Natural England Commissioned Report NECR330

# Development of a novel (DNA-based) method for monitoring inshore fish communities using a programmable large-volume marine eDNA sampler

First published November 2020

www.gov.uk/natural-england

NATURAL ENGLAND

# Foreword

Natural England commission a range of reports from external contractors to provide evidence and advice to assist us in delivering our duties. The views in this report are those of the authors and do not necessarily represent those of Natural England.

## Background

DNA – based methods offer a significant opportunity to change how we monitor and assess biodiversity. These techniques may provide cheaper alternatives to existing species monitoring or an ability to detect species that we cannot currently detect reliably.

However, for most species, there is still much development required before they can be used in routine monitoring. Natural England has been exploring the further use of these methods for environmental monitoring for several years, delivering a series of reports which focus on the development of DNA-based methods with potential in a particular area.

This report presents the development of a technique using eDNA collected by a large-volume marine eDNA sampler deployed on the seabed to detect inshore fish communities. It builds on previous work to explore the special and temporal variation required to detect changes in fish communities, and the use of haplotypes as a promising new technique for assessing the diversity of breeding fish populations.

This report should be cited as: Mynott, S., and Marsh, M., 2020. Development of a novel (DNA-based) method for monitoring inshore fish communities using a programmable large-volume marine eDNA sampler. *Natural England Commissioned Reports, Number NECR330*

**Natural England Project Manager** – Maija Marsh, Senior Specialist, Natural England: maija.marsh@naturalengland.org.uk

**Contractor** – Applied Genomics Ltd

**Keywords** – environmental DNA (eDNA), marine fish communities, programmable large-volume marine eDNA sampler, metabarcoding, haplotype

**Further information**
This report can be downloaded from the Natural England Access to Evidence Catalogue: **http://publications.naturalengland.org.uk/**. For information on Natural England publications contact the Natural England Enquiry Service on 0300 060 3900 or e-mail **enquiries@naturalengland.org.uk**.


ISBN 978-1-78354-673-2

Brixham Environmental Laboratory
Freshwater Quarry
Brixham, Devon
TQ5 8BA
+44 1803 611818
www.appliedgenomics.co.uk

8th October 2020

# Natural England Commissioned Report

Development of a Novel (DNA-based) Method for Monitoring Inshore Fish Communities Using a Programmable Large-Volume Marine eDNA Sampler.

Report prepared by

**Sebastian Mynott**

Principal Molecular Ecologist

Applied Genomics Ltd.

and

**Maija Marsh**

Marine Ecology Senior Specialist

Natural England

| Version | Date | Author | Comments |
|---|---|---|---|
| **Draft** | 30 Mar 2020 | Sebastian Mynott | Submission to NE for comments |
| **Draft 2** | 10 Apr 2020 | Sebastian Mynott | Response to comments from NE |
| **Draft 3** | 15 Apr 2020 | Sebastian Mynott | Submission to Defra for comments |
| **Draft 4** | 04 May 2020 | Sebastian Mynott, Maija Marsh | Response to comments from Defra & NE |
| **Draft 5** | 07 Aug 2020 | Sebastian Mynott, Maija Marsh | Final Review |
| **FINAL** | 18 Aug 2020 | Sebastian Mynott, Maija Marsh | Approved for submission to Defra |

## Acknowledgements

We are grateful to our collaborators at Cornwall, Devon & Severn, and Southern Inshore Fisheries and Conservation Authorities, who braved harsh sea conditions to undertake the sampling for this project. We are grateful also to Natural England and Defra for their support.

## Summary

There is a considerable lack of understanding of inshore fish community composition, community health and population status, as well as the scale and footprint of critical pressures and the efficacy of management measures. There is a clear evidential need for improved understanding of the nature and status of our inshore fish populations, be they commercial or non-commercial stocks. The overall aim of this project is to address these needs through validation of DNA-based methods for improved monitoring of inshore fish communities.

In 2018/19, Natural England led a pilot study to test the use of a large-volume marine eDNA sampler (inDepth eDNA sampler) developed by Applied Genomics for monitoring inshore fish communities at a whole-site and habitat-specific sub-feature level ("pilot study"; NECR287, 2018). Despite the small scale of this proof-of-concept study, it provided some promising results in the potential of this novel technique for monitoring inshore fish populations for marine conservation and fisheries management purposes. The pilot study report provided a number of recommendations on how to improve the sampling strategy and gain an understanding of the spatial and temporal capabilities of this sampling technology.

This study represents the next step in the process to developing a cost-effective method to monitor inshore fish communities, which in turn will underpin our ability to effectively conserve, manage and sustainably exploit these resources and allow the UK to develop a world class marine monitoring system. This project has been undertaken in the Southwest of England and it aligns with the pilot areas being used for a related, Defra funded (and Natural England managed) project entitled "Regional monitoring plan for inshore fish communities in the Southwest of England" (NERC 271, Franco *et al.,* 2020c).

The overall study aim was to identify spatial scales and frequency of sampling required to effectively monitor inshore fish (*Actinopterygii* and *Elasmobranch*) communities. We also endeavoured to undertake eDNA sample collection using paired deployments of a large-volume eDNA sampler to demonstrate the capability for developing an improved understanding of the relationship between fish communities and key protected habitats.

Fifteen large-volume eDNA samples were collected from six sampling locations across the south coast of England from October 2019 to February 2020. Recovered samples were processed, DNA purified, and three technical replicates were created for each of two mitochondrial barcode loci using fish-specific metabarcode primers. High-throughput DNA sequencing was undertaken on an Illumina MiSeq sequencer. The resulting sequence data were subjected to bioinformatic processing, taxonomic assignment and biostatistical analyses.

Our methods characterised a total of 74 fish species and two marine mammal species, of which 36 species were recorded as economically valuable from historical fisheries landings data (Table

9); 19 species were identified as UKBAP-listed protected species and one species was recognised as a GBNNSS-listed non-native species.

Differences in fish community composition were spatially correlated with larger geographical distances and temporally correlated with time between sampling events indicating that the sampling intensity and geographical scale chosen for this study were suitable for monitoring inshore fish populations and communities in the Southwest of England using the InDepth eDNA sampler. Our characterisation of haplotypes of each fish species within each sampling location could offer a promising new technique for assessing the diversity of breeding fish populations which, once sufficiently tested and validated, could in the future provide a valuable metric for monitoring inshore fish breeding population diversity and/or health.

We have shown that the data resulting from DNA-based monitoring of inshore marine habitats using large-volume eDNA samples can provide an unprecedented level of information about the state of inshore fish communities across large spatial scales and over time, resulting in data-driven insights for deeper due-diligence and more meaningful risk intelligence. Specifically,

1. We successfully demonstrated the operation of the inDepth eDNA samplers for consistent, reliable and automated *in-situ* filtration of large volume water samples as a method for monitoring the composition of inshore fish communities across a range of habitat-specific deployments;
2. This study clearly demonstrated that this technology could robustly and consistently assess inshore fish populations and delivered analytical outputs accompanied by detailed data confidence analyses which are crucial for end-user interpretation;
3. We gained important experience into developing our understanding of the temporal stochasticity of these eDNA sample data sufficient as a basis for planning the sampling frequency of future studies;
4. We gained a deeper understanding of the spatial requirements for effective sampler deployments based on the effect of current and tidal movements during the sampling period, which will support planning of sampler placement for future studies;
5. We demonstrated these methods are clearly suited to the task of monitoring the presence and distribution of protected species and show great potential for detection of non-native species (NNS) within MPAs and sensitive estuarine environments.

Further recommendations for monitoring inshore fish communities using large-volume eDNA data are provided in this report.



Applied Genomics Ltd is registered in England and Wales: 9180742 | VAT: GB198410583
Brixham Environmental Laboratory | Freshwater Quarry | TQ5 5AY

Brixham Environmental Laboratory
Freshwater Quarry
Brixham, Devon
TQ5 8BA
+44 1803 611818
www.appliedgenomics.co.uk

# Table of Contents

# 1 Table of Figures

# 2 Table of Tables

# 1 Background

Inshore fish communities are poorly understood. This was highlighted by the outputs of a workshop on the future requirements for the monitoring of inshore fish communities (Righton 2015), held by the Healthy Biologically Diverse Seas Evidence Group (HBDSEG) of the UK Marine Monitoring and Assessment Strategy (UKMMAS). There is a considerable lack of understanding of inshore fish community composition, community health and population status, as well as the scale and footprint of critical pressures and the efficacy of management measures. Filling these evidence gaps is a priority for UK regulators as, there is a realistic possibility that inshore fishing effort will increase, as a result of potential redistribution of any additional fishing opportunities (i.e. quotas) derived as a result of ceasing alignment with the Common Fisheries Policy (CFP) following the end of the implementation period (currently 31st December 2020). However, the UK is required to meet its national and international commitments related to fisheries management and marine conservation. When combined with commitments to improve the state of our marine environment identified in the 25 Year Environment Plan and the UK Marine Strategy as well as the Fisheries Objectives set out in the Fisheries Bill there is a clear evidential need to understand the nature and status of our inshore fish populations, be they commercial or non-commercial stocks. This recognised gap was the driver behind a Natural England (NE) European Union (EU) Exit project to investigate the costs of a holistic approach to inshore fish community monitoring in a Southwest England pilot (NECR 271), which advocated the use of eDNA as a crucial component of any future inshore fish community monitoring programme.

Environmental DNA (eDNA) is genetic material that is released by organisms living in their environment. Sources of genetic material may include, faeces, urine, skin cells, mucous or gametes. The DNA from these sources will degrade over time but persist in the environment for long enough that the presence of organisms in the environment may be detected without their being seen or captured. The marine environment is large, heterogeneous and dynamically variable. Validation studies undertaken by Applied Genomics have shown that large-volume eDNA sample collection is required to maximise the probability of detection, reduce inter-sample variability and ensure meaningful results. Applied Genomics have developed the inDepth eDNA Sampler for consistent large-volume environmental DNA sample collection, robust laboratory protocols for processing these large samples and a proven bioinformatics pipeline which delivers reliable results.

By its nature, DNA captured from the environment is normally associated with particulate matter, such as sloughed skin cells, faeces, etc. Whilst these particles may persist in the water column for some time, they will, eventually, precipitate out. We can therefore assume that the demersal layer may contain a slightly higher density of eDNA particulate matter than may be found nearer the surface. However, the rate at which this precipitation may occur in real terms is highly stochastic. Collins *et al.,* (2018) experimentally modelled the decay rate of eDNA in the marine environment, estimating an eDNA half-life of 21.2 hours in simulated marine inshore environments of the Western English Channel region.

The Applied Genomics inDepth eDNA Sampler was built to operate at a fixed depth, sampling the demersal layer currents. Demersal currents are generally slower moving due to laminar boundary

layer flows with less influence from surface fluctuations, such as Langmuir circulations (Barstow, 2003) and may have greater directional consistency due to Ekman layer effects (Ekman, 1905). For this study, the inDepth eDNA Samplers were programmed to collect each sample by filtering approximately 50 litres of seawater over two full tidal cycles (25 hours). This resulted in a Eulerian-transect sampling strategy, where the sampler itself remains in a fixed location whilst the environment being sampled moves past it over time. This sampling strategy was designed to mitigate sampling variability from tidal currents and longshore currents to provide a comprehensive and unbiased method for surveying local area marine life.

## 1.1 Aims & Objectives

To validate and operationalise the use of the inDepth eDNA sampler for use in routine monitoring of inshore fish communities, this project had the following objectives:

1. To test the feasibility of the inDepth eDNA sampler in monitoring the composition of fish communities across a range of habitat specific deployments in the south west.
2. To assess the eDNA data for its ability to assess the presence and scale of breeding populations, including Alpha, Beta and Gamma diversity statistics and haplotype diversity.
3. To statistically analyse the data to determine the sampling frequency needed when using a large-volume programmable marine eDNA sampler for measuring changes to the inshore fish communities in each habitat location.
4. To determine the spatial scale of sampling necessary to measure and infer changes to the inshore fish communities that would assist NE and IFCAs in developing appropriate advice.
5. To determine the presence and distribution of marine Species of Conservation Interest (SOCI) and non-native species (NNS), within the survey area.
6. Subsidiary objective: to test (in a restricted set of locations) the potential ability of eDNA data to assess species fidelity to key protected habitats by manipulation of the temporal sampling regime. For example, at a particular location comparing the eDNA signatures from samples collected during slack and peak tidal flow periods.

We considered the operational practicality of eDNA sampling using the inDepth eDNA sampling methodology to help define capabilities of DNA-based marine monitoring at a national scale.

# 2 Methodology

We built on the methodological approaches and analyses presented in the pilot study to further assess and validate the potential future utility of the large-volume eDNA sampling technique as a standard approach to monitoring inshore fish populations and communities. In order to move this approach from testing toward operationalisation, it is critical that elements of practical analysis deployment are investigated in detail. These are detailed in the aims and objectives, but essentially revolve around a deeper understanding of the power of this technique to detect changes in fish communities, intra species (haplotypic) diversity, and an effective assessment of the sampling frequency (both spatial and temporal) that would be required to incorporate these elements into an effective monitoring programme in the future. It is the intention of the following analyses to provide a clear steer on these questions to facilitate a further study which would test the operational efficacy of this technique for the monitoring of inshore fish species and communities.


*Figure 1: The Applied Genomics inDepth eDNA Sampler is capable of filtering 50 litres of seawater at greater than 150 metres depth over several tidal cycles.*

## 2.1 Field Sampling

### 2.1.1 Environmental DNA Sampling in the Marine Environment

The sampling methods applied in this study are largely similar to those described in the recent pilot study (NECR287) conducted over a similar time period in 2018/19 for Natural England with the support of Devon & Severn IFCA. Marine eDNA sampling was undertaken using the inDepth eDNA sampler, which has been designed specifically for eDNA sample collection in large, dynamic water environments such as ports, estuaries and coastal waters. The sampler consists of an encapsulated 1 µm polyester filter, attached to a pump, which is controlled by an on-board computer and powered by a rechargeable battery pack, all of which is mounted into a rugged waterproof housing.

Deployment and recovery from inshore waters was accomplished by one trained operator and an assistant. Once recovered, the sample material were captured within the enclosed filter capsule, ensuring sample integrity and ease of use by field personnel. An

ammonium sulphate based DNA preservative buffer supplied by Applied Genomics, which included an internal positive control synthetic DNA fragment to monitor DNA recovery, was added to the filter capsules in the field prior to return to the laboratory for sample processing. For this study, training on operation of the samplers was provided to Fisheries Conservation Officers from Cornwall, Devon & Severn, and Southern Inshore Fisheries Conservation Authority (CIFCA, D&SIFCA and SIFCA, respectively).

The inDepth eDNA Sampler works by periodic sampling of large water volumes over a number of tidal cycles. The final volume sampled was approximately 50 litres / sample, depending on turbidity. The sampler was programmed to undertake in-situ water filtration over a 25-hour period, covering two full tidal cycles. The volume of water sampled is at least an order of magnitude greater than other eDNA sampling methods (Bettey *et al*., 2020) whilst sampling over a long duration further increases the probability of detection. The inDepth eDNA Sampler sampling algorithm ensured even sampling over this time period by progressively increasing the sampling effort to compensate for reduced flow rates as the sample was collected. The filter inlet was positioned approximately 1m above the benthos (Figure 1), which enabled the detection of epibenthic species such as flatfish and rays. A non-return valve was fitted to the inlet of the filter capsule to prevent accidental loss of the sample. Each sampler was tagged with a coloured marker and randomly allocated to a sampling location.

### 2.1.2 Spatiotemporal Sampling Strategy

Expanding on the methods developed in the 2018/19 pilot study (NECR287), two inDepth eDNA Samplers were simultaneously deployed at each of three pairs of inshore locations along the open coastline of south west England (Table 1). Refer to Appendix 5 for a full list of MPAs within the areas sampled. Each of the sampling locations were selected to be in less than 20 metres water depth and preferably within a designated no-trawl zone to reduce the risk of sampler loss.

*Table 1: Sampling locations and the Marine Protected Areas fully or partially covered in the sampling regime.*

| *Location* | Lat | Lon | Marine Protected Area | IFCA |
|---|---|---|---|---|
| *Falmouth* | 50.11611 | -5.04167 | Fal and Helford SAC (UK0013112) | Cornwall |
| *Fowey* | 50.31222 | -4.65444 | Falmouth Bay to St Austell Bay SPA (UK9020323) | Cornwall |
| *Start Bay* | 50.26305 | -3.63725 | Skerries Bank and Surrounds MCZ (UKMCZ0015) | Devon & Severn |
| *Lyme Bay* | 50.66277 | -3.10103 | Lyme Bay and Torbay SAC (UK0030372) | Devon & Severn |
| *Studland* | 50.64200 | -1.91500 | Studland to Portland SAC (UK0030382) | Southern |
| *Poole Rocks* | 50.68000 | -1.87600 | Southbourne Rough MCZ (UKMCZ0071) | Southern |

Sampling locations are shown in Figure 2. The area of water sampled was estimated using tidal and longshore current excursion modelling over a simulated 25-hour deployment using the AVS Dev Tide Modelling Tool. Spring tidal ranges are shown, which represent the conditions in which the majority of samples were collected.

As shown in Figure 2, the area in green displays the water excursion at the sampling depth, 1 metre above the benthos; the area in yellow shows the excursion due to currents and tides at the sea surface. The accuracy of the models decreases closer to shore, therefore the areas displayed are for informative purposes only.

Each deployment location was selected along the south coasts of Cornwall, Devon and Dorset, such that eDNA samplers would be placed upon benthic substrates which could be characterised as soft (mud, muddy sand or sandy mud) or hard (mixed sediment or rock). The geographical distances between each pair of sampling sites varied, as shown in Table 2.

*Table 2: Haversine distances between sampling locations (km) with benthic substrate type.*

| | ***Falmouth*** | ***Fowey*** | ***Lyme Bay*** | ***Poole Rocks*** | ***Start Bay*** | ***Studland*** |
|---|---|---|---|---|---|---|
| ***Falmouth*** | *Soft* | | | | | |
| ***Fowey*** | 35.2 km | *Hard* | | | | |
| ***Lyme Bay*** | 150.6 km | 116.7 km | *Hard* | | | |
| ***Poole Rocks*** | 233.2 km | 201.0 km | 86.4 km | *Hard* | | |
| ***Start Bay*** | 101.4 km | 72.6 km | 58.5 km | 133.1 km | *Soft* | |
| ***Studland*** | 229.6 km | 197.5 km | 83.7 km | 5.0 km | 129.2 km | *Soft* |

The 2018/19 pilot study (NECR287) had suffered from a lack of statistical power which we intended to resolve by collecting 6 eDNA samples per month over 5 months. These 30 samples,

each with 3 technical replicates, would provide a statistical power greater than 0.95 for a balanced one-way analysis of variance calculation comparing two groups and power greater than 0.9 when comparing three groups. In other words, to compare differences in biological communities by collecting 45 technical replicate samples in 1 group and 45 technical replicate samples in another group, we would have a 95% chance of correctly interpreting a statistical significance of $p < 0.05$ when observing a large effect (Cohen, 2013) in the study data. Power analyses were calculated using R the package *pwr* (Champeley, 2020).


*Figure 2: Sampling locations showing estimated water movement at the sampling depth, one metre above the benthos (green) and at the surface (yellow) during the sampling period. Tidal and current excursion modelling was performed using the ASVDev Tide Modelling Tool (© AVS Developments Ltd., 2020; https://tide.avsdev.uk). Spring tidal ranges shown.*

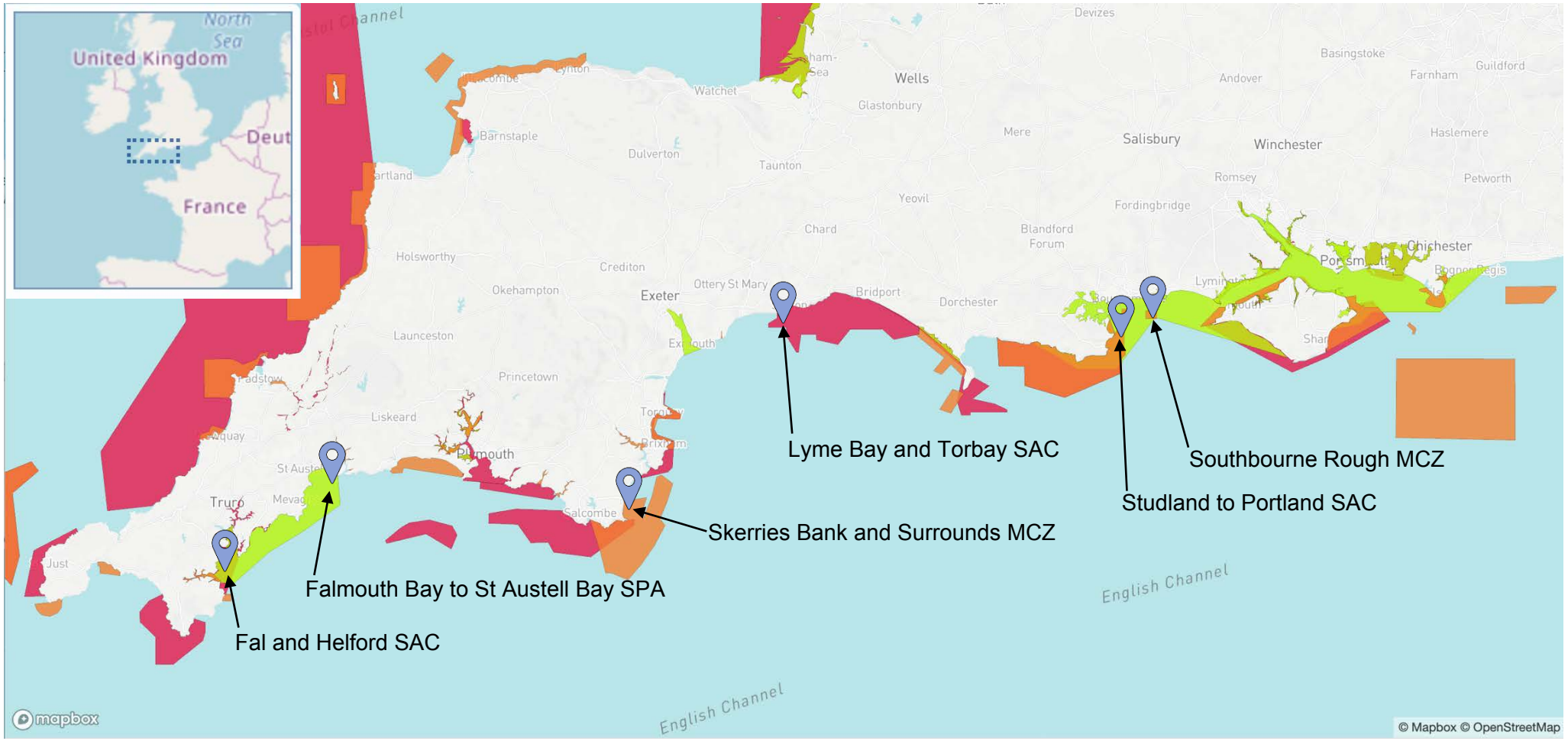



*Figure 3: Map of marine protected areas in Southwest England. Special Areas of Conservation (SACs) are in pink; Special Protection Areas (SPAs) in green and Marine Conservation Zones (MCZs) in orange.*

## 2.2 Sample Processing & Sequencing

To mitigate any risk of sample contamination, all sample processing was undertaken in a dedicated laboratory, free of PCR products and physically separated from all downstream processes.

For each sample, the filtrate was recovered from the filter membrane by vortexing the filter capsule for a minimum of ten minutes before elution of the preservative and filtrate. The organic matter was then concentrated by high-speed centrifugation. DNA was purified from each sample using QIAGEN PowerMax Soil kits and quantified by fluorimetry using Qubit 3.0 High Sensitivity kit following the manufacturer's instructions for double-stranded DNA. A negative extraction blank was included in the DNA purification step and processed alongside the eDNA samples throughout all subsequent steps. These steps were identical to those undertaken in the pilot study (NECR287). Sample integrity for each sample was confirmed through successful recovery of the internal positive control spike using qPCR. Environmental DNA recovery concentrations for each sample are given in Appendix 2: Environmental Sampling Metadata ("*eDNA conc*").

PCR (polymerase chain reaction), is a process commonly used in molecular biology to amplify, or make many copies of, a specific target region of DNA. Fish-specific 'barcode' regions of DNA were targeted using DNA primers for the ribosomal small subunit (12S; Miya *et al*., 2015) and Cytochrome B (CytB; Evans *et al*., 2016) mitochondrial loci to provide robust detection and broad coverage of fish species. The amplified products of the PCR reaction, amplicons, were created based on the conditions described in the literature and in Appendix 1: PCR Conditions.

Figure 4 illustrates the process by which, for each sample, technical replicates were created by resampling the DNA extract and making minor adjustments to the PCR conditions to improve the detection efficiency of each of the fish-specific universal primers (Doi *et al.,* 2019).

Each of each the steps in the PCR amplification process were undertaken in 12 PCR replicates which were then pooled to maximise the probability of detection. Purification of amplicons were achieved using AMPure XP magnetic beads according to the manufacturer's protocol and quality checked using capillary electrophoresis on a QIAGEN QIAxcel Advanced instrument using a DNA Screening kit.

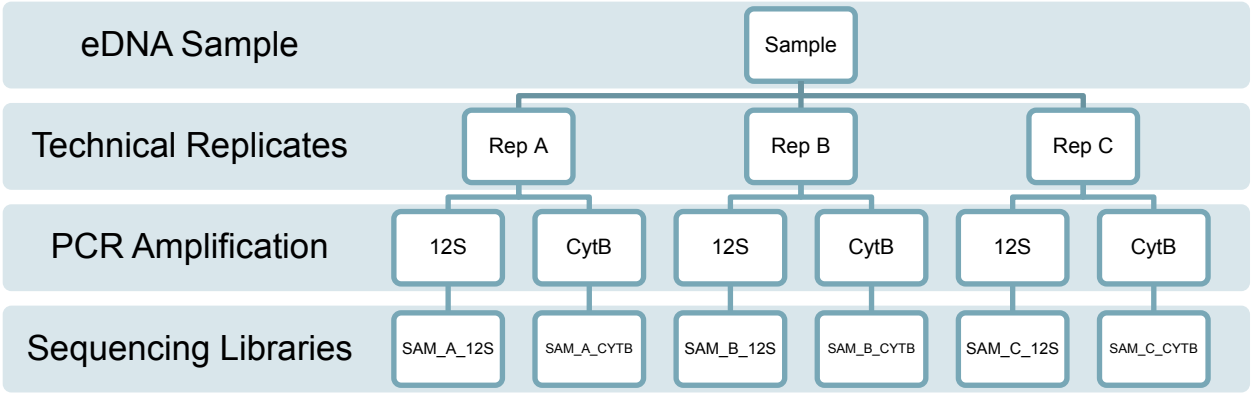



*Figure 4: Laboratory processing steps for generating technical replicates for each field sample, amplification of 12S and CytB barcode regions and sequencing library preparation.*

Amplicon sequencing libraries were prepared using a Nextera XTv2 index kit according to the methods developed by Illumina. Library purification was achieved using AMPure XP beads according to the manufacturer's protocol and quality checked using capillary electrophoresis.

The library concentrations were quantified by fluorimetry, adjusted to 4 nM using pure water, pooled, denatured with NaOH and PhiX control added according to published Illumina protocols. The eDNA amplicon library were sequenced on an Illumina MiSeq sequencer by the Exeter Sequencing Service.

## 2.3 Bioinformatics & Taxonomic Assignment

All analyses were undertaken in R (R Core Team, 2018) using core packages listed in Figure 5.

To ensure detected sequences could be reliably compared between studies, we applied the DADA2 algorithm which creates exact amplicon sequence variants (ASVs) based on modelling and correction of sequenced amplicon errors (Callahan *et al*., 2016). The principal advantage of this pipeline is that, unlike operational taxonomic units (OTUs), ASVs are directly comparable between studies, ensuring consistency and reproducibility of results. OTU methods attempt to eliminate false positive inferences by lumping similar sequences together based on picking a random sequence and applying an arbitrary similarity threshold (Callahan *et al.*, 2017; Tsuji *et al.*, 2018; Forster *et al.*, 2019). DADA2 computes a statistical model of amplicon errors to infer the underlying sample sequences directly and thus denoises the data with improved specificity and sensitivity to true sequence variance, enabling the detection of rare variants (Callahan *et al*., 2016).

Existing taxonomic databases are known to be incomplete, redundant and/or to contain errors. We have created custom DNA databases by downloading *Actinopterygii* and *Elasmobranch* sequence data from GenBank (Sayers *et al*., 2018, Leray *et al*., 2019) for both the 12S & CytB gene regions. Taxonomic ranks were standardised against records in the Global Biodiversity Information Facility (GBIF; Chamberlain, 2017). These data were then curated by applying a recursive machine-learning algorithm to identify and remove redundant or incorrectly classified sequences. The algorithmic curation of our reference databases was able to identify many putative mislabelling errors, enabling our training sets to be automatically corrected by eliminating spurious sequences (Murali *et al.*, 2018). This resulted in 37,955 CytB reference database sequences for 11,622 *Actinopterygii* and *Elasmobranch* species and 20,421 12S reference database sequences for 11,141 *Actinopterygii* and *Elasmobranch* species.

The taxonomic assignment algorithm we used applied a hybrid approach that combined features of phylogenetic, distance-based, and machine learning classification methods to avoid over classification and misclassification errors (Murali *et al*., 2018). For each of the 12S and CytB sequences, taxonomic assignments were performed by matching eDNA sequences, first against our curated custom-built fish databases; sequences that failed to assign to species level were then matched against similarly curated versions of much larger publicly available (Machida *et al*., 2017) sequence databases, containing more comprehensive lists of non-target taxa. Since identification to species-level is required for end-user analyses, sequences that failed to assign to species level after matching against both databases were removed from the data.

Due to unquantifiable biases incurred throughout metabarcoding workflows (Brooks *et al.,* 2015, Lamb *et al*., 2019), sequencing read 'abundance' values were transformed to sequence detection or non-detection (i.e. presence/absence). The resulting data were then applied to all downstream analyses.

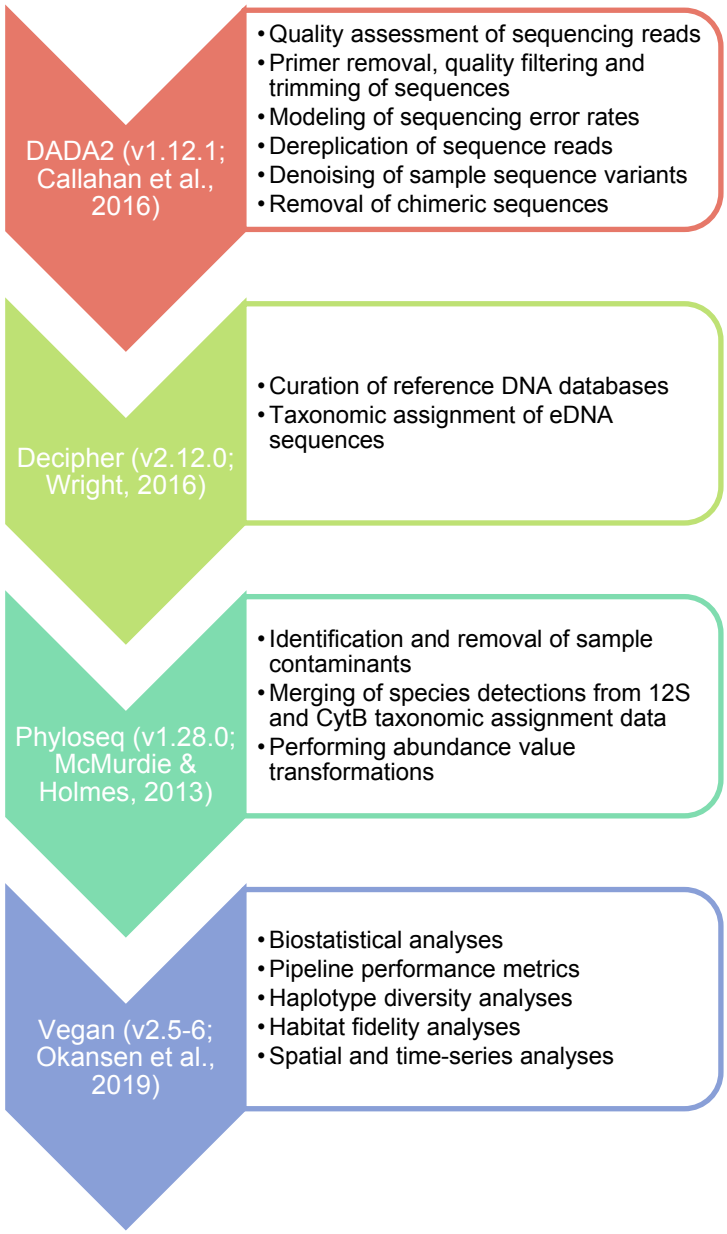



*Figure 5: Applied Genomics bioinformatics and analysis pipeline summary.*

Following taxonomic assignment, a number of sequences may be taxonomically matched to the same species. Where these occur, the detected variants may be regarded as haplotypic variants for the same genetic locus (Tsuji *et al.,* 2018, Callahan, 2019). Each individual within a population will have one genetic haplotype but several closely related individuals may share the same haplotype, usually through maternal inheritance (Billington & Hebert, 2011). So, whilst not an indication of numbers of individuals, haplotype diversity is an indication of the genetic diversity of breeding individuals within the sampled population (Wright, 1937; Ridley, 2003).

### 2.3.1 Inshore Fish Community Analyses

We built on the analyses presented in the pilot study to further assess the potential of large-volume eDNA sampling for monitoring inshore fish populations and communities by assessing species-habitat fidelity relationships and visualising spatial and time-series data to meet the stated project aims and objectives. We also explored the detected haplotype diversity of breeding individuals of each detected species over spatial and temporal distances (Elbrecht *et al*., 2018; Turon *et al*., 2020); and compared eDNA detections to Marine Management Organisation (MMO, 2019) historical landings data. Furthermore, to illustrate the potential of eDNA sampling in the detection of designated species within MPAs, we identified Marine Conservation Zone Species of Conservation Interest (MCZ SOCI) and designated species features (SACs) from the eDNA data. We also identified UK Biodiversity Action Plan (UKBAP, 2019) priority species using curated lists of marine and freshwater fishes from UK BAP listed species. Finally, we identified potential non-native species (NNS) using listed marine and freshwater fishes from the GB Non-Native Species Information Portal (GBNNSS, 2016).

The resulting eDNA data have enabled us to undertake whole community analyses of detected fishes using Alpha diversity (species richness at each site), Beta diversity (the difference in community composition between sites) and Gamma diversity (the total species richness among all sampled sites).

We evaluated the effectiveness of our eDNA approach for monitoring inshore fish communities using available MMO historical catch data, sourced from the UK sea fisheries annual statistics report 2018, (MMO, 2019) for recorded fisheries landings for the months of October to February, 2014 to 2018 by under-10-metre fishing vessels along the South West coast of England. Within these data, we selected International Council for the Exploration of the Sea (ICES) statistical rectangles 28E4, 29E4, 29E5, 29E6, 30E6, 30E7, 30E8 to enable eDNA diversity analyses to be compared to historical catch data for the same areas (Figure 6) at the same time of year. The under-10-metre vessel category was selected under the assumption that these vessels would be more likely to be fishing within the local area and near to the coast. These data allowed us to identify economically valuable species (EVS) detected in the eDNA data and we compared eDNA detected species against fisheries landings data, which are necessarily selective and biased but are nonetheless currently used to inform fisheries management decisions.

For detected fishes listed in EVS, UKBAP and INNS, we quantified within-population Beta diversity using detected haplotype data across all locations to investigate metapopulation distributions. Time-series data were investigated along various strata (i.e. habitat type, geographical distance) to investigate the capability of this sampling technique for detecting shifts in fish community distributions over time.

Purified DNA from each sample have been bio-banked at -20$^{o}$C and will remain archived at Applied Genomics for a period of one year, unless otherwise agreed. These samples will be made available to Natural England upon request. All raw sequencing data have been stored on our servers and will be made available to Natural England upon request.


*Figure 6: Map of the UK with ICES Statistical Rectangles used for selecting MMO catch data. Source: http://gis.ices.dk/sf/index.html?widget=StatRec; accessed on 25th March 2020.*

### 2.3.2 Performance Metrics

The performance of our bioinformatics and statistical analyses have been assessed and reported using a number of techniques. The range of confidence scores for taxonomic assignment of each ASV were visualised based on the calculated fraction of bootstrap replicates (that is, repeated random sampling of possible taxonomic matches with replacement) that were assigned to the species rank (Murali *et al*., 2018). The accuracy of each taxonomic assignment is, of course, dependent on the quality of our curated reference databases and the specificity of each barcode locus to allow precise discrimination between closely related species. Therefore, to classify the risk of having committed a false detection or misclassification error for each detected species, we performed false-positive detection analyses (Colquhoun, 2019). This was accomplished using a naïve Bayesian approach to modelling the uncertainties of the taxonomic assignments and incorporating historical data of prior detections (i.e. MMO catch data, UKBAP & GBNNSS species lists) for each species (Colquhoun, 2019).

Nonmetric multidimensional scaling (NMDS; Kruskal, 1964; Faith *et al*., 1987) were used to visualise the dissimilarities between inshore fish community groupings. Beta diversities of inshore fish communities were evaluated using permutational multivariate analyses of variance (PERMANOVA; Anderson, 2001).

In an effort to ensure credibility and reproducibility of our results, we have considered our results to be statistically significant at $p < 0.05$, as per convention and intrinsically credible at $p < 0.005$, as initially discussed by Matthews (2018) and further developed by Held (2019).

# 3 Results

## 3.1 Field Sampling

Field sampling was undertaken between October 2019 and February 2020. InDepth Samplers were deployed by The Inshore Fisheries and Conservation Agency (IFCA) Officers from Cornwall, Devon & Severn, and Southern regions. Note that in this report we will refer to the counties of Cornwall, Devon and Dorset as the regions served by each of these IFCAs respectively.

The dates of each sampler deployment by each of the collaborating IFCAs are given in Table 3. Weather became an issue for sampler deployments, particularly in the months from December 2019 to February 2020 when a number of large storms made sea conditions too dangerous to undertake fieldwork. The filtered seawater samples were preserved immediately upon returning to shore using the preservative buffer solution supplied by Applied Genomics, which allowed the samples to be stored at ambient temperatures until they were returned to Applied Genomics laboratories for processing.

## 3.2 Sample Processing

The concentrations of purified eDNA, 12S and CytB amplicons are provided alongside each technical replicate identifier (i.e. A, B or C) and other sample collection metadata in Appendix 2. Details are also given for the negative control blank which was processed alongside all biological

samples. Note that the eDNA and amplicon concentrations for the negative control sample blank were below the limit of detection (less than 10 pg/µl).

*Table 3: Deployment dates at each sampling location by each collaborating IFCA. In all cases, samplers were recovered 25 hours after deployment.*

| *Location* | Deployment Date | Tide Phase | Collaborator | Sample ID | Sample Group |
|---|---|---|---|---|---|
| *Start Bay* | 14 October 2019 | Spring | D&SIFCA | STRT_OCT | 1 |
| *Lyme Bay* | 14 October 2019 | Spring | D&SIFCA | LYME_OCT | 1 |
| *Poole Rocks* | 05 November 2019 | Neap | SIFCA | POOL_NOV | 1 |
| *Studland* | 05 November 2019 | Neap | SIFCA | STUD_NOV | 1 |
| *Falmouth* | 17 November 2019 | Spring | CIFCA | FALM_NOV | 1 |
| *Fowey* | 17 November 2019 | Spring | CIFCA | FOWY_NOV | 1 |
| *Start Bay* | 27 November 2019 | Spring | D&SIFCA | STRT_NOV | 2 |
| *Lyme Bay* | 27 November 2019 | Spring | D&SIFCA | LYME_NOV | 2 |
| *Poole Rocks* | 02 December 2019 | Spring | SIFCA | POOL_DEC | 2 |
| *Studland* | 02 December 2019 | Spring | SIFCA | STUD_DEC | 2 |
| *Falmouth* | 16 December 2019 | Spring | CIFCA | FALM_DEC | 2 |
| *Fowey* | 16 December 2019 | Spring | CIFCA | FOWY_DEC | 2 |
| *Start Bay* | 05 February 2020 | Neap | D&SIFCA | STRT_FEB | 3 |
| *Lyme Bay* | 05 February 2020 | Neap | D&SIFCA | LYME_FEB | 3 |
| *Falmouth* | 18 February 2020 | Neap | CIFCA | FALM_FEB | 3 |

## 3.3 Sequencing

DNA sequencing was undertaken on an Illumina MiSeq sequencer. Visualisations of the sequencing read quality profiles are given in Appendix 3: Quality Assessment, with examples of forward and reverse reads quality for 12S and CytB loci. As is typical in Illumina MiSeq sequencing, reverse sequence reads are normally of generally lower quality than forward reads. Trimming and quality filtering were undertaken on all sequence reads with details provided in Appendix 3.

## 3.4 Bioinformatics & Taxonomic Assignment

A summary of the bioinformatics pipeline throughput is given in Table 4 showing the mean and standard deviation for all samples and technical replicates. The full tables showing throughput results for each sample and technical replicate are given in Appendix 4: Pipeline Throughput. The variable *Input* gives the number of raw sequence reads produced by the MiSeq sequencer; *Filtered* are the number of sequences remaining after trimming and quality filtering of the sequences (trimming and filtering conditions are given in Appendix 3: Quality Assessment); *Denoised* show the number of inferred sequence variants calculated from modelled sequencing errors; *Merged* is the number of successful mergers of forward and reverse sequence paired

reads; *Nonchimeric* are the numbers of reads remaining after filtering of chimeric sequence reads, which may occur as artefacts of PCR amplification or sequencing misreads. The *% Reads* are then presented to provide an indication of the pipeline efficiency, where a target of greater than 60% reads retained is considered acceptable. Amplicon sequence variants (*ASVs*) are the number of unique amplicon sequences from each genetic locus. Many of these sequences were variants of the same species, whilst some could not be resolved to the species level. Those that could be characterised as species were then summarised as *ID'd Species*. Note that negative control blanks were omitted from the mean and standard deviation calculations.

*Table 4: Summary of bioinformatics pipeline throughput for 12S and Cytb barcode loci from sample eDNA.*

| | | *Input* | *Filtered* | *Denoised* | *Merged* | *Nonchimeric* | *% Reads* | *ASVs* | *ID'd Species* |
|---|---|---|---|---|---|---|---|---|---|
| 12S \| | *Mean* | 80793 | 71567 | 70181 | 55913 | 51574 | 65 | 44 | 14 |
| 12S \| | *St. Dev.* | 45117 | 40296 | 39575 | 31845 | 29482 | 12 | 16 | 3 |
| CytB \| | *Mean* | 77873 | 68878 | 67603 | 54929 | 50845 | 64 | 44 | 14 |
| CytB \| | *St. Dev.* | 42830 | 38104 | 37577 | 32094 | 29867 | 15 | 17 | 4 |

The nonchimeric reads were then matched against our curated taxonomic databases to assign taxonomic identities to each sequence. As described in the Methods Section 2.3, nonchimeric sequences for each genetic locus were matched against our custom fish-specific databases. Sequences that failed to assign to species were then matched against larger, more generalised databases. Following taxonomic assignment, we undertook manual quality assessment of sequences classified to species. Non-target organisms, such as terrestrial and avian species (eg. rabbit (*Oryctolagus cuniculus),* bank vole *(Myodes glareolus),* great northern diver *(Gavia immer)*, etc.), were identified and removed from the data. The numbers of eDNA sequences that were classified at each taxonomic rank are shown in Table 5.

*Table 5: Numbers of eDNA sequences assigned at each taxonomic rank.*

| *Locus* | *Phylum* | *Class* | *Order* | *Family* | *Genus* | *Species* | *Uniques* |
|---|---|---|---|---|---|---|---|
| *12S* | 360 | 350 | 270 | 287 | 233 | 205 | 48 |
| *CytB* | 1580 | 1553 | 1383 | 1372 | 1276 | 1175 | 52 |

The proportion of unique organisms characterised to species rank by 12S and/or CytB primers are shown in Figure 7. This figure illustrates the importance of a multi-locus approach in eDNA-based monitoring studies (Drummond *et al*., 2015).

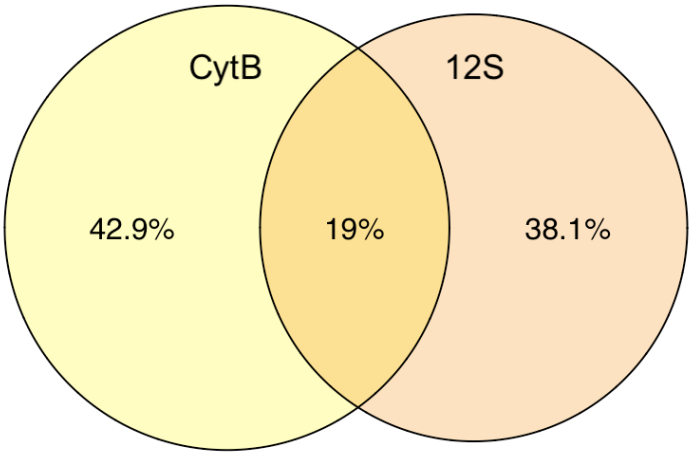



*Figure 7: Proportion of unique species characterised by 12S and/or CytB primers.*

## 3.5 Inshore Fish Community Analyses

This expanded study design with technical replicates was intended to achieve increased statistical power, sensitivity and specificity for the study. Whilst weather conditions interrupted sample collection, the addition of 3 technical replicates per sample provided a statistical power greater than 0.6 for a balanced one-way analysis of variance calculation comparing two groups and power greater than 0.5 when comparing three groups.

### 3.5.1 Alpha Diversity

Alpha diversity is the species richness of a small homogeneous community, within a given area (Whittaker, 1972). A community will have higher Alpha diversity when there is greater species richness and their abundances are very similar. In this study we characterised a total of 74 fish species and two species of marine mammals.

A permutational multivariate analysis of variance (PERMANOVA) indicated there was no statistically significant difference between replicates, even when stratified by IFCA region (i.e. county), substrate type (benthos) or sampling period (month). Similarly, there was not found to be any statistically significant difference in fish community data attributable to either of the two inDepth eDNA samplers used in this study.

Given that abundance estimates are, strictly speaking, not calculable due to compositional biases in the high-throughput sequencing workflows (Brooks *et al.,* 2015, Lamb *et al*., 2019), Alpha diversity estimates were calculated using pooled incidence coverage estimates (i.e. detection or non-detection). These calculations use the frequencies of species in a collection of sites. The variance of extrapolated richness with standard error were calculated using Chao's diversity index (Chao, 1987). Alpha diversity statistics are visualised in Figure 8.

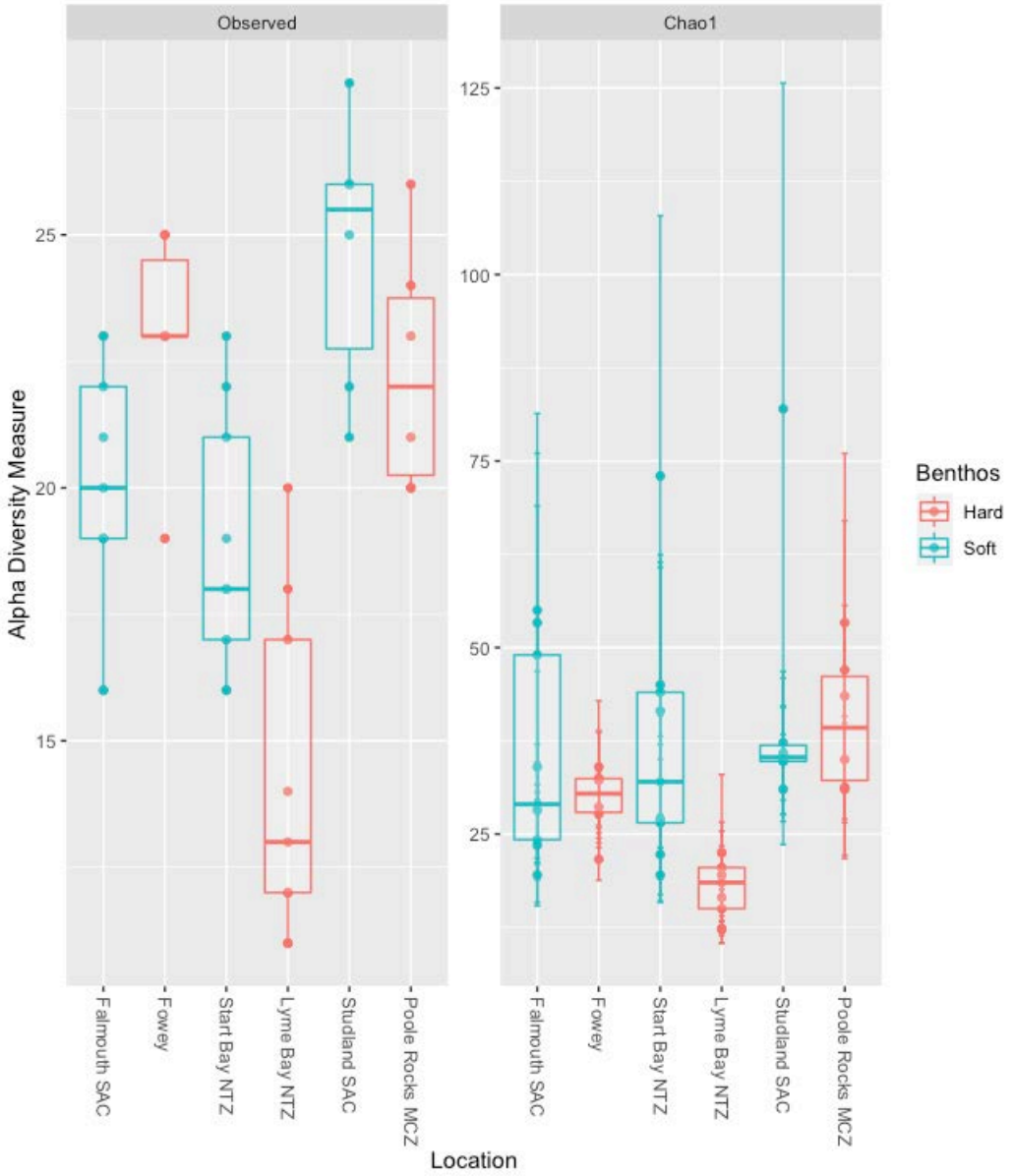



*Figure 8: Observed (left) and estimated (right) species richness at each sampling location stratified by substrate type. Points indicate observations of species richness for each technical replicate on each sampling occasion. Note there is a significant increase in scale for Chao1 estimates, as shown by the scale on the y-axis.*

### 3.5.2 Beta Diversity

Beta diversity estimation entails comparing one small and homogeneous community with another in the same general area. It is the number and composition of species unique to each community in the set of communities being compared. The greater the Beta diversity between a pair of communities, the greater the distinctiveness of the two communities (Whittaker, 1972).

Observed dissimilarities in species composition were found to be statistically significant ($p = 0.037$), though not intrinsically credible ($p > 0.005$; Held, 2019), for substrate type (“Benthos”)

using the PERMANOVA statistic (Anderson, 2001; Table 6); however these were not observed in non-metric multidimensional scaling (NMDS) ordinations (Kruskal, 1964; Faith *et al*., 1987; Figure 9), which may be explained by the low coefficient of determination ($R^2$ = 0.05) indicating the low variance in fish community data explained by benthic substrate habitats stratified by county. There were similarly no observed differences in the NMDS plots when the Beta diversity due to substrate type was decomposed into its components of species turnover or nestedness (Baselga, 2010).

*Table 6: PERMANOVA of fish community composition by substrate type, stratified by county.*

| | Df | SumOfSqs | R2 | F | Pr(>F) |
|---|---|---|---|---|---|
| sample_data(psFishes.spp)$Benthos | 1 | 0.08811119 | 0.0545307 | 2.48006 | 0.037 |
| Residual | 43 | 1.52769764 | 0.9454693 | NA | NA |
| Total | 44 | 1.61580883 | 1.0000000 | NA | NA |

Observed dissimilarities in fish community composition were found to be statistically significant and intrinsically credible (Held, 2019) by county (Table 7) using the PERMANOVA statistic. The NMDS of fish community diversity by county is shown in Figure 10.

*Table 7: PERMANOVA of fish community composition by county, stratified by substrate type.*

| | Df | SumOfSqs | R2 | F | Pr(>F) |
|---|---|---|---|---|---|
| sample_data(psFishes.spp)$County | 2 | 0.6127075 | 0.3791955 | 12.82708 | 0.001 |
| Residual | 42 | 1.0031013 | 0.6208045 | NA | NA |
| Total | 44 | 1.6158088 | 1.0000000 | NA | NA |

The spatial resolution required for eDNA sampling is illustrated in Figure 10 and Figure 11. Fish communities at sampling locations within each IFCA region, where the NMDS shows the overlap in the standard error ellipsoids of the mean of each group, showed greater similarity than fish communities between counties.

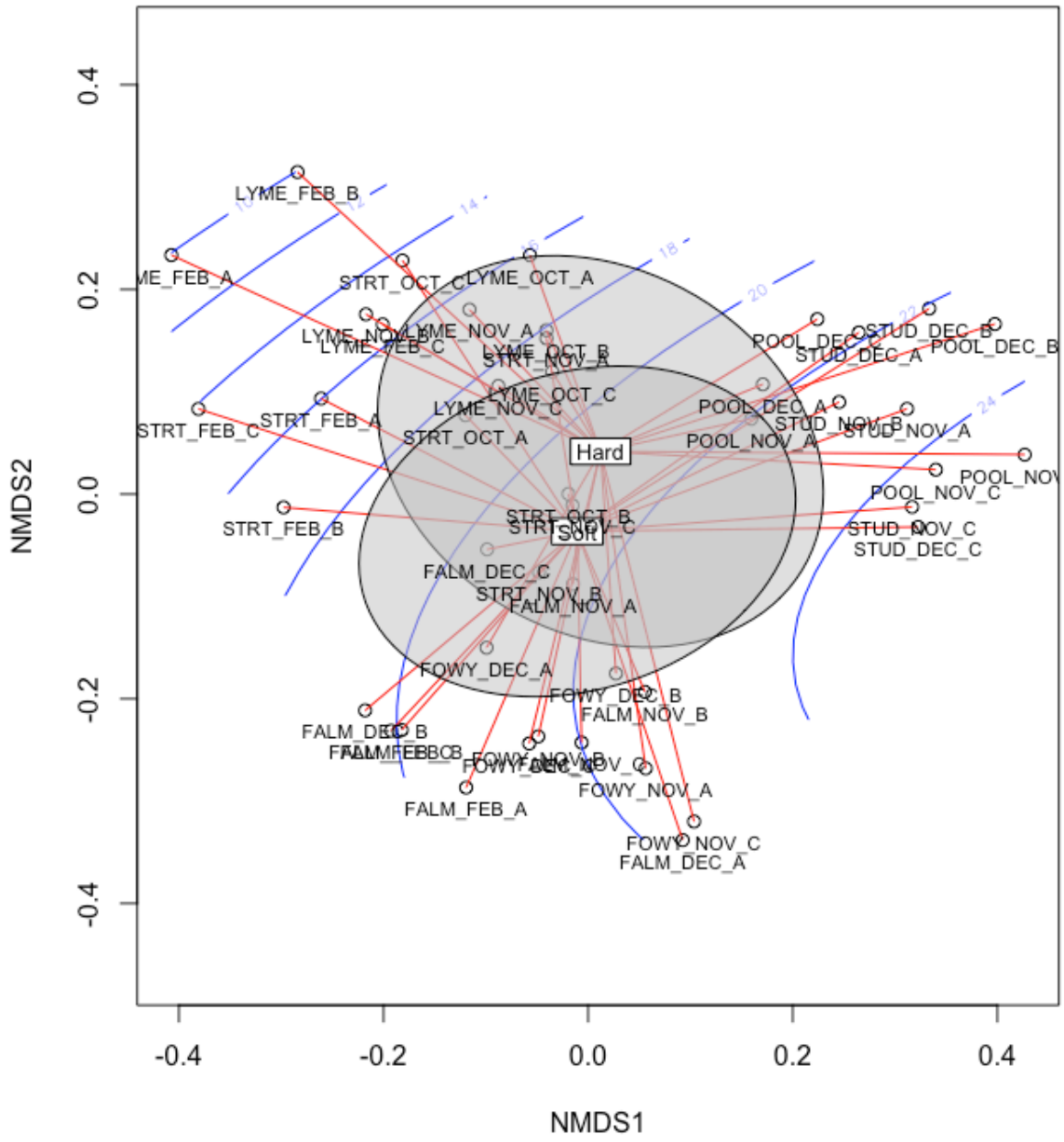



*Figure 9: Non-metric multidimensional scaling ordination of species occurrences for each sample and technical replicate by substrate type. Beta diversity was calculated using Jaccard dissimilarity matrix of species occurrences at sampling sites (points) along a richness gradient (blue lines). Points are connected (red lines) to the group centroid, indicating the mean richness, and the length of the segments indicate the variance of each point from the group mean. Benthic substrates were categorised as hard or soft. Ellipsoids display the standard error of the points and the weighted correlation defines the direction of the principal axis of the ellipse.*

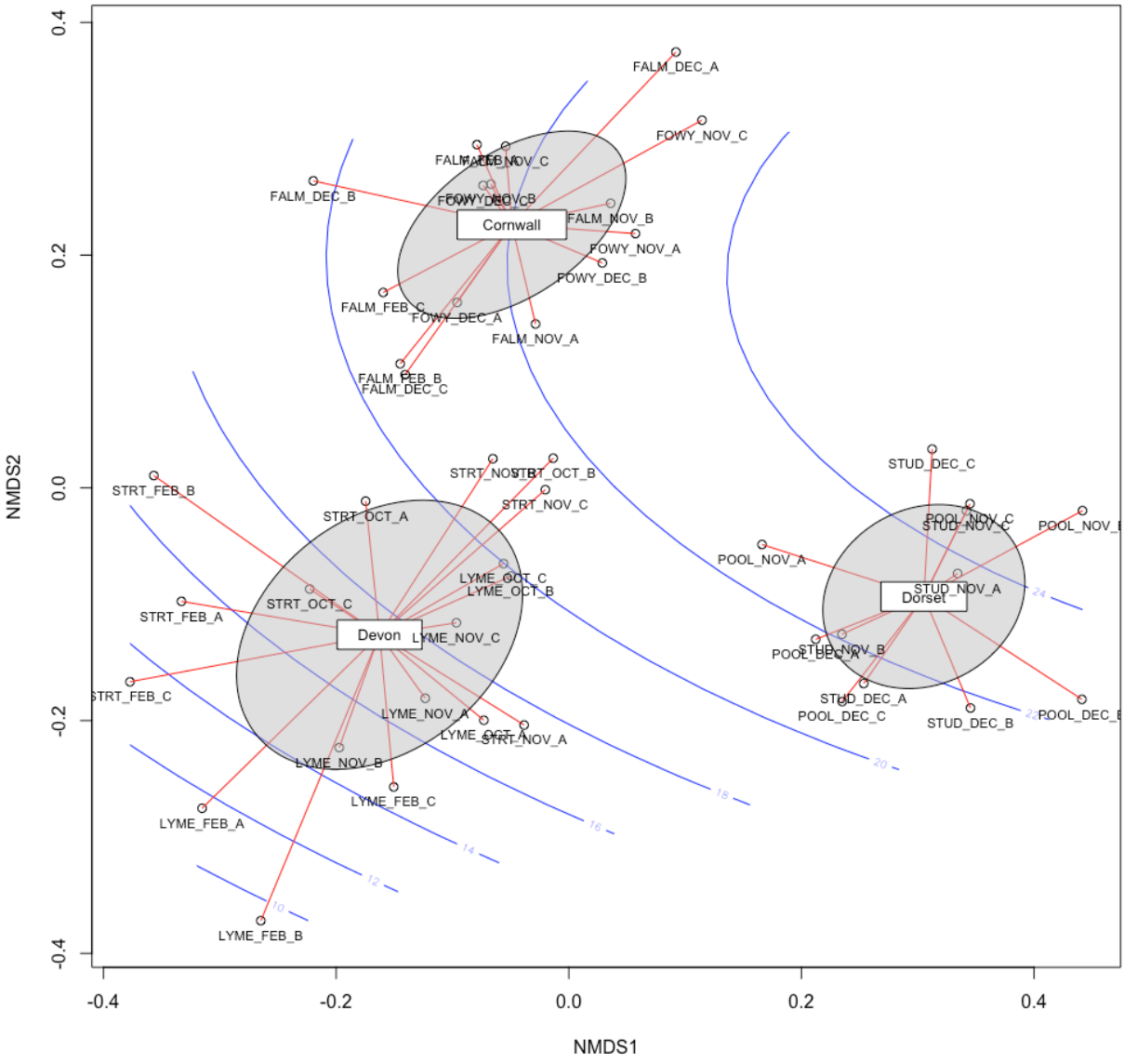



*Figure 10: Non-metric multidimensional scaling ordination of species occurrences for each sample and technical replicate by county. Beta diversity was calculated using Jaccard dissimilarity matrix of species occurrences at sampling sites (points) along a richness gradient (blue lines). Points are connected (red lines) to the group centroid, indicating the mean richness, and the length of the segments indicate the variance of each point from the group mean. Ellipsoids display the standard error of the points and the correlation defines the direction of the principal axis of the ellipse.*

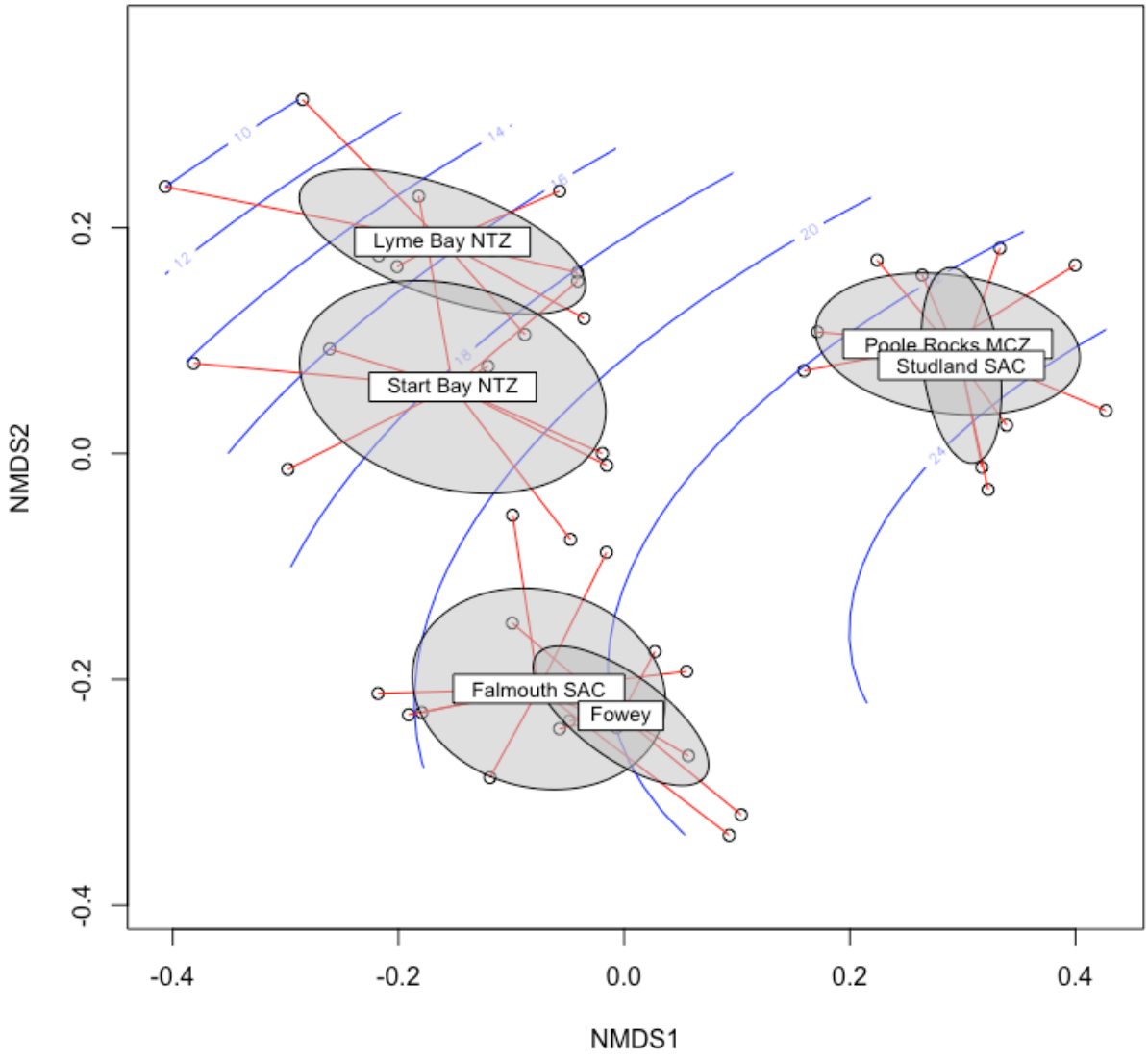



*Figure 11: Non-metric multidimensional scaling ordination of species occurrences for each sample and technical replicate by sampling location. Beta diversity was calculated using Jaccard dissimilarity matrix of species occurrences at sampling sites (points) along a richness gradient (blue lines). Points are connected (red lines) to the group centroid, indicating the mean richness, and the length of the segments indicate the variance of each point from the group mean. Ellipsoids display the standard error of the points and the correlation defines the direction of the principal axis of the ellipse. Point labels were omitted for clarity.*

### 3.5.3 Gamma Diversity

Gamma diversity is the total species diversity across the entire landscape (Whittaker, 1972). From the total 76 unique species characterisations, the estimated fish community diversity from all study samples, the Gamma diversity, was 81 ± 5 species (Chao ± Chao SE; Chao, 1987), indicating that eDNA analysis using the inDepth sampler detected an estimated 94% of fish community diversity. Species accumulation curves are shown in Figure 12, where the black line includes species detected in all technical replicates. Each of the technical replicates (A, B and C) are shown in red, green and blue lines, respectively.

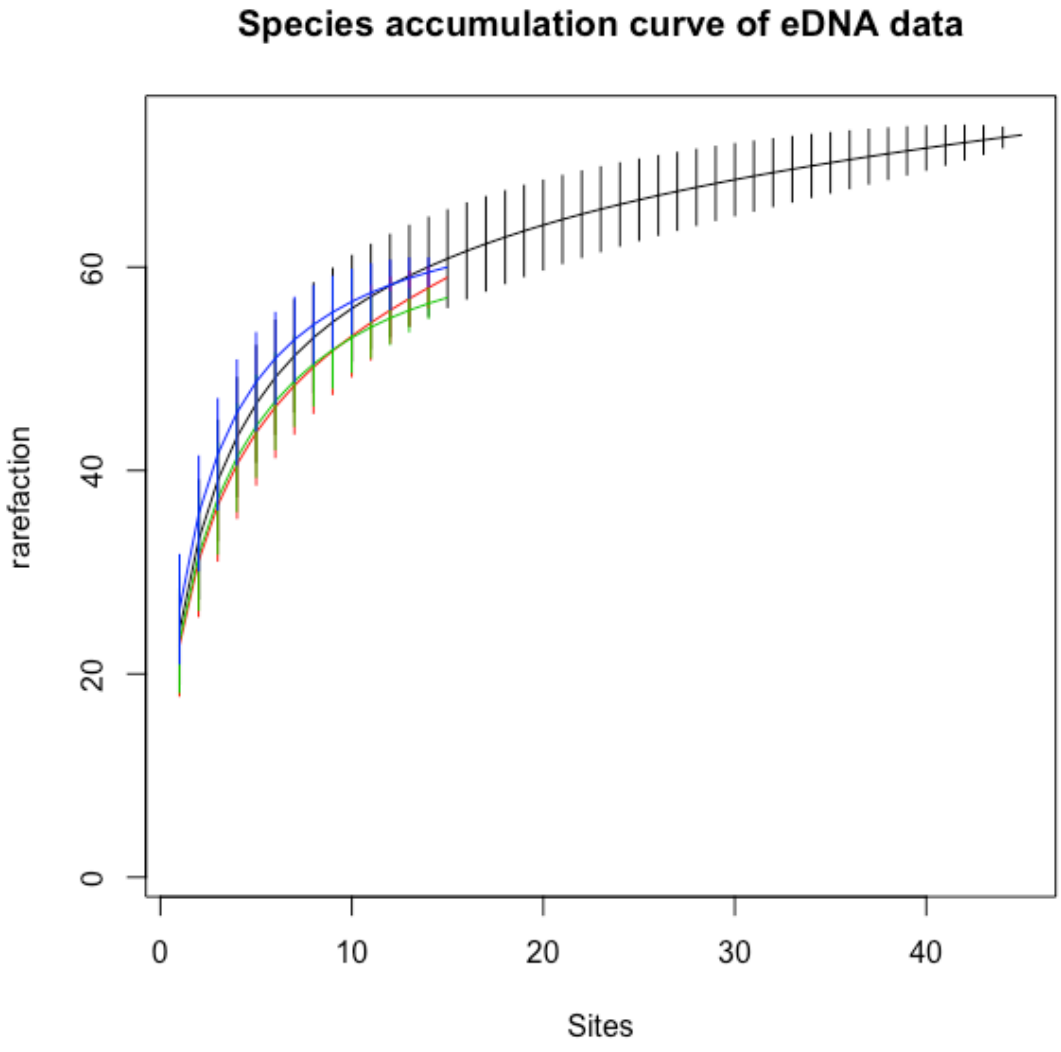



*Figure 12: Species accumulation curves; black line includes all technical replicates; red line is technical replicate A; green line shows technical replicate B; blue line shows technical replicate C.*

### 3.5.4 Identifying Designated Species in MPAs and Non-Native Species

The list of species detected using these fish-specific metabarcoding genes were cross referenced with MCZ SOCIs and SAC designated species features covered by the sampling regime (Figure 2, Table 20); the National Biodiversity Network (NBN, accessed via JNCC, 2019); the United Kingdom Biodiversity Action Plan (UKBAP, 2019) marine priority UKBAP species accounts spreadsheet; and the marine & freshwater non-native species (NNS) list (GBNNSS, 2016). The only SOCI fish species detected within MCZ sites where it is a protected feature was black seabream (*Spondyliosoma cantharus*) (Poole Rocks MCZ, Southbourne Rough MCZ, and Purbeck Coast MCZ (Table 20)). Atlantic salmon (*Salmo salar*) was detected at the Studland and Poole Rocks sampling locations, which according to the tidal and current excursion modelling partially covers the River Avon SAC within which the species is a designated feature (Figure 2; Table 20). One match was found for listed NNS and 19 species were returned as UKBAP priority listed species (Table 8). Not appearing on the UKBAP list but detected by this study were the great sand eel (*Hyperoplus lanceolatus*) and lesser sand eel (*Ammodytes marinus*) (Figure 13). Of the 76 detected species, 36 species were identified in the MMO landings data (2014-2018; MMO, 2019) and are listed in (Table 9).

*Table 8: Detected UKBAP-listed species (top, green bars) and GBNNSS-listed non-native species (bottom, red bars). Numbers indicate the detected haplotype variants for each species at each location.*

| species | Common.Names | Falmouth | Fowey | Start.Point | Lyme.Bay | Studland | Poole.Rocks |
|---|---|---|---|---|---|---|---|
| Alosa alosa | allis shad | 0 | 0 | 0 | 0 | 0 | 2 |
| Alosa fallax | twaite shad | 2 | 1 | 1 | 1 | 0 | 7 |
| Anguilla anguilla | European eel | 1 | 1 | 0 | 0 | 2 | 0 |
| Balaenoptera physalus | Fin whale | 1 | 0 | 0 | 0 | 0 | 0 |
| Clupea harengus | Atlantic herring | 16 | 39 | 16 | 4 | 13 | 15 |
| Gadus morhua | Atlantic cod | 2 | 4 | 2 | 0 | 0 | 0 |
| Lophius piscatorius | angler | 1 | 1 | 1 | 0 | 0 | 0 |
| Merlangius merlangus | whiting | 45 | 11 | 10 | 19 | 16 | 9 |
| Merluccius merluccius | European hake | 0 | 0 | 0 | 0 | 4 | 2 |
| Micromesistius poutassou | blue whiting | 1 | 0 | 0 | 0 | 0 | 0 |
| Molva molva | ling | 1 | 0 | 0 | 0 | 0 | 0 |
| Phocoena phocoena | harbor porpoise | 0 | 1 | 1 | 0 | 0 | 0 |
| Pleuronectes platessa | European plaice | 0 | 1 | 0 | 0 | 1 | 0 |
| Raja undulata | undulate ray | 0 | 0 | 0 | 2 | 11 | 1 |
| Salmo salar | Atlantic salmon | 0 | 0 | 0 | 0 | 1 | 1 |
| Scomber scombrus | Atlantic mackerel | 86 | 20 | 33 | 29 | 19 | 2 |
| Solea solea | common sole | 1 | 2 | 5 | 4 | 5 | 2 |
| Thunnus thynnus | Atlantic bluefin tuna | 1 | 0 | 0 | 0 | 0 | 0 |
| Trachurus trachurus | Atlantic horse mackerel | 1 | 1 | 1 | 1 | 1 | 1 |

| species | Common.Names | Falmouth | Fowey | Start.Point | Lyme.Bay | Studland | Poole.Rocks |
|---|---|---|---|---|---|---|---|
| Oncorhynchus mykiss | rainbow trout | 0 | 0 | 0 | 0 | 8 | 4 |



Applied Genomics Ltd is registered in England and Wales: 9180742 | VAT: GB198410583
Brixham Environmental Laboratory | Freshwater Quarry | TQ5 5AY

*Table 9: Identified fishes included in MMO landings data from 2014-2018 (MMO, 2019). Numbers indicate the detected haplotype variants for each species at each location.*

| species | MMO.Catch | Falmouth | Fowey | Start.Point | Lyme.Bay | Studland | Poole.Rocks |
|---|---|---|---|---|---|---|---|
| Anguilla anguilla | European eel | 1 | 1 | 0 | 0 | 2 | 0 |
| Ciliata septentrionalis | northern rockling | 1 | 15 | 24 | 5 | 12 | 0 |
| Clupea harengus | Atlantic herring | 16 | 39 | 16 | 4 | 13 | 15 |
| Ctenolabrus rupestris | goldsinny wrasse | 4 | 3 | 1 | 0 | 0 | 1 |
| Dicentrarchus labrax | European seabass | 5 | 3 | 5 | 3 | 14 | 5 |
| Engraulis encrasicolus | European anchovy | 60 | 19 | 16 | 18 | 6 | 20 |
| Eutrigla gurnardus | grey gurnard | 2 | 1 | 1 | 1 | 1 | 0 |
| Gadus morhua | Atlantic cod | 2 | 4 | 2 | 0 | 0 | 0 |
| Labrus bergylta | ballan wrasse | 0 | 3 | 0 | 0 | 5 | 1 |
| Liza aurata | golden grey mullet | 0 | 0 | 1 | 0 | 1 | 3 |
| Lophius piscatorius | angler | 1 | 1 | 1 | 0 | 0 | 0 |
| Melanogrammus aeglefinus | haddock | 1 | 2 | 0 | 0 | 0 | 0 |
| Merlangius merlangus | whiting | 45 | 11 | 10 | 19 | 16 | 9 |
| Merluccius merluccius | European hake | 0 | 0 | 0 | 0 | 4 | 2 |
| Molva molva | ling | 1 | 0 | 0 | 0 | 0 | 0 |
| Mullus surmuletus | striped red mullet | 2 | 0 | 2 | 0 | 9 | 1 |
| Pagrus pagrus | common seabream | 1 | 2 | 0 | 16 | 2 | 1 |
| Pegusa lascaris | sand sole | 0 | 0 | 0 | 0 | 0 | 0 |
| Platichthys flesus | European flounder | 0 | 0 | 0 | 0 | 2 | 0 |
| Pleuronectes platessa | European plaice | 0 | 1 | 0 | 0 | 1 | 0 |
| Pollachius pollachius | pollack | 3 | 5 | 1 | 2 | 2 | 3 |
| Pollachius virens | saithe | 0 | 0 | 0 | 1 | 0 | 0 |
| Raja brachyura | blonde ray | 0 | 0 | 8 | 1 | 0 | 0 |
| Raja clavata | thornback ray | 0 | 0 | 0 | 0 | 3 | 0 |
| Raja microocellata | small-eyed ray | 0 | 0 | 5 | 0 | 2 | 1 |
| Raja montagui | spotted ray | 0 | 0 | 0 | 0 | 1 | 0 |
| Raja undulata | undulate ray | 0 | 0 | 0 | 2 | 11 | 1 |
| Sardina pilchardus | sardine | 18 | 9 | 9 | 9 | 4 | 4 |
| Scomber scombrus | Atlantic mackerel | 86 | 20 | 33 | 29 | 19 | 2 |
| Scophthalmus maximus | turbot | 0 | 0 | 1 | 0 | 0 | 1 |
| Solea solea | common sole | 1 | 2 | 5 | 4 | 5 | 2 |
| Spondyliosoma cantharus | black seabream | 0 | 0 | 0 | 2 | 0 | 3 |
| Sprattus sprattus | European sprat | 20 | 11 | 25 | 28 | 4 | 24 |
| Trachurus trachurus | Atlantic horse mackerel | 1 | 1 | 1 | 1 | 1 | 1 |
| Trisopterus luscus | pouting | 47 | 37 | 37 | 49 | 52 | 72 |
| Zeus faber | John dory | 0 | 0 | 0 | 3 | 0 | 0 |

### 3.5.5 Haplotype Analysis

Effective population size is the number of individuals in a population who contribute offspring to the next generation (Wright, 1937; Ridley, 2003). Following taxonomic assignment, a number of amplicon sequence variants were assigned to the same species. Where these occurred, the detected variants may be regarded as haplotypic variants for the same genetic locus (Callahan, 2019). Each individual within a population will have one genetic haplotype but several closely related individuals may share the same haplotype, usually through maternal inheritance (Billington & Hebert, 1991). So, whilst not an indication of numbers of individuals, haplotype diversity is an indication of the genetic diversity of breeding individuals within the sampled population (Wright, 1937; Ridley, 2003).

Alongside haplotype diversity visualisations, Figure 13 displays quality assurance metrics for the taxonomic assignment and detection credibility. The range of confidence in taxonomic assignments are shown in the grid on the left, where confidence was calculated as the fraction of bootstrap replicates that were assigned to the species rank. A risk-based analysis of detection credibility is shown in the centre grid. This analysis examined the risk of having committed a false detection or misclassification error, known as a false-positive (Colquhoun, 2019). This was accomplished by modelling the uncertainties of the taxonomic assignments and incorporating historical data of previous detections of each species (as evidenced in species recorded in MMO landings, UKBAP or non-native species lists). The Credibility score was classified as “Low” (red triangle), “Moderate” (yellow diamond) and “High” (green circle). The threshold for “High” credibility was defined by the intrinsic credibility statistic (Held, 2019; Matthews, 2018). The threshold for the “Low” credibility category is a function of the statistical power of the study. Thus, in a sufficiently powered study, each detected species would have lower false-positive risk and higher credibility.

It is important to note that the credibility score is a decision-making tool that summarises the weight and quality of the evidence for the detection of each species. A classification at “Low” credibility does not imply the result is false; rather it indicates the amount of additional information that may be needed before basing decisions on the presence of a particular species. For example, the detection of a Fin whale (*Balenoptera physalus*) was the result of detecting a single haplotype variant in a single technical replicate from one sample at a single location. Because the species is listed in historical data records (a UKBAP-listed species), the species was previously observed in UK waters and thus this detection was scored as “moderately” credible. Incidentally, on the 14th February 2020, 3 days prior to deployment of the inDepth eDNA sampler at Falmouth, a Fin whale had beached itself at Nare Point on the Lizard Peninsula in Cornwall, as was reported in the local news. The location of the stranding was near to the sampler deployment location and within the modelled tidal excursion area, thereby providing evidence that the eDNA detection was credible.

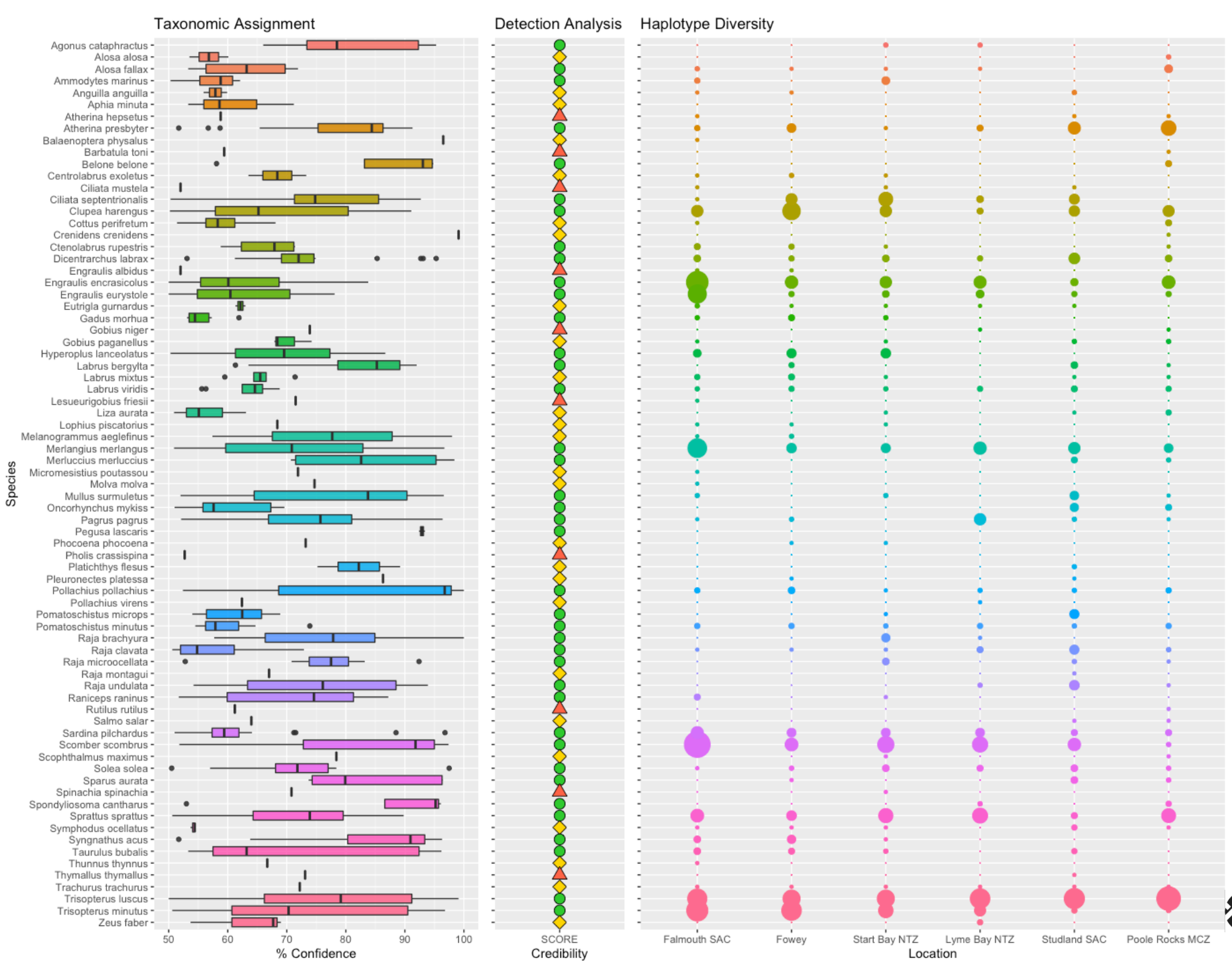



*Figure 13: Full list of eDNA detected species with range of confidence values for taxonomic assignments (left), risk-based analyses of detection credibility (centre) and haplotype diversity for each sampled location (right).*

Applied Genomics

#### 3.5.5.1 Within-Species Haplotype Diversity Analysis

We may also consider detected haplotypes as indicators of metapopulation dynamics which could uncover potential range shifts and gene flow patterns within species (De Jong *et al*., 2011; Hwang & Cho, 2018). For example, here we considered the species with the highest overall haplotype diversity detected in this study, *Trisopterus luscus*, where 184 unique haplotypes were detected across all samples. Though the species was found to occur in all sampled locations, we found evidence to suggest regional differences within the population (Table 10).

*Table 10: PERMANOVA of T. luscus haplotype diversity by county, stratified by benthic habitat type.*

| | Df | SumOfSqs | R2 | F | Pr(>F) |
|---|---|---|---|---|---|
| sample_data(T.luscus)$County | 2 | 0.6267339 | 0.1356373 | 3.295356 | 0.001 |
| Residual | 42 | 3.9939274 | 0.8643627 | NA | NA |
| Total | 44 | 4.6206613 | 1.0000000 | NA | NA |

Whilst constructing and mapping haplotype networks for multi-locus data were beyond the scope of this study, we visualised the haplotype diversity using non-metric multidimensional scaling (Figure 14).

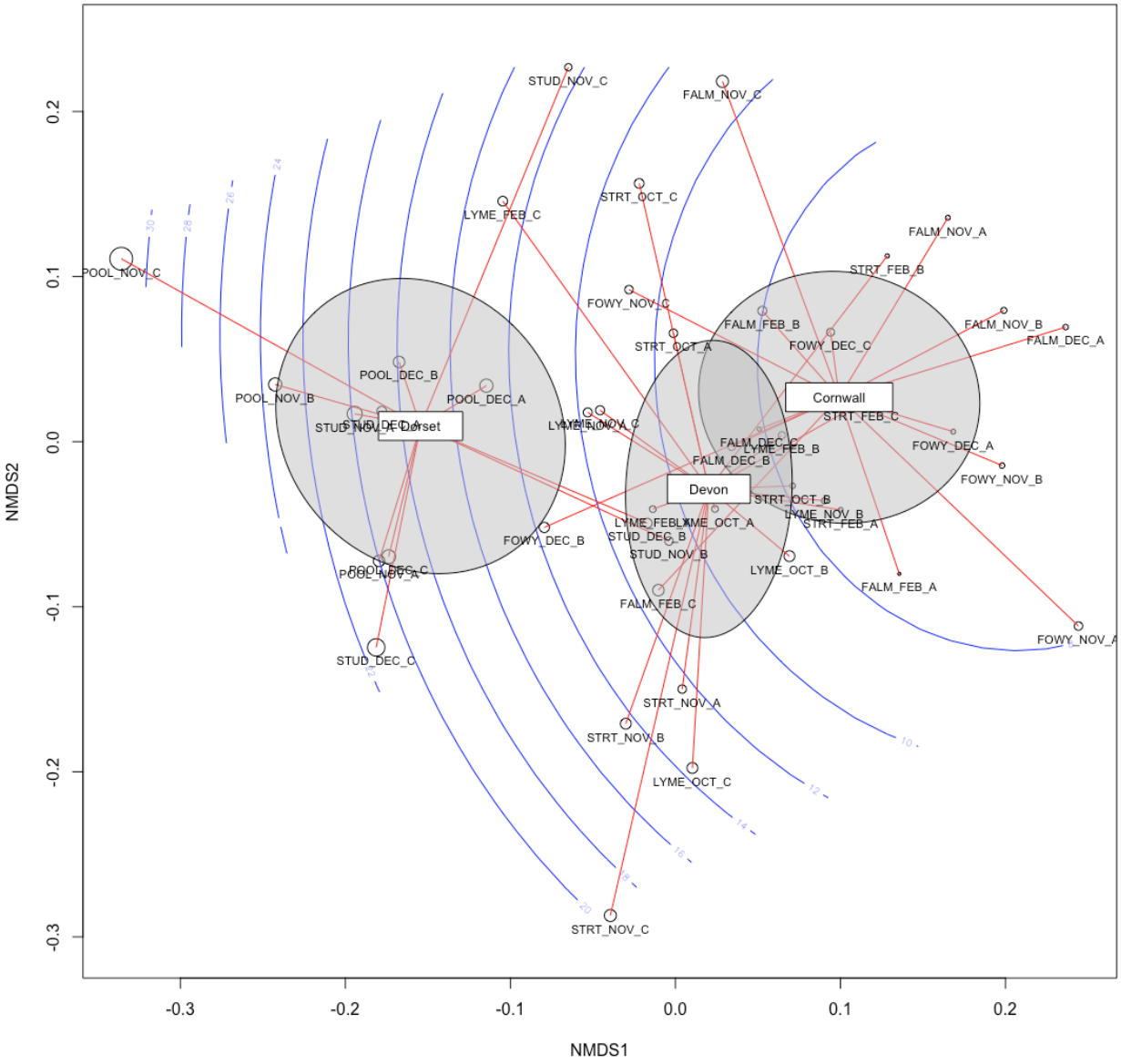



*Figure 14: Non-metric multidimensional scaling ordination of haplotype diversity of T. luscus detected in each replicate by sampling location. Beta diversity was calculated using the Jensen-Shannon Divergence matrix of sequence occurrences at sampling sites (points) along a diversity gradient (blue lines). Point sizes indicate the relative numbers of detected haplotypes within each replicate. Points are connected (red lines) to the group centroid, indicating the mean diversity, and the length of the segments indicate the variance of each point from the group mean. Ellipsoids display the standard error of the points and the correlation defines the direction of the principal axis of the ellipse.*

### 3.5.6 Time-Series Analysis

Changes in fish community compositions over time were investigated by comparing differences in species composition at each sampling location across monthly sampling rounds, which were found to be statistically significant and intrinsically credible (Table 11).

*Table 11: PERMANOVA of fish community composition by month, stratified by sampling location.*

| | Df | SumOfSqs | R2 | F | Pr(>F) |
|---|---|---|---|---|---|
| sample_data(psFishes.spp)$Month | 3 | 1.700498 | 0.1552772 | 2.512211 | 0.001 |
| Residual | 41 | 9.250869 | 0.8447228 | NA | NA |
| Total | 44 | 10.951367 | 1.0000000 | NA | NA |

Fish community data from eDNA samples collected during the 2018/19 pilot study and this 2019/20 study were visualised in a Venn diagram (Figure 15) with MMO landings data over the same monthly periods included.

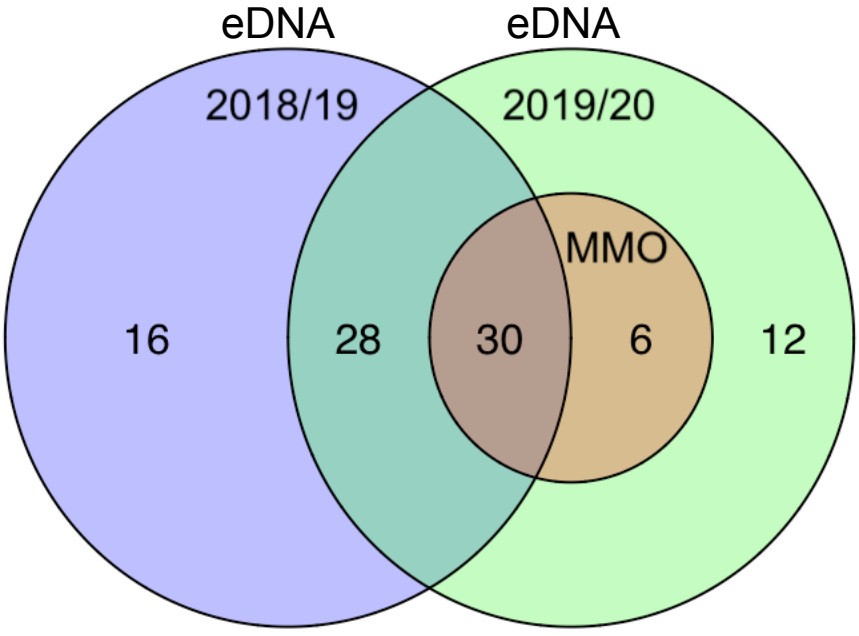
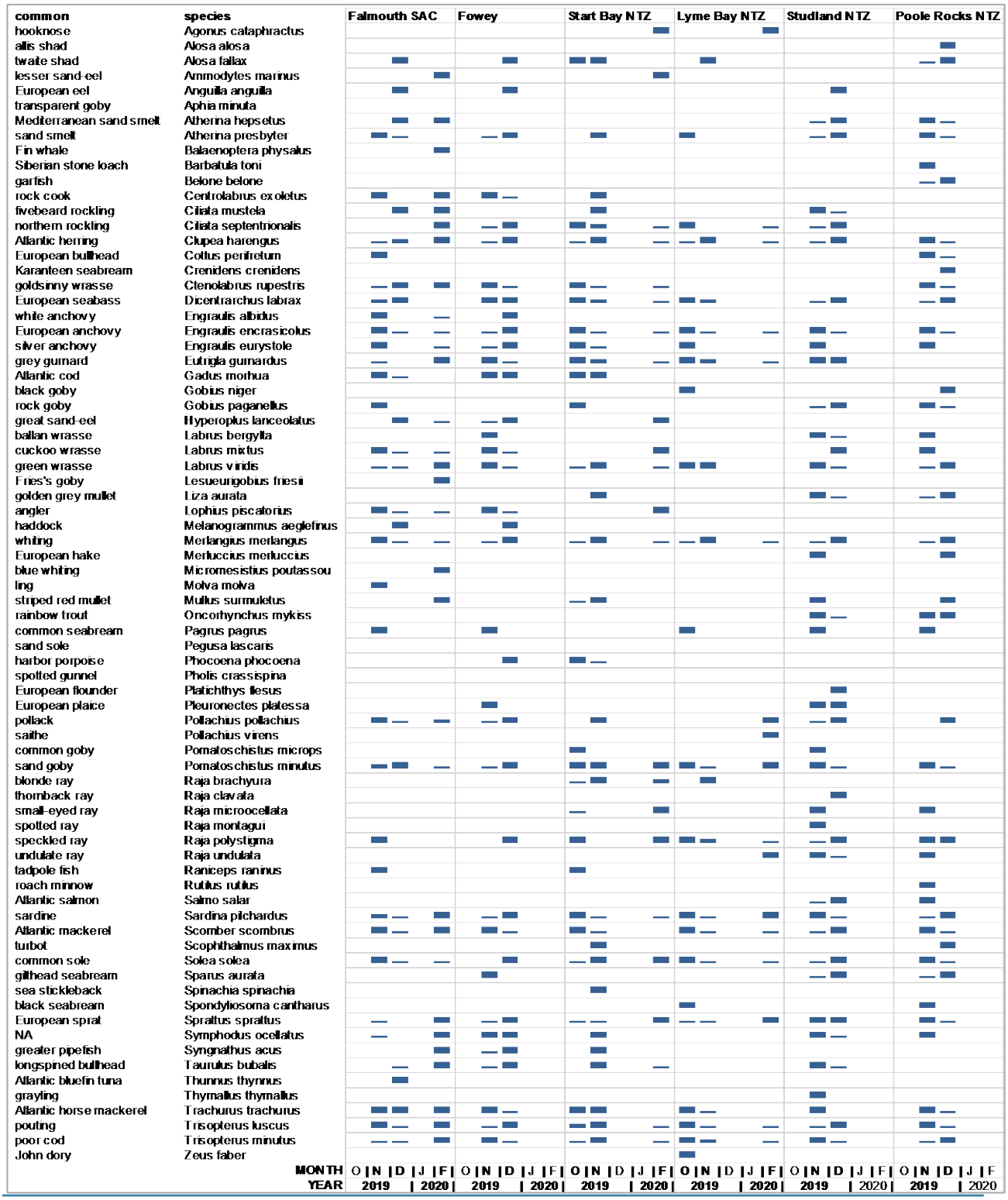



*Figure 15: Venn diagram showing numbers of species detected in 2018/19 and 2019/20 eDNA surveys with MMO landings data from 2014-2018 by under 10m vessels over the same monthly time periods in the same coastal areas.*

Table 12 provides a visual summary of increases and decreases in haplotype diversity for each species at each site over the course of the 5-month survey period.

*Table 12: Table of proportional changes to species haplotype diversity over the sampling period October 2019 to February 2020. Each bar plot indicates the proportional haplotype diversity for each species in each month at each sampling location. Sampled months in each location are in **bold**.*



### 3.5.6.1 Year-on-Year Comparison

Monthly sampling data from the Devon sampling locations, Start Bay and Lyme Bay, were combined with pilot study (NECR287) sample data collected near Torbay and Dartmouth over the same period in the previous year to create a year-on-year (YoY) dataset. Differences in year-on-year fish community compositions within these data were statistically significant and intrinsically credible (Held, 2019) (Table 13).

*Table 13: PERMANOVA showing year-on-year comparison of fish community composition by study year.*

| | Df | SumOfSqs | R2 | F | Pr(>F) |
|---|---|---|---|---|---|
| sample_data(YoY)$Study | 1 | 0.2801184 | 0.176345 | 7.279419 | 0.001 |
| Residual | 34 | 1.3083495 | 0.823655 | NA | NA |
| Total | 35 | 1.5884679 | 1.000000 | NA | NA |

Year-on-year survey results were analysed by selecting only samples collected in Devon in this 2019/20 study (samples collected from Lyme Bay (LYME) & Start Bay (STRT)) and comparing them with results from the 2018/19 pilot study (samples collected near Orestone Rock (ORE), Berry Head (BHD) and Dartmouth (DRT); NECR287). Figure 16 shows the NMDS for these samples. Both sets of samples were collected over similar calendar periods.

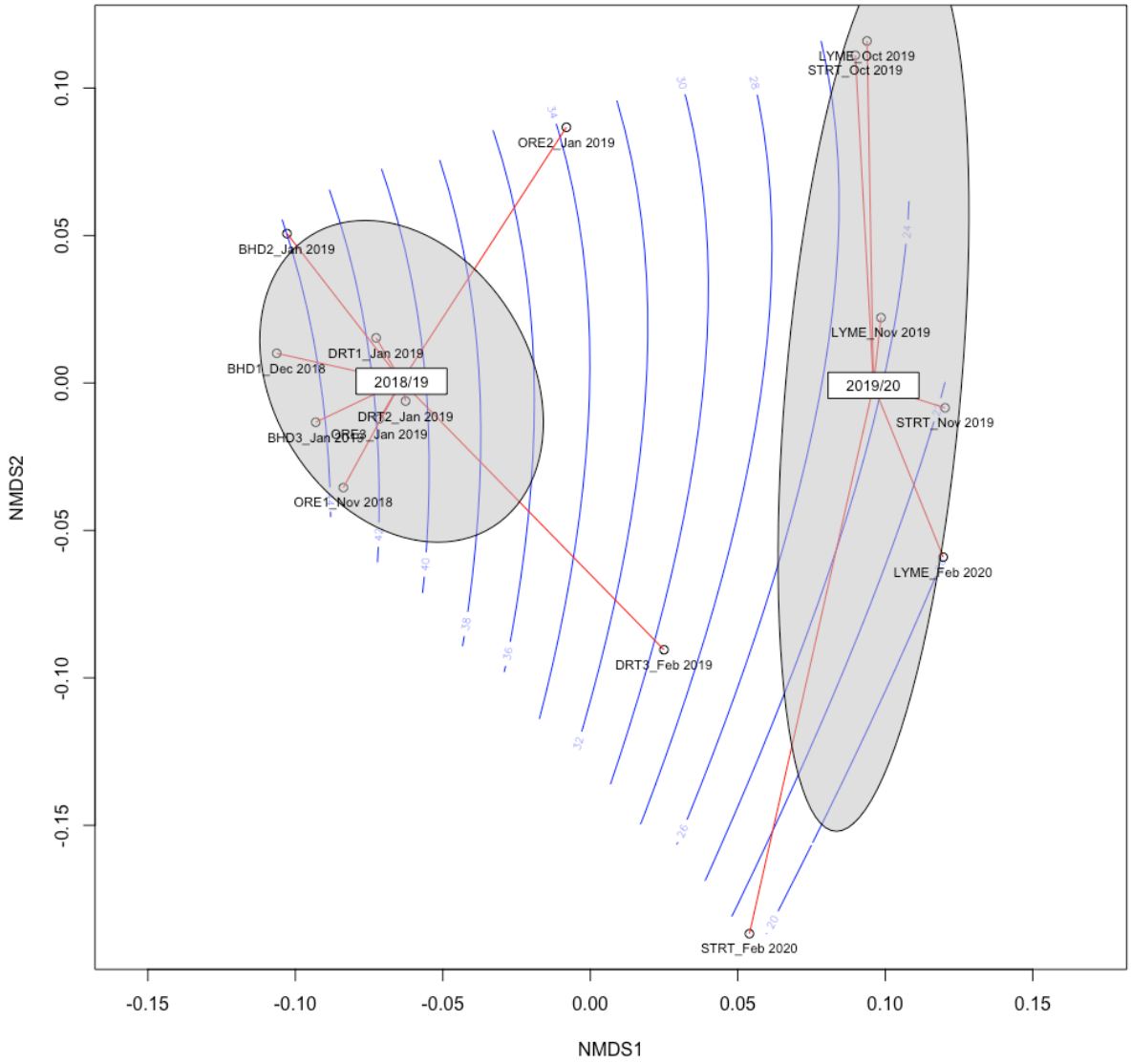



*Figure 16: Non-metric multidimensional scaling ordination of year-on-year beta diversity of inshore fish communities in south Devon. Beta diversity was calculated using the Jensen-Shannon Divergence matrix of species occurrences at sampling sites (points) along a richness gradient (blue lines). Points are connected (red lines) to the group centroid, indicating the mean richness, and the length of the segments indicate the variance of each point from the group mean. Ellipsoids display the standard error of the points and the correlation defines the direction of the principal axis of the ellipse. Point labels were omitted for clarity.*

The variation in detected haplotypes between studies may be seen in the top-10 most diverse species by numbers of haplotypes detected in each of the 2018/19 and 2019/20 studies (Table 14). Note that a total of 9 biological samples (3 samples from 3 locations) were collected in the 2018/19 study and are here compared to the 6 samples that were collected in Devon (3 samples from 2 locations) in this 2019/20 study.

*Table 14: Top 10 most diverse fishes by numbers of haplotypes detected in each of the 2018/19 pilot study and samples collected in Devon only in this 2019/20 study. Arrows indicate differences in detected haplotype diversity between studies.*

| **2018/19** | | **2019/20** | |
|---|---|---|---|
| *Sprattus sprattus* | 224 | *Trisopterus luscus* | 184 |
| *Merlangius merlangus* | 179 | *Scomber scombrus* | 101 |
| *Engraulis encrasicolus* | 162 | *Trisopterus minutus* | 77 |
| *Scomber scombrus* | 116 | *Sprattus sprattus* | 71 |
| *Clupea harengus* | 110 | *Merlangius merlangus* | 70 |
| *Trisopterus luscus* | 70 | *Sardina pilchardus* | 68 |
| *Ciliata septentrionalis* | 51 | *Engraulis encrasicolus* | 65 |
| *Trisopterus minutus* | 49 | *Ciliata septentrionalis* | 37 |
| *Belone belone* | 38 | *Clupea harengus* | 35 |
| *Engraulis eurystole* | 37 | *Dicentrarchus labrax* | 26 |

# 4 Discussion

In this study, we have investigated the effectiveness of the inDepth eDNA sampler for monitoring inshore fish communities over time, across varying geographical distances and between different substrate types. Table 15 outlines the original objectives of this study and highlights the relevant results and summary findings for each objective. In total, fifteen large-volume programmable eDNA water samples (approximately 50 litres per sample filtered *in situ* over 25 hours at 1 metre above the benthos) were collected at 6 sampling locations along the south coasts of Cornwall, Devon and Somerset over a period of 5 months. Once samples were processed and eDNA purified, three technical replicate amplicons for both 12S and CytB barcode regions were generated using PCR. Detections from both barcode loci were merged, producing a total of 45 technical replicate samples. There were an average 80,793 sequencing reads coverage for each 12S amplicon and 77,873 sequencing reads coverage for each CytB amplicon. Following bioinformatic processing and taxonomic assignment, there were 48 species detected by 12S amplicons and 52 species detected by CytB amplicons. Once merged, these resulted in a total of 74 detected fish species (Figure 14) and two marine mammal species.

*Table 15: Results sections and summary findings relevant for each study objective*

| Study objective | Relevant sections | Summary findings |
|---|---|---|
| 1. **To test the feasibility of the inDepth eDNA sampler in monitoring the composition of fish communities across a range of habitat specific deployments in the south west.** | 3.5 Inshore fish community analyses;<br>3.5.1 Alpha diversity;<br>3.5.2 Beta diversity;<br>3.5.4 Identifying protected and non-native species. | This study clearly demonstrated the operational practicality of the inDepth eDNA sampling system, particularly when compared to conventional marine scientific survey methods, ensuring that field teams would be able to reliably and cost-effectively collect consistent eDNA samples at regular intervals, enabling functional responsive monitoring of inshore fish communities.<br>The inDepth eDNA sampler characterised 74 fish species from all 6 sampling sites (Figure 8), which is more than twice as many than were detected through MMO landings data collected over five years from the same area (Table 9).<br>The fish communities were significantly different between the three deployment counties (Table 7) but the results were less conclusive for hard and soft habitat substrates within each county (Figure 9, though statistical differences were found; Table 6). Further work on habitat specificity is needed. |
| 2. **To assess the eDNA data for its ability to assess the presence and scale of breeding populations,** | 3.5.1 Alpha diversity; | This study demonstrates that this technology can robustly and consistently assess inshore fish populations and delivers analytical outputs accompanied by detailed data |

| Study objective | Relevant sections | Summary findings |
|---|---|---|
| **including Alpha, Beta and Gamma diversity statistics and haplotype diversity.** | 3.5.2 Beta diversity;<br>3.5.3 Gamma diversity;<br>3.5.5 Haplotype analysis;<br>3.5.5.1 within-species haplotype analysis | confidence analyses which are crucial for end-user interpretation.<br>By consistently detecting more than twice as many fishes as were reported in the MMO landings data (Figure 15), these results confirmed the findings of the 2018/19 pilot study on the suitability of large-volume eDNA sampling for monitoring inshore fish communities.<br>The expanded study design with increased number of samples and technical replicates significantly improved statistical power for community analyses, which was lacking in the pilot study. This allowed statistical comparison of diversity indices and haplotype diversity over time (Table 12; Table 14) and between sampling locations (Figure 13), as well as comparison of fish communities between substrate types (Figure 9). |
| **3. To statistically analyse the data to determine the sampling frequency needed when using a large-volume programmable marine eDNA sampler for measuring changes to the inshore fish communities in each habitat location.** | 3.5.6 Time-series analysis;<br>3.5.6.1 Year-on year comparison | The inDepth sampler deployment frequency used in this study enabled detection of statistically significant change in inshore fish community diversity over time (Table 12).<br>A comparison of the 2018/19 pilot study and this study showed significant difference in the inshore fish communities sampled within the same location (Figure 16, Table 14).<br>Follow-on studies should focus on regular monthly sampling over a longer time-period to detect seasonal phenology of inshore communities and provide increased confidence in results. |
| **4. To determine the spatial scale of sampling necessary to measure and infer changes to the inshore fish communities that would assist NE and IFCAs in developing appropriate advice.** | 3.5.2 Beta diversity;<br>3.5.4 Identifying protected and non-native species | The sampling regime used here allowed sufficient statistical power for detecting a significant difference in fish communities between the sampling locations at a larger geographical scale (i.e. for each IFCA region; Table 7, Figure 10). We also improved our understanding of the distances at which eDNA was able to resolve metapopulations for inshore fish communities (Figure 11, Figure 13).<br>The key finding is that we recommend sampler placement be informed by modelling of tides and currents such that there are |



Applied Genomics Ltd is registered in England and Wales: 9180742 | VAT: GB198410583
Brixham Environmental Laboratory | Freshwater Quarry | TQ5 5AY

| Study objective | Relevant sections | Summary findings |
|---|---|---|
| | | sufficient distances between sampler placements to avoid spatial autocorrelation and provide sufficient spatial resolution. |
| **5. To determine the presence and distribution of marine Species of Conservation Interest (SOCI), SAC species features, UKBAP priority species, and non-native species (NNS), within the survey area.** | 2.3.1: Inshore fish community analyses;<br>3.5.4: Identifying protected and non-native species | Whilst this study did not specifically focus on SOCI, SAC features, the data revealed that these techniques were clearly suited to detection of UKBAP priority species and NNS species, which were detectable using the targeted fish-specific primers and fish species already in the DNA reference libraries. Although sampling locations did not target SOCI of SAC species features, the results demonstrate the potential of species monitoring using this technique.<br>To further investigate the presence of other phyla within these protected areas, it is recommended that the data could be reanalysed using generic primers to focus on a broader range of priority species and NNS. |
| **6. Subsidiary objective: to test (in a restricted set of locations) the potential ability of eDNA data to assess species fidelity to key protected habitats by manipulation of the temporal sampling regime. For example, at a particular location comparing the eDNA signatures from samples collected during slack and peak tidal flow periods.** | 3.5.1 Alpha diversity;<br>3.5.2 Beta diversity. | This subsidiary objective had aimed to resolve this question by adjusting the inDepth sampler programme to improve targeting of these habitats. This could not be achieved due to the short duration and limited resources available for this study.<br>Although this study detected a statistical difference in fish communities between hard and soft substrates (Table 6), the results were inconclusive (Figure 8, Figure 9) due to tidal and current movements confounding the eDNA signal.<br>Further work is needed to investigate species habitat fidelity using eDNA. |

## 4.1 Objectives 1 and 2: Assessing the performance of large-volume eDNA sampling as a method for characterising inshore fish communities

We have demonstrated that large-volume eDNA sampling can provide a highly efficient method for characterising inshore fish communities. More than twice as many fish species were detected in the fifteen eDNA samples than in five years' of MMO landings data, which are necessarily limited to species of commercial value, collected from the same area during the same time of year (Figure 15). In addition to these economically valuable species, more than half of the species characterised in our study are fish species that may otherwise have been impacted by fishing

pressures but are not recorded in the landings data, thus indicating that eDNA sampling provides a more balanced method for monitoring inshore fish communities (see Figure 13 for a list of fish species detected). Traditional methods (e.g. netting and trapping) and visual sampling techniques (e.g. acoustic and observational surveys) for monitoring inshore fish community diversity, variability and condition, and population dynamics tend to be expensive, labour intensive, and selective to certain species, phenotypic morphologies, or life stages (Franco *et al.,* 2020a). Scientific survey methods which rely on conventional fishing techniques will also have the same unintended impacts as those activities, including bycatch, damage to seabed habitats, and mortalities. Our study shows that large-volume eDNA surveys could provide an efficient and reliable option for continuous and independent monitoring of inshore fish communities that avoids many of the problems associated with conventional survey techniques.

In the past, the techniques for eDNA sampling of aquatic ecosystems have varied particularly with regards to the volume of water sampled, even though this has been shown to be crucial for species detection. Cantera *et al.,* (2019) showed that the optimal DNA sampling effort consists of filtering 34 to 68 litres to characterise most of the fish species in highly diverse aquatic ecosystems. Use of the inDepth eDNA sampler allows automated collection of approximately 50 litres of moving seawater with little survey effort required, as all that is involved is deploying the sampler in the study location and collecting it when the predefined sampling period is over; thereby providing an optimal sample collection methodology for inshore fish communities that meets the estimated requirement for sampling effort. This strategy of collecting large-volume eDNA samples over multiple tidal cycles significantly reduces false-negative risk compared to more conventional eDNA survey techniques which have used smaller water volumes collected at a single point in time. Though it is possible that increasing sampling duration per deployment may have increased the false-negative risk of data loss due to DNA degradation, the current sample collection period of 25 hours is near to the modelled half-life for eDNA in inshore waters (21.2 hours; Collins *et al.,* 2018). Crucially, the ease-of-use of the inDepth eDNA sampling system, particularly when compared to conventional marine scientific survey methods, ensures that field teams would be able to reliably collect consistent eDNA samples which are representative of the wider marine environment.

Following the recommendations of the pilot report, the design of this study was expanded by increasing the number of geographical regions (i.e. IFCA regions) and technical replicates collected, as well as by extending the duration of sample collection to two full tidal cycles. The species accumulation curve (Figure 12) shows that the chosen study design allowed characterisation of approximately 94% of the fish species likely to be present in the entire study area. For all the species detected, we have calculated quality assurance metrics for the taxonomic assignment and detection credibility (Figure 13), which provides further assessment of data quality and transparency in the results. The adjustments in study design resulted in improved power for statistical analyses and allowed comparison of diversity indices and haplotype diversity over time (Table 7; Table 11; Table 13), and between sampling locations (Figure 10; Figure 11; Figure 13) and substrate types (Table 6).

We considered the operational practicality of eDNA sampling using the inDepth eDNA sampling methodology to help define the potential and capability of future DNA-based marine monitoring at

a national scale. The addition of technical replicates through the resampling of biological samples proved to be an economical technique for increasing the sensitivity and specificity of the data without distorting results. However, as shown by the year-on-year comparison (Table 13), biological replicates (i.e. replicate samples collected from the same location within the same time-period) outperformed technical replicates for characterising species richness. This agrees with Beentjes *et al.,* (2019). Biological replicates should therefore be used in preference to technical replicates, where resource constraints allow. Further fine-tuning and investigation of the optimal sampling regime for operational molecular monitoring of inshore fish communities will be possible once more large-volume eDNA data have been collected and analysed from different locations and environmental conditions and over multiple seasons.

## 4.2 Objectives 3 and 4: Spatial and temporal assessment of inshore fish communities

Our study detected dissimilarities in fish community composition that were spatially correlated with larger geographical distances, suggesting regional differences between inshore fish communities (Table 7; Figure 10). We also detected significant temporal changes in fish community composition within IFCA regions over the study period (Table 11) and within the same sampling locations between the 2018/19 pilot study and this study (Table 13; Figure 15). The key findings related to this are that we recommend sampler placement be informed by modelling of tides and currents such that there are sufficient distances between sampler placements to avoid spatial autocorrelation and to provide sufficient spatial resolution. Within the context of this study, this would amount to one sampler being deployed by each IFCA, focusing on regular monthly sampling over a longer time-scale to detect seasonal phenology of inshore communities and provide increased confidence in results.

These results highlight the exciting potential of eDNA surveys for filling some of the existing knowledge gaps around inshore fish community diversity and population dynamics that are outlined by Franco *et al.,* 2020c in the regional monitoring plan for inshore fish communities in the southwest of England. However, due to the short duration of this project, the interruptions in sampling due to adverse weather and the logistical challenges of sharing two inDepth samplers between the three collaborating IFCAs, there were insufficient data to determine whether observed differences in fish community compositions over time and distance were cyclical or stochastic. Investigating potential phenological patterns in fish communities could be accomplished by a study undertaken over a larger spatial scale with a sampling campaign of at least one calendar year.

Analyses of haplotype diversity have been well understood for some time (Wright, 1937; Ridley, 2003) but have mainly focussed on single species and targeted meta-populations (Billington & Hebert, 1991; De Jong *et al*., 2011; Hwang & Cho, 2018). The application of haplotype diversity analyses using eDNA biodiversity data is a relatively novel technique (Tsuji *et al*., 2018; Elbrecht *et al.,* 2018; Turon *et al*., 2020) made possible by using exact amplicon sequence variants (Callahan *et al.,* 2017; Callahan, 2019). This study has provided data and visualisation of species haplotype diversity as an indication of the diversity of breeding individuals within each sampled

population (Figure 13). We also detected changes in haplotype diversity over time for each species at each sampling location over the duration of this study (Table 12) and between the 2018/19 and 2019/20 datasets collected from south Devon (Table 14).

In their recent mesocosm study, Tsuji *et al.,* (2020) found that some denoising algorithms risked producing false positive haplotypes and other more stringent algorithms risked generating false-negatives. Clearly, false positives are less problematic compared to false negatives. The former can be weeded out of the data through any number of downstream quality control steps, from analysis of credibility to cross-checking of taxonomic assignments (as were undertaken in this study); whereas the sequences erroneously omitted from the data would be virtually impossible to correct. Whilst this haplotype method provides an indication of metapopulation genetic diversity, the technique requires  further research and validation.

There appeared to be a significant change in beta diversity between year-on-year inshore fish communities (Figure 16; Table 14), where haplotypes for detected species within the 2019/20 study appeared to have lower genetic diversity than haplotypes captured for similar species within the 2018/19 study. However, there were not enough data to determine whether this was an actual trend or a stochastic observation. Regular monitoring over a longer period of time would be required to be able to make a reliable determination of trends in within-species haplotype variation of these populations over time.

Environmental DNA techniques cannot reliably determine population abundance, sizes, life stages or sex ratios, at this time, the differences in haplotype diversity offer the potential for deeper investigation into the population dynamics of individual species over time, which has implications for delivering quantifiable, data-driven insights for simultaneous monitoring of the sustainability of multiple species metapopulations across large spatial scales.

The level of granularity available from haplotype analyses is of key importance to fisheries managers and Statutory Nature Conservation Bodies wishing to better understand the genetic diversity of local fish communities and population dynamics over time with the potential to inform adaptive fisheries management and conservation strategies. Once sufficiently tested and validated, haplotype diversity could provide a useful new metric for monitoring the breeding population diversity and/or health of inshore fish communities.

## 4.3 Objective 5: Using eDNA to monitor MCZ SOCI, SAC and UKBAP species and marine non-native species

One of the major applications of eDNA technology for marine monitoring is the detection of cryptic and/or less common species of interest, such as marine protected species (MCZ SOCI, SAC designated species features), other priority species in the wider seas (UKBAP species, Table 8), and non-native species (Table 8). Some studies have explored the detection of rare species (i.e. NECR252 (2018), development of eDNA methods for detection of seahorses and other lagoon protected species) and non-native species (Holman *et al*., 2020) in the marine environment using molecular methods. However, systematic application of molecular monitoring in the detection of marine species is still under development.

Because the focus of this study was inshore fish species diversity, we used fish-specific primers only. The detected MCZ SOCI and SAC species features (Table 20), UKBAP species (Table 8) and non-native species (Table 8) therefore only describe the detection and distribution of fish species of interest. However, as each of the purified eDNA extracts from each sample in this study have been bio-banked, the existing samples provide an opportunity to test the approach further by investigating the distribution and range of a broader range of species of interest. With the application of species targeted primers or other pan-specific universal primer sets the eDNA contained in each sample could enable early detection of potential non-native species invasions which could be investigated for little additional investment. Other groups of organisms could also be targeted, such as marine mammals or invertebrates, including parasites and even certain bacterial and viral pathogens.

The selection of sampling locations in this study was mainly influenced by proximity to areas visited by the vessels of the participating IFCAs and did not specifically target designated (or other priority) species or potential invasion hotspots. As a result, only a small number of samples provided positive identification of such species. Nevertheless, with a careful selection of sampling sites in the future, large-volume eDNA sample collection could provide a platform for a relatively simple method for regular collection of standardised data that is suitable for multiple purposes, including monitoring of protected and/or non-native species presence, distribution, trends and range shifts. Once sufficiently tested and validated, the haplotype method could provide a metric for population health of protected species within MPAs (or HPMAs), or insights into the invasion stage of non-native species. Furthermore, species detection could be linked to the diversity of the entire community and species assemblage. For example, time-series analyses may allow us to understand how non-native species alter the host community composition after arrival. If applied at large scale, our sampling technique could also detect longer-term changes, hinting at adaptation occurring both at the species and community level, linking into the effects of climate change on coastal habitats.

## 4.4 Objective 6: The potential of eDNA data for assessing habitat fidelity

The ability of eDNA sampling to detect habitat fidelity of fish species and populations is a key question for habitat and species conservation and management. The potential of eDNA data to distinguish such patterns remains unknown and may be limited due to the inability of eDNA data to infer abundance or biomass (although techniques such as the haplotype diversity method could, once validated, provide some insights in population abundance). The original subsidiary objective of this study was to test the ability of eDNA data to assess species fidelity to key protected habitats by manipulation of the temporal sampling regime in a restricted set of locations (Objective 6, Table 15). Unfortunately, due to the unforeseen delays in the sampler deployment caused by continuously severe weather conditions during the winter months, this subsidiary objective was not filled.

However, we did explore the impact of habitat type on fish community composition through the paired sampling regime of the main study. The effect of benthic substrate on fish community composition was found to be statistically significant ($p < 0.05$, Table 6; Figure 8); however, the

NMDS of beta diversity indicated no observable effect of benthic type on community composition (Figure 9). This result is unsurprising, as the tide and current excursion modelling showed water flowing across multiple habitat types over the 25-hour sampling cycle, thus confounding any habitat-specific eDNA signal.

Figure 11 shows how community differences between samples are better explained by the distance between sampling locations rather than substrate type, where the standard error given by the ellipsoids have greater overlap between the means of each sampling location that were nearer to each other despite these locations having different substrata (see also Table 2). Further investigation in the potential for eDNA surveys to characterise differences in fish community compositions attributable to different benthic habitats are required to fully understand habitat preferences and fidelity of inshore fish communities.

## 4.5 Implications for inshore fish conservation and management

Strategic planning for biodiversity conservation requires monitoring programmes that provide data on species distributions and population trends to assess their structural and functional characteristics and dynamics, as well as the impact of anthropogenic pressures or management measures to these species and populations (Elliott and Hemingway, 2002). Due to the inherent difficulties in monitoring all the components of inshore fish communities and assemblages that consist of a diversity of species, life stages, morphologies, functional groups and life strategies (e.g. Elliott and Dewailly, 1995; Franco *et al*., 2008), no single survey technique has so far been able to provide a representative picture of the whole fish community (NECR 269, 2020a). Instead, several methods have been required to capture all components of a fish assemblage. This labour-intensive strategy has been widely accepted and is currently used for example in the WFD fish monitoring programme for UK transitional waters (Coates *et al*., 2007; WFD-UKTAG, 2014). A coordinated and standardised monitoring programme for inshore fish populations that includes all species and not only those of commercial interest has been lacking for the UK waters. This monitoring gap has hindered the efficacy of biodiversity conservation and management efforts (NECR 271, 2020c).

Following a thorough review of all existing methods for monitoring inshore fish communities (NECR 269, 2020a) and identifying a selection of those techniques that are potentially viable for use in the inshore areas of the SW of England (NECR 270, 2020b), a Regional Monitoring Plan for inshore fish communities in the southwest of England was developed (NECR 271, 2020c). This monitoring plan aims to assess the inshore fish communities, detect current and future changes and determine the value of management measures. It proposes a combination of monitoring techniques to be applied:

(i) broadscale monitoring that allows the basic characterisation of the inshore fish assemblages by targeting key habitats and associated fish assemblage components in the study region;

(ii) the use of eDNA monitoring to provide a wider species coverage for integration and cross-validation;

(iii) the use of additional targeted (reactive or strategic) monitoring to address specific interests and questions as they arise (e.g. to assess the effect of a pressure or a management measure on the fish assemblage).

While the proposed broadscale monitoring entails the use of a combination of traditional fish monitoring techniques such as fyke netting, seine netting, beam trawling and scuba diving, the report suggests that eDNA monitoring could be used for collection of additional data on fish biodiversity to identify possible gaps in the broadscale monitoring strategy, for example in poorly sampled species (e.g. rare species, more cryptic species) that may be under-represented by the core broadscale monitoring (NECR 271, 2020c). eDNA monitoring is also thought to allow for cross-validation of the level to which the proposed core methods are reflective of the inshore fish community as a whole, and therefore to allow identification of possible additional monitoring priorities and needs that will guide the development of targeted monitoring plans in the future.

This study has emphasised the considerable potential of large-volume eDNA-based monitoring, particularly at detecting changes in fish communities over large spatial scales while highlighting the importance of consistent monthly sampling to produce robust time-series data. The inDepth eDNA sampler has proven itself as a practical tool for the collection of eDNA samples over a number of tidal cycles with minimal risk of sample contamination, which can often be a problem in eDNA surveys, particularly in challenging marine environments. It provides a simple and relatively quick method for collecting large volumes of water, shown to be sufficient for species detection in other large aquatic ecosystems (Cantera *et al*., 2019). We have therefore demonstrated that large-volume eDNA sampling can deliver an efficient and sensitive method for collecting representative data that could be used to fulfil the second objective of the monitoring plan (Franco *et al.,* 2020c).

We also propose that the eDNA element of the monitoring strategy could become more significant and substitute at least some of the core broadscale monitoring elements described by Franco *et al.,* (2020c) that do not require physical measurements of the individuals (e.g. weight, size, etc). Franco *et al.,* (2020a) state that because fish assemblages include different taxonomical, morphological and functional groups representing different niches and inhabiting different habitats (e.g. Potter *et al*., 2015), the monitoring of fish communities may require the use of different complementary methods.

With careful deployment of inDepth samplers in the correct habitats and at the right time, all of these structural and functional groups could be captured by eDNA monitoring as opposed to multiple conventional sampling techniques that tend to be costly, time consuming, and often damaging to the underlying habitat and/ or the individuals captured. Furthermore, our risk-based analysis of detection credibility is able to highlight potential issues in species detection and population data which may require further investigation into changes in fish communities or assemblages. As a result, traditional and visual methods could be used more effectively to target the regions, locations, habitats, seasons, or populations that may require adaptive management or conservation measures to be put in place.

Finally, we have demonstrated the exciting potential of large-volume eDNA sampling as a reliable tool for monitoring marine biodiversity, protected species and non-native species within protected sites, including MPAs (and HPMAs) within the UK coastal waters. The use of tidal excursion modelling (Figure 2) enables estimation of the extent of the MPA "surveyed" during the sampling period. Regular sampling over time would be able to detect any changes in fish communities and assemblages following introduction of management measures, thereby demonstrating the effectiveness of the MPA management measures compared to another unprotected area with similar characteristics. Similarly, large-volume eDNA sampling could be used for monitoring anthropogenic impacts on marine biodiversity. For example, regular eDNA collection over time could be utilised for monitoring the recovery from a pollution incident or characterisation of biodiversity changes in the presence of anthropogenic pressures in marine benthic invertebrate communities for applications in the extractives industries and aquaculture (research in area is already ongoing). This would require data on community composition before a pollution incident as well as during the impact and recovery phases, thereby obviating the need for ongoing eDNA monitoring.

The key is careful selection of eDNA sampling sites: with correct placement of inDepth samplers in or near MPAs, pollution source, or non-native species invasion hotspots, the long-term datasets resulting from these samples could serve the remit of multiple government agencies, thereby providing an unprecedented level of information about the sustainability of our fisheries, the success of conservation efforts of protected species and sites, and the health of the wider coastal ecosystem, resulting in data-driven insights for deeper due-diligence and more meaningful risk intelligence.

## 4.6 Recommendations for future work

This study has highlighted a number of opportunities to improve our understanding of the benefits of eDNA analysis techniques for adaptive management of marine biological resources.

### 4.6.1 Development of a marine biodiversity monitoring programme using large-scale eDNA time-series data for adaptive management of marine biological resources

The primary recommendation for follow-on work to come from this study is to conduct a nationwide study where large-volume eDNA samples would be collected from multiple sites around Great Britain (GB) on a monthly basis for a period not less than a full calendar year.

The inDepth eDNA samplers would be distributed to and deployed by fisheries conservation officers and other stakeholders, such as fishers or aquaculture farmers. We have estimated at least 24 sites could be surveyed around GB each month; 10 sites in England, 8 in Scotland and 6 in Wales to provide adequate coverage of inshore waters.

The data resulting from this expanded study would provide a large-scale overview of the distribution of fishes in inshore waters around GB and provide insights into potential phenological patterns in fish communities. The return on investment may be improved by the analysis of multiple primer sets which would be able characterisation of a broader range of taxa, enabling simultaneous monitoring of multiple species groups and serving the remit of multiple stakeholders.

We currently have in-principle support for this project from Natural England, the Inshore Fisheries and Conservation Authorities, Marine Scotland Science and Welsh Government Marine and Fisheries Division. The project has been fully costed and we are currently in the process of seeking funding.

The results from this proposed project would be distributed to all participants on a monthly basis, delivering actionable insights that would enable informed and evidence-based adaptive decision making in establishment of conservation and management measures. We would build on lessons learned over the course of our previous studies to deliver a reliable service with robust results. Monthly deliverables would include:

- A site by species matrix (in .csv format) of organisms detected at all locations for the past month;
- An overview of the detected haplotype diversity for each species at each location;
- Monthly updated time-series analyses.

This project would provide excellent value for money as it would significantly contribute to filling the current evidence gaps outlined in this study, allowing UK regulators to meet their statutory commitments for fisheries management and provide timely, implementable data to fisheries resource managers and conservation authorities.

### 4.6.2 Complementing inshore marine biodiversity monitoring by targeting estuarine and offshore environments

It is important to note that the sampling strategy applied in this study is one of several potential applications which have been trialled using the inDepth eDNA Samplers. The samplers may be easily reprogrammed for deployment in different scenarios to address specific questions and target particular environments. For example, samplers have been programmed to collect samples only during certain tidal phases to enable targeting of particular water flows or during slack tidal periods, for deployment in estuaries or maritime shipping ports. Samplers could equally be programmed for continuous large-volume sampling, for deployment in non-tidal aquatic environments, such as rivers or when towed behind vessels. The inDepth eDNA samplers are capable of sampling to depths in excess of 150 metres, meaning that several samplers could be deployed on a shot-line to capture marine biodiversity along depth gradients from the surface to the bottom of the North Sea, for example. The resulting data from each of these activities would be complementary to a nationwide coastal monitoring project, as proposed in section 4.6.1.

The key is to obtain clear commitment to collecting large-volume eDNA samples at multiple priority locations (i.e. marine protected areas, e.g.: Table 20, and known hotspots at risk of non-native species introductions; e.g.: Tidbury *et al.,* 2014), consistently and frequently. Once DNA has been isolated and purified, it may be bio-banked and queried multiple times over and for any number of research questions.

### 4.6.3 Validating the use of haplotype analysis in metabarcoding data sets

Currently, many survey methods (WFD included) rely on abundance measures to determine the ecological status of a population. Haplotype analysis is arguably of greater value because it

provides a measure of the genetic diversity of a population, which is a more direct indicator of population health and resilience. Whilst haplotypes are an established method for targeted metapopulation diversity analyses (Wright, 1937; Billington & Hebert, 1991; De Jong *et al*., 2011; Hwang & Cho, 2018), their application using eDNA biodiversity data is a relatively novel technique (Tsuji *et al*., 2018; Elbrecht *et al.,* 2018; Turon *et al*., 2020) which has been made possible by using exact amplicon sequence variants (Callahan *et al.,* 2017; Callahan, 2019) as a method for denoising sequencing read data.

A recent mesocom study by Tsuji *et al.,* (2019) showed that false positives may be an issue with haplotype methods, however false-positives are less problematic than false-negatives. The former can be weeded out of the data through downstream quality control steps, such as those applied in this study; whereas sequences erroneously omitted from the data would be far more difficult to correct.

We propose to validate the application of haplotype analyses by first conducting a review of the scientific literature which would then inform a mesocosm experiment and field-based study aimed at answering particular knowledge gaps specific to eDNA-based monitoring of fishes and other priority species in the marine environment, followed by further data collection at regional or national scales over several months. Field data collection could be paired with recommendations for fieldwork from the Regional Monitoring Plan for inshore fish communities in the southwest of England (NECR 271, 2020c).

### 4.6.4 Investigating the effectiveness of marine protected areas and habitat preference of inshore fishes using eDNA monitoring

Forty percent of England's seas are designated as Marine Protected Areas (MPAs). However, the UK Marine Strategy updated report (2019) indicates the environment is not in a healthy state.

To investigate the effectiveness of MPAs, we propose to undertake large-volume eDNA sampling surveys inside and outside of designated MPA waters. We would model tidal excursions under different conditions to ensure adequate separation of the two sampling sites. Two inDepth eDNA samplers would be deployed, ideally at the same time to provide a robust comparison between the two locations. Selection of appropriate sampling sites would be informed by work already undertaken to develop a Regional Monitoring Plan for inshore fish communities in the southwest of England (NECR 271, 2020c).

This project would also be designed to answer questions of species' habitat fidelity (Subsidiary Objective 6), which we weren't able to fully address in this study.

### 4.6.5 Investigating other phyla detectable by large-volume marine sampling using existing eDNA samples

The eDNA samples collected as part of this study will be bio-banked at Applied Genomics laboratories for a period of one year, unless otherwise agreed. They are currently an untapped resource for information on the distribution of other species which were not targeted as part of this study.

To improve the return on investment from the samples collected for these surveys, we recommend that the samples be reanalysed using pan-specific universal primer sets to investigate their potential for early detection of potential non-native species invasions, including parasites and even certain bacterial and viral pathogens and/or other groups of protected species, such as marine mammals or invertebrates. The relevance of the information contained in these samples is, of course dependent on the date they were collected; so, these additional analyses should be prioritised to maximise their efficacy and value.

*Table 16: Summary of recommendations for future project opportunities. NB: Timeframes and Costs are indicative estimates only.*

| Project Recommendation | Relevance | Timeframe | Cost |
|---|---|---|---|
| **Development of a marine biodiversity monitoring programme using large-scale eDNA time-series data for adaptive management of marine biological resources** | Towards operationalisation of DNA-based monitoring the status of inshore fish communities; development of GB-wide eDNA time-series dataset; detection of non-native species | 18+ months | ~ £300k |
| **Complementing inshore marine biodiversity monitoring by targeting estuarine or offshore environments** | Towards development of an environmental biosecurity surveillance network for early detection of non-native species at maritime ports. | 18+ months | ~ £120k |
| | DNA-based monitoring of biodiversity near offshore marine structures. | 12+ months | ~ £80k |
| **Validating the use of haplotype analysis in metabarcoding data sets** | Towards validation of eDNA data as an indicator of population genetics and metapopulation dynamics. | 12+ months | ~ £60k |
| **Investigating the effectiveness of marine protected areas and habitat preference of inshore fishes using eDNA monitoring** | Towards an understanding of the relationship between MPA networks and habitat preferences of inshore fish populations. | 18+ months | ~ £100k |
| **Investigating other phyla detectable by large-volume marine sampling using existing eDNA samples** | Improving return-on-investment for eDNA samples collected as part of this and other studies. | 3 months | ~ £20k |

# 5 Works Cited


Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26: 32–46.

Barstow, S.F. (2003). The ecology of Langmuir circulation: A review. *Marine Environmental Research*. Pp. 211-236.

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. Global Ecology and Biogeography 19:134-143

Bessey, C., Jarman, S. N., Berry, O., Olsen, Y. S., Bunce, M., Simpson, T., ... & Keesing, J. (2020) Maximizing fish detection with eDNA metabarcoding. *Environmental DNA*.

Beentjes, K. K., Speksnijder, A. G., Schilthuizen, M., Hoogeveen, M., & van der Hoorn, B. B. (2019). The effects of spatial and temporal replicate sampling on eDNA metabarcoding. *PeerJ*, *7*, e7335.

Billington, N., & Hebert, P. D. (1991). Mitochondrial DNA diversity in fishes and its implications for introductions. *Canadian Journal of Fisheries and Aquatic Sciences*, *48*(S1), 80-94.

Brooks, J. P., D. J. Edwards, M. D. Harwich, M. C. Rivera, J. M. Fettweis, *et al*. (2015). The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiology* 15.1, p. 66.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods*, *13*(7), 581.

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME journal*, *11*(12), 2639.

Callahan, B. J. (2019). “I think it is right to think of the sequence variants as haplotypes in the population of that species.” *Personal Communication via eMail*.

Cantera, I., Cilleros, K., Valentini, A., Cerdan, A., Dejean, T., Iribar, A., ... & Brosse, S. (2019). Optimizing environmental DNA sampling effort for fish inventories in tropical streams and rivers. *Scientific reports*, *9*(1), 1-11.

Chamberlain, S. (2017). rgbif: Interface to the Global 'Biodiversity' Information Facility 'API'. R package version 2.2.0.

Champely, S. (2020). pwr: Basic Functions for Power Analysis. R package version 1.3-0.

Coates, S., Waugh, A., Anwar, A., & Robson, M. (2007). Efficacy of a multi-metric fish index as an analysis tool for the transitional fish component of the Water Framework Directive. *Marine Pollution Bulletin*, *55*(1-6), 225-240.

Cohen, J. (2013). Statistical power analysis for the behavioral sciences. *Routledge*.

Collins, R. A., Wangensteen, O. S., O’Gorman, E. J., Mariani, S., Sims, D. W., & Genner, M. J. (2018). Persistence of environmental DNA in marine systems. *Communications biology*, *1*(1), 185.

Colquhoun, D. (2019). The false positive risk: a proposal concerning what to do About p-values. *The American Statistician*, *73*(sup1), 192-201.

Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics,* 43, 783–791.

De Jong, M. A., Wahlberg, N., Van Eijk, M., Brakefield, P. M., & Zwaan, B. J. (2011). Mitochondrial DNA signature for range-wide populations of Bicyclus anynana suggests a rapid expansion from recent refugia. *PloS one*, *6*(6).

Department for Environment, Food & Rural Affairs (2019). Marine strategy part one: UK updated assessment and Good Environmental Status. Available at: www.gov.uk/government/publications/marine-strategy-part-oneuk-updated-assessment-and-good-environmental-status

Doi, H., Fukaya, K., Oka, S. I., Sato, K., Kondoh, M., & Miya, M. (2019). Evaluation of detection probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a multispecies site occupancy model. *Scientific reports*, *9*(1), 3581.

Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C., ... & Park, D. (2015). Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience*, *4*(1), s13742-015.

Ekman, V.W. (1905). On the influence of the earth's rotation on ocean currents. *Arkiv för matematik, astronomi och fysik*, 2 (11): 1–52.

Elbrecht, V., Vamos, E. E., Steinke, D., & Leese, F. (2018). Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ*, *6*, e4644.

Elliott M, Dewailly F (1995) The structure and components of European estuarine fish assemblages. *Neth J Aquat Ecol* 29:397 – 417

Elliott, M. & Hemingway, K. (Editors) (2002) Fishes in Estuaries. *Blackwell Science,* London.

Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., ... & Lodge, D. M. (2016). Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Molecular ecology resources, 16(1), 29-41.

Faith, D. P, Minchin, P. R. and Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69, 57–68.

Forster, D., Lentendu, G., Filker, S., Dubois, E., Wilding, T. A., & Stoeck, T. (2019). Improving eDNA-based protist diversity assessments using networks of amplicon sequence variants. *Environmental microbiology*, *21*(11), 4109-4124.

Franco, A., Elliott, M., Franzoi, P., & Torricelli, P. (2008). Life strategies of fishes in European estuaries: the functional guild approach. *Marine Ecology Progress Series*, *354*, 219-228.

GB Non-native Species Secretariat (GBNNSS, 2016). GB Non-native Species Information Portal. http://www.nonnativespecies.org/factsheet/index.cfm, downloaded April 2019.

Held, L. (2019) The assessment of intrinsic credibility and a new argument for p < 0.005. *Royal Society open science*, 6(3), 181534.

Holman, L. E., de Bryun, M., Creer, S., Carvalho, G., Robidart, J. & Rius, M. (2019). Detection of resident and introduced marine species using environmental DNA metabarcoding of sediment and water. *Scientific Reports*, 9:11559.

Hwang, J. Y., & Cho, G. J. (2018). Identification of novel haplotypes and interpretation of gene flow of mitochondrial DNA control region of Eurasian otter (Lutra lutra) for the effective conservation. *Journal of Veterinary Medical Science*, 17-0678.

Kruskal, J.B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29, 115–129.

Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., & Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular ecology*, *28*(2), 420-430.

Leray, M., Knowlton, N., Ho, S. L., Nguyen, B. N., & Machida, R. J. (2019). GenBank is a reliable resource for 21st century biodiversity research. *Proceedings of the National Academy of Sciences*, *116*(45), 22651-22656.

Machida, R. J., Leray, M., Ho, S. L., & Knowlton, N. (2017). Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Scientific data*, *4*, 170027.

Matthews, R. A. (2018). Beyond 'significance': Principles and practice of the Analysis of Credibility. *Royal Society Open Science*, *5*(1), 171047.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one*, *8*(4), e61217.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., ... & Kondoh, M. (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*(7), 150088.

Marine Management Organisation (2019); https://www.gov.uk/government/statistics/uk-sea-fisheries-annual-statistics-report-2018; downloaded 13 March 2020.

Murali, A., Bhargava, A., & Wright, E. S. (2018). IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome*, *6*(1), 140.

NECR252, Tang, C.Q., Crampton-Platt, A., Townend, S., Bruce, K., Bista, I. & Creer, S. (2018). Development of DNA applications in Natural England 2016/2017. *Natural England Commissioned Report*, NECR 252.

NECR 269, Franco, A., Nunn, A., Smyth, K., Hänfling, B. and Mazik, K.(2020a). A review of methods for the monitoring of inshore fish biodiversity. *Natural England Commissioned Report*, NECR 269.

NECR 270, Franco, A., Barnard, S. and Smyth, K. (2020b). An assessment of the viability of fish monitoring techniques for use in a pilot approach in SW England. *Natural England Commissioned Report*, NECR 270.

NECR 271, Franco, A., Hänfling, B., Young, M. and Elliott, M. (2020c). Regional monitoring plan for inshore fish communities in the Southwest of England. *Natural England Commissioned Report*, NECR 271.

NECR287, Mynott S. (2020). Pilot study to validate an environmental DNA sampler for monitoring inshore fish communities. Natural England Commissioned Reports. *Natural England Commissioned Report*, NECR 287.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H. (2019). vegan: Community Ecology Package. R package version 2.5-5.

Potter, I. C., Tweedley, J. R., Elliott, M., & Whitfield, A. K. (2015). The ways in which fish use estuaries: a refinement and expansion of the guild approach. *Fish and Fisheries*, *16*(2), 230-239.

Ridley, M (2003). Evolution; 3rd ed., Wiley ISBN 978-1-4051-0345-9. *Blackwell Publishing*

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2018). GenBank. *Nucleic acids research*, *47*(D1), D94-D99.

Tidbury, H., Taylor, N., Copp, G., Garnacho, E., & Stebbing, P. (2014). Introduction of marine non-indigenous species into Great Britain and Ireland: hotspots of introduction and the merit of risk based monitoring. *Cefas contract report C*, *5995*.

Tsuji, S., Miya, M., Ushio, M., Sato, H., Minamoto, T., & Yamanaka, H. (2018). Evaluating intraspecific diversity of a fish population using environmental DNA: An approach to distinguish true haplotypes from erroneous sequences. *bioRxiv*, 429993.

Tsuji, S., Miya, M., Ushio, M., Sato, H., Minamoto, T. & Yamanaka, H. (2020). Evaluating intraspecific genetic diversity using environmental DNA and denoising approach: A case study using tank water. *Environmental DNA,* 2*:*42-52.

Turon, X., Antich, A., Palacín, C., Præbel, K., & Wangensteen, O. S. (2020). From metabarcoding to metaphylogeography: separating the wheat from the chaff. *Ecological Applications*, *30*(2), e02036.

UKBAP (United Kingdom Biodiversity Action Plan) priority species accounts spreadsheet, http://jncc.defra.gov.uk/Docs/UKBAP_SpeciesIndexPageLIVE-V2-20111608.xls, downloaded March 2019.

WFD-UKTAG (2014). UKTAG Transitional Water Assessment Method Fish Fauna. Transitional Fish Classification Index. July 2014. ISBN: 978-1-906934-32-3

Wright E. S. (2016). Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *_The R Journal_*, 8(1), 352-359.

Wright, S. (1937). The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences of the United States of America*, *23*(6), 307.

Whittaker, RH (1972), Evolution and measurement of species diversity. *Taxon*, 213–251.

# 6 Appendix 1: PCR Conditions

## 6.1 CytB

CytB gene amplification were run in 12 replicate 15 µl total reaction volume containing 3.67 µl of $H_2O$, 2.5 µl of 5x reaction buffer, 1 µl of dNTPs (10 mM), 0.75 µl of $MgCl_2$ (25 mM), 2.5 µl of each F/R primer (10 µM) and 0.08 µl of 5 U/µl All Taq polymerase (All Taq PCR Core Kit, QIAGEN, Hilden, Germany). PCR follows these conditions: initial denaturation at 98 °C for 2 min followed by 9 cycles at 98 °C for 10 s, 60 °C for 20 s and 72 °C for 30 s; then 9 cycles at 98 °C for 10 s, 58 °C for 20 s and 72 °C for 30 s; finally, 29 cycles at 98 °C for 10 s, 55 °C for 20 s and 72 °C for 30 s, with a final extension at 72 °C for 10 min.

Technical replicates were created by repeated sampling of the purified DNA extract using the same PCR conditions.

## 6.2 12S

12S gene amplification were run in 12 replicate 15 µl total reaction volume containing 4.0 µl of $H_2O$, 0.3 µl of DMSO (100%), 3 µl of 5x reaction buffer, 1.2 µl of dNTPs (4 mM), 1 µl of $MgCl_2$ (25 mM), 1.5 µl of each F/R primer (5 µM) and 0.07 µl of 5 U/µl All Taq polymerase (All Taq PCR Core Kit, QIAGEN, Hilden, Germany). PCR were run following these conditions: initial denaturation at 95 °C for 3 min, then 45 cycles of 98 °C for 30 s, 65 °C for 15 s, 72 °C for 15 s, and final extension at 72 °C for 5 min.

Technical replicates were created by repeated sampling of the purified DNA extract and increasing or decreasing the annealing temperatures by 2 °C, as described in Doi *et al.,* (2019).

# 7 Appendix 2: Environmental Sampling Metadata

These tables contain field observations for the month, sampling location, county and collaborating IFCA. The position of each sampling location is given in decimal degrees of latitude and longitude. Dates of each sampler deployment and recovery are recorded. Concentrations for purified eDNA, 12S and CytB amplicons are provided alongside technical replicate identifiers (i.e. A, B or C) for each biological sample and sample negatives.

*Table 17: Environmental sampling metadata. DNA amplicon concentrations are provided in ng/ul.*

| SampleID | Month | Location | Lat | Lon | Dep_Date | Rec_Date | Benthos | County | Collaborator | Replicate | eDNA conc | 12S conc | CytB conc | neg |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CTRL_NEG_X | NA | Control | NA | NA | NA | NA | NA | Control | NEG | X | 0.0 | 0.0 | 0.0 | TRUE |
| FALM_DEC_A | December | Falmouth | 50.11611 | -5.04167 | 16/12/2019 | 17/12/2019 | Soft | Cornwall | C IFCA | A | 10.3 | 6.8 | 6 | FALSE |
| FALM_DEC_B | December | Falmouth | 50.11611 | -5.04167 | 16/12/2019 | 17/12/2019 | Soft | Cornwall | C IFCA | B | 10.3 | 28.2 | 16.8 | FALSE |
| FALM_DEC_C | December | Falmouth | 50.11611 | -5.04167 | 16/12/2019 | 17/12/2019 | Soft | Cornwall | C IFCA | C | 10.3 | 25.3 | 3 | FALSE |
| FALM_FEB_A | February | Falmouth | 50.11611 | -5.04167 | 18/02/2020 | 19/02/2020 | Soft | Cornwall | C IFCA | A | 6.5 | 10.3 | 14.5 | FALSE |
| FALM_FEB_B | February | Falmouth | 50.11611 | -5.04167 | 18/02/2020 | 19/02/2020 | Soft | Cornwall | C IFCA | B | 6.5 | 11.7 | 18.7 | FALSE |
| FALM_FEB_C | February | Falmouth | 50.11611 | -5.04167 | 18/02/2020 | 19/02/2020 | Soft | Cornwall | C IFCA | C | 6.5 | 11.5 | 17.7 | FALSE |
| FALM_NOV_A | November | Falmouth | 50.11611 | -5.04167 | 17/11/2019 | 18/11/2019 | Soft | Cornwall | C IFCA | A | 14.9 | 4 | 10.6 | FALSE |
| FALM_NOV_B | November | Falmouth | 50.11611 | -5.04167 | 17/11/2019 | 18/11/2019 | Soft | Cornwall | C IFCA | B | 14.9 | 20.2 | 21.3 | FALSE |
| FALM_NOV_C | November | Falmouth | 50.11611 | -5.04167 | 17/11/2019 | 18/11/2019 | Soft | Cornwall | C IFCA | C | 14.9 | 26.5 | 26.4 | FALSE |
| FOWY_DEC_A | December | Fowey | 50.31222 | -4.65444 | 16/12/2019 | 17/12/2019 | Hard | Cornwall | C IFCA | A | 9.9 | 5.2 | 7.1 | FALSE |
| FOWY_DEC_B | December | Fowey | 50.31222 | -4.65444 | 16/12/2019 | 17/12/2019 | Hard | Cornwall | C IFCA | B | 9.9 | 23.8 | 21.6 | FALSE |
| FOWY_DEC_C | December | Fowey | 50.31222 | -4.65444 | 16/12/2019 | 17/12/2019 | Hard | Cornwall | C IFCA | C | 9.9 | 9.7 | 37.3 | FALSE |
| FOWY_NOV_A | November | Fowey | 50.31222 | -4.65444 | 17/11/2019 | 18/11/2019 | Hard | Cornwall | C IFCA | A | 11.5 | 3.4 | 8.8 | FALSE |
| FOWY_NOV_B | November | Fowey | 50.31222 | -4.65444 | 17/11/2019 | 18/11/2019 | Hard | Cornwall | C IFCA | B | 11.5 | 26.8 | 3.8 | FALSE |
| FOWY_NOV_C | November | Fowey | 50.31222 | -4.65444 | 17/11/2019 | 18/11/2019 | Hard | Cornwall | C IFCA | C | 11.5 | 25.9 | 38.1 | FALSE |
| LYME_FEB_A | February | Lyme Bay | 50.66277 | -3.101033 | 05/02/2020 | 06/02/2020 | Hard | Devon | D&S IFCA | A | 9.3 | 11.4 | 17.9 | FALSE |
| LYME_FEB_B | February | Lyme Bay | 50.66277 | -3.101033 | 05/02/2020 | 06/02/2020 | Hard | Devon | D&S IFCA | B | 9.3 | 9.8 | 10 | FALSE |
| LYME_FEB_C | February | Lyme Bay | 50.66277 | -3.101033 | 05/02/2020 | 06/02/2020 | Hard | Devon | D&S IFCA | C | 9.3 | 9.4 | 16.6 | FALSE |
| LYME_NOV_A | November | Lyme Bay | 50.66277 | -3.10103 | 27/11/2019 | 28/11/2019 | Hard | Devon | D&S IFCA | A | 41.1 | 7.4 | 8.7 | FALSE |
| LYME_NOV_B | November | Lyme Bay | 50.66277 | -3.10103 | 27/11/2019 | 28/11/2019 | Hard | Devon | D&S IFCA | B | 41.1 | 10.8 | 3.1 | FALSE |
| LYME_NOV_C | November | Lyme Bay | 50.66277 | -3.10103 | 27/11/2019 | 28/11/2019 | Hard | Devon | D&S IFCA | C | 41.1 | 16.7 | 8.4 | FALSE |

| SampleID | Month | Location | Lat | Lon | Dep_Date | Rec_Date | Benthos | County | Collaborator | Replicate | eDNA conc | 12S conc | CytB conc | neg |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LYME_OCT_A | October | Lyme Bay | 50.66277 | -3.10103 | 14/10/2019 | 15/10/2019 | Hard | Devon | D&S IFCA | A | 10.3 | 2.5 | 11 | FALSE |
| LYME_OCT_B | October | Lyme Bay | 50.66277 | -3.10103 | 14/10/2019 | 15/10/2019 | Hard | Devon | D&S IFCA | B | 10.3 | 16.7 | 7.8 | FALSE |
| LYME_OCT_C | October | Lyme Bay | 50.66277 | -3.10103 | 14/10/2019 | 15/10/2019 | Hard | Devon | D&S IFCA | C | 10.3 | 16.5 | 23.7 | FALSE |
| POOL_DEC_A | December | Poole Rocks | 50.68 | -1.876 | 02/12/2019 | 03/12/2019 | Hard | Dorset | S IFCA | A | 11.5 | 6 | 8.4 | FALSE |
| POOL_DEC_B | December | Poole Rocks | 50.68 | -1.876 | 02/12/2019 | 03/12/2019 | Hard | Dorset | S IFCA | B | 11.5 | 6.7 | 30.1 | FALSE |
| POOL_DEC_C | December | Poole Rocks | 50.68 | -1.876 | 02/12/2019 | 03/12/2019 | Hard | Dorset | S IFCA | C | 11.5 | 17.4 | 30.5 | FALSE |
| POOL_NOV_A | November | Poole Rocks | 50.68 | -1.876 | 05/11/2019 | 06/11/2019 | Hard | Dorset | S IFCA | A | 25.8 | 3.7 | 10.6 | FALSE |
| POOL_NOV_B | November | Poole Rocks | 50.68 | -1.876 | 05/11/2019 | 06/11/2019 | Hard | Dorset | S IFCA | B | 25.8 | 5.1 | 14.4 | FALSE |
| POOL_NOV_C | November | Poole Rocks | 50.68 | -1.876 | 05/11/2019 | 06/11/2019 | Hard | Dorset | S IFCA | C | 25.8 | 20 | 40.6 | FALSE |
| STRT_FEB_A | February | Start Bay | 50.26305 | -3.63725 | 05/02/2020 | 06/02/2020 | Soft | Devon | D&S IFCA | A | 7.3 | 14.1 | 12.6 | FALSE |
| STRT_FEB_B | February | Start Bay | 50.26305 | -3.63725 | 05/02/2020 | 06/02/2020 | Soft | Devon | D&S IFCA | B | 7.3 | 13.2 | 24.1 | FALSE |
| STRT_FEB_C | February | Start Bay | 50.26305 | -3.63725 | 05/02/2020 | 06/02/2020 | Soft | Devon | D&S IFCA | C | 7.3 | 17.5 | 15.9 | FALSE |
| STRT_NOV_A | November | Start Bay | 50.26305 | -3.63725 | 27/11/2019 | 28/11/2019 | Soft | Devon | D&S IFCA | A | 44.1 | 4.9 | 7.2 | FALSE |
| STRT_NOV_B | November | Start Bay | 50.26305 | -3.63725 | 27/11/2019 | 28/11/2019 | Soft | Devon | D&S IFCA | B | 44.1 | 12 | 17.5 | FALSE |
| STRT_NOV_C | November | Start Bay | 50.26305 | -3.63725 | 27/11/2019 | 28/11/2019 | Soft | Devon | D&S IFCA | C | 44.1 | 7.5 | 26.6 | FALSE |
| STRT_OCT_A | October | Start Bay | 50.26305 | -3.63725 | 14/10/2019 | 15/10/2019 | Soft | Devon | D&S IFCA | A | 20.2 | 2.9 | 10.3 | FALSE |
| STRT_OCT_B | October | Start Bay | 50.26305 | -3.63725 | 14/10/2019 | 15/10/2019 | Soft | Devon | D&S IFCA | B | 20.2 | 15.4 | 4 | FALSE |
| STRT_OCT_C | October | Start Bay | 50.26305 | -3.63725 | 14/10/2019 | 15/10/2019 | Soft | Devon | D&S IFCA | C | 20.2 | 15.3 | 11.2 | FALSE |
| STUD_DEC_A | December | Studland | 50.642 | -1.915 | 02/12/2019 | 03/12/2019 | Soft | Dorset | S IFCA | A | 9.3 | 7.4 | 9.4 | FALSE |
| STUD_DEC_B | December | Studland | 50.642 | -1.915 | 02/12/2019 | 03/12/2019 | Soft | Dorset | S IFCA | B | 9.3 | 19.1 | 22.3 | FALSE |
| STUD_DEC_C | December | Studland | 50.642 | -1.915 | 02/12/2019 | 03/12/2019 | Soft | Dorset | S IFCA | C | 9.3 | 18.4 | 35.5 | FALSE |
| STUD_NOV_A | November | Studland | 50.642 | -1.915 | 05/11/2019 | 06/11/2019 | Soft | Dorset | S IFCA | A | 13.8 | 2.4 | 9.5 | FALSE |
| STUD_NOV_B | November | Studland | 50.642 | -1.915 | 05/11/2019 | 06/11/2019 | Soft | Dorset | S IFCA | B | 13.8 | 14.2 | 36.3 | FALSE |
| STUD_NOV_C | November | Studland | 50.642 | -1.915 | 05/11/2019 | 06/11/2019 | Soft | Dorset | S IFCA | C | 13.8 | 14.9 | 6.9 | FALSE |

# 8 Appendix 3: Quality Assessment

Sequencing reads quality profiles are presented for a subset of samples, showing Forward and Reverse reads quality for 12S and CytB loci. In grey-scale is a heat map of the frequency of each quality score at each base position. The mean quality score at each position is shown by the green line, and the quartiles of the quality score distribution by the orange lines.

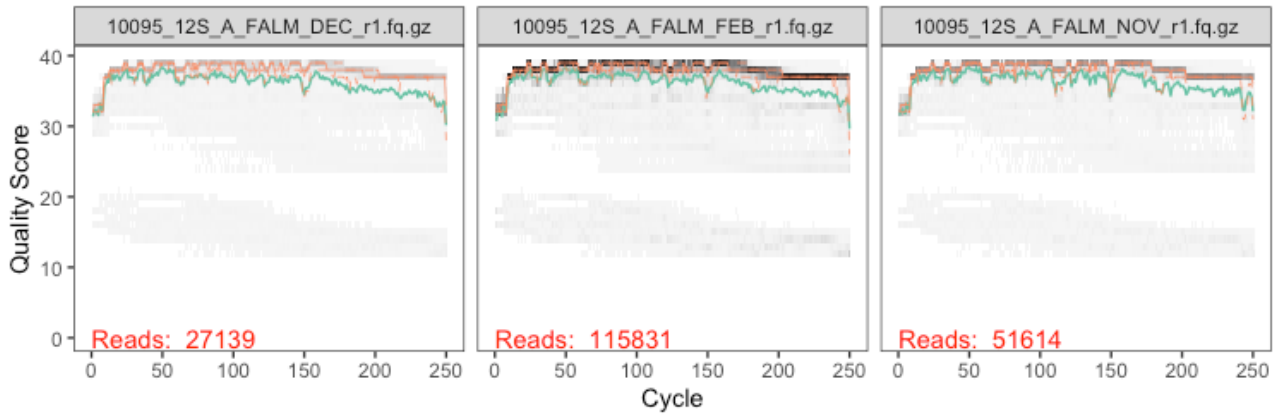



*Figure 17: Example sequencing reads quality profiles for 12S Forward reads.*

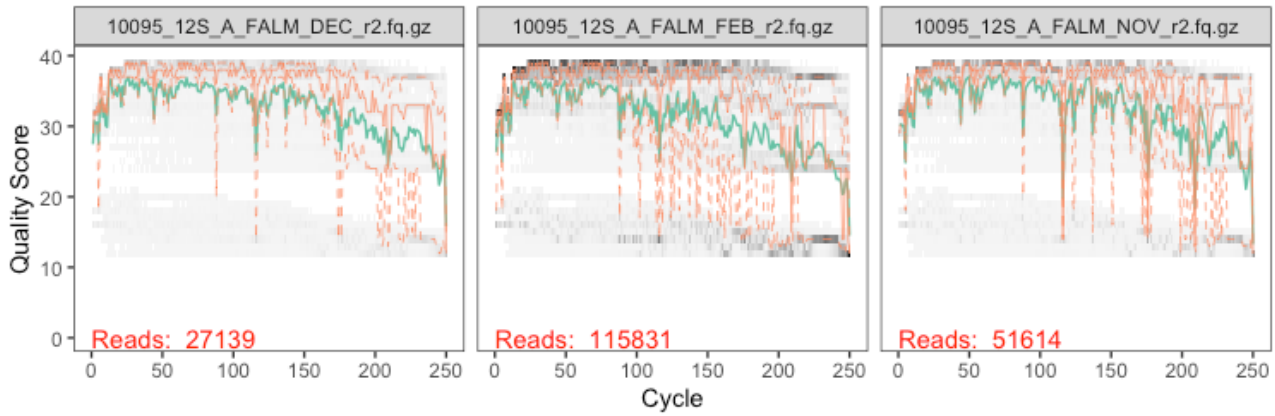



*Figure 18: Example sequencing reads quality profiles for 12S Reverse reads.*

**Trimming and Quality Filtering:** All sequences with ambiguous base calls ("*N*"s) were removed from the data. The forward and reverse primer sequences were trimmed from each of the 5' ends of 12S sequence reads. The 3' ends were truncated to 240 base-pairs (bp) for forward reads and 200 bp for reverse reads. All sequences were truncated at the first instance of a quality score (Q) less than or equal to 2. Following truncation, reads with cumulative expected errors greater than 2 were discarded. Expected errors were calculated from the nominal definition of the quality score: EE = sum(10^(-Q/10)).

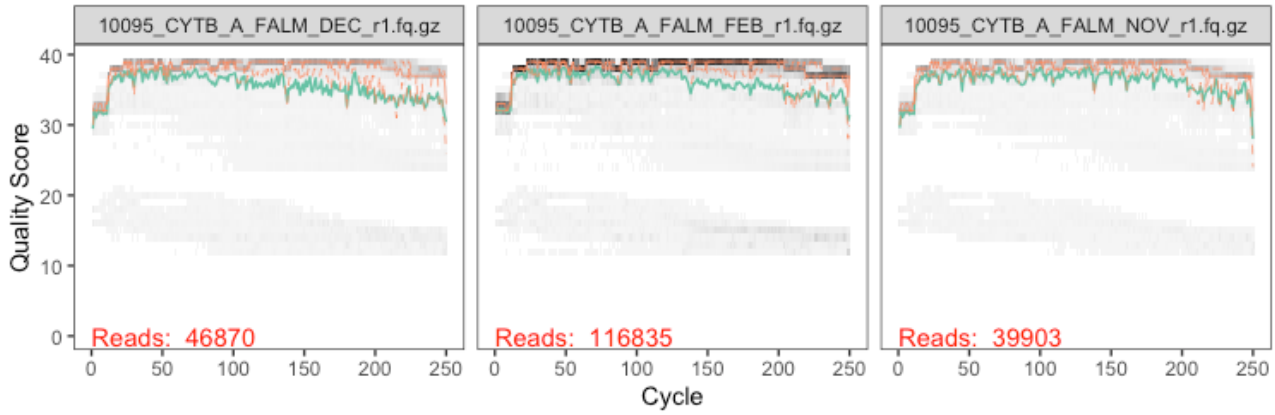



*Figure 19: Example sequencing reads quality profiles for CytB Forward reads.*

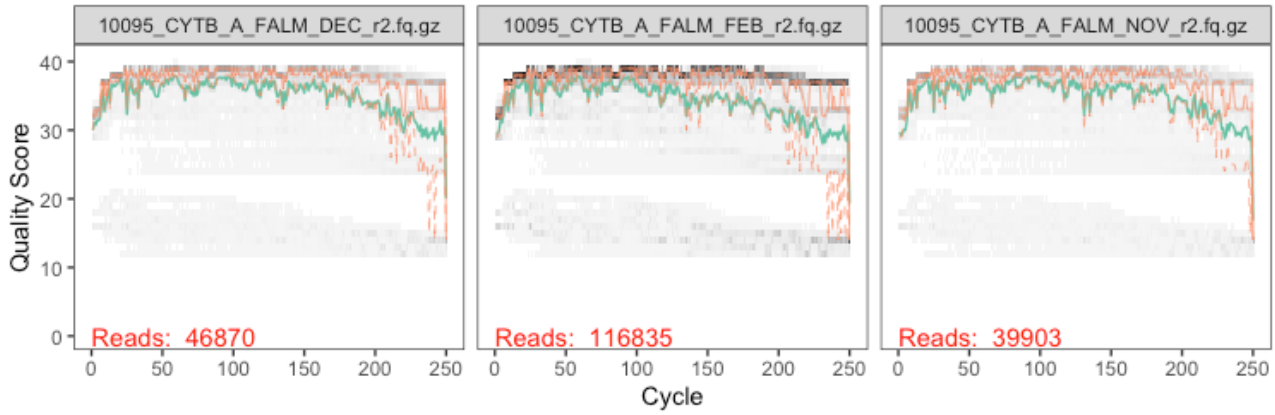



*Figure 20: Example sequencing reads quality profiles for CytB Reverse reads.*

**Trimming and Quality Filtering:** All sequences with ambiguous base calls ("*N*"s) were removed from the data. The forward and reverse primer sequences were trimmed from each of the 5' ends of CytB sequence reads. The 3' ends were truncated to 240 base-pairs (bp) for forward reads and 220 bp for reverse reads. All sequences were truncated at the first instance of a quality score (Q) less than or equal to 2. Following truncation, reads with cumulative expected errors greater than 2 were discarded. Expected errors were calculated from the nominal definition of the quality score: EE = sum(10^(-Q/10)).

# 9 Appendix 4: Pipeline Throughput

## 12S Bioinformatics throughput

*Table 18: Bioinformatics throughput for 12S amplicons for each eDNA sample and technical replicate.*

| *12S Samples* | *input* | *filtered* | *denoised* | *merged* | *nonchimeric* | *% reads* | *ASVs* | *speciesIDs* |
|---|---|---|---|---|---|---|---|---|
| *CTRL_NEG_X_12S* | 885 | 18 | 1 | 0 | 0 | 0.0 | 0 | 0 |
| *FALM_DEC_A_12S* | 27139 | 24976 | 24724 | 23145 | 20938 | 77.2 | 39 | 16 |
| *FALM_DEC_B_12S* | 80270 | 72064 | 71476 | 63844 | 60076 | 74.8 | 37 | 12 |
| *FALM_DEC_C_12S* | 162170 | 147480 | 146216 | 128652 | 122445 | 75.5 | 69 | 14 |
| *FALM_FEB_A_12S* | 115831 | 105800 | 103766 | 76851 | 70910 | 61.2 | 47 | 15 |
| *FALM_FEB_B_12S* | 70111 | 63074 | 61774 | 45003 | 41727 | 59.5 | 38 | 13 |
| *FALM_FEB_C_12S* | 53412 | 49008 | 48126 | 36944 | 34215 | 64.1 | 53 | 17 |
| *FALM_NOV_A_12S* | 51614 | 47862 | 47634 | 41687 | 40372 | 78.2 | 30 | 14 |
| *FALM_NOV_B_12S* | 78091 | 72594 | 72148 | 58509 | 56607 | 72.5 | 30 | 17 |
| *FALM_NOV_C_12S* | 168849 | 159013 | 157808 | 125243 | 122244 | 72.4 | 44 | 18 |
| *FOWY_DEC_A_12S* | 52892 | 48422 | 47902 | 42288 | 38388 | 72.6 | 45 | 16 |
| *FOWY_DEC_B_12S* | 70361 | 59721 | 58876 | 49440 | 44810 | 63.7 | 53 | 18 |
| *FOWY_DEC_C_12S* | 100522 | 92919 | 91786 | 79726 | 70439 | 70.1 | 73 | 20 |
| *FOWY_NOV_A_12S* | 44624 | 40562 | 40100 | 37049 | 32131 | 72.0 | 39 | 14 |
| *FOWY_NOV_B_12S* | 72908 | 62016 | 61064 | 54494 | 48046 | 65.9 | 57 | 18 |
| *FOWY_NOV_C_12S* | 174977 | 154231 | 152087 | 131362 | 116630 | 66.7 | 72 | 19 |
| *LYME_FEB_A_12S* | 193153 | 174709 | 168923 | 90461 | 75985 | 39.3 | 24 | 7 |
| *LYME_FEB_B_12S* | 75232 | 65649 | 63004 | 35122 | 30812 | 41.0 | 27 | 7 |
| *LYME_FEB_C_12S* | 53336 | 47127 | 45322 | 23639 | 20252 | 38.0 | 29 | 7 |
| *LYME_NOV_A_12S* | 16979 | 15208 | 14706 | 8785 | 7682 | 45.2 | 25 | 10 |
| *LYME_NOV_B_12S* | 65515 | 55264 | 53050 | 29250 | 26225 | 40.0 | 30 | 11 |
| *LYME_NOV_C_12S* | 143585 | 127075 | 122770 | 69297 | 61726 | 43.0 | 80 | 13 |
| *LYME_OCT_A_12S* | 46391 | 42167 | 41176 | 31454 | 29059 | 62.6 | 33 | 12 |
| *LYME_OCT_B_12S* | 96464 | 84220 | 81550 | 55770 | 52106 | 54.0 | 60 | 15 |
| *LYME_OCT_C_12S* | 72738 | 66898 | 64555 | 43683 | 40149 | 55.2 | 50 | 15 |
| *POOL_DEC_A_12S* | 29733 | 26999 | 26398 | 20855 | 19737 | 66.4 | 29 | 10 |
| *POOL_DEC_B_12S* | 69375 | 55440 | 53761 | 41845 | 38484 | 55.5 | 38 | 13 |
| *POOL_DEC_C_12S* | 145941 | 126657 | 124051 | 107914 | 98727 | 67.6 | 65 | 16 |
| *POOL_NOV_A_12S* | 54154 | 49608 | 48648 | 41621 | 39915 | 73.7 | 48 | 15 |
| *POOL_NOV_B_12S* | 134643 | 105697 | 101625 | 85715 | 81545 | 60.6 | 51 | 15 |
| *POOL_NOV_C_12S* | 36990 | 27481 | 25804 | 21218 | 20778 | 56.2 | 31 | 13 |
| *STRT_FEB_A_12S* | 42291 | 38905 | 38832 | 33613 | 33535 | 79.3 | 24 | 10 |
| *STRT_FEB_B_12S* | 111080 | 101090 | 100876 | 84947 | 84757 | 76.3 | 24 | 10 |

| 12S Samples | input | filtered | denoised | merged | nonchimeric | % reads | ASVs | speciesIDs |
|---|---|---|---|---|---|---|---|---|
| STRT_FEB_C_12S | 33107 | 30524 | 30437 | 26506 | 26472 | 80.0 | 20 | 9 |
| STRT_NOV_A_12S | 30867 | 27652 | 26736 | 17398 | 15896 | 51.5 | 31 | 13 |
| STRT_NOV_B_12S | 127331 | 114308 | 111580 | 83246 | 75091 | 59.0 | 48 | 15 |
| STRT_NOV_C_12S | 53961 | 45000 | 43654 | 32654 | 30449 | 56.4 | 51 | 13 |
| STRT_OCT_A_12S | 44492 | 40900 | 40062 | 32702 | 30790 | 69.2 | 35 | 13 |
| STRT_OCT_B_12S | 62587 | 56433 | 55706 | 48626 | 46122 | 73.7 | 41 | 12 |
| STRT_OCT_C_12S | 124257 | 109984 | 108646 | 94481 | 88540 | 71.3 | 75 | 13 |
| STUD_DEC_A_12S | 18943 | 17354 | 17076 | 16081 | 14957 | 79.0 | 34 | 15 |
| STUD_DEC_B_12S | 112760 | 97340 | 96218 | 91852 | 82618 | 73.3 | 40 | 15 |
| STUD_DEC_C_12S | 126956 | 107400 | 106119 | 101678 | 90869 | 71.6 | 60 | 17 |
| STUD_NOV_A_12S | 43504 | 39955 | 39322 | 35713 | 32702 | 75.2 | 44 | 16 |
| STUD_NOV_B_12S | 79586 | 64649 | 63800 | 60820 | 55862 | 70.2 | 44 | 14 |
| STUD_NOV_C_12S | 65956 | 59082 | 58252 | 54924 | 49024 | 74.3 | 79 | 22 |

## CytB Bioinformatics throughput

*Table 19: Bioinformatics throughput for CytB amplicons for each eDNA sample and technical replicate.*

| CytB Samples | input | filtered | denoised | merged | nonchimeric | % reads | ASVs | speciesIDs |
|---|---|---|---|---|---|---|---|---|
| CTRL_NEG_X_CYTB | 427 | 1 | 1 | 1 | 1 | 0.2 | 1 | 0 |
| FALM_DEC_A_CYTB | 46870 | 39423 | 39182 | 38944 | 26702 | 57.0 | 47 | 10 |
| FALM_DEC_B_CYTB | 167514 | 134841 | 134676 | 134516 | 111303 | 66.4 | 41 | 9 |
| FALM_DEC_C_CYTB | 30724 | 24695 | 24672 | 24656 | 21032 | 68.5 | 18 | 6 |
| FALM_FEB_A_CYTB | 116835 | 99092 | 98966 | 98863 | 90938 | 77.8 | 52 | 15 |
| FALM_FEB_B_CYTB | 196531 | 183437 | 182878 | 182336 | 130953 | 66.6 | 108 | 16 |
| FALM_FEB_C_CYTB | 98808 | 92067 | 91741 | 91421 | 63594 | 64.4 | 108 | 15 |
| FALM_NOV_A_CYTB | 39903 | 36507 | 36425 | 36346 | 27810 | 69.7 | 108 | 12 |
| FALM_NOV_B_CYTB | 83823 | 77127 | 76978 | 76835 | 60080 | 71.7 | 134 | 16 |
| FALM_NOV_C_CYTB | 219985 | 202948 | 202340 | 201743 | 141220 | 64.2 | 202 | 16 |
| FOWY_DEC_A_CYTB | 47239 | 43639 | 43502 | 43375 | 32982 | 69.8 | 64 | 16 |
| FOWY_DEC_B_CYTB | 169288 | 157333 | 157004 | 156683 | 96350 | 56.9 | 109 | 19 |
| FOWY_DEC_C_CYTB | 110540 | 102598 | 102390 | 102191 | 79370 | 71.8 | 57 | 13 |
| FOWY_NOV_A_CYTB | 73093 | 69795 | 69669 | 69557 | 49593 | 67.8 | 53 | 14 |
| FOWY_NOV_B_CYTB | 98033 | 75626 | 75542 | 75492 | 60299 | 61.5 | 40 | 12 |
| FOWY_NOV_C_CYTB | 75452 | 64735 | 64630 | 64535 | 50560 | 67.0 | 62 | 11 |
| LYME_FEB_A_CYTB | 52143 | 50044 | 49960 | 49881 | 43448 | 83.3 | 26 | 7 |
| LYME_FEB_B_CYTB | 36703 | 27091 | 27040 | 27011 | 20642 | 56.2 | 17 | 6 |
| LYME_FEB_C_CYTB | 96830 | 92928 | 92861 | 92802 | 88271 | 91.2 | 36 | 7 |
| LYME_NOV_A_CYTB | 42824 | 40268 | 40209 | 40151 | 35917 | 83.9 | 35 | 10 |

| CytB Samples | input | filtered | denoised | merged | nonchimeric | % reads | ASVs | speciesIDs |
|---|---|---|---|---|---|---|---|---|
| LYME_NOV_B_CYTB | 42010 | 35362 | 35320 | 35287 | 27904 | 66.4 | 24 | 8 |
| LYME_NOV_C_CYTB | 46467 | 43819 | 43761 | 43706 | 39171 | 84.3 | 28 | 8 |
| LYME_OCT_A_CYTB | 33661 | 31313 | 31222 | 31132 | 27622 | 82.1 | 58 | 11 |
| LYME_OCT_B_CYTB | 98640 | 93448 | 93352 | 93261 | 85919 | 87.1 | 52 | 11 |
| LYME_OCT_C_CYTB | 65639 | 61553 | 61392 | 61232 | 53227 | 81.1 | 84 | 13 |
| POOL_DEC_A_CYTB | 40146 | 37493 | 37446 | 37401 | 35512 | 88.5 | 52 | 14 |
| POOL_DEC_B_CYTB | 203679 | 193451 | 193284 | 193121 | 173989 | 85.4 | 48 | 13 |
| POOL_DEC_C_CYTB | 172897 | 166372 | 166332 | 166293 | 160435 | 92.8 | 50 | 18 |
| POOL_NOV_A_CYTB | 48921 | 45874 | 45822 | 45780 | 44625 | 91.2 | 47 | 16 |
| POOL_NOV_B_CYTB | 101325 | 94876 | 94774 | 94677 | 89817 | 88.6 | 74 | 13 |
| POOL_NOV_C_CYTB | 231044 | 217947 | 217764 | 217589 | 202666 | 87.7 | 99 | 15 |
| STRT_FEB_A_CYTB | 29033 | 23745 | 23696 | 23660 | 22961 | 79.1 | 33 | 12 |
| STRT_FEB_B_CYTB | 50458 | 46138 | 46034 | 45933 | 39308 | 77.9 | 35 | 11 |
| STRT_FEB_C_CYTB | 41668 | 34983 | 34880 | 34785 | 32012 | 76.8 | 31 | 9 |
| STRT_NOV_A_CYTB | 42982 | 41210 | 41138 | 41075 | 29925 | 69.6 | 47 | 12 |
| STRT_NOV_B_CYTB | 133053 | 127249 | 126964 | 126684 | 100844 | 75.8 | 60 | 14 |
| STRT_NOV_C_CYTB | 83976 | 80565 | 80458 | 80356 | 63472 | 75.6 | 57 | 16 |
| STRT_OCT_A_CYTB | 46396 | 43582 | 43510 | 43442 | 36914 | 79.6 | 53 | 14 |
| STRT_OCT_B_CYTB | 22215 | 17247 | 16980 | 16936 | 13925 | 62.7 | 37 | 12 |
| STRT_OCT_C_CYTB | 58256 | 48143 | 48058 | 47981 | 41400 | 71.1 | 59 | 12 |
| STUD_DEC_A_CYTB | 46345 | 42018 | 41980 | 41947 | 40504 | 87.4 | 46 | 12 |
| STUD_DEC_B_CYTB | 72323 | 68306 | 68198 | 68094 | 59455 | 82.2 | 48 | 11 |
| STUD_DEC_C_CYTB | 234945 | 218530 | 218172 | 217830 | 203439 | 86.6 | 105 | 19 |
| STUD_NOV_A_CYTB | 43163 | 38971 | 38858 | 38749 | 36925 | 85.5 | 72 | 16 |
| STUD_NOV_B_CYTB | 34785 | 32546 | 32530 | 32515 | 31098 | 89.4 | 30 | 15 |
| STUD_NOV_C_CYTB | 61631 | 51665 | 51517 | 51433 | 48654 | 78.9 | 70 | 15 |


Applied Genomics Ltd is registered in England and Wales: 9180742 | VAT: GB198410583
PO Box 104 | Brixham Environmental Laboratory | Freshwater Quarry | TQ5 5AY

# 10 Appendix 5: Marine Protected Areas

*Table 20: Marine Protected Areas covered in the sampling regime and the associated MCZ SOCI and SAC species features relevant for this study.*

| Sampling location | MPA within area sampled | Fish SOCI/ designated SAC feature |
|---|---|---|
| Falmouth | Helford Estuary MCZ | |
| | Lower Fal and Helford Intertidal SSSI | |
| | Fal and Helford SAC | |
| | Lizard Point SAC | |
| Fowey | Upper Fowey to Pont Pill MCZ | |
| Start Bay | Salcombe to Kingsbridge Estuary SSSI | |
| | Start Point to Plymouth Sound and Eddistone SAC | |
| | Skerries Bank and Surrounds MCZ | |
| Lyme Bay | Lyme Bay and Torbay SAC | |
| Studland & Poole Rocks | Studland to Portland SAC | |
| | Chesil Beach to Stennis Ledges MCZ | |
| | Southbourne Rough MCZ | Black seabream |
| | South of Portland MCZ | |
| | South Dorset MCZ | |
| | Purbeck Coast MCZ | Black seabream |
| | Studland Bay MCZ | Long-snouted seahorse |
| | Poole Rocks MCZ | Black seabream, Couch's goby |
| | Poole Harbour SSSI | |
| | Albert Field MCZ | |
| | The Needles MCZ | |
| | Yarmouth to Cowes MCZ | |
| | South of Wight Maritime SAC | |
| | Solent Maritime SAC | |
| | Hurst Castle and Lymington River Estuary SSSI | |
| | Yar Estuary SSSI | |
| | North Solent SSSI | |
| | River Avon SAC | Sea lamprey, Brook lamprey, Atlantic salmon, Bullhead |