LJMU Research Online

Dawnay, N, Hughes, R, Court, DS and Duxbury, N

Species detection using HyBeacon(®) probe technology: Working towards rapid onsite testing in non-human forensic and food authentication applications.

http://researchonline.ljmu.ac.uk/id/eprint/2716/

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Dawnay, N, Hughes, R, Court, DS and Duxbury, N (2015) Species detection using HyBeacon(®) probe technology: Working towards rapid onsite testing in non-human forensic and food authentication applications. Forensic Science International: Genetics, 20. pp. 103-111. ISSN 1878-0326

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

http://researchonline.ljmu.ac.uk/

Title: Species Detection using HyBeacon® Probe Technology: Working Towards Rapid Onsite Testing in Non-Human Forensic and Food Authentication Applications.

Nick Dawnay[ab*], Rebecca Hughes[c], Denise Syndercombe Court[c], Nicola Duxbury[a],

[a]*Product Development Group, LGC Forensics, Culham Science Centre, Abingdon, OX14 3ED.*

[b]*School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF*

[c]*Department of Pharmacy and Forensic Science, King's College London, Faculty of Life Sciences and Medicine, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH*

*Corresponding author: Tel +44 844 2641 999.

E-mail address: paradna@lgcforensics.com

Abstract

Identifying individual species or determining species' composition in an unknown sample is important for a variety of forensic applications. Food authentication, monitoring illegal trade in endangered species, forensic entomology, sexual assault case work and counter terrorism are just some of the fields that can require the detection of the biological species present. Traditional laboratory based approaches employ a wide variety of tools and technologies and exploit a number of different species specific traits including morphology, molecular differences and immuno-chemical analyses. A large number of these approaches require laboratory based apparatus and results can take a number of days to be returned to investigating authorities. Having a presumptive test for rapid identification could lead to savings in terms of cost and time and allow sample prioritisation if confirmatory testing in a laboratory is required later. This model study describes the development of an assay using a single HyBeacon® probe and melt curve analyses allowing rapid screening and authentication of food products labelled as Atlantic cod (*Gadus morhua*). Exploiting melt curve detection of species specific SNP sites on the COI gene the test allows detection of a target species (Atlantic cod) and closely related species which may be used as substitutes. The assay has been designed for use with the Field Portable ParaDNA system, a molecular detection platform for non-expert users. The entire process from sampling to result takes approximately 75 minutes. Validation studies were performed on both single source genomic DNA, mixed genomic DNA and commercial samples. Data suggests the assay has a lower limit of detection of 31 pg DNA. The specificity of the assay to Atlantic cod was measured by testing highly processed food samples including frozen, defrosted and cooked fish fillets as well as fish fingers, battered fish fillet and fish pie Ninety-six (92.7 %) of all Atlantic cod food products tested provided a correct single species result with the remaining samples erroneously identified as containing non-target species. The data shows that the assay was quick to design and characterise and is also capable of yielding results that would be beneficial in a variety of fields, not least the authentication of food.

51 Key Words: Species identification, food authentication, non-human forensics, on-site testing, food
52 crime.

Introduction.

The application of forensic DNA techniques to non-human species is increasingly prevalent in today's legal system. It is used to support or refute prosecution or defence hypotheses in areas as wide ranging as murder [1], food safety [2], sexual assault [3] and illegal animal killing [4]. The forensic analysts in this field are routinely tasked with answering four broad questions. Firstly, *what species is present in the unknown sample?* (species identification); secondly, *how much of the species is present in the unknown sample?* (species or species' quantification); thirdly, *what area did the species originate from?* (species provenance) and finally; *what is the probability that another individual member of the same species could have left the crime scene stain?* (individual identification) [5].

Species identification is the most common question asked in non-human forensics. Techniques used to identify an individual organism to the species level are broad and include enzyme-linked immunosorbent assays (ELISA) [6], Raman spectroscopy [7], matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI TOF) [8] and DNA-based methods [5]. DNA-based approaches are often preferred as they can offer a more robust approach. They tend to have high sensitivity due to the use of polymerase chain reaction (PCR), high specificity to the chosen target and can be used on highly processed samples, many of which have been exposed to high temperatures [9]. Common DNA techniques using PCR include PCR-restriction fragment length polymorphisms (PCR-RFLP), amplified fragment length polymorphisms (AFLP), forensically informative nucleotide sequencing (FINS), random amplified polymorphic DNA (RAPD), melt curve analyses [10, 11] and DNA sequencing [12]. A lack of governance and standardisation relating to species identification in food standards means that each laboratory often develops ad-hoc approaches, many of which have been phased out of use in routine forensic applications. However, DNA sequencing is often considered the gold standard due to the ability to detect and clearly identify a large number of species specific single nucleotide polymorphisms (SNPs) in the gene regions tested [12,13].

Perhaps one of the best known uses of non-human forensic genetic techniques today is in the detection of food fraud, defined by Europol and Interpol as *'the deliberate placing on the market, for financial gain, foods which are falsely described or otherwise intended to deceive the consumer'* [14]. Food authenticity and food safety testing is carried out on an international scale by a number of government and private testing laboratories [15, 16, 17]. Of recent concern is fisheries food fraud of which there are 7 distinct forms: species substitution; fishery substitution; illegal, unreported and unregulated (IUU) substitution; species adulteration; chain of custody abuse; catch method fraud; and undeclared product extension [14]. An increase in global seafood consumption has led to a rise in product mislabelling, and its effects are far reaching [17]. In 2009, a fifth of the annual global seafood catch came from IUU fishing [18]. Not only of interest to food standards groups, IUU fishing is also of major concern in conservation with 60 % of marine stocks at their maximum sustainable threshold [19]. With limitations in the current regulatory systems it is suspected that much of the illegal fishing is going undetected.

Molecular techniques have been developed to enable the detection of illegally fished species [6, 15, 16, 17]. The detection of fishery and stock provenance is the subject of ongoing work [20, 21]. The large majority of these applications remain laboratory based which can be time consuming, costly and require expert analysis of the results. The development of a fast, reliable, user friendly testing kits with the capability of being taken out of the lab and into the field for rapid deployment, monitoring and sample prioritisation may allow an increased rate of screening and detection. There are a large number of recently launched portable devices that allow testing of forensic samples outside the laboratory. Many of these systems are based on familiar laboratory platforms, either utilising capillary electrophoresis (CE) based detection [22], microarray [23] or utilise melt curve detection of PCR amplicons [24, 25]. The ParaDNA System is one of these detection platforms. It exists as a standalone instrument with attached laptop, or as a battery operated field portable unit. Template material is collected using a plastic sample collector (analogous to a traditional swab) and inserted directly into PCR wells containing the required assay mix pre-loaded and ready for use (Figure 1). The system utilises a direct PCR approach meaning there is no need to purify template material and offers automatic identification for non-expert users which involves characterising the change in Relative Fluorescence Units (ΔRFU) as the

HyBeacon probe melts away from the target [26]. The basis for PCR product differentiation using HyBeacon melt curve analyses is that greater homology between target and HyBeacon probe confers greater stability requiring a higher melting temperature before the target and probe disassociate. Therefore an application such as forensic species identification may also be performed using this approach providing there is high intra-species homology for the probe sequence and a degree of inter-species variability at the probe site between target and non target species.

Here we present the development of an Atlantic Cod (*Gadus morhua*) specific assay for use in conjunction with the ParaDNA System. High value atlantic cod is often suspected of being replaced with species of a lesser value. A rapid means of detecting these species substitution events would be beneficial. The mitochondrial COI gene was selected and a species specific HyBeacon probe was designed to bind to the target species. Experiments were designed to characterise the test sensitivity, accuracy and robustness. The data presented in this study aims to show the utility of HyBeacon technology for species identification using the ParaDNA system.

Materials and methods

Sample authentication – DNA Sequencing

Samples were sourced from local stores and markets and authenticated for use in the HyBeacon assay development work by DNA sequencing. DNA was extracted from 42 fish samples, representing 15 different species (Table 1) using the QIAGEN DNeasy Blood and Tissue kit. Extracts were then quantified using Nanodrop and amplified using COI universal primers obtained from Ward et al 2007 [27] and Ivanova et al 2005 [28];

VF2_t1 (TGTAAAACGACGGCCAGTCAACCAACCACAAAGACATTGGCAC),

FishF2_t1 (TGTAAAACGACGGCCAGTCGACTAATCATAAAGATATCGGCAC),

FishR2_t1: (CAGGAAACAGCTATGACACTTCAGGGTGACCGAAGAATCAGAA),

FR1d_t1: (CAGGAAACAGCTATGACACCTCAGGGTGTCCGAARAAYCARAA).

PCR amplification was performed on a BioRad CFX96 Real-Time PCR Detection System. After PCR, samples were visualised on gels (1 x TBE Buffer & EtBr, Lonza reliant gel), purified using the

QIAquick PCR purification kit and re-run out on an agarose gel. Purified samples were amplified using BigDye® Terminator v3.1 Cycle Sequencing Kit (LifeTechnologies) using the LabCycler (SensoQuest). CE was performed on an ABI 3730xl DNA Analyzer XL (LifeTechnologies) using polymer pop7. All successfully sequenced COI amplicons were visualised using Chromas 2.4.3 (Technelysium Pty Ltd) to verify base calling. Species identification was then confirmed and performed by sequence similarity searches using GenBank BLASTn (http://www.ncbi.nlm.nih.gov/BLAST/) for all 42 samples (Table 1).

HyBeacon Assay Design

Atlantic cod and closely related species' sequence data were downloaded from both BOLD and NCBI databases and aligned using Clustal 2.0 [29]. Hybeacon probes are generally 20-30bp in length and short regions of COI homology were identified in Atlantic cod with putative species specific SNP sites identified for key target species (Table 2) using Mega 6.0 [30]. Multiple rare haplotypes were observed in Alaskan Pollock (*Gadus chalcogrammus*) suggesting the melt curve for individuals with these sequences would show variation in melt temperature within the species, while a single haplotype was observed in Pacific cod (*Gadus macrocephalus*). However, there remained a large number of individuals (93 %) showing the common haplotype. The predicted melt Tm suggests that there exists the potential for multiple non-target species showing the same melt curve transition. Therefore, each of the three predicted melt curves observed in this study were labelled A (Atlantic cod specific), B (including adulterants Alaskan Pollock, Norway Pollock, Arctic cod) and C (including adulterants Pacific cod, Greenland cod and Alaskan Pollock) (see Table 2).

Sequences were queried against GenBank using the Blastn search tool to identify if any other known species show homology at the selected Atlantic cod probe binding site. Aside from two unrelated species (South Mandarin dogfish, *Cirrhigaleus australis,* and a bee, *Hylaeus strenuus*) no other 100 % matching regions were observed. However it was considered unlikely that these two species will amplify given the multiple mis-matches observed at the primer binding sites (data not shown). Primer sites were identified that showed cross-species amplification in the closely related species. To generate

an excess of the DNA strand complementary to the probe, an asymmetric PCR approach was used with the reverse primer (5'-CCAGAGGATGCTAAAAGGAGCAGGAAA-3') in excess of the forward primer (5'-TGGAGGCTTTGGGAACTGACTCATT-3'). This assay design results in a 126bp COI amplicon. The thermocycling parameters are not end user configurable and were based on those developed for the ParaDNA System.

Training data was generated in the laboratory by PCR amplification of a large and diverse panel of samples that should be representative of the wider population of samples encountered. The assay successfully produced three species-specific melt curves (Figure 2) with the following melt temperatures (Melt curve A Tm=60.04, SD=0.49; Melt curve B Tm=57.77, SD=0.45, Melt curve C Tm=53.90, SD=0.46). A t-test showed each species mean melt temperature differs significantly from that nearest to it ($p < 0.001$) suggesting that each melt curve is highly differentiated. Modelling each sample melt curve against positive and negative data allows the specific melt curve ΔRFU to be calculated and plotted. The ParaDNA software provides a ΔRFU measure for all three melt targets (A, B, C) for every sample analysed. Due to the specific SNP differences located in each species' target sequence it is expected that a single source sample would generate a large ΔRFU at its predicated melt temperature and also have very low ΔRFU values at the other melt temperatures (categorised as noise). Thresholds were set to differentiate signal to noise at the three different melt temperatures allowing further characterisation of the training data.

HyBeacon Assay Development

For the development of the cod specific assay all training data was generated using quantified genomic DNA. Sensitivity and specificity studies were performed to generate an appropriate range of training data to model each representative melt curve (A, B, C). Sensitivity data used to measure the limit of detection was collected by performing a serial dilution of genomic DNA from Atlantic cod (Melt curve A), Alaskan Pollock (melt curve B) and Pacific cod (melt curve C). DNA was added to the reaction mix prior to PCR in the following amounts; 5 ng, 2.5 ng, 1.0 ng, 750 pg, 500 pg, 250 pg, 125 pg, 62.5 pg, 31 pg, 15 pg and 7.5 pg. These inputs indicate the total amount of DNA added to the assay. Eight

replicates were performed for each of the 11 DNA input level for each of the three tested species, including 8 no-template control samples in each species group (288 measurements in total). Species specificity data was also collected by amplifying 500 pg genomic DNA from four replicates of each of the 42 samples listed in table 1 with 12 no-template control samples (180 measurements). Data were analysed using ParaDNA Batch Processor software, a software tool used to set data analyses parameters and calling thresholds. Key metrics, including melt temperature and width of the melt transition were set to allow the software to differentiate each species melt curve. Thresholds for automatic calling were subsequently set by observing plots of the ΔRFU values derived from each sample. A measure of the test sensitivity and specificity for each melt curve peak was calculated following the method of Altman and Bland [31]. This involves characterising all results as either true positive (target species detected in sample known to contain target), true negative (target species not detected in sample known to be absent of target), false positive (target species detected in sample known to be absent of target) and false negative (target species not detected in sample known to contain target). Instances of single source samples erroneously identified as mixtures were also reported.

HyBeacon Assay Application

The sensitivity of the HyBeacon assay to sample adulterants was investigated through the analyses of mixed genomic DNA samples. Extracted Atlantic cod DNA was mixed with both Alaskan Pollock or Pacific cod DNA in the following ratios; 100:0, 90:10, 70:30, 50:50, 30:70, 10:90 and 0:100. Four replicates at three different total DNA template concentrations (1 ng, 500 pg and 250 pg total DNA) were tested. To investigate the everyday application of the assay fish samples purchased from local stores and markets were tested using the ParaDNA platform. Eight samples of Atlantic cod and Alaskan Pollock fillets were sampled from when frozen, defrosted and when cooked following the manufacturer's instructions (oven at 200°C for 30 minutes). Frozen and defrosted Pacific cod fillets were also analysed. Additional fish products representing different levels of processing were also tested (battered fillets, fish fingers and fish pie). All were defrosted and uncooked at time of sampling. In total, 192 samples were collected.

The ParaDNA system offers two methods of sample collection and both methods were assessed in this study. Firstly direct sampling, which involved the sample collector being scratched directly across the fish fillet for 30 seconds. Secondly indirect sampling, which involved recovery of template material using a cotton swab before the swab itself underwent sub-sampling for 1 minute. Care was taken during sampling to access the centre of the fish tissue and thereby avoid the possibility of surface contamination. ΔRFU values for each species were extracted and statistical tests were then performed. Data were tested for normality (Anderson-Darlings test) and variance (Levene's test). Analysis of variance (ANOVA), t-tests and Mann Whitney U tests were performed. The test sensitivity and specificity was again calculated based on the performance with complex food samples and compared to the results obtained from genomic DNA.

Results and discussion

HyBeacon Assay Development

The assay displayed a high level of sensitivity with detection of each species down to 7.5 pg of template DNA. The lower limit of detection (LOD) for each melt peak was determined when a significant difference (t-test $p<0.05$) was observed between the ΔRFU of the non-template control (NTC) samples and the samples containing DNA. This analysis shows the assay has a LOD of 7.5 pg for melt curve B (Alaskan Pollock and other adulterants), 15 pg for melt curve B (Pacific cod and other adulterants) and 31 pg for melt curve A (Atlantic cod specific) suggesting all targets have a high level of sensitivity. Analyses of the primer annealing sites in the other adulterants detected by melt curves B and C show a high degree of similarity suggesting they could also show a similar level of sensitivity.

The accuracy of identification was assessed by observing the total number of correct calls from the sensitivity data (after removal of samples below the identified LOD) and the species specificity data combined (Table 3). Sensitivity {true positive/(true positive + false negative)} and specificity {true negative/(true negative + false positive)} were calculated [31]. This showed that the assay has a test *sensitivity* of 93.4 % for Atlantic cod (melt curve A). Four of the 127 true positive samples observed (3 %) were identified as an Atlantic cod/Alaskan Pollock mixture. If these samples were categorised as

false negatives (due to the lack of a single species result) the sensitivity drops to 90.4 %. The detection of Atlantic cod also showed a *specificity* of 96.3 % i.e. when Atlantic cod was absent in an unknown source sample the test correctly identified it as absent 96 % of the time. Melt curve B (Alaskan Pollock and adulterants) has a *sensitivity* of 97.8 % and a *specificity* of 98.9 % with seven of the 133 true positive samples (5.3 %) identified as an Atlantic cod/Alaskan Pollock mixture. If these samples were categorised as false negatives the sensitivity drops to 92.7 %. Melt curve C (Pacific cod and adulterants) has a *sensitivity* of 95.5 % and a *specificity* of 99.7 % with one of the 126 single source Pacific cod samples (0.8 %) identified as an Atlantic cod/Pacific cod mixture. If this sample was categorised as false negatives the sensitivity drops to 94.7 %.

The identification of mixtures in single source Atlantic cod and Alaskan Pollock samples is likely due to the close Tm values seen between melt curve A and melt curve B. The single instance of mixture detection seen in melt curve C was due to an occurrence of unexplained noise in one sample. None of the non-target species tested showed repeatable amplification or melt curve detection in the same regions as the three targets characterised in this study. This is due to a lack of homology between the primers and the species tested here. A diagnostic peak for Haddock (*Melanogrammus aeglefinus*) amplified and produced a melt peak however it was seen at a lower T_m (45°C) and was therefore not miscalled. ANOVA showed no significant difference between the ΔRFU of each melt curve (A, B, C) between the non-target species tested and the NTC samples (melt curve A, $p=0.667$; B, $p=0.409$; C, $p=0.737$). Pooling of this datum revealed a significant difference ($p<0.001$) in the comparison to the ΔRFU when the three target species were tested. This shows that the single HyBeacon probe can identify the presence of unknown samples as containing Atlantic cod with a high level of accuracy.

As the ParaDNA software provides a ΔRFU measure for all three melt targets (A, B, C) more than one species can be detected in a sample at the same time. The ability to detect numerous species compositions in a single sample may allow identification of food adulteration, separate from the species replacement application outlined above. However, it is important to recognise that adulteration can occur during food processing simply due to the handling of multiple species in a production line.

Therefore adulterants have to be set below a statutory limit to differentiate between minor adulteration due to the use of common processing equipment and the deliberate placing of lesser value species into the product together with the true species. Mixed samples of both Pacific cod/Atlantic cod and Alaskan Pollock/Atlantic cod were analysed using the ParaDNA assay and software with threshold set based on the training datum described. The automatic output provided by the ParaDNA system does not currently provide a measure of each melt curve specific ΔRFU, although the data is accessible using the training software. Analyses of the Pacific cod/Atlantic cod mixtures show a strong relationship between the ΔRFU of the major and minor contributors (figure 3a). ANOVA revealed significant differences in the ΔRFU of melt curve A at different mixture ratios ($p<0.001$) suggesting the assay can detect different levels of Atlantic cod in each mixture level tested. ANOVA also showed a significant difference in the ΔRFU of melt curve C at different mixture ratios ($p<0.001$) suggesting the assay can detect different levels of Pacific cod in each mixture level tested. While still significant (ANOVA $p=<0.001$) the same relationship was not as clearly defined in the Atlantic cod:Alaskan Pollock mixed samples (Figure 3b) as all the contributions from each species show little difference between the 70:30 0:100 ratios. This is due to the smaller Tm difference seen between these two species melt curves which increases the difficulty of attributing the detectable fluorescence between each of the species. It is likely that the identification of SNP sites that allow a greater difference in Tm between these species would allow more discrimination between different mixture levels. It is the author's opinion that further development and optimisation is needed before this approach could be used to accurately identify the presence of multiple species in a food source. Such work should include a concordance study looking to correlate the results of the described Hybeacon approach against a more traditional approach such as cloning or Massively Parallel Sequencing to understand the true composition of the amplified product and thereby measure the accuracy of this novel approach.

HyBeacon Assay Application

Detecting instances of food fraud is currently performed by individuals from a variety of different sectors including food standards agencies and industry groups with those involved utilising a variety of different schemes and methods to monitor and police these illegal activities. These include routine

sample testing in the laboratory [32], chain of custody documentation and product certification [33], all of which make it more difficult for substitution and food-mislabelling to take place. The recent creation of a Food Crime Unit in the UK will further support and strengthen the activities of these groups. However, food substitution events may remain difficult to detect due to the difficulty in identifying processed food samples [14] and the lack of basic monitoring equipment. We believe the development of a rapid portable method of verifying food claims will be a useful tool. Not only would it allow seized items to be screened onsite prior to confirmatory testing in a laboratory it also has the potential to allow a greater number of samples to be investigated allowing for better monitoring. This technology for a non-expert user should be easy to use, offer simple analyses and provide savings either in terms of money and/or time. Traditional laboratory analyses often take a number of days and weeks to return a result due to the number of steps required to process a sample. Sometimes the result may not confirm the original suspicion leading to lost time while waiting for the laboratory result and the wasted cost of running the analyses. In instances of food monitoring, quality control and large scale food fraud it is often necessary to collect a large number of samples for analyses. The ability to prioritise which samples to send for confirmatory testing is something that can lead to a greater number of food samples being tested and has the potential to lead to a higher instance of positive detection [34]. Such applications will be required to work on a number of different sample types which may be highly processed or degraded.

The ability of the HyBeacon assay to directly amplify the template DNA from a variety of different seafood products was assessed using the ParaDNA System. Our data shows template material was successfully recovered from all samples types using both the direct and indirect sample collection methods. There was weak statistical support for improved sample recovery when using the indirect sampling method. The same improved performance was observed in each species (Atlantic cod $p=0.011$, Alaskan Pollock $p=0.013$, Pacific cod $p=0.029$) suggesting the result is genuine. During the sampling process it was observed that the direct sampling approach often removed visible tissue from the sample which was then inserted directly into the assay for PCR. Such overloading of template could lead to reduced amplification if impurities and PCR inhibitors are also transferred. However, given the weak statistical support and the fact that direct sampling also provided good quality results, both

sampling strategies are considered an appropriate method for recovering DNA from the food samples tested. PCR amplification was observed across all samples types (Figure 4, direct collection data not shown). Of the 96 Atlantic cod samples analysed four were erroneously identified as Alaskan Pollock (4.2 %) - which is higher than that measured in the training data (Table 3). However, of the 64 Alaskan Pollock samples, nine returned results that did not conform to those expected (1 %). Five samples were identified as Atlantic cod, three samples were identified as Alaskan Pollock/Atlantic cod mixtures and one sample failed to amplify. Of the 32 Pacific cod samples all were correctly identified with one sample failing to amplify. The erroneous samples were spread through the sample types tested, although there was a higher number seen in the Pollock fish pie (50 % of erroneous Pollock calls) and the battered cod (57 % of erroneous Atlantic cod calls). This suggests that there may be an artefact within the matrix that subtly alters the melting Tm or that optimisation of the sampling procedure is necessary. Possible strategies to correct for small variations in melt temperature include the inclusion of an Internal Positive Control (IPC). Such a mechanism would allow the melt curve Tm to be set based on a HyBeacon probe dissociating from the amplified IPC product theoretically allowing the software to correct for shifts in Tm brought about through the addition of unknown sample additives. Despite the observed erroneous calls, direct amplification of template material and rapid identification of correct species was observed in a large number of samples. The robust amplification maybe in part due to the short 126 bp amplicon withstanding some of the degradation and inhibition processes likely to occur during sample processing. Given that HyBeacon probes are typically 20-30 bp long it is possible to design extremely short amplicons (~60-80 bp) which would be expected to confer further advantages if degradation or sample inhibition is expected [35].

The required accuracy of a test is largely dependent on its application and whether for presumptive or confirmatory testing. Presumptive tests often display limitations in accuracy but remain useful either due to their speed, ease of use or cost. Therefore the authors feel that in its current form the assay would be suitable for on-site presumptive testing, field monitoring and also sample prioritisation when a large number of food items need investigating. If increased test sensitivity and specificity is required, the addition of multiple probes detecting additional sequence variation within the same or an alternative

gene would be a straightforward development. This will increase the discriminatory ability of the system as it would not only confirm the single marker result but also allow a greater number of non-target species to be differentiated. The required power is therefore configurable depending on the application, with a minimum of two gene probes being sufficient for most applications as recommended in [36]. While this approach provides certain benefits to non-expert users in its speed and ease of use, the data generated is a simple confirmation of the presence of a SNP in a sequence and caution should be used when designing assays and interpreting the outcome. Depending on the intended application further confirmatory species testing may be required though traditional approaches such as DNA sequencing. The bespoke PCR consumables utilised by the ParaDNA System has four wells for independent amplification and four channel detection capabilities allowing up to 16 different HyBeacon probe combinations. The study described here only utilised a single channel and well allowing three additional wells and channels for increasing the discriminatory power if required.

Conclusions

Product mislabelling and species substitution is known to occur on an international scale, however the extent of the practice is currently unknown. The use of a field portable non-expert user device could allow improved monitoring and sample prioritisation in large food fraud investigations. The application detailed here not only allows identification of the target species, but can also detect adulterants in the form of closely related species. The use of a field portable technology would allow better monitoring and detection of species in a wide variety of food and forensic investigations including the detection of illegally traded endangered species in traditional Asian medicines at airports, the identification of blowfly species at a crime scene to infer time of death, and also the detection of food fraud as highlighted here. The ParaDNA system is used by non-expert users outside a laboratory. The analysis requires the collection of training data based on the relevant species and possible substitutions or alternatives to optimise the automatic species detection criteria. As such, discussion with non-expert user groups interested in employing such technology is essential to ensure that the correct forensic assay is developed. The data presented here show it is possible to rapidly design and characterise a field portable assay that can distinguish between three closely related species using HyBeacon probe technology. Low

limits of detection were recorded together with the ability to recover DNA from a variety of sample types suggesting the assay is both sensitive and robust. We have also shown there is some potential for misidentification between target species where only a single SNP is required for differentiation. However, there are a number of optimisation strategies open for increasing the discriminatory power of the assay suggesting that future work on this and other assays may lead to a truly diagnostic result rather than a presumptive screen as detailed here.

Acknowledgments

This work was carried out as part of a project placement by Rebecca Hughes in partial fulfilment of her MSc in Forensic Science, Department of Pharmacy and Forensic Science, King's College London. Thanks to the ParaDNA team at LGC Forensics Culham for all their help and support throughout this project. Special thanks go to Paul Rendell for software support and to Simon Wells and Victoria Moore for reviewing the manuscript.

References.

[1] Detectives solved murder case using dog DNA.
http://news.bbc.co.uk/1/hi/england/london/8559319.stm. Accessed on 11.09.2014

[2] Severgnini M, Cremonesi P, Consolandi C, De Bellis G, Castiglioni B (2011) Advances in DNA microarray technology for the detection of food borne pathogens. *Food and Bioprocess Technology* **4**: 936-953.

[3] Giampaoli S, Berti A, Valeriani F, Gianfranceschi G, Piccolella A, Buggiotti L, Rapone C, Valentini A, Ripani L, Spica VR (2012) Molecular identification of vaginal fluid by microbial signature. *Forensic Sci Int Genet* **6**: 559–564.

[4] Caniglia R, Fabbri E, Greco G, Galaverni M, Randi E (2010) Forensic DNA against wildlife poaching: Identification of a serial wolf killing in Italy. *Forensic Sci Int Genet* **4**: 334-338.

[5] Johnson RN, Wilson-Wilde L, Linacre A (2014) Current and future directions on DNA in wildlife forensic science. *Forensic Sci Int Genet* **10**: 1-11.

[6] Asensio L, Gonzalez I, Garcia T, Martin R (2008) Determination of food authenticity by enzyme-linked immunosorbent assay (ELISA). *Food Control* **19**: 1-8.

[7] McLaughlin G, Doty KC, Lednev IK (2014) Discrimination of human and animal blood traces via Raman spectroscopy. *Forensic Sci Int* **238**: 91–95.

[8] De Bruyne K, Slabbinck B, Waegeman W, Vauterin P, De Baets B, Vandamme P (2011) Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. *Systematic and Applied Microbiology* **34**: 20–29.

[9] Lenstra JA (2003) DNA methods for identifying plant and animal species in food. *Food authenticity and traceability*. Cambridge, U.K.: Woodhead Publishing Ltd. p 34–53.

[10] Berry O, Sarre SD (2007) Gel-free species identification using melt-curve analysis. *Molecular Ecology Notes* **7**: 1–4.

[11] Winder L, Phillips C, Richards N, Ochoa-Corona F, Hardwick S, Vink C.J, Goldson S (2011) Evaluation of DNA melting analysis as a tool for species identification. *Methods in Ecology and Evolution* **2**: 312–320.

[12] Dawnay N, Ogden R, McEwing R, Carvalho GR, Thorpe RS (2007) Validation of the barcoding gene COI for use in forensic genetic species identification. *Forensic Sci Int* **173**: 1-6.

[13] Branicki W, Kupiec T, Pawlowski R (2003) Validation of cytochrome b sequence analysis as a method of species identification. *J Forensic Sci* **48**: 83-7.

[14] HM Government (2013) Elliott review into the integrity and assurance of food supply networks – interim report.

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

[16] Cawthorn DM, Steinman HA, Corli Witthuhn RC (2011) DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. *Food Research International*. **46**: 30-40.

[17] Rasmussen R, Morrissey MT (2008) DNA-Based methods for the identification of commercial fish and seafood species. *Comprehensive reviews in food science and food safety* **7**: 280-295.

[18] Agnew DJ, Pearce J, Pramod G, Peatman T, Watson R, Beddington JR, Pitcher TJ (2009) Estimating the worldwide extent of illegal fishing. *PLoS ONE* **4**; e4570.

[19] FAO Fisheries and Aquaculture Department (2007). *The state of world fisheries and aquaculture 2006*. Food and Agriculture Organisation of the United Nations, Rome.

[20] Ogden R (2008) Fisheries forensics: the use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. *Fish and Fisheries* **9**: 462–472.

[21] Nielsen E, Cariani A, Mac Aoidh E, Maes GE, Milano I, Ogden R et al (2011) Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications* **3**: Article number 851.

[22] Hennessy LK, Mehendale N, Chear K, Jovanovich S, Williams S, Park C, Gangano S (2014) Developmental Validation of the GlobalFiler® Express Kit, a 24-marker STR Assay, on the RapidHIT® System. *Forensic Sci Int*: **In Press.**

[23] Grubaugh ND, Petz LN, Melanson VR, McMenamy SS, Turell MJ, Long LS, Pisarcik SE, Kengluecha A, Jaichapor B, O'Guinn ML, Lee JS (2013) Evaluation of a Field-Portable DNA Microarray Platform and Nucleic Acid Amplification Strategies for the Detection of Arboviruses, Arthropods, and Bloodmeals. *Am J Trop Med Hyg* **88**: 245-253.

[24] Dawnay N, Stafford-Allen B, Moore D, Blackman S et al (2014) Developmental validation of the ParaDNA® Screening system – A presumptive test for the detection of DNA on forensic evidence items. *Forensic Sci Int Genet:* **11**: 73-79.

[25] Ball G, Dawnay N, Stafford-Allen R, Panasiuk M, Rendell P, Blackman S, Duxbury N, Wells S (2014) Concordance study between the ParaDNA® Intelligence Test, a Rapid DNA profiling assay, and a conventional STR typing kit (AmpFlSTR® SGM Plus®) *Forensic Sci Int Genet:* **16**: 48-51.

[26] French DJ, Archard CL, Brown T, McDowell DG (2001) HyBeacon probes: a new tool for DNA sequence detection and allele discrimination. *Molecular and Cellular Probes.* **15**: 363–374.

[27] Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*. **360**: 1847-1857.

[28] Ivanova N, Zemlak TS, Hanner RH, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes* **7**: 544-548.

[29] Larkin MA, Blackshields G, Brown MP, Chenna R, McGettigan PA, McWilliam H et al (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: *2947-8.*

[30] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular biology and evolution.* **30 (12)**: 2725-2729.

[31] Altman DG, Bland JM (1994) Diagnostic tests: sensitivity and specificity. *BMJ*. **308:** 1552.

[32] Rasmussen RS, Morrissey MT (2009) Application of DNA-Based Methods to Identify Fish and Seafood Substitution on the Commercial Market. *Comprehensive Reviews in Food Science and Food Safety* **8**: 118–154.

[33] Marine Stewardship Council Certification Requirements http://www.msc.org/about-us/standards/certification-requirements/msc-certification-requirements. Accessed 22.09.2014.

[34] Dawnay N, Ahmed R, Naif S (2014) The ParaDNA Screening System - A case study in bringing forensic R&D to market. *Science and Justice*. **54**: 481-486.

[35] Pionzio AM, McCord BR (2014) The effect of internal control sequence and length on the response to PCR inhibition in real-time PCR quantitation. *Forensic Sci Int Genet:* **9**: 55–60.

[36] Linacre A, Gusmão L, Hecht W, Hellmann AP, Mayr WR, Parson W, Prinz M, Schneider PM, Morling N (2011) ISFG: Recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. *Forensic Sci Int Genet: 5*: 501-505.

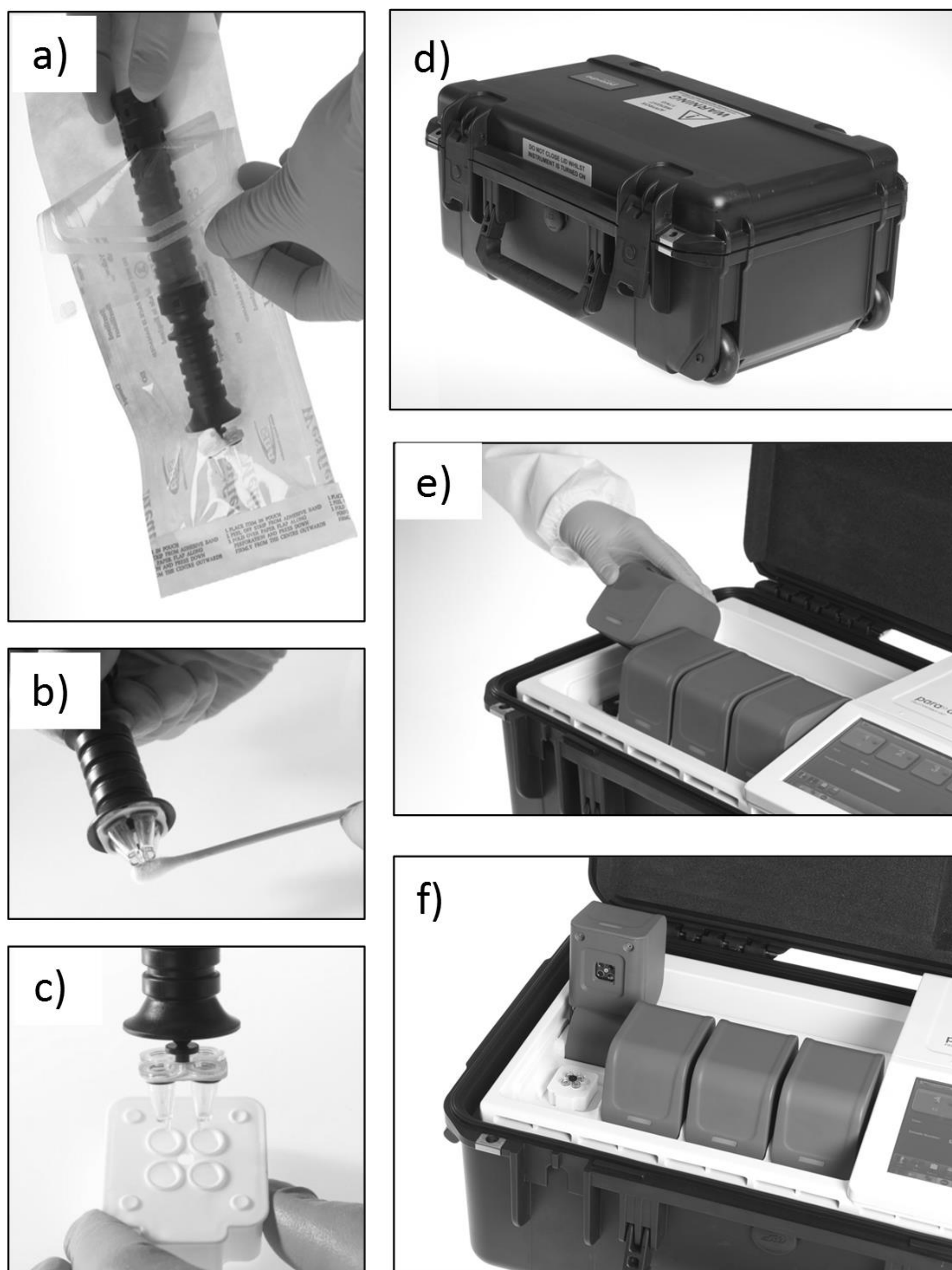


Figure 1. To use the ParaDNA System simply, (a) open the disposable Sample Collector; (b) recover the cellular material from an evidence item; and (c) introduce the template material into the PCR plate containing the assay mix. To load the sample on to the field portable unit (d) simply, open the independent head (e) and place the PCR plate onto the heating block. The process is finished by labelling the sample, closing the head and pressing start on the touch screen.

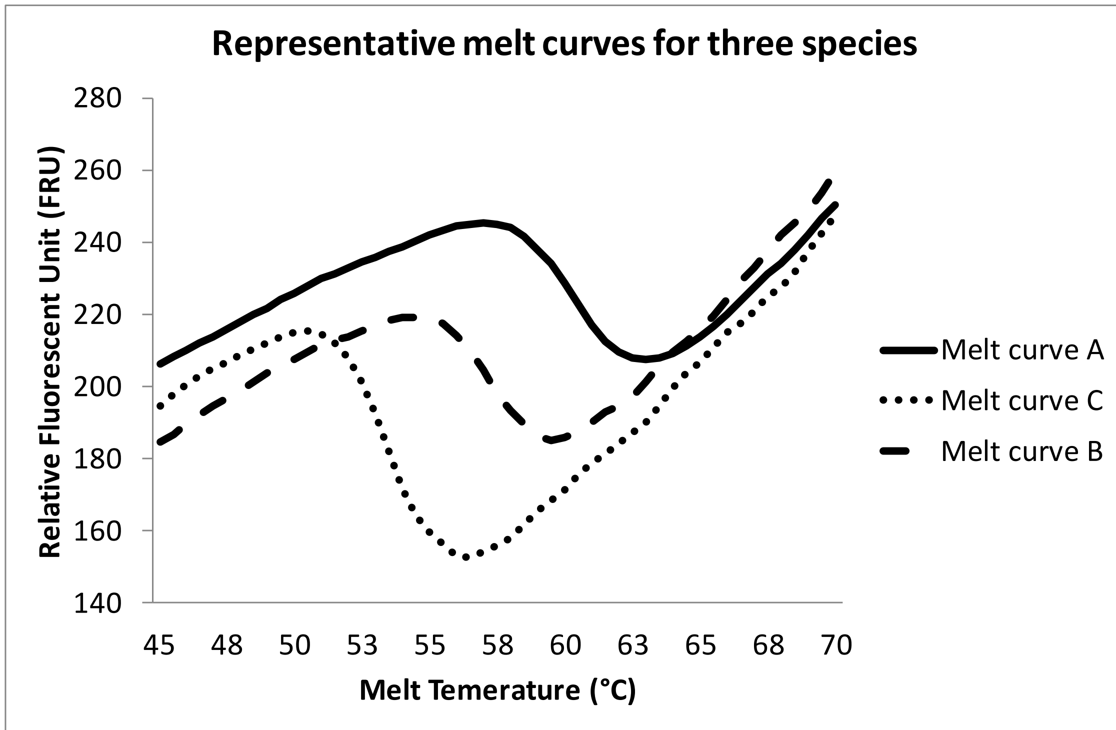
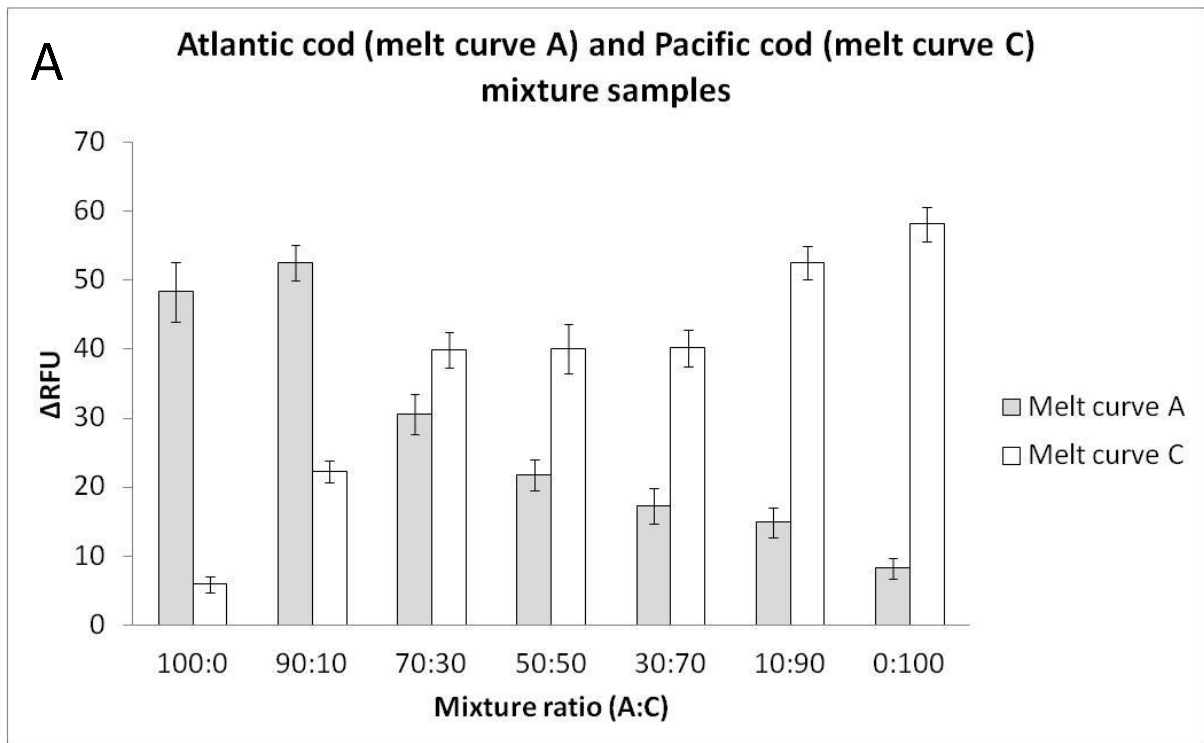


Figure 2. Representative melt curves for Atlantic cod (curve A), Alaskan Pollock (curve B) and Pacific cod (curved C).



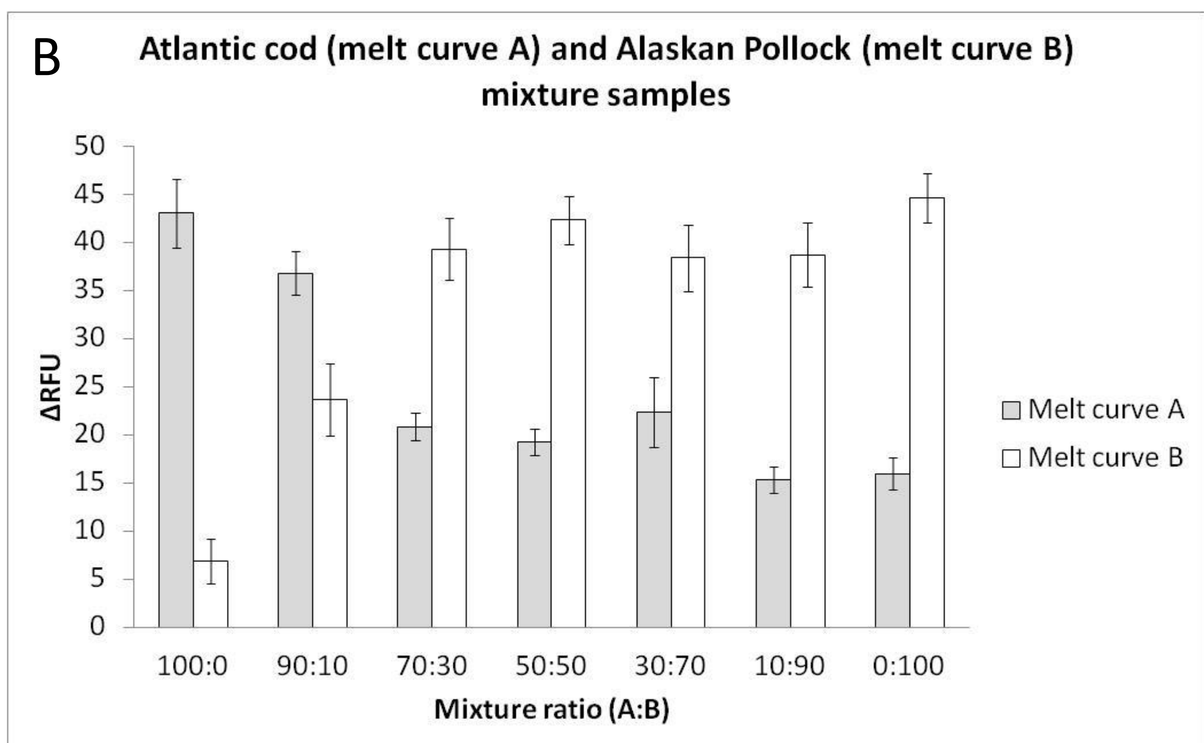


Figure 3. Melt curve measures of ΔRFU for different species mixtures. Transition in RFU in different complex mixtures shows potential ability to detect components at various levels. Error bars represent SEM.

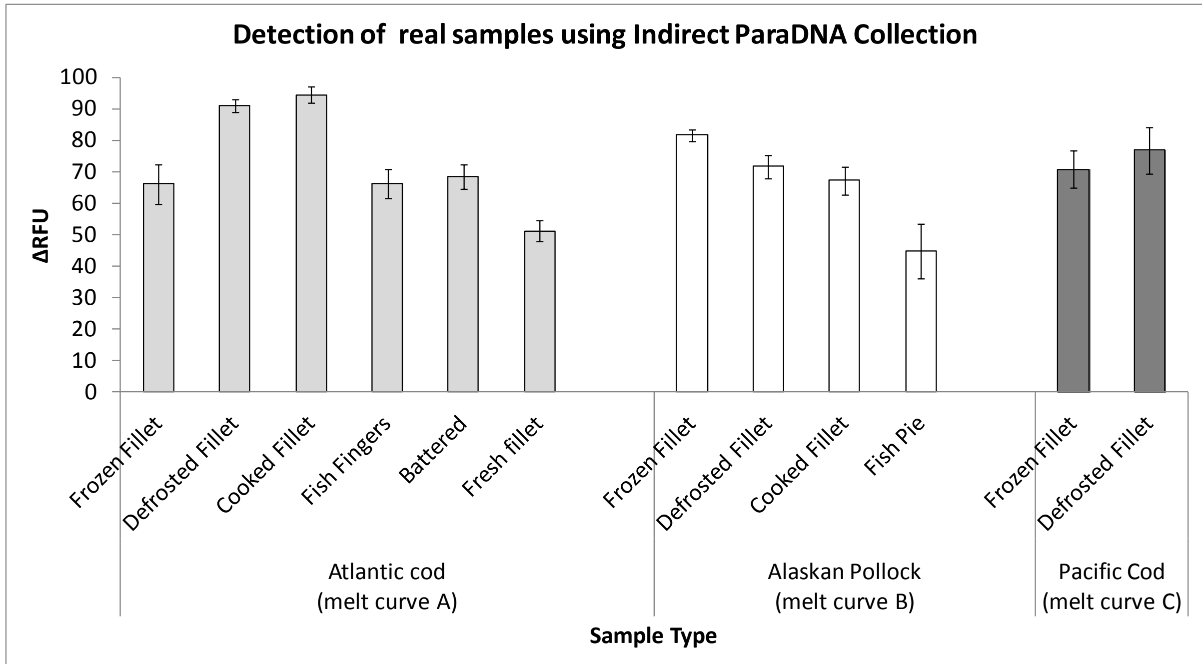


Figure 4. Real sample validation data sets showing the average ΔRFU (n=8) for each sample type tested. Error bars represent SEM.

Table 1. Description of all extracted and analysed samples.

Common name	Number	Expected match	Blast Match	Sequence Length (COI region)	Similarity (%)
Atlantic Cod	10	*Gadus morhua*	*Gadus morhua*	374 – 642	99-100
Alaskan Pollock	10	*Gadus chalcogrammus*	*Gadus chalcogramma*	489 – 651	99-100
			Gadus finnmarchica		
Pacific Cod	10	*Gadus macrocephalus*	*Gadus macrocephalus*	510 – 660	98-100
			Gadus ogac		
Plaice	1	*Pleuronectes platessa*	*Pleuronectes platessa*	464	100
Sea bream	1	*Sparus auratus*	*Sparus auratus*	246	100
Haddock	1	*Melanogrammus aeglefinus*	*Melanogrammus aeglefinus*	500	100
Mackerel	1	*Scomber scombrus*	*Scomber scombrus*	655	100
Salmon	1	*Salmo salar*	*Salmo salar*	563	100
Nile Tilapia	1	*Oreochromis niloticus*	*Oreochromis niloticus*	438	100
			Oreochromis mossambicus		
Coley	1	*Pollachius virens*	*Pollachius virens*	657	100
Hake	1	*Merluccius merluccius*	*Merluccius paradoxus*	482	97
Vietnamese River Cobbler	1	*Pangasius bocourti*	*Pangasianodon hypophthalmus*	504	100
Sea bass	1	*Dicentrachus labrax*	-	-	-
Lemon Sole	1	*Microstomus kitt*	-	-	-
Herring	1	*Clupea harengus*	-	-	-

-' is used to show non-amplification or an unreadable sequence data. Similarity = percentage of query sequence that overlaps the subject sequence.

Table 2. HyBeacon probe primer binding locations in study species, close taxonomic species and unrelated species identified through sequence similarity searches.

Test category	Scientific Name	HyBeacon Haplotype	Target Sequence	Predicted Tm	Melt curve detected	Downloaded sequences (n)	Species coverage
Probe	HyBeacon probe		*A T C G G T G C A C C A G A* ***F*** *A T A G C T* ***F*** *T C*				
Target species	*G. morhua*	H1	T A G C C A C G **T** G G **T** C T A T A T C G A A A G	59°c	A	29	100%
Key non target species tested	*G. chalcogrammus*	H2	. . . . **G**	54.5°C	B	121	92.4%
		H3	. . **A** . . . . . **G**	48.5°C	–	5	3.8%
		H4	. . **A** . . . . . **G** . . . . **C**	42.5°C	–	1	0.8%
		H5	. . . . **G** . . . . **C**	48.5°C	–	2	1.5%
		H6	. . . . **G** . . . . **A**	51.5°C	–	1	0.8%
		H7	. . . . . **G** . . **G**	49.9°C	C	1	0.8%
	G. macrocephalus	H8	. . . . **G** . . **G**	49.9°C	C	44	100.0%
Key non target species untested	*B. saida*	H2	. . . . **G**	54.5°C	B	44	100%
	M. tomcod	H2	. . . . **G**	54.5°C	B	7	100%
	M.proximus	H5	. . . . **G** . . . . **C**	48.5°C	–	19	100%
	G. ogac	H8	. . . . **G** . . **G**	49.9°C	C	2	100%
	M. merlangus	H9	. . . . **A**	53.2°C	–	38	92.70%
		H10	. . **A** . . . . . **A**	47.2°C	–	3	7.30%
Non target species tested	*P. virens*	–	. . . . G . . . . C . . G . . .	–	–	–	–
	M. aeglefinus	–	. . A . . . . . G . . . . C	–	–	–	–
	M. merluccius	–	. . . . . T . . G . . G . . G . . C . . G . . .	–	–	–	–
	P. bocourti	–	. . A . . T . . . . C . . T . . A	–	–	–	–
	C. harengus	–	. . . . . T . . C . . . . C . . T . . .	–	–	–	–
	M. kitt	–	. . A . . C . . G . . A . . G . . . . A	–	–	–	–
	S. scombrus	–	. . . . . T . . G . . G . . . . C . . .	–	–	–	–
	P. platessa	–	. . A . . C . . G . . G . . . . . C . . G . . .	–	–	–	–
	S. salar	–	. . . . . C . . G . . G . . G . . . . . T . . .	–	–	–	–
	D. labrax	–	. . A . . C . . C . . A . . . . T . . A	–	–	–	–
	S. auratus	–	. . . . G . . A . . G . . . . . T . . .	–	–	–	–
	O. niloticus	–	. . A . . . . . G . . . . C . . G . . .	–	–	–	–

Table 3. Accuracy data assessed by sensitivity and specificity for each species under study.

Species	Melt Curve	Total Sample Number	Sample type	True Positives (n)	True Negatives (n)	False Positives (n)	False Negatives (n)	Sensitivity	Specificity
Atlantic cod	A	488*	Genomic DNA	127	339	13	9	93.38%	96.31%
Alaskan pollock	B		Genomic DNA	133	348	4	3	97.79%	98.86%
Pacific cod	C		Genomic DNA	126	355	1	6	95.45%	99.72%
Atlantic cod	A	192*	Tissue	89	88	8	7	92.71%	91.67%
Alaskan pollock	B		Tissue	55	124	4	9	85.94%	96.88%
Pacific cod	C		Tissue	31	157	3	1	96.88%	98.13%

*Total number of measurements taken across all species. True Positive and False Negative results were only obtained in species under study. True Negative and False Positive results were obtained by looking at results from negative control samples and also non-target species data.