From DNA to biomass: opportunities and challenges in species quantification of bulk fisheries products

From DNA to biomass: species quantification of bulk fisheries products

Brian Klitgaard Hansen[1], Gregory Kevin Farrant[2], Rob Ogden[3, X], Emily Humble[3], Guðbjörg Ólafsdóttir[2], Dorte Bekkevold[1], Steen Wilhelm Knudsen[4], Peter Rask Møller[4] & Einar Eg Nielsen[1]

[1]Section for Marine Living Resources, National Institute of Aquatic Resources, Technical University of Denmark, Vejlsøvej 39, 8600 Silkeborg, Denmark

[2]Matis, Vinlandsleid 12, 113 Reykjavik, Iceland

[3]TRACE Wildlife Forensics Network, 16 Corstorphine Hill Avenue, Edinburgh, EH12 6LE, UK

[4]Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, Easter Bush Campus, EH25 9RG, UK.

[5]University of Copenhagen, Universitetsparken 15, 2100 Copenhagen Ø, Denmark

Corresponding author: tel: +45 35883105 ;fax: +45 35883150 ;email: bkha@aqua.dtu.dk

Keywords: DNA quantification, multi-species fisheries products, quantitative PCR, MinION, metabarcoding

1. Abstract

Fisheries enforcement relies on visual catch identification and quantification at sea or when landed. Silage (fish dissolved in acid) and fish blocks (block frozen fish) are promising methods for on-board processing and storage of low-value catches. We examined the use of two DNA based methods, quantitative PCR (qPCR) and metabarcoding, for assessing species composition and quantity in industrial grade experimental fish blocks and silage. Using non-destructive sampling and standard DNA analysis, we demonstrate the ability to identify and quantify DNA from fish species in both products. qPCR analysis of small silage samples collected over 21 days, detected all target control species. DNA from one species (wolf fish) was consistently overrepresented while for three species of gadoids (cod, haddock and whiting), DNA content matched input tissue proportions with high accuracy. Both qPCR and metabarcoding of fish blocks, sampled as run-off water and exterior swabs, provided consistent species detection, with the highest variance observed in quantification from swab samples. Our analysis show that DNA based methods have significant potential as a tool for species identification and quantification of complex onboard-processed seafood products and are readily applicable to taxonomically and morphologically similar fish. There is, however; a need for establishing DNA/weight calibration factors for primary fisheries species.

2. Introduction

Historically, discards of fish have represented a considerable fraction of the total catch in many fisheries (Heath *et al.*, 2014; Guillen *et al.*, 2018). Discarding generally occurs when fish are undersized, represent no commercial value or are outside existing quotas (Guillen *et al.*, 2018). However, discarding is widely regarded as a waste of resources and has many undesirable ecological and socio-economical side effects (Heath *et al.*, 2014; Guillen *et al.*, 2018). As a result, a number of countries are starting to implement regulations obliging fishermen to land all caught species, including the most recent European Landing Obligation under the Common Fisheries Policy (European Commission, 2013). However, on board handling, sorting and storing of low-value species or specimen is not economically attractive for fisheries (Batsleer *et al.*, 2015). Hence, alternatives to minimize handling and storage of unwanted catches have been suggested, including

development of bulk products (Larsen *et al.*, 2013). These include for example block-frozen fish or skinless fillets, minced fish and silage. On board silage production entails dissolving fish in strong acid. The silage process both concentrates and conserves the raw material for further processing and importantly saves valuable space for storage of high-value fish. Another bulk handling and storage option is to pack and freeze low-value species on board into so-called "fish blocks" (Larsen *et al.*, 2013). However, on board production of silage and fish blocks makes Monitoring Control and Surveillance (MCS) by fisheries inspectors virtually impossible, as it effectively prohibits crosscheck reporting in the mandatory landing manifest with the content of the landed bulk product. Currently therefore, such mixed products cannot be landed in the EU due to the principle of control throughout the food chain - from the catch of fish to landing in the European Union (Reg. EC 1005/2008). Hence, finding reliable and cost-efficient alternatives to visual identification methods for assessing contents of these complex fish products could potentially benefit both the MCS practitioners and commercial fisheries by allowing landing of analytically certified bulk products.

Over the last decade, genetic methods for species identification have gone through extensive development, from being restricted to identifying a single species from specimen based samples, to the present state of the art allowing both qualitative and quantitative information to be derived from complex multi-species samples (e.g. Floren *et al.*, 2015; Thomas *et al.*, 2016; Ushio *et al.*, 2018). Quantifying input biomass from DNA analysed in fish products is an emerging area of research and has a large unexplored potential in comparison to traditional visual identification, especially in samples where the morphological characteristics are sparse or absent (Nagase *et al.*, 2010; Bojolly *et al.*, 2017; Sánchez *et al.*, 2019). Further, DNA, particularly mitochondrial DNA (mtDNA), has been found suitable for species detection and quantification even when products are highly processed (e.g. Nagase *et al.*, 2010; Wen *et al.*, 2015; Piskata *et al.*, 2017). Quantitative real-time PCR (qPCR) is the gold standard of DNA quantification and has been utilized in food fraud detection of raw and processed meat for over a decade (e.g. Lopez and Pardo, 2005; Tanabe *et al.*, 2007). More recent metabarcoding approaches using high-throughput sequencing (HTS) can also potentially be used to quantity target DNA but have not yet been developed to the same extent. One advantage of metabarcoding is that it can provide information on the entire DNA biodiversity within a sample, without *a priori* knowledge of which species to assess, as required for qPCR (Miya *et al.*, 2015; Menegon *et al.*, 2017; Stat *et al.*, 2017; Srivathsan

et al., 2018). One particularly promising device for metabarcoding is the miniaturised nanopore based DNA sequencing platform, the MinION (Oxford Nanopore Technologies, UK), which offers several advantages over traditional HTS technologies, including portability, low initial start-up costs and real-time analysis (Mikheyev and Tin, 2014).

Our aim was to test DNA based methods for species identification and quantification in mixed fish species products using silage and fish block products as examples of products high on the agenda in relation to practical use in fisheries under the European landing obligation. We focused analyses on three gadoids, Atlantic cod (*Gadus morhua* – hereafter cod), whiting (*Merlangius merlangus*) and haddock (*Merlanogrammus aeglefinus*), as they are all important fisheries species subject to quotas, but also often unintentionally caught as bycatch, which historically has led to discarding (Heath *et al.*, 2014). In addition, we used wolffish (*Anarhichas lupus*), which represents a taxonomically different and less common species, that due to its demersal behaviour is a bycatch in gadoid fisheries (Grant and Hiscock, 2014). Using experimental mixes of tissue from individuals of these species we calculated and compared relative estimates of DNA abundance and input tissue using both qPCR and MinION metabarcoding. Our focus was on experimentally mixed silage and fish block samples, but we developed and test a series of control samples of known species mixtures using normalized tissue and DNA, in order to understand variation in the relationships between levels of input tissue,DNA and resulting species measurements.

3. Materials and methods

3.1. Tissue and DNA normalized samples

Three whole individuals of each species were purchased at a local fish store and fin tissue was collected using sterile forceps and scalpels. The tissue was stored at room temperature in sterile tubes filled with 99% EtOH until sample preparation. Single- and mixed species samples for both tissue and DNA normalized samples were prepared (See Supplementary Table S1). Tissue samples were weighed on a Mettler AT460 (Mettler-Toledo, Slovenia) using a maximum of 30 mg tissue/sample to minimize the risk of saturating extraction yield. Forceps and scalpels were changed between each individual for both collection and preparation of tissue

samples. DNA extractions were conducted using the Omega Biotek E.Z.N.A. Tissue DNA kit (Omega Biotek, USA) applying the Tissue DNA protocol. Final elution volume was 2 x 100 µl in elution buffer. DNA concentration was measured using a Qubit 3.0 fluorometer (dsDNA BR Assay Kit, Thermo Fisher Scientific, USA). DNA normalized samples was made with DNA extracted from single species samples and was normalized using nuclease free water to 10 ng/µl. After dilution, the samples were measured again in order to verify that the final concentration was close to the targeted 10 ng/µl. Tissue and DNA normalized samples are hereafter referred to as control samples.

3.2. **Silage samples**

The silage acid solution consisted of 1.5% formic acid, 0.1% of potassium sorbate and 200 ppm ethoxyquin. The pH was adjusted with sodium hydroxide to reach a target pH of 3.5. In total 6.9 kg whole gutted fish representing all four species, i.e. cod, whiting, haddock and wolffish were added to the silage solution on day 0. The silage was kept at RT during the entire experiment. 250 µl silage samples were collected at day 1, 2, 3, 4, 5, 7, 8, 9, 14 and 21 from the centre of the silage container, roughly 1 cm below the surface. Before sampling (except for day 1), the silage was stirred to homogenize and improve the decomposition of the fish. DNA extraction followed the standard Omega Biotek E.Z.N.A. tissue DNA kit protocol, with one hour incubation and final elution in 2 x 100µl elution buffer. Extracted DNA was measured on a Qubit and with a Bioanalyzer (Agilent Technologies, USA) using the High Sensitivity DNA assay. Samples from day 2 and day 21 were analysed on MinION and samples from all days were analysed with qPCR. Two different laboratories carried out parallel sample analysis, each with their own MinION (termed MinION-1 and MinION-2).

3.3. **Fish block samples**

A fish block was prepared from fresh cod (83%) and wolffish (17%). The fish were weighed, put into a box and frozen at -24 °C, identical to standard fish block operating procedures. Before sampling, collection tools were cleaned with a 0.5% bleach solution and rinsed with nuclease free water. Two approaches for fish block content assessment were tested, referred to as “swab” (SW) and “runoff water” (RO). Three swab samples (SW1-3) were collected from fish block surfaces using a sterile cotton swab (806-WC, Puritan, USA) swiped

across the surface of the fish block. For each sample, we altered the sampling pattern, i.e. front (SW1), back (SW2) and edges of the fish block (SW3). DNA was extracted from the swabs using the QIAamp DNA Mini Kit (QIAGEN, Germany). All swab sample replicates were analysed using qPCR and MinION-2, while PCR products, using a universal primer set (see section 3.5), for all three samples were pooled (SWp) for the MinION-1 analysis. For the RO method 8 litres of demineralized water were poured over the fish block, repeating it three times reusing the same water and each time collecting run-off in a tray. Subsequently, triplicate 200-300 ml RO water samples (RO1-3) were subsampled and filtered using a sterile 60 ml syringe and a 0.22 µm Sterivex filter (SVGPL10RC, Merck, USA). DNA was extracted with a modified protocol of QIAGEN's blood and tissue kit (QIAGEN, Germany) following Spens and colleagues (2017). The protocol was modified to include a 2-hour incubation period at 56°C and with final elution carried out in 2 x 100µl AE buffer. DNA concentration was assessed using Qubit. All three RO samples were analysed with both qPCR and MinION-2 metabarcoding, while RO1 was also analysed on MinION-1.

3.4. **Quantitative PCR**

We used species-selective assays targeting cod, whiting, haddock and wolffish (Gm, Mm, Ma and Al, respectively). Mm, Ma and Al assays were developed by aligning sequences of the mtDNA cytochrome b (cytb) or NADH dehydrogenase subunit 4 (ND4) genes and identifying gene regions with maximum interspecific sequence difference between target and non-target species. The Gm assay was developed by Knudsen and colleagues (2018). As DNA is susceptible to degradation in processed samples (Piskata *et al.*, 2017), we developed assays to produce short PCR products (70-150 bp). Sequences were obtained from Genbank (https://www.ncbi.nlm.nih.gov/genbank/) using an integrated search tool in Geneious v. 9.1.6. where candidate primers and probes were found using the built-in Primer3 v.2.3.4 search engine (Kearse *et al.*, 2012). Distance matrices of nucleotide differences between target and non-target species were developed for primers and probes (Supplementary Table S2). Finally, the assays were tested *in silico* using Genbank's online BLAST function, Primer-BLAST and Nucleotide BLAST to assess global specificity of primers and probes. Assays were tested and evaluated *in vitro* using the control samples.

Prior to qPCR analysis, control and RO samples were diluted (between 1:10 and 1:100) with nuclease free water to avoid PCR inhibition. qPCR reactions were conducted in 10 µl volumes with 4 µl TaqMan Universal PCR Master Mix (ThermoFisher Scientific, USA) and 1 µl template DNA. Final PCR reaction volume was 10 µl, varying volumes of primers, probes and nuclease free water were used to obtain optimal qPCR concentrations for each assay (see Supplementary Table S3). Assays used a double-quencher probe, 5'FAM/ZEN/3'IBFQ (Integrated DNA technologies, USA) to improve delta fluorescence. Assay sequences, qPCR concentrations and standard curve parameters are found in Supplementary Table S3. Thermal cycling cycle conditions were: 50°C for 2 minutes and 95°C for 10 min followed by 40 cycles at 95°C for 15 seconds and 60°C for 1 minute. All samples were run in triplicate on a StepOnePlus Real-Time PCR System (Life Technologies, USA). Each run had minimum three negative plate controls and a standard curve of 10-fold dilutions ranging from 10 to $1x10^7$ copies/reaction.

3.5. **MinION and bioinformatics**

For the MinION based metabarcoding approach, we used universal primers (Hereafter COIP, forward primer 5'-ACAAATCAYAARGAYATYGG-3' and reverse 5'-TTCAGGRTGNCCRAARAAYCA-3') (Mikkelsen *et al.*, 2006), which target a 699bp fragment of the Cytochrome oxidase subunit I gene (COI). PCR was carried out in 50 µl reaction volumes containing 3 µl DNA, 0.4 µl Taq DNA polymerase (New England BioLabs, UK), 5 µl 10× standard buffer (New England BioLabs, UK), 5 µl of 10 mM dNTP, 0.3 µl of the forward primer (100 µM) and 0.3 µl of the reverse primer (100 µM). The PCR thermal profile was: 4 min at 94°C followed by 35 cycles of 50 s at 94°C, 40 s at 48°C, 1 min at 68°C, with a final elongation step of 7 min at 68°C. The DNA was sequenced on the MinION with the 2D Amplicon sequencing protocol (SQK-LSK208, Oxford Nanopore Technologies, UK). To extract the nucleotide sequences from the raw data generated by MinKNOW, we used Albacore v.2.3.1 for base-calling and de-multiplexing. The de-multiplexed fastq files were converted to fasta files using bash scripts. The resulting sequences were blasted to a database of 5220 COI sequences using blastn with an e-value cut-off of $1e^{-4}$, minimum percent identity of 50 and a maximum number of target sequences of two. Best blast hits (highest bit score) were selected when queries were assigned to multiple sequences. The BLAST database included 5004 full-length sequences downloaded from NCBI, plus additional

216 sequences of Gadiformes and Anarhichadidae that were not already included. This ensured that our assignment approach would be robust in the presence of closely related species. Database sequences were identical to the region of COI targeted by the primers. The composition of species within each sample was determined with an identity cut-off threshold of <85%. Besides the four target species, species representing <2% of sequences from a sample were categorized to 'other'.

4. Results

4.1. **Cross reactivity and false positives**

Throughout the study, we employed a rigorous system of controls for monitoring potential contamination, including DNA extraction blanks and at minimum triplicate PCR blanks for each qPCR run. Contamination was observed in a few negative controls (Supplementary Table S4). However, levels of contamination in negative controls were extremely low compared to qPCR results from positive samples, with a maximum of ≤0.86% (wolffish, in silage day 3) and overall average of ≤0.0008% contaminant DNA within samples (Supplementary Table S4). We assessed potential cross-amplification of non-target species using single species samples. Minute cross-amplification was observed when using high template concentrations, but the target species always amplified over four orders of magnitude better than non-target species (Supplementary Table S4 and S5), and non-target samples always amplified below the limit of quantification (<10 copies/reaction; Supplementary Table S3). False positive species were detected with the MinION (Supplementary Table S6). Across all samples, the cumulated read counts assigned to false-positive species was on average 1.39%, with the highest single species average of 0.75% of the reads. One pure haddock sample, HExt1a (see Supplementary Table S1), showed an unexpectedly high contribution of cod reads (9.2%). This was likely caused by insufficient cleaning of the MINION flow cell since it had previously been used for a pure cod analysis (CExt1a) sample, and inference on cod was therefore omitted for this sample. With the exclusion of this sample, averages of false positives were down to 1.05% and 0.38%, for all species and for the most quantitatively dominant species, respectively.

4.2. **Basic inferences**

Estimation of tissue proportions from DNA copies (qPCR) or reads (MinION), commonly build on the assumption that there is a linear relationship between proportions of DNA and tissue, and thus that there is little or no variation in copies/reads per weight unit tissue. To assess intra- and inter-specific variation in DNA/tissue ratios we compared estimates of DNA copies per mg tissue (copies/mg tissue), DNA concentration per tissue weight (ηgDNA/mg tissue) and copies per total DNA concentration (copies/ηgDNA) among individuals and species. Average DNA concentrations in tissue were in the same range (10^7 copies/mg); only one individual, H3, showed a slightly lower concentration (See Supplementary Table S1 and Figure S1). Still, it is apparent that cod and whiting generally had slightly higher average DNA copies/mg tissue, than wolffish and haddock (Supplementary Figure S1). Cod also had the highest ηgDNA/mg tissue ratio whereas whiting, haddock and wolffish, has more similar ηgDNA/mg tissue ratios (Supplementary Figure S1). For the copies/ηg DNA, we found average concentrations in the order 10^4 copies/ηg DNA across all species (Supplementary Figure S1). The highest variation within species was found for cod, having both the highest 2.3 x 10^4 copies/ng DNA, and lowest, 1.4 x 10^4 copies/ng DNA estimated across individuals. Individual C2 showed a lower ratio of copies/ng DNA of any individual fish (see Supplementary Table S1 and Figure S1), despite having relatively high copies/mg tissue, potentially illustrating DNA degradation.

4.3. **qPCR analysis of control samples**

Single species control samples showed the expected 100% target species proportions (Supplementary Table S7). Still, qPCR analysis revealed minute proportions of non-target species averaging at 0.0075% (Max = 0.03%). The mixed species tissue samples, with controlled weight of starting tissue, showed modest power for DNA-based estimation of initial tissue proportions (Figure 1(A-D), Supplementary Table S7 and S8). Normalized mixed species tissue samples (hereafter mixed tissue samples) showed overall deviation from the expected proportions of 12% ±15 (Cod, C), -17% ±8 (Haddock, H) and 5% ±16 (Whiting, W). The systematic errors, i.e. percentage deviation of the determined value from the expected proportion, were estimated to 49% ±67, -54% ±23 and 12% ±49 for C, H, W. In contrast, qPCR results for normalized mixed species DNA samples, with controlled input DNA concentration (hereafter mixed DNA samples), showed better

correspondence with DNA input proportions, as seen from the lower level of deviation, 9% ±13, -6% ±5 and -3% ±10, and systematic error, 28% ±55, -11% ±13, -8% ±26, for C, H and W, respectively. For all samples, qPCR had an average accuracy of 11% ±15, -14% ±9 and 3% ±14 for C, H and W, respectively. These results suggest that haddock mtDNA copies were under-represented, in particular for the mixed tissue samples, as seen in sample CHW1, CHW2 and CHW3. However, samples CHW299 and CHW929 were exceptions with less underrepresentation and deviation (average -5% and -7%, respectively). Variation among replicate samples was on average 11% ±5 for the mixed tissue samples.

4.4. **qPCR analysis of silage and fish block**

Tissue from fish in the acid solution dissolved and liquefied within days after submergence. From observations on day 2 it was estimated that >80% of fish tissue was dissolved beyond visual recognition. Still, all four species were detected by DNA even after 21 days at room temperature. Comparison of DNA based qPCR to expected proportions from tissue in the silage showed initial average systematic errors of -74% ±10, -68% ±9, -44% ±23 and 279% ±34 for cod, haddock, whiting and wolffish, respectively, with average deviations of 37% ±5, -17% ± 2, -2% ±1 and 56% ±7 (Figure 2A, Supplementary Table S9). The primary reason for deviations between observed and expected contents was that wolffish was considerably overrepresented. Therefore, we then assessed silage composition excluding information for wolffish, which considerably lowered the systematic errors for the gadoids, -13% ±16, 9% ±32, 83% ±47 for C, H, W and significantly increased accuracy (-8% ±10, 3% ±10 and 6% ±3) for the qPCR methodology, (Figure 2B, Supplementary Table S10).

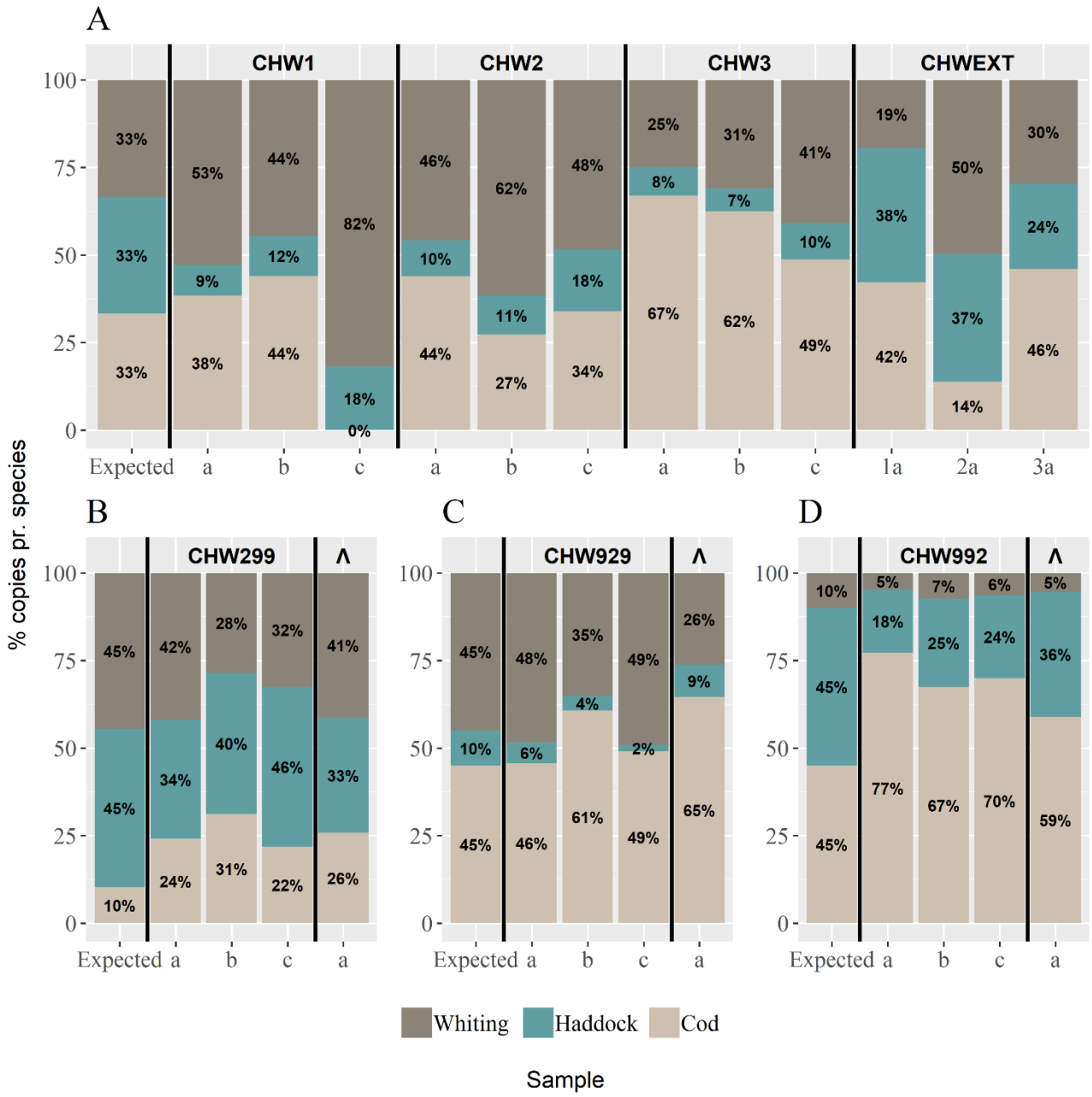

Figure 1: Comparison of tissue input weight percentages (Expected) with qPCR estimated DNA proportions for control samples. Tissue and DNA mixture percentages for cod, haddock, whiting and wolffish are 33:33:33:0 (A), 10:45:45:0 (B), 45:10:45:0 (C) and 45:45:10:0 (D). Tissue mixed samples are denoted as CHW1, CHW2, CHW3, CHW299, CHW929 and CHW992. In (A) analysis of DNA normalized samples are shown as CHWEXT and in B-D denoted as Λ. See text and Supplementary Table S1 for explanation.

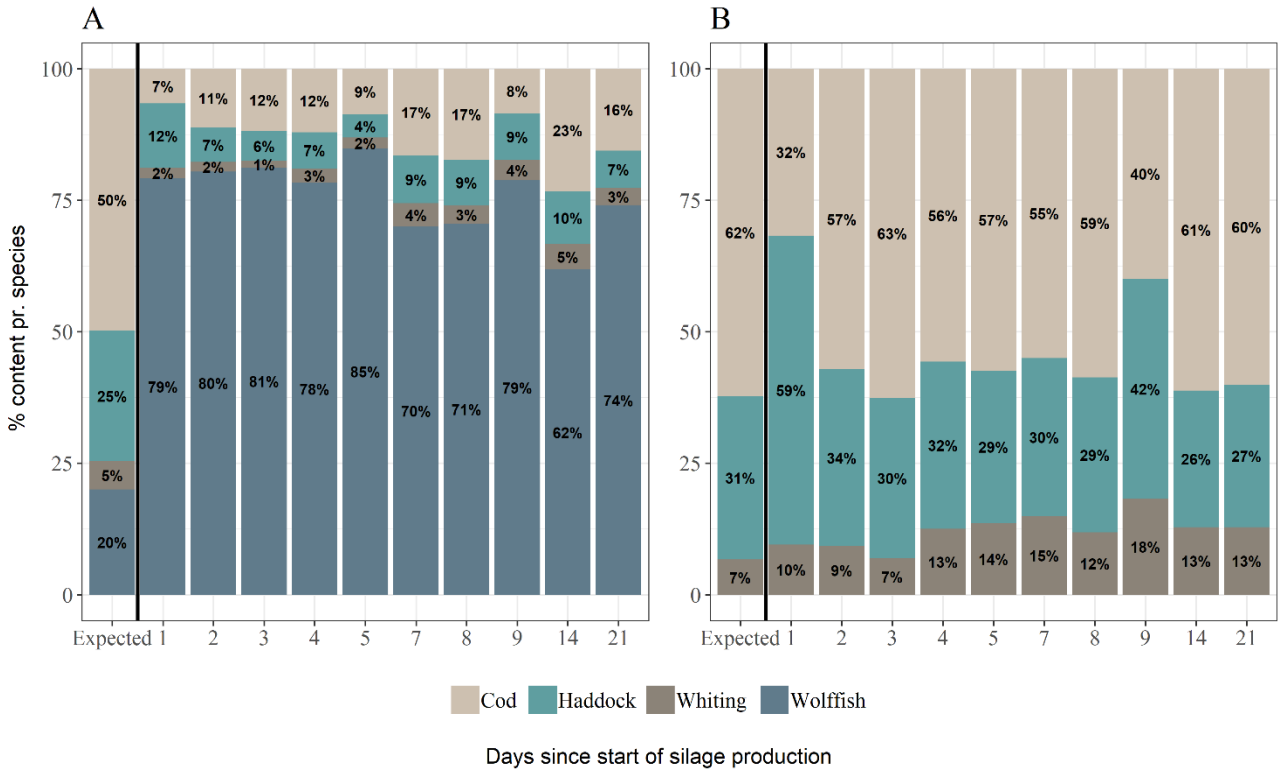

Figure 2: Comparison of tissue input weight percentages (Expected) with qPCR estimated DNA proportions for fish silage. (A) shows results for all four species, while (B) only includes analysis of gadoids (cod, haddock and whiting). Numbers on x-axis show day of sampling.

Further, it was evident that DNA copy concentration increased on day 7, 14 and 21, where silage stirring had not been conducted in days leading up to the sampling (Supplementary Figure S2). However, this did not obscure the proportional DNA estimates among species (Figure 2). Surprisingly, the highest copy concentrations for any sample were found on day 21 for whiting and cod.

The qPCR method consistently detected both species in the fish blocks for all RO and swab samples. As for the silage analysis, wolffish DNA copies were significantly overrepresented (except for sample SW3) with an average deviation of 20% ±1 and 23% ± 27 in RO and swab samples, respectively (Figure 3A; Supplementary Table S11). Systematic errors were estimated to be -25% ±1 and -28% ± 32 for cod and 119% ±5 and 136% ± 158 for wolffish in RO and swab samples, respectively. There was little variation (SD±1) between estimates from the RO samples, illustrating their homogeneity. As expected, the swab method showed higher variation (SD±27), as the samples originated from different non-replicated swabbing patterns.

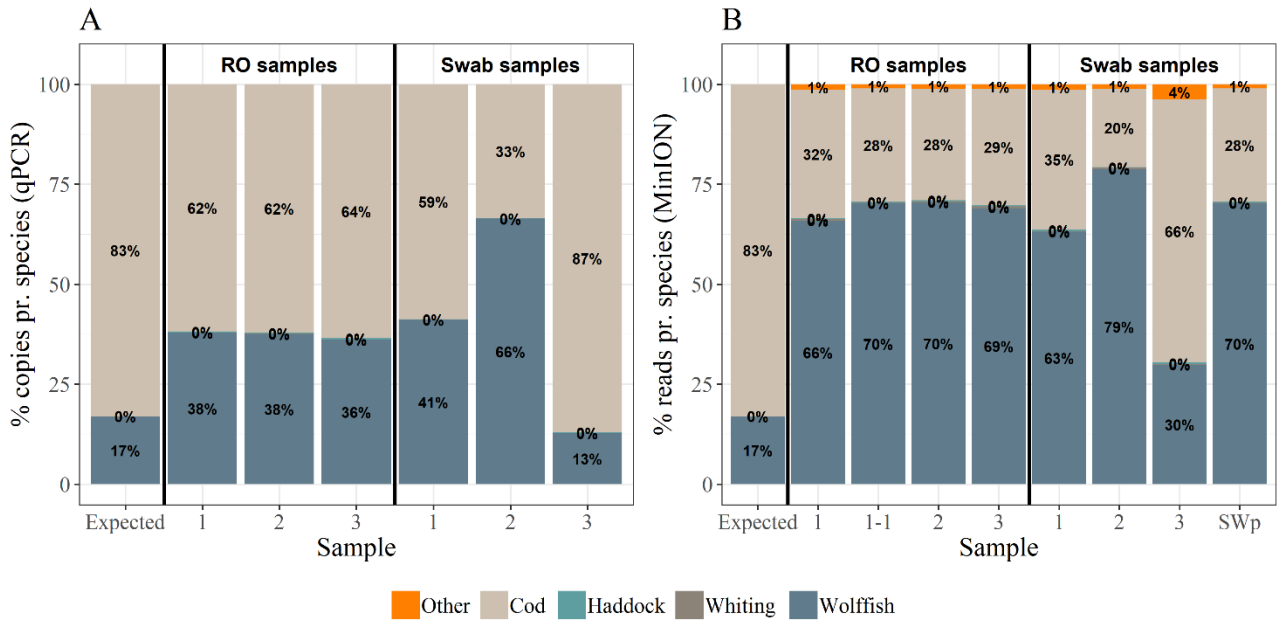

Figure 3: Input tissue weight percentages ('Expected') and estimated DNA proportions for fish block samples collected through run-off water (RO) and external surface swabs (swab). (A) qPCR estimates, (B) MinION metabarcoding estimates. In (B) sample 1-1 refers to analysis of sample replicate 1 on MinION-1.

4.5. **Metabarcoding of samples with MinION**

Selected samples were analysed with the metabarcoding approach on the MinION (Supplementary Table S1). Single species samples (see section 4.1) showed close to 100% read assignment to the sampled species. The highest proportion of reads assigned to other species was 0.42%; likely due to random sequencing errors (Jain *et al.*, 2016; Quick *et al.*, 2016). The mixed tissue samples showed deviations from input proptions of of -8% ±12, -13% ± 10 and 21% ± 17 with systematic errors averaging -17% ±42, -31% ±35 and 93% ±59 for C, H, W (Figure 4A). Similar to the qPCR analysis, the MinION also showed improved accuracy, 0% ± 3, 7% ±14 and -7% ±12, and lower systematic error, 9% ±24, 72% ± 121 and -18% ±26, for C, H, W when analysing mixed DNA samples (Figure 4B). All target species were detected in all mixed samples and were represented by a considerable proportion of the generated reads (>2% of the accumulated reads). Averaging across all

mixed control samples, the MinION metabarcoding approach showed deviations of -3% ± 10, -4% ± 14 and 7% ± 18 for C, H, W (Figure 4; Supplementary Table S12).

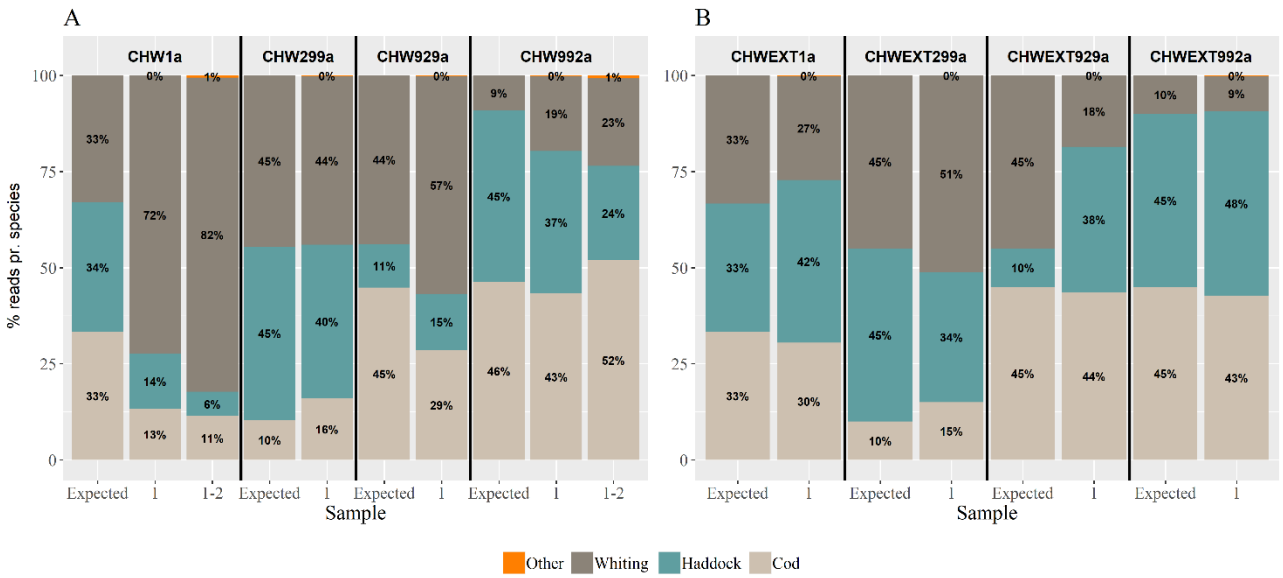

Figure 4: Comparison of tissue (A) and DNA (B) input percentages (Expected) with DNA proportions (reads) estimated with MinION based metabarcoding. All samples were analysed on MinION-1 (1), while some samples were analysed on both MinION-1 and MinION-2 (1-2).

Silage samples (day 2 and 21) analysed on the MinION, showed overrepresentation of wolffish DNA reads, similar to the qPCR copy number analysis (Supplementary Figure S3 and Table S13). The day 2 sample was analysed on both MinION-1 and MinION-2, revealing quantitatively different results from the same sample. For MinION-2 analysis, all gadoids were equally underrepresented, while only haddock and whiting were underrepresented for MinION-1. Altogether, MinION-2 metabarcoding provided proportion estimates similar to those of qPCR with a relatively lower inaccuracy, -10%, 13% and -3%, and systematic error, -16%, 41% and -47%, for C, H, W respectively.

Wolffish was also overrepresented in run-off and swab samples with an average deviation of 52% ± 2 and 43% ± 19 and a systematic error of 305% ± 11 and -27% ± 22 for RO and swab samples, respectively (Figure 3B; Supplementary Table S14). Consequently, cod was underestimated with deviations -54% ±2 and -46 ±17 and systematic errors of -65% ±2 and 119% ±102 for RO and swab samples, respectively. Similar to

for qPCR, the MinION metabarcoding also returned low sample variance (SD ±2) in analysis of the replicated RO samples.

5. Discussion

The present study demonstrates that DNA is a powerful tool for detecting and quantifying species contributions in complex fish samples, which can supplement or even replace visual inspection for MSC. DNA based methods are versatile and robust allowing quick and easy sample collection and analysis for a broad range of samples and species. The successful demonstration of high-throughput sequencing using the ONT MinIon suggests that such platforms can yield equivalent semi-quantitative results to those generated using traditional qPCR approaches, raising the possibility of developing diagnostic, laboratory-free testing of fish discard products. However, the results also revealed significant species-specific quantification bias and further development would be needed to prior to routine implementation of DNA methods for particular fisheries and products.

Relative quantification is a tug-of-war between DNA contributions among species. Hence, for relative quantification to be directly applicable, all individuals and species should contain similar numbers of DNA copies per weight of tissue. However, we found that control tissue samples showed relatively weak relationships between tissue weight and DNA copies available for both qPCR and MinION metabarcoding. Accordingly, for the single tissue type investigated (fin tissue), the relationship varied substantially among individuals and species. This variation may reflect natural variations in tissue mtDNA content among species and individuals, but could also reflect multiple technical factors, such as different DNA extraction and qPCR/metabarcoding efficiencies, as well as sample variation and degradation. The many potential sources of variance are highlighted by the finding of more accurate and precise estimates of contributions to mixed samples from mixing DNA than from mixing tissue by weight, for both qPCR and MinION. The improved accuracy from DNA mixed samples suggests that the ratio between mtDNA and nuclear DNA (nDNA) is relatively stable in the samples analysed, while variation in DNA content (mtDNA) among even relatively homogenous tissue samples, is a potentially important source of intra- and interspecific variation. We found

DNA copies of wolffish to be considerably overrepresented in silage and fish block samples for both qPCR and MinION metabarcoding, while the wolffish control samples derived from fin tissue had the lowest estimated number copies/mg of tissue. Thus, there was no straightforward link between tissue DNA content and DNA results in mixed samples. Other studies have found 5-10 fold variation in DNA content between tissue types (Hartmann *et al.*, 2011; Cole, 2016), suggesting that other tissue types, or more likely, proportions of different tissue types among species, can explain the disproportional number/weight of mtDNA copies in wolffish compared to gadoids. We speculate that wolffish in general contain more "active" tissue with higher respiratory needs, e.g. thick skin, explaining the elevated mtDNA copy number per unit weight of whole fish. Noticeably, DNA copies/reads among the gadoids varied much less and were more proportional to tissue input, especially in the silage samples. This may suggest that species that are closely related and morphologically similar also contain more similar mtDNA copies per weight tissue than unrelated species. Other studies of sister-species generally demonstrate proportional estimates for relative quantification (Lopez and Pardo, 2005; Bojolly *et al.*, 2017), whereas disproportional relationships are seen in mixtures with more distantly related species (Thomas *et al.*, 2014; Floren *et al.*, 2015). There is generally little evidence that unrelated species contain the same amount of mtDNA per weight tissue (Hartmann *et al.*, 2011; Floren *et al.*, 2015; Cole, 2016), which supports our observations of interspecific variance. A potential way to minimise this difference is by targeting nuclear DNA, instead of mtDNA, as each cell only contains one nDNA copy, but can contain many and variable numbers of mtDNA copies (Cole, 2016). Still, cell number per tissue weight may vary considerably (Kozłowski *et al.*, 2010). A more robust approach to this challenge would be to implement assay-specific correction factors, as has previously been successfully applied to minimize biasing factors (Thomas *et al.*, 2016; Vasselon *et al.*, 2018). Noticeably, a correction factor can account for all biases in concert, regardless of biological or technical origin. For the silage analysis, wolffish contributions appear to be estimated approximately four times higher than expected, suggesting a specific correction factor of ~0.25. Further studies are warranted in order to determine the local and global robustness of correction factors; our analysis suggests that appropriate correction factors would need to be calculated across specific analysis types and different species.

Despite reservations regarding estimation of tissue proportions from DNA suggested by the mixed tissue control samples, silage results for gadoids were encouraging, with high precision for determining relative proportions of starting tissue weight used in silage sample production. We hypothesize that the lower precision for mixed tissue samples may be due to the relatively higher sampling stochasticity. For the silage, the dissolved fish contributed to a homogeneous DNA pool, which provides a robust integrated DNA signal for the entire pool of organisms in the sample. Thus, we expect well-mixed commercial scale silage production to be more stable regarding DNA/tissue ratios than smaller mixtures. Similarly, the run-off samples provide an integrated signal of all fish from the frozen block, which have been in contact with the water, thereby likely providing a better representation the full content than the swab samples. Still, if the content of the fish block is not homogeneous, e.g. with different species compositions in centre and on surface, both run-off and swab samples are likely to provide only a crude assessment of species content. On the other hand, these methods may prove highly valuable as they are non-invasive and can be conducted without spoiling the content, in contrast to visual inspection, where fish blocks have to be thawed, fish identified and weighed for MCS purposes.

The MinION and qPCR analyses provided highly similar results for the qualitative and quantitative analysis of the control samples. The main difference was the additional occurrence of false positive species (other), associated with high read abundance (Supplementary Table S6). The MinION uses 3rd generation sequencing technology, which besides a number of positive aspects, also currently has a relatively high error rate (~10-20%) (Quick *et al.*, 2016). Thus, by chance a number of sequences may show higher affinity to species not present in the sample. In particular, higher throughput increases the risks of random sequencing errors (Jain *et al.*, 2016; Quick *et al.*, 2016), and a high error rate coupled with a relatively low identity threshold (85%) will generate false positive species identification. However, true- and false positive proportions are expected to be stabile regardless of throughput, thus false positives will remain at a very low rate.

For control and fish block samples the sequence error rate did not seem to be problematic, as long as species detection was based on a 2% minimum threshold of cumulated reads. In contrast, the MinION struggled with reliable detection of some included species (i.e. haddock and whiting represented <2% of the cumulated

reads) and for obtaining accurate biomass proportions in silage. We suspect that the difference between the two approaches is due to the targeted DNA fragment sizes. qPCR targeted DNA fragments between 72 and 129 bp whereas the MinION based metabarcoding approach targeted a 699 bp mtDNA fragment. Longer DNA fragments are in general rare in processed material, such as silage, and the analysis will be more stochastic when only few long molecules (>200bp) remain (Deiner *et al.*, 2017; Jo *et al.*, 2017; Piskata *et al.*, 2017). Smaller fragment, so-called "minibarcodes" would likely improve detection power, but there is a trade-off between minimizing amplicon length and taxonomic resolution (Shokralla *et al.*, 2015; Thomsen *et al.*, 2016). Clear taxonomic sequence distinction provided by long DNA barcodes is still vital for MinION based species assignment due to the system's high sequencing error rate and relative poor sequence performance with short target amplicons. Future application of "direct sequencing" that is independent of an initial taxon-specific PCR represents an appealing approach to avoiding both the targeted sequence analysis and amplification biases (Thomas *et al.*, 2016; Stat *et al.*, 2017; Fonseca, 2018). This approach can also be combined with MinION analysis known as "selective sequencing", where only predefined target sequences are processed (Loose *et al.*, 2016).

The general issues of contamination and sensitivity are important to address before implementing the techniques in fisheries MCS. Catching, handling and processing related to commercial fishing practices are far from sterile procedures with many possibilities for both natural and "technical" contamination. For example, many commercial fish species are predators, potentially with stomach contents including other MCS target species. Likewise, all exterior fish surfaces have potentially been in contact with other species likely leaving false-positive DNA traces. However, low level contamination in the the samples utilized here did not generally approach the normal limits of detection or quantification in the assays and we expect that fish present in the silage and fish blocks to swamp-out any trace contaminant species. Hence, setting a quantitative threshold (e.g. at ≤30 PCR cycles) and/or a ratio threshold (e.g. proportion of target to contaminant species DNA), while decreasing the effective sensitivity of an assay, would also provide more robust estimates of the true species composition of bulk fish products.

In conclusion, this study yielded some very encouraging results for the use of DNA-based product analysis to estimate the initial relative biomass of different fish species in processed discard products. It

represents a “proof of concept” rather than an exhaustive evaluation of all parameters of importance for robust species quantification relevant to all fisheries and products. Any future practical application would require significant refinement and calibration to be conducted, and methods would need to be formally validated to fully characterise the sensitivity, specificity and robustness of diagnostic tests. However, the DNA analysis platforms assessed here potentially form the basis of robust, standardized and cost effective methods to verify the species composition of complex bulk fish products, which would be of significant interest to the industry, wherever visual identification and quantification are not possible.

6. Acknowledgements

We are grateful to the following people for assistance and support for the laboratory experiments: Maj-Britt Jacobsen, Britta Sønderskov Pedersen, Dorte Meldrup, Lars Magnus Wulf Jacobsen & Clara Ulrich from DTU Aqua. Further, we thank Rasmus Bach Ebert from Copenhagen University for help with the Gm assay design. This work has been funded by the European Union’s Horizon 2020 research and innovation programme under grant no. 633680 (DiscardLess).

7. References

Batsleer, J., Hamon, K. G., van Overzee, H. M. J., Rijnsdorp, A. D., and Poos, J. J. 2015. High-grading and over-quota discarding in mixed fisheries. Reviews in Fish Biology and Fisheries, 25: 715–736.

Bojolly, D., Doyen, P., Le Fur, B., Christaki, U., Verrez-Bagnis, V., Grard, T., Périne, P., *et al.* 2017. Development of a qPCR Method for the Identification and Quantification of Two Closely Related Tuna Species, Bigeye Tuna (Thunnus obesus) and Yellowfin Tuna (Thunnus albacares), in Canned Tuna. Journal of Agricultural and Food Chemistry, 65: 913–920.

Cole, L. W. 2016. The Evolution of Per-cell Organelle Number. Frontiers in cell and developmental biology, 4: 85.

Deiner, K., Renshaw, M. A., Li, Y., Olds, B. P., Lodge, D. M., and Pfrender, M. E. 2017. Long-range PCR

allows sequencing of mitochondrial genomes from environmental DNA. Methods in Ecology and Evolution, 8: 1888–1898.

European Commission. 2013. REGULATION (EU) No 1380/2013 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 December 2013.

Floren, C., Wiedemann, I., Brenig, B., Schütz, E., and Beck, J. 2015. Species identification and quantification in meat and meat products using droplet digital PCR (ddPCR). Food Chemistry, 173: 1054–1058.

Fonseca, V. G. 2018. Pitfalls in relative abundance estimation using edna metabarcoding. Molecular Ecology Resources, 18: 923–926.

Grant, S. M., and Hiscock, W. 2014. Post-capture survival of Atlantic wolffish (Anarhichas lupus) captured by bottom otter trawl: Can live release programs contribute to the recovery of species at risk? Fisheries Research, 151: 169–176.

Guillen, J., Holmes, S. J., Carvalho, N., Casey, J., Dörner, H., Gibin, M., Mannini, A., *et al.* 2018. A review of the European union landing obligation focusing on its implications for fisheries and the environment. Sustainability, 10: 900.

Hartmann, N., Reichwald, K., Wittig, I., Dröse, S., Schmeisser, S., Lück, C., Hahn, C., *et al.* 2011. Mitochondrial DNA copy number and function decrease with age in the short-lived fish Nothobranchius furzeri. Aging Cell, 10: 824–831.

Heath, M. R., Cook, R. M., Cameron, A. I., Morris, D. J., and Speirs, D. C. 2014. Cascading ecological effects of eliminating fishery discards. Nature communications, 5: 1–8.

Jain, M., Olsen, H. E., Paten, B., and Akeson, M. 2016. The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. Genome Biology, 17: 239.

Jo, T., Murakami, H., Masuda, R., Sakata, M. K., Yamamoto, S., and Minamoto, T. 2017. Rapid degradation of longer DNA fragments enables the improved estimation of distribution and biomass using environmental DNA. Molecular Ecology Resources, 17: e25–e33.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., *et al.* 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28: 1647–1649.

Knudsen, S. W., Ebert, R. B., Hesselsøe, M. , Kuntke, F., Hassingboe, J., Mortensen, P. B., Thomsen, P. F., *et al.* 2018. Species-specific detection and quantification of environmental DNA from marine fishes in the Baltic Sea. Journal of Experimental Marine Biology and Ecology, 510: 31–45.

Kozłowski, J., Czarnołęski, M., François-Krassowska, A., Maciak, S., and Pis, T. 2010. Cell size is positively correlated between different tissues in passerine birds and amphibians, but not necessarily in mammals. Biology Letters, 6: 792–796.

Larsen, E. P., Dalskov, J., Nielsen, E. E., Kirkegaard, E., Nielsen, J. W. N., Tørring, P., and Schou, M. 2013. Dansk fiskeris udnyttelse af discardforbuddet.

Loose, M., Malla, S., and Stout, M. 2016. Real-time selective sequencing using nanopore technology. Nature Methods, 13: 751–754.

Lopez, I., and Pardo, M. A. 2005. Application of Relative Quantification TaqMan Real-Time Polymerase Chain Reaction Technology for the Identification and Quantification of Thunnus alalunga and Thunnus albacares. Journal of Agricultural and Food Chemistry, 53: 4554–4560.

Menegon, M., Cantaloni, C., Rodriguez-Prieto, A., Centomo, C., Abdelfattah, A., Rossato, M., Bernardi, M., *et al.* 2017. On site DNA barcoding by nanopore sequencing. PLOS ONE, 12: e0184741.

Mikheyev, A. S., and Tin, M. M. Y. 2014. A first look at the Oxford Nanopore MinION sequencer. Molecular Ecology Resources, 14: 1097–1102.

Mikkelsen, P. M., Bieler, R., Kappner, I., and Rawlings, T. A. 2006. Phylogeny of Veneroidea (Mollusca: Bivalvia) based on morphology and molecules. Zoological Journal of the Linnean Society, 148: 439–521.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., *et al.* 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. Royal Society Open Science, 2: 150088.

Nagase, M., Yi, R., Hidaka, F., Maeta, K., Aimi, T., Yamaguchi, T., Suginaka, K., *et al.* 2010. Quantification of relative flying fish paste content in the processed seafood ago-noyaki using real-time PCR. Fisheries Science, 76: 885–892.

Piskata, Z., Pospisilova, E., and Borilova, G. 2017. Comparative study of DNA extraction methods from fresh

and processed yellowfin tuna muscle tissue. International Journal of Food Properties, 20: S430–S443.

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., *et al.* 2016. Real-time, portable genome sequencing for Ebola surveillance. Nature, 530: 228–232.

Sánchez, A., Quinteiro, J., Vázquez, J. A., Perez-Martín, R. I., and Sotelo, C. G. 2019. Comparison of real-time PCR methods for quantification of European hake (Merluccius merluccius) in processed food samples. Food Chemistry, 272: 279–285.

Shokralla, S., Hellberg, R. S., Handy, S. M., King, I., and Hajibabaei, M. 2015. A DNA Mini-Barcoding System for Authentication of Processed Fish Products. Scientific Reports, 5: 15894.

Spens, J., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S. T., Sigsgaard, E. E., *et al.* 2017. Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. Methods in Ecology and Evolution, 8: 635–645.

Srivathsan, A., Baloğlu, B., Wang, W., Tan, W. X., Bertrand, D., Ng, A. H. Q., Boey, E. J. H., *et al.* 2018. A MinION^{TM}-based pipeline for fast and cost-effective DNA barcoding. Molecular Ecology Resources, 18: 1035–1049.

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., Harvey, E. S., *et al.* 2017. Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. Scientific Reports, 7: 12240.

Tanabe, S., Hase, M., Yano, T., Sato, M., Fujimura, T., and Akiyama, H. 2007. A Real-Time Quantitative PCR Detection Method for Pork, Chicken, Beef, Mutton, and Horseflesh in Foods. Bioscience, Biotechnology, and Biochemistry, 71: 3131–3135.

Thomas, A. C., Jarman, S. N., Haman, K. H., Trites, A. W., and Deagle, B. E. 2014. Improving accuracy of DNA diet estimates using food tissue control materials and an evaluation of proxies for digestion bias. Molecular Ecology, 23: 3706–3718.

Thomas, A. C., Deagle, B. E., Eveson, J. P., Harsch, C. H., and Trites, A. W. 2016. Quantitative DNA metabarcoding: improved estimates of species proportional biomass using correction factors derived from control material. Molecular Ecology Resources, 16: 714–726.

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., and Willerslev, E. 2016.

Environmental DNA from Seawater Samples Correlate with Trawl Catches of Subarctic, Deepwater Fishes. PLOS ONE, 11: e0165252.

Ushio, M., Murakami, H., Masuda, R., Sado, T., Miya, M., Sakurai, S., Yamanaka, H., *et al.* 2018. Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. Metabarcoding and Metagenomics, 2: e23297.

Vasselon, V., Bouchez, A., Rimet, F., Jacquet, S., Trobajo, R., Corniquel, M., Tapolczai, K., *et al.* 2018. Avoiding quantification bias in metabarcoding: Application of a cell biovolume correction factor in diatom molecular biomonitoring. Methods in Ecology and Evolution, 9: 1060–1069.

Wen, J., Zeng, L., Sun, Y., Chen, D., Xu, Y., Luo, P., Zhao, Z., *et al.* 2015. Authentication and traceability of fish maw products from the market using DNA sequencing. Food Control, 55: 185–189.