa Silva [b], Dafni Maria Kagkli [a], Maddalena Querci [a], Alex Patak [a], Alexandre Angers-Loustau [a, *]

[a] European Commission, Joint Research Centre (JRC), via E. Fermi 2749, 21027 Ispra, Italy
[b] European Commission, Joint Research Centre (JRC), Retieseweg 111, 2440 Geel, Belgium

A R T I C L E I N F O

Article history:
Received 7 November 2016
Received in revised form
5 April 2017
Accepted 6 April 2017
Available online 7 April 2017

Keywords:
Bioinformatics
DNA barcoding
Next-generation sequencing
Seafood identification

A B S T R A C T

The development of an efficient seafood traceability framework is crucial for the management of sustainable fisheries and the monitoring of potential substitution fraud across the food chain. Recent studies have shown the potential of DNA barcoding methods in this framework, with most of the efforts focusing on using mitochondrial targets such as the *cytochrome oxidase 1* and *cytochrome b* genes. In this article, we show the identification of novel targets in the nuclear genome, and their associated primers, to be used for the efficient identification of flatfishes of the *Pleuronectidae* family. In addition, different *in silico* methods are described to generate a dataset of barcode reference sequences from the ever-growing wealth of publicly available sequence information, replacing, where possible, labour-intensive laboratory work. The short amplicon lengths render the analysis of these new barcode target regions ideally suited to next-generation sequencing techniques, allowing characterisation of multiple fish species in mixed and processed samples. Their location in the nucleus also improves currently used methods by allowing the identification of hybrid individuals.
© 2017 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The ongoing global population growth, tied to a general increase of income and urbanization, has led to a steady increase of fish consumption, resulting in a strong expansion of fish production and the development of complex distribution patterns. This is paralleled by an extensive, and often unsustainable, exploitation of fish populations and an ever increasing aquaculture production (FAO, 2014).

In this context, the development of an efficient seafood traceability framework is crucial for the correct management of fisheries and the protection of fish biodiversity. For example, a recent study classified 7.5% of European marine fish species as having an elevated risk of extinction in European waters, while the insufficient availability of data prevented the full assessment for more than 20% of the studied species (Fricke, 2015). Species identification is also needed to ensure a correct labelling through the whole food distribution chain, in order to protect the consumers' right to choose what they eat and to protect them against fraud such as the substitution of some species for cheaper, hard to distinguish alternatives.

The importance of seafood traceability is reflected in the current European Union legal framework. In addition to the established general rules on the provision of food information to consumers (FIC) (EU, 2011) that brought together the existing EU rules on general food and nutrition labelling into one piece of legislation, recent amendments were made to the Common Organisation of the Markets (CMO) in fishery and aquaculture products (EU, 2013). For unprocessed and some processed fish products, these requirements mandate that both the commercial and scientific names are displayed, matching those on the official lists drawn up and published by each EU country (D'Amico, Armani, Gianfaldoni, & Guidi, 2016).

Fish species identification based on external morphological features is, at best, only applicable to whole (and ideally adult) specimens and sometimes is hampered by the presence of sibling species (Knowlton, 1993). Due to its ubiquity and relatively high resilience to processing, DNA has established itself as the target of choice for the analysis of unknown fish samples (Rasmussen & Morrissey, 2009; Teletchea, 2009). One of the associated techniques is DNA barcoding, which involves the amplification of specific DNA regions by the polymerase chain reaction (PCR), followed

* Corresponding author.
E-mail address: Alexandre.Angers@ec.europa.eu (A. Angers-Loustau).

http://dx.doi.org/10.1016/j.foodcont.2017.04.009
0956-7135/© 2017 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

by sequencing of the produced amplicon (Hebert, Ratnasingham, & Waard, 2003). Since the species is identified at the sequencing step, this technique has the advantage (over, for example, real-time PCR taxon-specific methods) that it does not require *a priori* knowledge of the potential species in the test sample. On the other hand, this approach depends on two main conditions: the selection of a suitable barcode target and the identification of associated primers, as well as knowledge about the reference barcode sequence for each of the species of interest.

For animals, the mitochondrial *cytochrome oxidase 1* (*COI*) gene has become the *de facto* target of choice, with the development of a whole community dedicated to generate and compile the reference sequences for known species (Stockle & Hebert, 2008). For fish, the FISH-BOL campaign (Ward, Hanner, & Hebert, 2009) has already compiled the sequence of a 648 bp region of the *COI* gene for more than 100 000 specimens representing more than 10 000 species. With these resources, COI DNA barcoding has been used to trace fish species in Egyptian aquafeed formulations (Galal-Khallaf, Osman, Carleos, Garcia-Vazquez, & Borrell, 2016), differentiate between members of the *Sparidae* species (Armani, Guardone et al., 2015; Armani, Tinacci et al., 2015), and verify the labels on seafood products commercialized in Southern Brazil marketplaces (Carvalho, Palhares, Drummond, & Frigo, 2015), to name a few recent examples.

Recent studies aiming to evaluate the extent of fish product mislabelling, due to either fraud or negligence, have shown that the issue remains a serious concern, both in Europe (Armani et al., 2013; Di Pinto et al., 2013; Helyar et al., 2014; Vandamme et al., 2016) and the rest of the world (Cunha et al., 2015; Khaksar et al., 2015; Yan et al., 2016). In 2015, a control plan coordinated at the EU level was launched to assess the prevalence of mislabelling of white fish products at all stages of the food chain. The results report a high number of non-compliances in products declared as Atlantic cod (*Gadus morhua*), in part due to the fact that it was one of the most sampled species. For this species, the observed mislabelling rate was 4%, which is in line with the values obtained in another recent survey (Mariani et al., 2015), although higher rates have also been reported (Filonzi, Chiesa, Vaghi, & Nonnis Marzano, 2010; Herrero, Madriñán, Vieites, & Espiñeira, 2010). When adjusting for the number of samples tested, the flatfishes were among the highest in terms of mislabelling frequencies, including the common sole (*Solea solea*, 24%), the yellowfin sole (*Limanda aspera*, 15%) and the halibut (*Hippoglossus hippoglossus*, 8%). Previous studies in Germany and Spain suggested substitution rates for sole as high as 30%–50% (Herrero, Lago, Vieites, & Espiñeira, 2012; Kappel & Schröder, 2015).

Although the use of DNA barcoding targeting mitochondrial regions such as the *COI* or *cytochrome b* genes for species identification is well established and documented, some limitations have been identified that justify the development of additional targets. Closely related species are sometimes difficult to differentiate with mitochondrial sequences; in these cases, the efficiency of barcoding can be improved by targeting additional genomic positions. For this purpose, the nuclear rhodopsin gene has been used (Sevilla et al., 2007). Still, the length of the sequenced fragments required (usually >600 bp) is above that of the average DNA fragment lengths resulting from certain processing techniques (Armani et al., 2013; Chapela et al., 2007), which renders the amplification problematic. Moreover, those fragment lengths tend to be incompatible with the current Next Generation Sequencing technologies which rely on shorter fragments. This difficulty in assessing highly processed food products consisting of a mixture of species in unknown quantities motivated the development of strategies producing shorter amplicons (Armani, Guardone et al., 2015; Armani, Tinacci et al., 2015; Muñoz-Colmenero, Martínez, Roca, & Garcia-Vazquez, 2017; Shokralla, Hellberg, Handy, King, & Hajibabaei, 2015).

In order to address this need for powerful and efficient DNA-based fish species identification methods with a wider application scope, we describe in this article the identification of novel nuclear barcoding targets, tailored to the identification of flatfish group members. We started by using an automated screening approach to identify short candidate barcode regions within the nuclear genome. We then tested samples of various flatfish species with a selected set of the generated primer pairs, demonstrating the advantages of these novel targets over the mitochondrial ones, such as the capability to identify hybrid individuals, as well as multiple fish species in complex mixtures.

2. Materials and methods

2.1. Primers design

A set of scripts was used to automatically screen a set of candidate ("seed") barcode regions, and to then produce primer pairs on regions matching the requirements of a DNA barcode. This was achieved through the following steps:

2.1.1. Seeds selection

A first strategy to identify seed sequences involved comparing two fish genome sequences using BLAST. The two genomes chosen were *G. morhua* and *T. rubripes* (both from Ensembl, release 78). These genomes were chosen due to the relative evolutionary distance between the two species. The *T. rubripes* genome sequences, in stretches of 100.000 bases, were BLASTed against the *G. morhua* genome. Positive hits with at least 90% identity over a region of at least 200 base pairs were extracted. The final set of seeds were then numbered (1–867) and saved in a FASTA file.

The second source of seed sequences were UCEs identified in humans and available at the UCNE, the Ultra-Conserved Non-coding Element base web portal (ccg.vital-it.ch/UCNEbase). The latest file for humans UCEs (hg19_UCNEs.fasta) was downloaded and potential duplicates eliminated. The resulting sequences were saved in a FASTA file, numbered from 868 to 5052.

The last set of seeds was obtained from the publication by Faircloth et al. (2013). The fish contigs FASTA files from the different species were pooled into a single file and duplicates eliminated. The resulting sequences were saved in a FASTA file, numbered from 5053 to 11836.

To identify potential barcode primers, these seeds were then expanded and their sequences obtained in a small subset of genomes as previously described for plants (Angers-Loustau et al., 2016). The extension size of 2 kb from these analyses was used for the current work. The two genomes selected for the extension were *Danio rerio* and *Oryzias latipes*, and the screen was done using all the 11 fish genomes from Ensembl (release 78), i.e. *Astyanax mexicanus* (Mexican tetra), *Danio rerio* (Zebrafish), *Gadus morhua* (Atlantic cod), *Gasterosteus aculeatus* (Stickleback), *Lepisosteus oculatus* (Spotted gar), *Oreochromis niloticus* (Nile tilapia), *Oryzias latipes* (Medaka), *Poecilia formosa* (Amazon molly), *Takifugu rubripes* (Japanese pufferfish), *Tetraodon nigroviridis* (Green puffer fish) and *Xiphophorus maculatus* (Southern platyfish).

2.1.2. Primer generation

For the seeds giving at least one "valid" hit in each of the eleven genomes, the corresponding sequences were extracted and aligned using MAFFT (Katoh, Misawa, Kuma, & Miyata, 2002). From the aligned sequences, custom scripts were used to identify potential primer pairs (primer length: 20 bp), matching the following criteria: 1) each position is conserved in at least 8 of the 11 species

Download English Version:

https://daneshyari.com/en/article/5767464

Download Persian Version:

https://daneshyari.com/article/5767464

Daneshyari.com