Developing eDNA approaches for the detection of European smelt (*Osmerus eperlanus*) on the River Wyre

January 2024

Natural England Commissioned Report NECR516

www.gov.uk/natural-england

NATURAL ENGLAND

About Natural England

Natural England is here to secure a healthy natural environment for people to enjoy, where wildlife is protected and England's traditional landscapes are safeguarded for future generations.

Further Information

This report can be downloaded from the Natural England Access to Evidence Catalogue. For information on Natural England publications or if you require an alternative format, please contact the Natural England Enquiry Service on 0300 060 3900 or email enquiries@naturalengland.org.uk.

Copyright

Catalogue code: NECR516

Report details

Author(s)

Barbara Morrisey, Bernd Hänfling, Demetra Andreou, Graham Sellers, Victoria Pritchard.

Natural England Project Manager

Rosie Horner.

Contractor

UHI Inverness, 1 Inverness Campus, Inverness, IV2 5NA.

Keywords

Osmerus eperlanus, European Smelt, River Wyre, Wyre MCZ, eDNA.

Citation

Morrisey, B., Hänfling, B., Andreou, D., Sellers, G., Horner, R., Pritchard, V. (2022) Developing eDNA approaches for the detection of European smelt (*Osmerus eperlanus*) on the River Wyre. UHI Inverness report for Natural England. Natural England.

UHI Inverness

1 Inverness Campus

Inverness

IV2 5NA

Email: bernd.haenfling.ic@uhi.ac.uk

Foreword

The University of Highlands and Islands (UHI) were commissioned by Natural England to develop an eDNA based sampling methodology for detection of European smelt in the River Wyre.

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.

Summary

The Wyre-Lune Marine Conservation Zone (MCZ) was designated in 2019 and has the fish species European smelt (*Osmerus eperlanus*) as its only feature. The intention of designating the MCZ was to help protect remaining populations of smelt and prevent further loss or decline of populations. Historically, very little is known about the smelt population in the River Wyre, which is part of the Wyre-Lune MCZ. Therefore, this study contributes to the ongoing monitoring of smelt populations in the Wyre-Lune MCZ to better understand their spatio-temporal distributions.

The University of Highlands and Islands (UHI) were commissioned by Natural England to develop an eDNA based sampling methodology for detection of European smelt in the River Wyre.

The main aims of the study were to obtain a DNA barcode of smelt, develop primers for single species eDNA smelt detection and analyse water samples for Smelt DNA presence. From this paper, Natural England and UHI were able to establish evidence that eDNA-based approaches are highly suitable for assessing spatio-temporal distribution of UK smelt populations in estuaries and coastal freshwater.

Additionally, the species-specific assay developed here is highly promising, providing at comparable costs, a slightly higher detection probability for European smelt. However, the eDNA metabarcoding was shown to have potential in providing information on other species in the fish community which could lead to a deeper understanding of interspecific ecological interactions.

Contents

Introduction

The population decline of *Osmerus eperlanus*

The European smelt*, Osmerus eperlanus* is a small anadromous fish which largely occupies coastal and brackish waters and moves into freshwater for spawning. Historically, *Osmerus eperlanus* was an important and abundant key species in many estuarine fisheries across the UK and Northern Europe. In the 1600's it was not uncommon for 50,000 fish per day to be landed at Deptford Creek on the Thames. The fish was one of the first to receive legal protection in 1630 after Sir Robert Dulcie forbade smelt fishing to protect the smelt as they made their spawning run up the River Thames and throughout the heart of London. Unfortunately, this could not prevent population decline and by the 1870's the species was becoming scarce in the Thames. This trend was mirrored across the country as pollution and overfishing decimated rivers and the species which had resided within them.

Smelt ecology and reproductive behaviour

Densities of smelt may be affected by changes to the physical, chemical or hydrological coherence of the site, and by potential exploitation in freshwater or marine and coastal waters. Such impacts may reduce the number of adults returning to spawn. As a short-lived species, smelt has relatively few year classes, therefore high catches repeated over a few years can lead to localised extinctions. During the summer recently hatched smelt are limited to certain parts of the estuary and restricted by salinity. This makes them very vulnerable to pollution events, which can wipe out a whole year class. Cooling water intakes can kill considerable numbers of European smelt (Maitland 1997; Maes et al. 2004).

Smelt form large shoals in the lower reaches of estuaries and move up to spawn in the freshwater reaches in the early spring, usually just above the zone of saline influence. The spawning run tends to last only a few days, and disturbance during this period can have a significant negative impact on reproductive success. The thermal regime in the lower river during preceding weeks is considered the main factor in initiating spawning, but prevailing temperature, tide and flow conditions determine exactly where and when. Normally, spawning takes place during the highest spring tides when water has reached at least 5 °C. Smelt use stone, sand and gravel substratum, as well as soft vegetation, for spawning. Spawning usually takes place at night.

Osmerus eperlanus in the Wyre Estuary

There are few historical records of a smelt fishery in the River Wyre, however it is likely that the fish were probably caught as by-catch and sold on in a fishery that was primarily focussed on marine species such as Hake and Cod. Until recently, little was known about the smelt population in the Wyre, but during the last 10 years evidence of *Osmerus eperlanus* populations had been found by Environment Agency TraC (Transitional and Coastal Waters) surveys and from recreational catches either by boat or rod and line. Therefore, there a reason to believe that the Wyre Estuary might be home to a viable population of *Osmerus eperlanus.*

In 2016 the Wyre Rivers Trust were commissioned by Natural England to undertake a small-scale survey of smelt within the Wyre Estuary (Myerscough, 2017). This project was unsuccessful in identifying smelt at the tidal limit of the river Wyre. Following the conclusion of the project a number of smelt were caught at Fleetwood in October and November 2017 by the Wyre Rivers Trust. Following this work the creation of a marine conservation zone (MCZ) was recommended for the rivers Wyre and Lune, this was in order to allow local smelt populations to recover to a favourable condition. The site became an MCZ in May 2019 and the Wyre Rivers Trust were approached by Natural England in July 2020 with a view to formulating a two-year, robust monitoring proposal for smelt within the river Wyre. The aim of the monitoring would be to gather information on the makeup of smelt populations in the Wyre, their temporal and spatial distribution in the estuary, their spawning habits, the location/s of spawning, identification of the key habitats for all life stages of the smelt and whether there are any barriers to the Wyre smelt population recovering to a favourable condition. Following this work the creation of a marine conservation zone (MCZ) was recommended for the rivers Wyre and Lune, this was in order to allow local smelt populations to recover to a favourable condition.

eDNA based approaches as a method for monitoring fish populations

Environmental DNA (eDNA) approaches are increasingly used for biodiversity monitoring of aquatic habitats. These include targeted approaches to monitor individual species using species specific assays (Blackman et al., 2020; Harper et al., 2019) and metabarcoding approaches which allow the simultaneous characterization of entire biological communities. eDNA approaches are usually more effective at detecting elusive fish species than established invasive surveying techniques such as electric fishing or fyke netting, and in both lentic and lotic habitats (Griffiths et al., 2020; Hänfling et al., 2016; Pont et al., 2018; Wang et al., 2021). The potential downstream transport of eDNA complicates the spatial interpretation of eDNA data from the lotic habitats. Water samples taken from rivers integrate the fish community over a certain distance upstream of the sample location. This distance depends on river width and flow volume, ranging from hundreds of metres in small brooks to tens of kilometres in large streams such as the River Rhone in France (Pont et al. 2018). Recently, modelling approaches have been developed to take these factors into account to predict the spatial distribution of species

more accurately from eDNA records, but this research is not yet at a stage to be used in an applied context (Carraro, Mächler, Wüthrich, & Altermatt, 2020).

Despite these current limitations, eDNA metabarcoding has already been used to reconstruct riverine fish communities accurately on a catchment scale (Pont et al. 2018; Griffiths et al. 2020; Antognazza et al., 2021). Using three replicate samples per site consistently recovered the majority of species expected at a particular site based on long-term conventional data, ii) the results are very consistent among replicates, but the detection of rare species is greatly enhanced by replication; iii) eDNA profiles can reflect distinct, habitat specific fish communities for sites 10km apart; and iv) the Citizen Science approach used in this project produced high quality data. Previous studies have shown that the eDNA signal can be successfully used to determine the location and timing of spawning events in fish as spawning activity is causes a considerable spike in eDNA concentration (Antognazza et al. 2021b**;** Di Muri et al., 2022; Tsuji & Shibata, 2021).

Project aims and Objectives

This study aims to use and further develop eDNA techniques for the detection of European smelt (*Osmerus eperlanus*) on the River Wyre during the spawning season and has the following objectives:

1. Obtain DNA barcode of UK *Osmerus eperlanus* specimens, if no high confidence records available through public barcode libraries eg BOLD
2. Develop primers for single species eDNA smelt detection if, following literature search, none is available or only have low confidence
3. Analyse water samples (obtained by NE) for DNA of *Osmerus eperlanus* (single species assay).
4. Use COASTER to understand confidence in the single species assay
5. Metabarcode a subset of water samples (samples obtained by NE) to see if detection of *Osmerus eperlanus* is possible through this technique along with detecting other fish present

The outcomes of this study will inform the development an eDNA based monitoring programme of European smelt in five Marine Conservation Zone (MCZ): The Solway, Wyre-Lune, Ribble, Tamar and Medway. Four out of the five sites have no conservation advice to underpin management. These MCZ's are large and are located in estuaries where Natural England will be required to provide statutory advice on the potential impacts of plans projects and activities.

Methods

Development of a species-specific assay for European smelt

***In silico* validation:** *O. eperlaus* and *O. mordax* COI sequences were downloaded from BOLD (Ratnasingham and Hebert, 2007) and Genbank (Benson et al., 2013) and aligned in Mega 11 (Tamura et al., 2021). No sequences from *O. eperlanus* from the UK were available, so four samples from the UK were Sanger sequenced (Eurofins) for a fragment of COI and added to the alignment. This alignment was used to check the suitability of available *O. mordax* qPCR TaqMan assays targeting the COI region for use in *O. eperlanus* (Hulley et al., 2019; Berger et al., 2020; Hernandez et al., 2020). It was decided that these assays were not suitable for use in *O. eperlanus* as there were several sequence mismatches therefore new primers and probes were designed.

Primer design: Primer 3 (Untergasser et al., 2012) was used to design a suitable species-specific set of primers and probe, Primer Blast (Ye et al., 2012) was used to check for cross-amplification with other species. Primers were then checked against an alignment of 51 UK fish species (See Appendix 7) COI sequences to check for cross-amplification in species that are likely to co-occur with *O. eperlanus*. Sequences were downloaded from Genbank (Benson et al., 2013).

***In vitro* validation:** The selected assay was tested using a conventional gradient PCR to determine the optimal annealing temperature. The gradient PCR was run with target DNA extracted from tissue. The reaction was carried out in a final volume of 12µl containing 6.25µl goTaq green mastermix (Promega) 0.25µl of each primer (10µM) and 2µl DNA. The reaction was run on an Agilent Surecycler 8800 with the following conditions: 95°C for 3 minutes then 35 cycles at 95°C for 30 seconds, 53°C to 65°C for 30 seconds and 72°C for 1 minute followed by a final elongation at 72°C for 5 minutes. A qPCR was then run to test for amplification of 16 non-target species from tissue samples and 3 non-target species from eDNA samples. Because there are no species that are closely related to *O. eperlanus* in UK coastal water, the species with the most sequence similarity to primers were tested (Table 1). *O. eperlanus* scales and tissue samples from 13 non-target species were extracted using a hot sodium hydroxide and Tris (HotSHOT) protocol. Briefly, a small piece of tissue was placed in a 1.5ml microtube with 50ul alkaline lysis buffer (25mM NaOH, 0.2mM Na_2EDTA, pH12.0) and vortexed. The sample was heated to 95°C for 1.5 hours. 50µl Neutralising buffer (40mM Tris-HCL, pH 5.0) was then added and was mixed via vortexing. Samples were quantified on either a Qbit fluorometer or a QIAxpert spectrophotomer and stored at -20°C until amplification. Tissue samples were not available for European eel (*Anguilla anguilla*), european flounder (*Platichthys flesus*) and sand goby (*Pomatoschistus minutus).* Therefore, eDNA samples that had previously been tested positive for these species through eDNA metabarcoding and negative for *O. eperlanus* were used for the in vitro test instead. Additionally, four eDNA samples which had previously tested positive for with *O. eperlanus* through eDNA metabarcoding were also included.

Table 1. Fish species tested with the *O. eperlanus* assay and sequence mismatches to primers and probe (in red)

Species	Common name	Forward sequence (3'-5')	Reverse sequence (5'-3')	Probe sequence (3'-5')	No of samples	Sample type
Osmerus eperlanus	European smelt	TTATCGTCACTGCGCACGC	CCTGACATAGCCTTTCCCCG	TGGCTCATCCCCCTTATGATTGGGGCC	4	eDNA
Ameiurus nebulosus	brown bullhead	TTATTGTTACTGCCCACCGC	CCGGACATGGCTTTTCCCCG	TGACTCGTGCCCCTTATGATTGGGGCA	4	tissue
Anguilla anguilla	European eel	TCATCGTCACAGCGCATGC	CCAGACATAGCATTCCCCCG	TGACTTGTGCCATTAATAATCGGCGCC	3	eDNA
Barbus barbus	common barbel	TTATTGTTACTGCTCACCGC	CCAGACATAGCATTCCCCCG	TGACTTGTACCGCTAATAATTGGAGCC	5	tissue
Esox lucius	northern pike	TTATCGTTACAGCCCATGC	CCCGACATAGCCTTCCCCCG	TGATTAATTCCCCTAATGATTGGTGCC	6	tissue
Cyprinus carpio	common carp	TTATCGTGACTGCCCACGC	CCAGACATAGCATTCCCACG	TGACTTGTACCACTAATAATCGGAGCC	4	tissue
Leuciscus leuciscus	common dace	TTATCGTTACCGCCCACGC	CCTGACATGGCATTCCCGCG	TGACTCGTCCCACTAATGATTGGCGCA	4	tissue

Species	Common name	Forward sequence (3'-5')	Reverse sequence (5'-3')	Probe sequence (3'-5')	No of samples	Sample type
Pseudorasbora parva	stone moroko	TTATTGTTACCGCCCACGC	CCTGACATAGCATTCCCCCG	TGACTGGTCCCTTTAATGATTGGGGCC	3	tissue
Rhodeus amarus	European bitterling	TCATTGTTACCGCCCACGC	CCTGATATGGCCTTTCCCCG	TGACTTGTCCCACTAATAATTGGGGCA	3	tissue
Umbra pygmaea	eastern mudminnow	TTATCGTCACCGCCCACGC	CCAGACATGGCATTCCCTCG	TGACTAATTCCCCTAATAATTGGAGCC	2	tissue
Abramis brama	common bream	TCATCGTTACTGCCCACGC	CCTGATATGGCATTCCCACG	TGACTCGTCCCACTAATAATTGGTGCG	3	tissue
Barbatula barbatula	stone loach	TTATTGTCACCGCACATGC	CCAGACATGGCGTTCCCACG	TGACTTGTGCCACTAATGATTGGAGCC	4	tissue
Cobitis taenia	spined loach	TTATTGTTACTGCACATGC	CCTGATATAGCATTTCCGCG	TGACTTATTCCACTAATAATTGGTGCA	2	tissue
Coregonus oxyrinchus	houting	TAATCGTCACGGCCCACGC	CCCGACATGGCATTTCCCCG	TGATTAATCCCACTTATAATCGGGGCC	2	tissue

Species	Common name	Forward sequence (3'-5')	Reverse sequence (5'-3')	Probe sequence (3'-5')	No of samples	Sample type
Gobio gobio	gudgeon	TAATCGTTACTGCCCACGC	CCAGACATGGCATTCCCACG	TGGCTTGTACCACTAATAATTGGGGCC	3	tissue
Rutilus rutilus	roach	TCATCGTTACCGCCCACGC	CCTGACATAGCATTCCCACG	TGACTCGTCCCACTAATAATTGGTGCA	3	tissue
Salmo salar	atlantic salmon	TAATTGTTACAGCCCATGC	CCCGACATAGCATTCCCCCG	TGATTAATTCCTCTTATAATCGGGGCC	2	tissue
Salmo trutta	brown trout	TAATTGTTACAGCCCATGC	CCCGACATAGCATTCCCCCG	TGATTAATCCCTCTCATAATCGGAGCC	4	tissue
Pomatoschistus minutus	sand goby	TGATCGTAACAGCTCATGC	CCCGACATGGCCTTTCCTCG	TGACTCATCCCCCTTATGATCGGAGCC	2	eDNA
Platichthys flesus	European flounder	TAATCGTCACCGCACACGC	CCCGATATGGCCTTCCCTCG	TGACTTATTCCATTGATAATTGGGGCC	3	eDNA

Water samples were collected at two sites in the vicinity of known smelt spawning areas plus one upstream and one downstream ‘control’ site (Fig. 1). All water samples were collected by Natural England staff between 18th and 28th March 2022. For each sampling point, three replicate water samples (250- 700ml) were collected from mid-river at each location and individually filtered on site using Sterivex filters. Collectors recorded the volume of water filtered for each replicate sample. Filters were stored in ? in individual containers and were transported to the University of Bournemouth for DNA extraction. Individual sampling replicates underwent DNA extraction at the University of Bournemouth. The extracted DNA samples were transported to UHI Inverness for genetic analysis (see Appendix for detailed descriptions of filtration and DNA extraction).

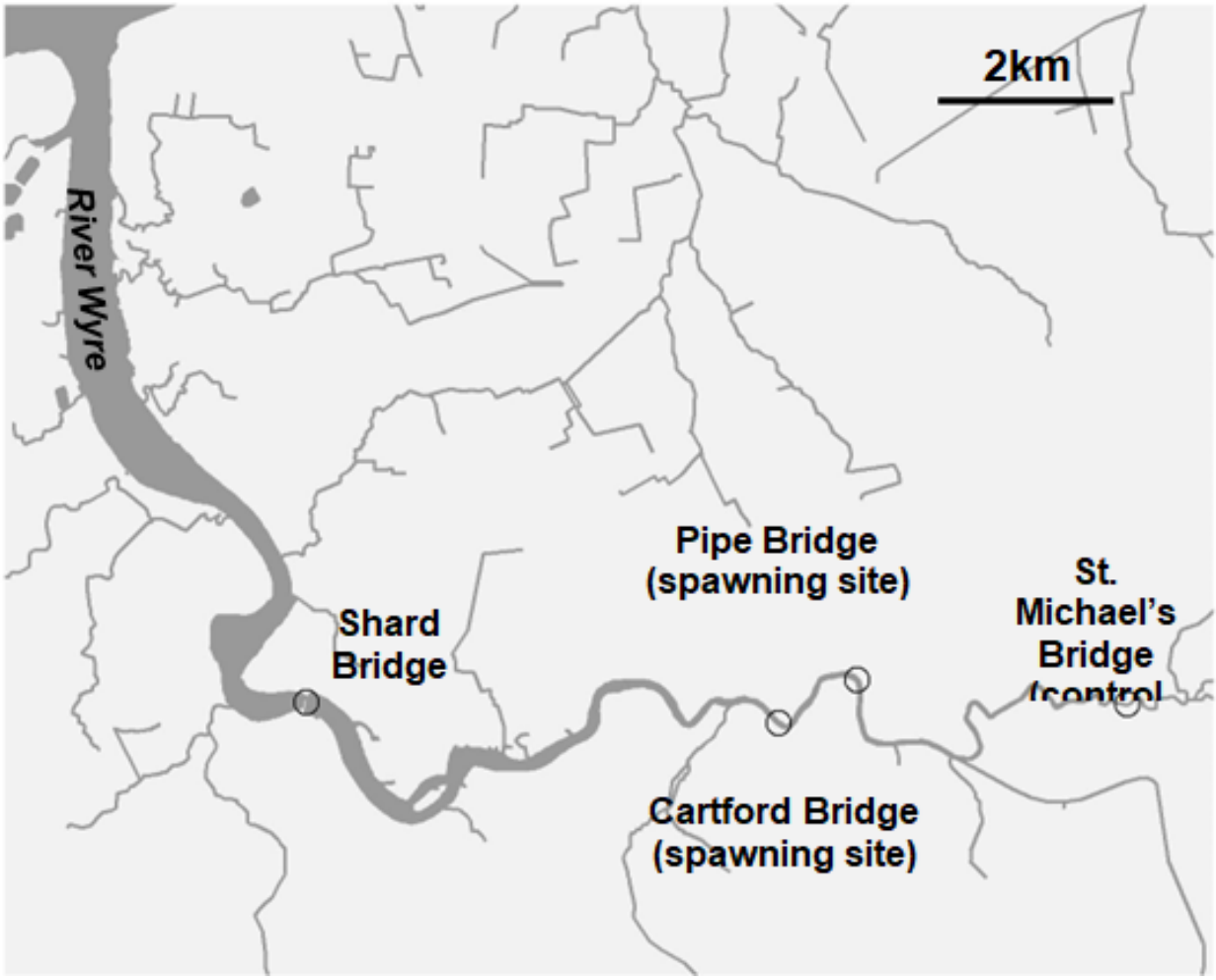


Figure 1. eDNA sampling locations at putative spawning sites and control sites above and below spawning grounds

Screening of eDNA samples using qPCR assay

The final TaqMan assay which was selected and optimised during this project amplifies a 121 bases region within the COI mitochondrial barcode marker (forward primer: 5’-TTATCGTCACTGCGCACGC-3’; reverse primer: 5’-CGGGGAAAGGCTATGTCAGG-3'; probe 5’TGGCTCATCCCCCTTATGATTGGGGCC-3’).

DNA tissue extraction of *O. eperlanus* were used to create DNA standard curves for calculating DNA concentrations of the eDNA samples. DNA concentrations of tissue extractions were measured on a Qbit fluorometer and a 5 in 1 dilution series were created starting at 6ng/µl.

Samples were prepared for qPCR in a dedicated eDNA room inside a UV cabinet to reduce the risk of contamination. Assays were carried out in a final volume of 15µl reactions containing 7.5µl of TaqMan environmental mastermix (Applied Biosystems), 1.35µl of each primer (10µM), 0.375µl of probe (10 µM), 2.45µl molecular grade water and 2µl of DNA. The reactions were run on an Agilent AriaMx qPCR machine following these conditions: 50°C for 2 minutes then 95°C for 10 minutes followed by 50 cycles of 95°C for 15 seconds and 63°C for 1 minute. All samples, including extraction blanks, were run with 10 replicates and each plate contained triplicate standard curves and non-template controls.

eDNA metabarcoding

We amplified a region of the 12S mitochondrial with vertebrate-specific PCR primers (described in Kelly et al. 2014) using a protocol previously optimized at the University of Hull. A nested two-step polymerase chain reaction (PCR) was used which adds two unique identifying tags to the sequences generated from each sample (Kitson et al. 2019); this minimizes the risk of so-called 'tag hopping' during sequencing which is a possible cause of false positives in genetic metabarcoding studies (Schnell et al. 2015). Briefly, samples (including field, filtration, extraction and PCR blanks) were first amplified using the 12S specific primers with the identifying tags attached. Each sample was amplified three times to compensate for random variation that arises when target DNA concentrations are low. The triplicate PCR products were pooled and Illumina sequencing tails added via a second PCR. PCR products from all samples were then pooled into a single sequencing library. The final library was paired-end sequenced on an Illumina MiSeq® using 2 x 300 bp V3 chemistry (Illumina Inc., CA, USA) with 10% PhiX Control added.

Raw sequencing data were analysed using a reproducible metabarcoding bioinformatic workflow, *Tapirs* (https://github.com/EvoHull/Tapirs). Sequencing reads underwent a *BLAST* (Zhang et al. 2000) taxonomic assignment against a curated UK vertebrate reference database (Harper et al. 2018). Following taxonomic assignment, a noise threshold of 0.1% of total reads per sample was applied to remove taxa with low frequency reads (Hänfling et al. 2016), and read counts were adjusted to reads per 1200ml to account for varying volumes of water filtered. Most reads were assigned to the species level. However, as the molecular marker used here cannot distinguish all vertebrate species, some taxon are assigned to a higher category. Specifically for this study, we are unable to discriminate between European perch (*Perca fluviatilis*) and zander (*Sander lucioperca*), and between brook and river lamprey (*Lampetra fluviatilis* and *L. planeri*).

Reads assigned to positive controls, reads which could not be assigned to any taxon and samples with no taxonomically assignable reads were also removed from the data set.

Data analysis was performed in the statistical programming environment R v.3.6.3 (R Core Team, 2020).

The fish community composition was summarised using two different metrics. First, we used site occupancy (the number of samples with positive detections for a given species), which is commonly used to demonstrate spatial abundance across a site. Previous studies have shown strong correlations with rank abundance of fish estimated from direct catch methods. However, the relationship with total abundance is not linear and the most common species can be underrepresented. Second, we show the relative proportion of sequences assigned to each species, which provides a better estimate for the difference in total abundance between the common and rare species but can be less accurate in differentiating the relative abundance of the rarer species.

A full detailed methodology can be found in the appendix.

Results

qPCR assay – in vitro validation

No cross-species amplification was detected in either the tissue or smelt-free eDNA samples that were tested. When screening target eDNA samples no amplification was detected in any non-template controls or extraction negatives. Amplification efficiencies ranged between 79.27-90% which is lower than the generally accepted level of above 90% (Svec et al., 2015). The efficiency of a qPCR is based on the fraction of target molecules that are amplified in one cycle. The theoretical maximum of 100% efficiency would mean that the polymerase was working at maximum capacity however it is possible to get efficiencies over 100% due to factors such as PCR inhibition. The low efficiencies in this study are likely due to the low concentration of DNA that was used for the standard curves and could be improved by using synthetic DNA for amplification curves in future. R^2 ranged between 0.98-1.0, R^2 is the coefficient of correlation of the standard curve and acceptable scores are >0.98.

eDNA metabarcoding quality control

No evidence of PCR inhibition was found in any of the eDNA samples when using the 12S metabarcoding assay. The reads assigned to each sample after initial Quality control ranged from 13,688 – 14,6578 (mean 58,283), which is comparable to that in the data set used in Willby et al. (2019). No fish eDNA was detected in any filtration/extraction blanks and PCR negative controls indication that no significant contamination has occurred at any part of the workflow.

Smelt detection

qPCR screening of eDNA samples: A total of 30 out of 54 samples tested positive for smelt eDNA with the number of positive replicates per positive sample ranging from 1-10 out of 10. The smelt DNA concentrations in the eDNA extractions ranged from 0.00006 to 0.023805 ng/ul per sample, as an average across the 10 replicates (Table S1). Taking into account variations in filtration volume this translated into 0.2ng – 4.0ng per litre of water. Smelt eDNA concentrations were highest at the putative spawning sites and peaked at the last day of sampling, 28th March 2022 (Figure 2). No positive detection was recorded at the control side upstream of the spawning locations.

eDNA metabarcoding: A total of 23 out of 54 samples tested positive for smelt eDNA with the number of smelt sequence reads per sample ranging from 14 – 8868 (Table S1). The relative proportion of smelt sequence reads from all fish sequence reads ranged from 0.001 to 0.2 per sample. Smelt eDNA concentrations were highest at the putative spawning sites and peaked at the last day of sampling, 28th March 2022 (Figure 2). No positive detection was recorded at the control side upstream of the spawning locations.

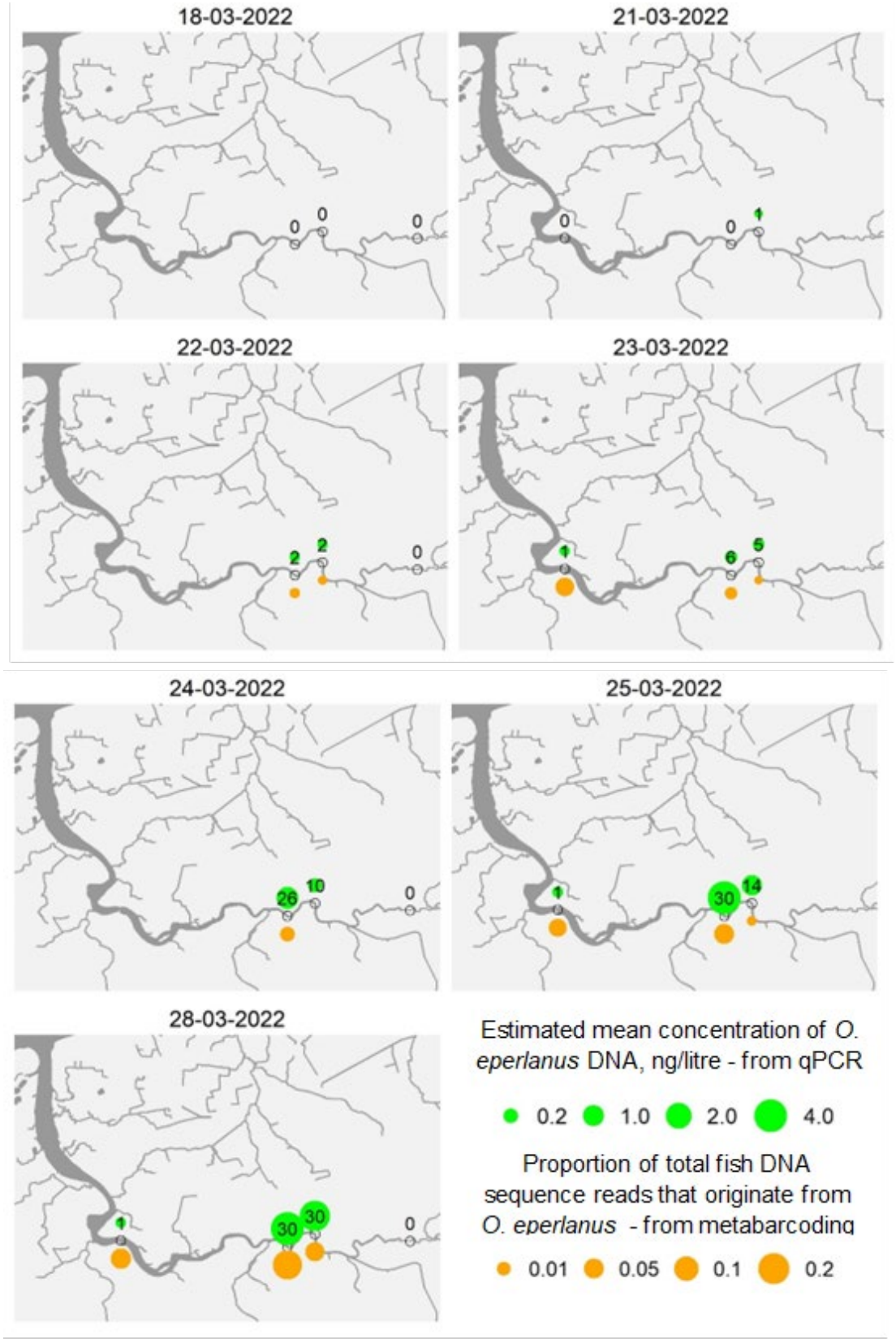


Figure 2. Detection of *Osmerus eperlanus* DNA from metabarcoding (yellow) and qPCR(green) across different sampling sites and time points. See figure legend overleaf for more information. Figure also shows the number of qPCR replicates that detected *O. eperlanus* DNA, out of a total of 30 (3 sample replicates x 10 qPCR replicates)

Comparison between methods

The two approaches for smelt eDNA detection showed a high level of congruency. On a sample basis 45 (83%) of all 54 samples had the same outcome (smelt positive or smelt negative), whereas 8 samples (15%) where positive for qPCR but not for metabarcoding and a single sample (2%) tested positive for metabarcoding only. When all three replicates at each site for a specific date where pooled congruency was even higher with 19 sites/date (90%) having the same outcome for both methods and 2 sites/date (10%) positive for qPCR only. There was also a strong correlation between metabarcoding read counts and smelt eDNA concentration estimated from qPCR (spearman rank correlation, R^2 = 0.65, P<0.001). The congruence between method in terms of detection positive detections increased with increasing DNA concentration. At smelt DNA concentrations >0.001 ng/ul (approximately half of the positive samples) both methods showed 100% congruence whereas at concentrations < 0.001 ng/ul the qPCR approach had twice as many positive detections compared to eDNA metabarcoding.

Fish community structure based on eDNA metabarcoding

A total of 19 fish taxa were detected in eDNA samples from the river Wyre respectively. The α-diversity ranged from 13 to 18 species among sites (Figure 4). The fish diversity was lowest in the site located furthest downstream in the catchment (Shard). Generally, there the fish community structure of the four sites was very similar with most species showing similar relative abundances. The most pronounced differences were shown in two species associated with the estuarine environment, flounder (*Platichthys flesus*) and sand goby (*Pomatoschistus minutus*) which were more abundant at the estuarine site Shard.

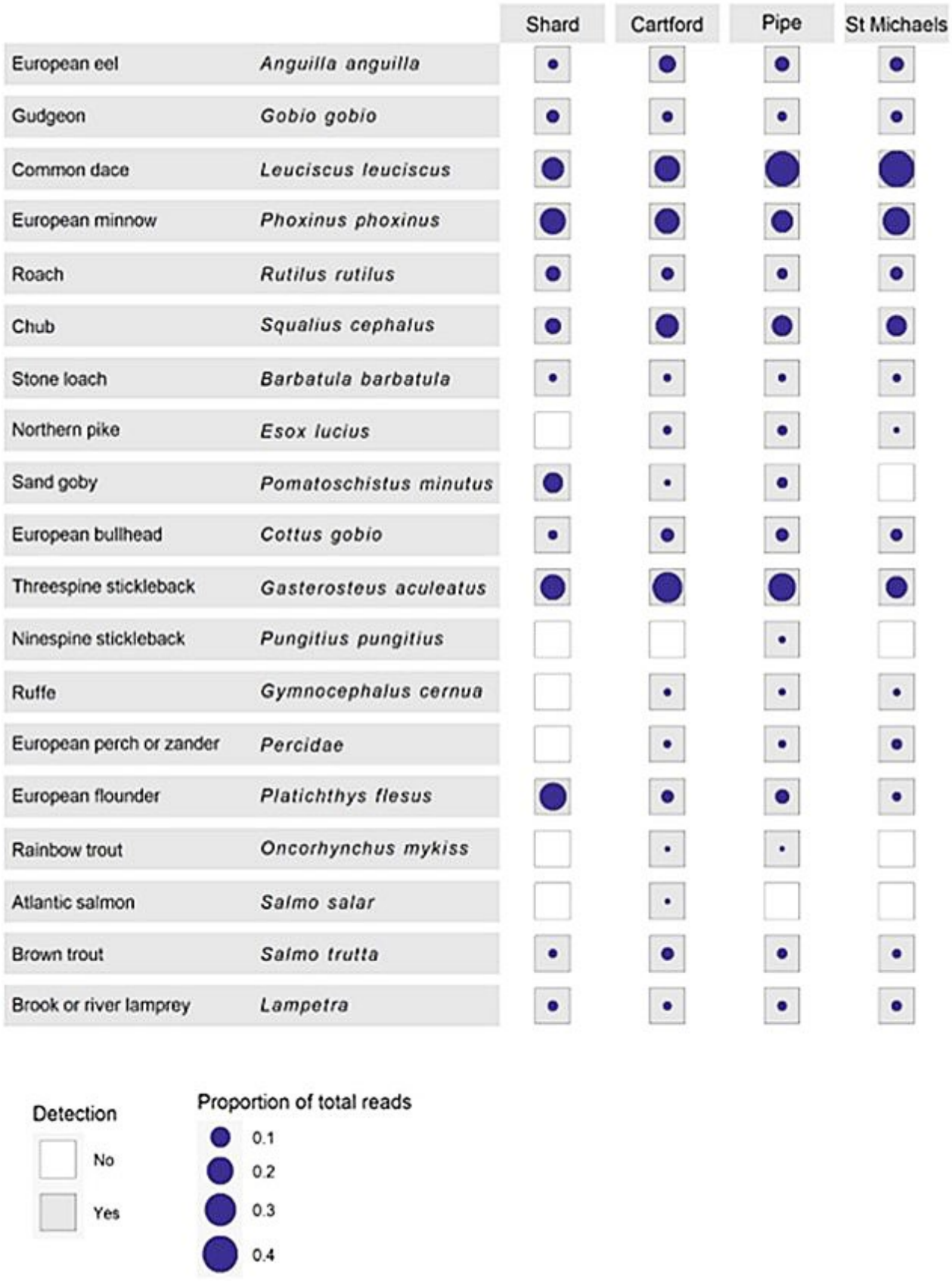


Figure 3. Other fish species detected via DNA metabarcoding. Circle sizes indicate the proportion of total fish sequence reads detected at a sampling site that originated from each taxon. Results are pooled across all sampling time points

Discussion

Performance of the qPCR assay

The in-silico and in-vitro analysis showed that the newly designed TaqMan qPCR assay is highly specific to *O. eperlanus.* The extensive testing regime included all species which are likely to co-occur with smelt in a UK context and therefore false positive detection due to cross-amplification with other species is extremely unlikely when using this assay. This was additionally verified through the inclusion of eDNA samples from sites were smelt was known to be absent and which all tested negative for smelt eDNA using this assay. Unfortunately, we were unable to determine the limit of detection (LOD) with available smelt tissue samples as concentrations and quantities were low. Without the LOD, the assay has not reached level four in the 5-level validation scale determined by Thalinger et al (2021). However, following Hernandez et al (2020) it would be a simple next step to determine the LOD using synthetic DNA, briefly synthetic DNA based on the region of interest was diluted in a nine level dilution series from 2000-1 copy per reaction with 10 replicates at each level to determine the lowest level with amplification >95% (Klymus et al., 2020). The relatively low efficiency values for the qPCR amplification during screening of the eDNA target samples suggests that further improvements in sensitivity might be achievable through using synthetic oligonucleotides for the standard curves. Alternatively, digital PCR technology which allows template DNA quantification without using standard curves could be explored.

eDNA based detection of Smelt

The detection probability of smelt was high for both qPCR and eDNA metabarcoding across the study indicating that both approaches are useful for monitoring this rare and elusive species in UK waters. This is consistent with other studies which have shown that eDNA based approaches are highly sensitive to detecting fish at low abundance and have generally lower false negative rates compared with conventional sampling methods. However, without independent data on the presence of the species in the study system at the time of eDNA sampling it is not possible to verify this here.

Despite highly congruent results between approaches the species-specific qPCR approach used here was slightly more sensitive and detected smelt DNA more often at low DNA concentrations compared to the eDNA metabarcoding approach. A higher detection probability of species-specific approaches has also been confirmed by other studies which compared the two different methodological approaches (Harper et al., 2018; Wood et al., 2019; Yu et al., 2022). However, the reasons for this can be multifaceted and there as to date been no systematic investigation of this issue (Yu et al., 2022). At least some aspect of lower DNA detection probability could be that generally lower number of PCR replicates are used in DNA metabarcoding approaches compared to species-specific assays. For example, 10 and three PCR replicates for qPCR and metabarcoding respectively were used in this study. Given that only 1-4 out of 10 qPCR replicates were positive in samples

which showed only positive result for qPCR it is possible that a higher number of PCR replicates for metabarcoding could have achieved similarly high detection probabilities.

The spatio-temporal pattern of eDNA distribution showed the highest detection probability and eDNA abundance at the putative spawning sites and no detection at the control sites above the spawning grounds. This suggests that the eDNA approach was able to detect spawning aggregations and/or activity, but verification with independent data is required to confirm this. A number of previous studies have demonstrated a clear signal of elevated eDNA concentration associated with spawning events (Di Muri et al., 2022; Tsuji & Shibata, 2021).

qPCR assay validation using the COASTER framework

Environmental DNA is increasingly used for the detection of aquatic species along with qPCR for single -species detection. There have recently been a number of tools developed to improve validation of these assays as well to improve consistency between laboratories and standardise reporting. Thalinger et al (2021) created an eDNA validation scale to determine if an assay is ready to be used for routine species monitoring, the scale is based on 122 variables and is arranged on a 5-level validation scale from 1 “incomplete” to 5 “operational”. The Confidence assessment Tool for eDNA qPCR Results (COASTER) (Harper et al 2021), has been developed to move towards the standardisation and reporting of qPCR performance metrics across eDNA studies (Harper et al., 2021). The *O. eperlanus* assay was designed following the COASTER validation checklist (see table 2) and reaches the first step of the medium confidence level category but LOD would need to be determined before the assay could be considered to have a medium confidence level.

Table 2. COASTER Assay validation checklist to assess the level of confidence in an eDNA qPCR assay from Harper et al. (2021)

Validation step	Confidence level category	Yes/No
Was in silico testing conducted and primers shown to amplify the target species?	Low	Yes
Were the primers tested on tissue from the target species?	Low	Yes
Was in silico testing conducted and potential cross amplification of non-target species shown to be low?	Low	Yes

Validation step	Confidence level category	Yes/No
Were primers tested on non-target tissue of closely related potentially co-occurring species?	Low	Yes
Did the assay successfully detect the target species at a site of known presence?	Low	Yes
Did the assay return negative results for the target species at multiple sites of known absence?	Medium	Yes
Has assay sensitivity (Limit of Detection and/or Limit of Quantification) been assessed?	Medium	No
Has site occupancy modelling (or equivalent) been conducted?	High	No
Has the probability of detecting a target species at a site been calculated?	High	No
Has the number of water samples needed to achieve reliable detection from a site been calculated?	High	No
Has the number of water samples needed to estimate probability of species absence given negative results from a site been calculated?	High	No
Has the number of qPCR replicates needed to achieve reliable detection in an eDNA sample been calculated?	High	No

Conclusion and recommendations

Based on the results of this small-scale study eDNA-based approaches appear a highly suitable tool to monitor the spatio-temporal distribution of smelt in UK estuarine and freshwaters. The species-specific assay developed here is a very promising and cost-effective approach which provides at comparable costs/sample a slightly higher detection probability for European smelt. On the other hand eDNA metabarcoding can provide information on other species in the fish community which can potentially lead to a deeper understanding of ecological interactions between species and might be informative for management decision. In order to fully operationalise the method we recommend the following further steps:

1. Fully develop the species-specific assay by completing the steps necessary to reach the highest confidence level including LOD determination, site occupancy modelling, and confidence in absence estimation.

2. Explore the use of digital PCR as a potentially, faster, cheaper, and more sensitive approach compared to qPCR.

3. A systematic and “fair” comparison of metabarcoding with the species-specific approach using equal numbers of PCR replicates.

4. Extending the sampling season for further surveys to ensure that the full length of the spawning period is covered.

References

Alberdi, A., Aizpurua, O., Gilbert, M.T.P. and Bohmann, K. (2018), “Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods Ecol Evol 9: 134--147”.

Antognazza, C. M., Britton, R. J., Read, D. S., Goodall, T., Mantzouratou, A., De Santis, V., … Andreou, D. (2021). Application of eDNA metabarcoding in a fragmented lowland river: Spatial and methodological comparison of fish species composition. *Environmental DNA*, *3*(2), 458–471.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013) GenBank. *Nucleic Acids Res* **41**: D36–D42.

Berger, C.S., Bougas, B., Turgeon, S., Ferchiou, S., Ménard, N., and Bernatchez, L. (2020) Groundtruthing of pelagic forage fish detected by hydroacoustics in a whale feeding area using environmental DNA. *Environmental DNA* **2**: 477–492.

Blackman, R. C., Ling, K. K. S., Harper, L. R., Shum, P., Hänfling, B., & Lawson-Handley, L. (2020). Targeted and passive environmental DNA approaches outperform established methods for detection of quagga mussels, Dreissena rostriformis bugensis in flowing water. *Ecology and Evolution*, *10*(23), 13248–13259.

Bolger, A.M., Lohse, M. and Usadel, B. (2014), “Trimmomatic: a flexible trimmer for Illumina sequence data”, Bioinformatics , Vol. 30 No. 15, pp. 2114–2120.

Carraro, L., Mächler, E., Wüthrich, R., & Altermatt, F. (2020). Environmental DNA allows upscaling spatial patterns of biodiversity in freshwater ecosystems. *Nature Communications*, *11*(1), 3585.

Deakin, C.T., Deakin, J.J., Ginn, S.L., Young, P., Humphreys, D., Suter, C.M., Alexander, I.E., et al. (2014), “Impact of next-generation sequencing error on analysis of barcoded plasmid libraries of known complexity and sequence”, Nucleic Acids Research, Vol. 42 No. 16, p. e129.

Di Muri, C., Lawson Handley, L., Bean, C. W., Benucci, M., Harper, L. R., James, B., … Hänfling, B. (2022). Spatio-temporal monitoring of lake fish spawning activity using environmental DNA metabarcoding (p. 2022.02.07.478003). doi: 10.1101/2022.02.07.478003

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. and Knight, R. (2011), “UCHIME improves sensitivity and speed of chimera detection”, Bioinformatics , Vol. 27 No. 16, pp. 2194–2200.

Griffiths, N. P., Bolland, J. D., Wright, R. M., Murphy, L. A., Donnelly, R. K., Watson, H. V., & Hänfling, B. (2020). Environmental DNA metabarcoding provides enhanced detection of the European eel Anguilla anguilla and fish community structure in pumped river catchments. *Journal of Fish Biology*, *97*(5), 1375–1384.

Hänfling, B., Lawson Handley, L., Read, D.S., Hahn, C., Li, J., Nichols, P., Blackman, R.C., et al. (2016), “Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods”, Molecular Ecology, Vol. 25 No. 13, pp. 3101–3119.

Harper, K.J., Tang, C.Q., Bruce, K., Ross-Gillespie, A., Ross-Gillespie, V., and Egeter, B. (2021) A framework for assessing confidence in environmental DNA qPCR assays and results. Natural England Commissioned Reports, Number NECR359

Harper, L. R., Griffiths, N. P., Lawson Handley, L., Sayer, C. D., Read, D. S., Harper, K. J., … Hänfling, B. (2019). Development and application of environmental DNA surveillance for the threatened crucian carp (Carassius carassius). *Freshwater Biology*, *64*(1), 93–107.

Harper, L. R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H. C., Gough, K. C., … Hänfling, B. (2018). Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (Triturus cristatus). *Ecology and Evolution*, *8*(12), 6330–6341.

Hernandez, C., Bougas, B., Perreault-Payette, A., Simard, A., Côté, G., and Bernatchez, L. (2020) 60 specific eDNA qPCR assays to detect invasive, threatened, and exploited freshwater vertebrates and invertebrates in Eastern Canada. *Environmental DNA* **2**: 373–386.

Hulley, E.N., Tharmalingam, S., Zarnke, A., and Boreham, D.R. (2019) Development and validation of probe-based multiplex real-time PCR assays for the rapid and accurate detection of freshwater fish species. *PLOS ONE* **14**: e0210165.

Kitson, J.J.N., Hahn, C., Sands, R.J., Straw, N.A., Evans, D.M. and Lunt, D.H. (2019), “Detecting host–parasitoid interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding”, Molecular Ecology, Vol. 28 No. 2, pp. 471–483.

Klymus, K.E., Merkes, C.M., Allison, M.J., Goldberg, C.S., Helbing, C.C., Hunter, M.E., et al. (2020) Reporting the limits of detection and quantification for environmental DNA assays. *Environmental DNA* **2**: 271–282.

Magoč, T. and Salzberg, S.L. (2011), “FLASH: fast length adjustment of short reads to improve genome assemblies”, Bioinformatics , Vol. 27 No. 21, pp. 2957–2963.

Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A., … Dejean, T. (2018). Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific Reports*, *8*(1), 10361.

Ratnasingham, S. and Hebert, P.D. (2007) BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular ecology notes* **7**: 355–364.

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P. and Coissac, E. (2011), “ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis”, Nucleic Acids Research, Vol. 39 No. 21, p. e145.

Rognes, T., Flouri, T., Nichols, B., Quince, C. and Mahé, F. (2016), “VSEARCH: a versatile open source tool for metagenomics”, PeerJ, Vol. 4, p. e2584.

Schnell, I.B., Bohmann, K. and Gilbert, M.T.P. (2015), “Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies”, Molecular Ecology Resources.

Sellers, G.S., Di Muri, C., Gómez, A. and Hänfling, B. (2018), “Mu-DNA: a modular universal DNA extraction method adaptable for a wide range of sample types”, Metabarcoding and Metagenomics, Pensoft Publishers, Vol. 2, p. e24556.

Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M.W., and Kubista, M. (2015) How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. Biomolecular Detection and Quantification 3: 9–16.

Tamura, K., Stecher, G., and Kumar, S. (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* **38**: 3022–3027.

Thalinger, B., Deiner, K., Harper, L.R., Rees, H.C., Blackman, R.C., Sint, D., et al. (2021) A validation scale to determine the readiness of environmental DNA assays for routine species monitoring. *Environmental DNA* **3**: 823–836.

Tsuji, S., & Shibata, N. (2021). Identifying spawning events in fish by observing a spike in environmental DNA concentration after spawning. *Environmental DNA*, *3*(1), 190–199.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012) Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**: e115.

Wang, S., Yan, Z., Hänfling, B., Zheng, X., Wang, P., Fan, J., & Li, J. (2021). Methodology of fish eDNA and its applications in ecology and environment. The Science of the Total Environment, 755(Pt 2), 142622.

Wood, S. A., Pochon, X., Laroche, O., von Ammon, U., Adamson, J., & Zaiko, A. (2019). A comparison of droplet digital polymerase chain reaction (PCR), quantitative PCR and metabarcoding for species-specific detection in environmental DNA. *Molecular Ecology Resources*, *19*(6), 1407–1419.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**: 134.

Yu, Z., Ito, S.-I., Wong, M. K.-S., Yoshizawa, S., Inoue, J., Itoh, S., … Hyodo, S. (2022). Comparison of species-specific qPCR and metabarcoding methods to detect small pelagic fish distribution from open ocean environmental DNA. *PloS One*, *17*(9), e0273670.

Zhang, Z., Schwartz, S., Wagner, L. and Miller, W. (2000), “A greedy algorithm for aligning DNA sequences”, Journal of Computational Biology: A Journal of Computational Molecular Cell Biology, Vol. 7 No. 1-2, pp. 203–214.

Appendices

Appendix 1: detailed methods

Sample collection

Below is the citizen science sampling protocol as used for all sites in this study.

Sampling protocol

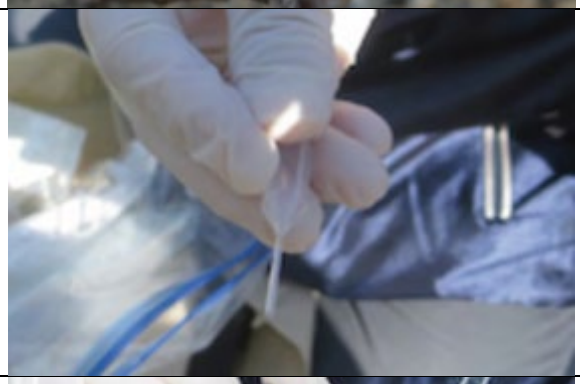
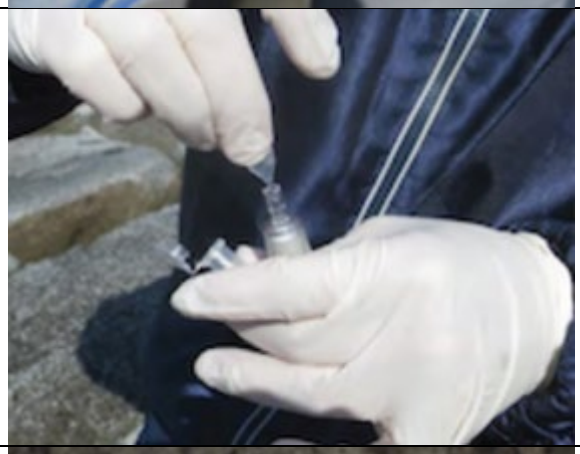
GLOVES must be work at all times of sampling

You will be working with bleach - wear protective eye gear and take care of your clothes.

DNA degrades when in sunlight and warmth. Store Sterivex filter units in cool bag immediately upon completion of the steps outlined below.

Step 1: Make 10 % bleach. Measure out 100 ml of thick bleach (using the 50ml falcon tubes) and fill up the 1 l bottles with tap water. Tighten and shake to mix.	**You can prepare the 10 % bleach the day before, but do take all the equipment with you as you will also need to make some bleach solution for soaking the sampling equipment in between sites.**
Step 2: Wipe the white buckets with 10 % bleach. Clean the whole surface (inside and out) and wipe to dry carefully. Place sampling pack within the bucket.	
Step 3: Set-up sampling equipment by attaching the whirl-pak bag to the sampler using a cable tie. Attach chain to sampler, and attach the cain to the rope. Lower sampler from bridge and fill up with water.	

Step 4: Place filled whirl-pak bag in white bucket (the bag can stand). Carefully remove the cable tie (the cable ties are releasable). Ensure that the sampling unit is held high as the water will spill if it is not. Place sampler in clean white bucket to prevent contamination.	
Step 5: Use the syringe to draw up water	
Step 6: Attach syringe to Sterivex filter unit Take care NOT to overtighten. Perform pressure filtration until <u>all</u> the sampled water (approximately 1 l) has been filtered. If the Sterivex unit is blocked and you cannot filter the 1 l, then please note the approximate volume filtered on the table provided.	
Step 7: When the filtration is finished, remove the Sterivex filter unit and fill the syringe with air. Reattach the Sterivex filter unit to the syringe, and push out the residual moisture from the filter unit. Repeat this procedure several times until no water comes out of the filter unit.	

Step 8: Seal the outlet post of the Sterivex filter unit with parafilm, while the Sterivex filter unit remains attached to the syringe.	
Step 9: Pipette RNAlater from the microcentrifuge tube using a disposable pipette.	
Step 10: Inject the RNAlater to the Sterivex filter unit from the inlet port using the disposable pipette.	
Step 11: Seal the inlet port with parafilm.	

Step 12: Wipe the surface of the Sterivex filter unit using paper towel and write the necessary information, using a felt pen: Site Sample number Date Place in the prelabelled plastic bag (the one which contained the RNAlater, disposable pipette and parafilm) Place all individual Sterivex filter units in the Site sample bag and store immediately in cool box/bag.	

*Steps 8-12 are from the Japanese eDNA society; Environmental DNA Sampling and Experiment Manual Version 2.1 (published April 25, 2019)

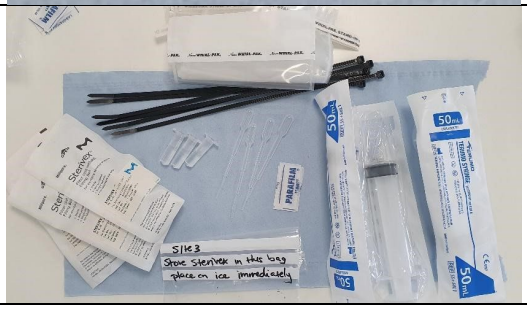
Disinfecting the water sampler and chain between sites

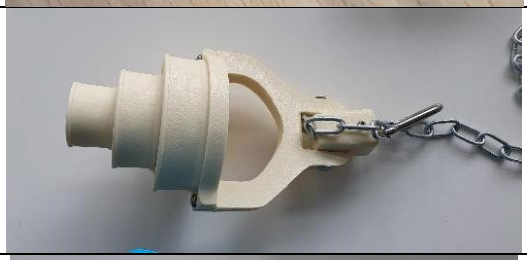
Place the sampler and chain in a white bucket. Prepare enough 10 % bleach solution to fully immerse them. Soak for a minimum of 10 minutes. You can leave to soak whilst travelling to the next site (a lid has been provided to prevent bleach solution from splashing).

Once disinfected, rinse with tap water or river water (collected form the site to be sampled and collected DOWNSTREAM from the sampling site). Wipe with blue towel (check that towel is not turning pink, if it does, please rinse again).

GLOSSARY

Water sampler	
Syringe	

Sterivex filter unit	
Parafilm (you must peel the paper, leaving you with the pliable parafilm)	
Disposable pipettes	
Microcentrifuge tube	
Sampling kit An example of all the material that are supplied per sampling site	

Attaching sampler to rope (a loop has been created for each of the supplied ropes)	
Attaching sampler to chain. 2 types of sampler are provided, one (pictured here) is to be used with the pole so the chain has to the threaded through as shown)	
Attaching sample to chain	

eDNA capture and extraction

The DNA extraction has been adapted from the detailed protocol published by The eDNA Society (2019).

The DNA was extracted in a specialist room that is dedicated to DNA extraction of eDNA samples preventing contamination.

DNA extraction negatives were included at the start of each extraction session.

1. Discharge the RNAlater on a tabletop ultracentrifuge by placing a 2 ml tube within a 50 ml conical tube, followed by adding the Sterivex filter with the inlet placed inside the 2 ml tube. This is then centrifuged at 6000 g for 1 minute.

2. Prepare the premixes using the DNeasy Blood & Tissue kit and PBS. Mix proteinase-K, AL, and PBS (–) at a ratio of 20 µl, 200 µl, and 220 µl, respectively, per Sterivex filter unit. One more premix should be prepared for the extraction blank for detecting contamination during DNA extraction.

3. Open the inlet port of the Sterivex filter unit and fill the filter unit with the above premix using a micropipette (P-1000) and a 1000 µl filter tip. (Caution: there is a ledge at the junction between inside the inlet port and the cartridge; the liquid may overflow if the tip is not properly inserted.

4. Cut the parafilm to a size of about 1 cm ´ 5 cm, and tightly seal the inlet port of the Sterivex filter unit with the parafilm.

5. Insert the Sterivex filter unit into the tube holder of the rotator and attach the tube holder to the rotator body in a manner to make the Sterivex filter unit parallel to the ground.

6. Place the rotator with Sterivex filter units in a fan oven, rotate at 10 rpm, and heat at 56°C for 20 minutes.

7. While warming the Sterivex filter unit to 56°C, prepare a 2.0 ml tube for DNA recovery (low DNA adsorption) and a 50 ml conical tube and put the 2.0 ml tube into the 50 ml conical tube. (Note: Label the cap of the 2.0 ml tube; do not push the tube deeply into the conical tube)

8. After completion of warming, carefully remove the parafilm or the luer fitting on the inlet port of the Sterivex filter unit, while preventing liquid inside from leaking.

9. Insert the inlet port of the Sterivex filter unit into the 2.0 ml tube contained in the conical tube and lightly push it down to the bottom of the 50 ml conical tube. Then, close the cap of the conical tube firmly.

10. Centrifuge the conical tube containing the Sterivex filter unit at 6,000 g for 1 minute and collect the extracted DNA in a 2 ml tube.

11. Remove the 50 ml conical tube from the centrifuge and remove the Sterivex filter unit and 2.0 ml tube in order using tweezers. (Note: The 2.0 ml tube is uncapped; handle it carefully.)

12. Discard the used Sterivex filter unit and firmly cap the 2.0 ml tube.

DNA purification took place using the Qiagen DNeasy Blood and Tissue kit by following the steps outlined below:

13. Use as many columns attached to the DNeasy Blood & Tissue kits (DNeasy) as the Sterivex filter unit filter units plus one extraction blank available. (Note: Label the necessary information on the column cap.)

14. Add 200 µl ethanol (96% to 100%) to the 2.0 ml tube containing the extracted DNA and mix thoroughly with a pipette.

15. Set the suction volume of the pipette (P-1000) at 700 µl and pipet the extracted DNA into the column. (Note: The solution may reach a larger volume than 640 µl because of a small amount of residual RNAlater. The extraction blank is obtained by adding 200 µl ethanol 96% to 100%) to 440 µl of the mixture prepared in Step 2 (see above) and mixing the mixture with a pipette.

16. Centrifuge the column containing the solution at 6000 g for 1 minute.

17. After centrifuging, remove the column collection tube and place the column on a new 2 ml collection tube. Discard the used collection tube.

18. Add 500 µl Buffer AW1 to the column and centrifuge at 6000 g for 1 minute.

19. After centrifuging, place the column to a new 2 ml collection tube. Discard used collection tubes.

20. 500 µl Buffer AW2 to the column and centrifuge at 20,000 g for 3 minutes to dry the DNeasy membrane.

21. Prepare a new 1.5 ml tube with low DNA adsorption and write the necessary information on the cap.

22. After centrifuging, place the column in the new 1.5 ml tube. Discard the used collection tubes.

23. Pipet 200 µl Buffer AE (elution buffer) directly onto the DNeasy membrane. Incubate at room temperature for 1 minute and then centrifuge at 6000 g for 1 minute to elute.

24. After centrifuging, remove the column and tightly cap the tube. Discard the used column.

25. The purified DNA can be stored stably at -80°C.

eDNA metabarcoding library preparation

Dedicated rooms were available for pre-PCR and post-PCR processes. Pre-PCR processes were performed in the IBFC eDNA facility, which has separate rooms for water filtration, DNA extraction, and PCR preparation of sensitive environmental samples. PCR reactions were set up in an ultraviolet (UV) and bleach sterilised laminar flow hood. Post-PCR processes were performed in the IBFC Genetics Laboratory.

Libraries were prepared for sequencing using a nested metabarcoding workflow with a two step PCR protocol, where Multiplex Identification (MID) tags (unique 8-nucleotide sequences) were included in the first and second PCR for sample identification (Kitson et al., 2019). DNA extracts were PCR-amplified using vertebrate-specific primers that target a 106 bp fragment of the mitochondrial 12S ribosomal RNA (rRNA) region in fish (Riaz et al., 2011). The primers were modified for the present study to include MID tags, heterogeneity spacers, sequencing primers, and pre-adapters. There were 24 unique MID tags for the forward and 24 unique MID tags for the reverse primers. This allowed each sample per library to each be labelled with a unique forward and a unique reverse primer to reduce barcode misassignment and tag jumps (Deakin et al., 2014; Schnell et al., 2015). During the first PCR each library included one extraction blank, one negative control and one positive control. The PCR positive control was zebra mbuna (*Maylandia zebra*) DNA (0.05

ng/µl). *M. zebra* is an exotic cichlid which is not found in UK freshwater habitats, thus reducing risk of positive contamination in samples.

The first PCR was performed in triplicate for each sample/control to combat stochasticity arising from low target DNA concentrations. PCR replicates for each sample/control had the same tag combination. Eight-strip PCR tubes with individually attached lids were used for PCR reactions. PCR reactions were performed in 25 µl volumes, consisting of: 12.5 µl of Q5® High-Fidelity 2x Master Mix (New England Biolabs), 0.5 µl of Thermo Scientific Bovine Serum Albumin (Fisher Scientific UK Ltd.), 7 µl of MGW (Fisher Scientific UK Ltd.), 1.5 µl of each 10 µM tagged primer, and 2 µl of template DNA. PCR reactions were sealed with mineral oil (Sigma-Aldrich) droplets. PCR was performed on an Applied Biosystems Veriti Thermal Cycler (Life Technologies) with the following thermocycling profile: 98°C for 5 mins, 35 cycles of 98°C for 10 s, 58°C for 20 s and 72°C for 30 s, 72°C for 7 mins then held at 4°C.

PCR products were stored at 4°C until PCR technical replicates for each sample/control were pooled. 2 µl of each pooled PCR product was visualised on 2% agarose gels. PCR products were deemed positive where there was amplification at the expected size (200-300 bp) on the gel. PCR products were stored at -20°C until they were pooled according to lake and band strength (no/very faint band = 20 µl, faint band = 15 µl, bright band = 10 µl, very bright band = 5 µl) on gel (Alberdi et al., 2018) to create sub-libraries for a double-size selection bead purification protocol. Ratios of 0.9x and 0.15x Mag-BIND RxnPure Plus magnetic beads (Omega Bio-tek) to 100 µl of each sub-library were used for purification. Eluted DNA (25 µl) was stored at 4°C until second PCR amplification. Sites that failed and showed signs of inhibition were cleaned and diluted, then PCR was repeated. The treatment which showed best amplification of our target region was then taken forward.

The second PCR bound pre-adapters, MID tags, and Illumina adapters to the purified sub-libraries. 10 unique forward and reverse MID tag combinations were selected and applied to sub-libraries. Two replicates were performed for each sub-library in 50 µl volumes, consisting of: 25 µl of Q5 High-Fidelity 2x Master Mix (New England Biolabs), 13 µl of MGW (Fisher Scientific UK Ltd.), 3 µl of each 10 µM tagged primer (final concentration 0.6 µM; Integrated DNA Technologies), and 4 µl of template DNA. PCR was performed on an Applied Biosystems Veriti Thermal Cycler (Life Technologies) with the following thermocycling profile: 95°C for 3 mins, 10 cycles of 98°C for 20 s and 72°C for 1 min, 72°C for 5 mins then held at 4°C. PCR duplicates for each sub-library had the same tag combination.

PCR products were stored at 4°C until duplicates for each sub-library were pooled. 2 µl of each pooled PCR product was visualised on 2% agarose gels. PCR products were deemed positive where there was amplification at the expected size (300-400 bp) on the gel. Sub-libraries were stored at 4°C until double-size selection bead purification. Ratios of 0.9x and 0.15x Mag-BIND RxnPure Plus magnetic beads (Omega Bio-tek) to 50 µl of each sub-library were used for purification. Eluted DNA (25 µl) was stored at 4°C until normalisation and final purification.

Sub-libraries were quantified on a Qubit 3.0 fluorometer using a dsDNA HS Assay Kit (Invitrogen) and normalised by pooling according to sample size and library concentration. The pooled library was purified using the same ratios, volumes, and protocol as the second PCR purification. Based on the Qubit™ concentration, the library was diluted to 4 nM. The library was checked with an Agilent 2200 TapeStation using High Sensitivity D1000 ScreenTape (Agilent Technologies) to verify a fragment of the expected size (315 bp) remained. The library was then quantified by qPCR using the NEBNext Library Quant Kit for Illumina (New England Biolabs). Based on the qPCR concentration, the library was adjusted to 4 nM and denatured following the Illumina MiSeq library denaturation and dilution guide. The final library was sequenced at 13 pM with 10% PhiX Control on an Illumina MiSeq using 2 x 300 bp V3 chemistry (Illumina).

eDNA metabarcoding bioinformatics

Sequencing data was automatically demultiplexed to separate (forward and reverse) fastq files per library using the onboard Illumina MiSeq Reporter software. Library sequence reads were further demultiplexed to sample using a custom Python script. *Tapirs*, a reproducible workflow for the analysis of DNA metabarcoding data (https://github.com/EvoHull/Tapirs), was used for taxonomic assignment of demultiplexed sequencing reads. *Tapirs* uses the *Snakemake* workflow manager (Köster & Rahmann 2012) and a *conda* virtual environment to ensure software compatibility.

Raw reads were quality trimmed from the tail with a 5 bp sliding window (qualifying phred score of Q30 and an average window phred score of Q30) using *fastp* (Chen et al. 2018), allowing no more than 40% of the final trimmed read bases to be below Q30. Primers were removed by trimming the first 21 and 27 bp of forward and reverse reads respectively. Reads were then tail cropped to a maximum length of 170 bp and reads shorter than 90 bp were discarded.

Sequence read pairs were merged into single reads using *fastp*, provided there was a minimum overlap of 20 bp, no more than 5% mismatches and no more than 5 mismatched bases between pairs. Only forward reads were kept from read pairs that failed to be merged. A final length filter removed any reads longer than 190 bp to ensure sequence lengths approximated the expected fragment size (~170 bp) and removed any non-specific bacterial ribosomal RNA product (~255 bp) known to be amplified by the MiFish primers.

Redundant sequences were removed by clustering at 100% read identity and length (--derep_fulllength) in *VSEARCH* (Rognes et al. 2016). Clusters represented by less than three sequences were omitted from further processing. Reads were further clustered (--cluster_unoise) to remove redundancies due to sequencing errors (retaining all cluster sizes). Retained sequences were screened for chimeric sequences with *VSEARCH* (--uchime3_denovo).

The final clustered, non-redundant query sequences were then compared against a curated UK vertebrate reference database (Harper et al. 2018) using *BLAST* (Zhang et al. 2000). Taxonomic identity was assigned using a custom majority lowest common ancestor

(MLCA) approach based on the top 2% query *BLAST* hit bit-scores, with at least 90% query coverage and a minimum identity of 98%. Of these filtered hits, 80% of unique taxonomic lineages therein had to agree at descending taxonomic rank (domain, phylum, class, order, family, genus, species) for it to be assigned a taxonomic identity. If a query had a single *BLAST* hit it was assigned directly to this taxon only if it met all previous MLCA criteria. Read counts assigned to each taxonomic identity were calculated from query cluster sizes. Lowest taxonomic rank was to species and assignments higher than order were classed as unassigned.

Appendix 2: Additional tables and figures

Table S1. Sample details and raw data from screening of eDNA samples with qPCR and metabarcoding approach. Sample codes used for the analysis, lab codes for DNA sample storage and collectors code, date of collection, descriptive site name, number of qPCR replicates positive for smelt, average Cq value across qPCR replicates, estimated concentration of smelt DNA as average across replicates, number of metabarcoding sequence reads assigned to smelt

Sample code	Lab code	Collectors code	Site name	Date of collection	Plate number	qPCR No of positive reactions /10	qPCRAve Cq	qPCR ng/ul	Metabarcoding No of *O. eperlanus* reads
SM-01	AK250	extraction blank	n/a	n/a	1	0	No Cq	0	0
SM-02	AK251	CO-1	St Michaels	18/03/22	1	0	No Cq	0	0
SM-03	AK252	CO-2	Shard	21/03/22	1	0	No Cq	0	0
SM-04	AK253	CO-3	St Michaels	24/03/22	1	0	No Cq	0	0

Sample code	Lab code	Collectors code	Site name	Date of collection	Plate number	qPCR No of positive reactions /10	qPCRAve Cq	qPCR ng/ul	Metabarcoding No of *O. eperlanus* reads
SM-05	AK254	CO-4	St Michaels	22/03/22	1	0	No Cq	0	0
SM-06	AK255	CO-5	Shard	23/03/22	1	1	35.53	0.000131438	148
SM-07	AK256	CO-6	Shard	25/03/22	2	1	37.95	0.000105766	358
SM-08	AK257	CO-7	St Michaels	28/03/22	2	0	No Cq	0	0
SM-09	AK258	CO-8	Shard	28/03/22	2	1	36.78	0.000118949	185

Sample code	Lab code	Collectors code	Site name	Date of collection	Plate number	qPCR No of positive reactions /10	qPCRAve Cq	qPCR ng/ul	Metabarcoding No of *O. eperlanus* reads
SM-10	AK259	S3-1	Cartford bridge	21/03/22	2	0	No Cq	0	0
SM-11	AK260	S3-2	Cartford bridge	21/03/22	2	0	No Cq	0	0
SM-12	AK261	S3-3	Cartford bridge	21/03/22	2	0	No Cq	0	0
SM-13	AK262	S7-1	Cartford bridge	23/03/22	3	3	36.80	0.000457409	0
SM-14	AK263	S7-2	Cartford bridge	23/03/22	3	3	36.04	0.000368323	352
SM-15	AK264	S7-3	Cartford bridge	23/03/22	3	0	No Cq	0	0

Sample code	Lab code	Collectors code	Site name	Date of collection	Plate number	qPCR No of positive reactions /10	qPCRAve Cq	qPCR ng/ul	Metabarcoding No of *O. eperlanus* reads
SM-16	AK265	extraction blank	n/a	n/a	3	0	No Cq	0	0
SM-17	AK266	S1-1	Cartford bridge	18/03/22	3	0	No Cq	0	0
SM-18	AK267	S1-2	Cartford bridge	18/03/22	3	0	No Cq	0	0
SM-19	AK268	S1-3	Cartford bridge	18/03/22	3	0	No Cq	0	0
SM-20	AK269	S2-1	Pipe Bridge	18/03/22	3	0	No Cq	0	0
SM-21	AK270	S2-2	Pipe Bridge	18/03/22	4	0	No Cq	0	0

Sample code	Lab code	Collectors code	Site name	Date of collection	Plate number	qPCR No of positive reactions /10	qPCRAve Cq	qPCR ng/ul	Metabarcoding No of *O. eperlanus* reads
SM-22	AK271	S2-3	Pipe Bridge	18/03/22	4	0	No Cq	0	0
SM-23	AK272	S4-1	Pipe Bridge	21/03/22	4	1	36.64	6.11E-05	0
SM-24	AK273	S4-2	Pipe Bridge	21/03/22	4	0	No Cq	0	0
SM-25	AK274	S4-3	Pipe Bridge	21/03/22	5	0	No Cq	0	0
SM-26	AK275	S5-1	Cartford bridge	22/03/22	5	0	No Cq	0	0
SM-27	AK276	S5-2	Cartford bridge	22/03/22	5	2	36.46	0.000359144	287
SM-28	AK277	S5-3	Cartford bridge	22/03/22	5	0	No Cq	0	0

Sample code	Lab code	Collectors code	Site name	Date of collection	Plate number	qPCR No of positive reactions /10	qPCRAve Cq	qPCR ng/ul	Metabarcoding No of *O. eperlanus* reads
SM-29	AK278	S6-1	Pipe Bridge	22/03/22	6	0	No Cq	0	33
SM-30	AK279	S6-2	Pipe Bridge	22/03/22	6	1	35.86	0.000133199	0
SM-31	AK280	S6-3	Pipe Bridge	22/03/22	6	1	36.23	0.000105804	0
SM-32	AK281	S8-1	Pipe Bridge	23/03/22	6	2	36.22	0.000229869	16
SM-33	AK282	S8-2	Pipe Bridge	23/03/22	6	2	35.98	0.000267124	61
SM-34	AK283	S8-3	Pipe Bridge	23/03/22	6	1	37.06	6.21E-05	0
SM-35	AK284	S9-1	Cartford bridge	24/03/22	6	7	34.48	0.002557334	337

Sample code	Lab code	Collectors code	Site name	Date of collection	Plate number	qPCR No of positive reactions /10	qPCRAve Cq	qPCR ng/ul	Metabarcoding No of *O. eperlanus* reads
SM-36	AK285	S9-2	Cartford bridge	24/03/22	7	10	34.06	0.004624621	269
SM-37	AK286	S9-3	Cartford bridge	24/03/22	7	9	35.18	0.002303494	423
SM-38	AK287	S10-1	Pipe Bridge	24/03/22	7	3	35.64	0.000557594	0
SM-39	AK288	S10-2	Pipe Bridge	24/03/22	7	3	35.30	0.000628878	0
SM-40	AK289	S10-3	Pipe Bridge	24/03/22	7	4	35.83	0.000857943	0
SM-41	AK290	extraction blank	n/a	n/a	7	0	No Cq	0	0
SM-42	AK291	S11-1	Pipe Bridge	25/03/22	7	6	35.66	0.001034048	36

Sample code	Lab code	Collectors code	Site name	Date of collection	Plate number	qPCR No of positive reactions /10	qPCRAve Cq	qPCR ng/ul	Metabarcoding No of *O. eperlanus* reads
SM-43	AK292	S11-2	Pipe Bridge	25/03/22	8	7	34.49	0.007001373	14
SM-44	AK293	S11-3	Pipe Bridge	25/03/22	8	1	35.06	0.000643434	148
SM-45	AK294	S12-1	Cartford bridge	25/03/22	8	10	33.51	0.01848264	1523
SM-46	AK295	S12-2	Cartford bridge	25/03/22	8	10	33.21	0.022663664	2881
SM-47	AK296	S12-3	Cartford bridge	25/03/22	4	10	32.99	0.006285301	1517
SM-48	AK297	S13-1	Cartford bridge	28/03/22	4	10	31.71	0.014375977	6357
SM-49	AK298	S13-2	Cartford bridge	28/03/22	4	10	31.71	0.014112572	2057

Sample code	Lab code	Collectors code	Site name	Date of collection	Plate number	qPCR No of positive reactions /10	qPCRAve Cq	qPCR ng/ul	Metabarcoding No of *O. eperlanus* reads
SM-50	AK299	S13-3	Cartford bridge	28/03/22	5	10	33.83	0.009144294	5985
SM-51	AK300	S14-1	Pipe Bridge	28/03/22	5	10	34.89	0.004834975	875
SM-52	AK301	S14-2	Pipe Bridge	28/03/22	5	10	32.47	0.019077323	8868
SM-53	AK302	S14-3	Pipe Bridge	28/03/22	8	10	33.17	0.023804965	1063
SM-54	AK303	extraction blank	n/a	n/a	8	0	No Cq	0	0

Table S2. fish species included in alignment used in the in-silico analysis to check qPCR assay specificity

Latin name	common name
Abramis brama	common bream
Alburnoides bipunctatus	bleak
Alburnus alburnus	common bleak
Ameiurus nebulosus	brown bullhead
Ammodytes tobianus	sand eel
Anguilla anguilla	European eel
Babka gymnotrachelus	racer goby
Barbatula barbatula	stone loach
Barbus barbus	common barbel
Blicca bjoerkna	white bream
Carassius gibelio	Prussian carp
Clupea harengus	Atlantic herring
Cobitis taenia	spined loach
Coregonus oxyrinchus	houting
Cottus gobio	european bullhead
Ctenopharyngodon idella	grass carp
Cyprinus carpio	common carp
Dicentrarchus labrax	european bass
Echiichthys vipera	lesser weever
Esox lucius	northern pike
Gasterosteus aculeatus	three spined stickleback
Gobio gobio	gudgeon
Gymnocephalus cernua	ruffe
Lampetra planeri	brook lamprey
Lepomis gibbosus	pumpkin seed
Leucaspius delineatus	sunbleak
Leuciscus aspius	asp
Leuciscus idus	ide
Leuciscus leuciscus	common dace
Lota lota	burbot
Misgurnus fossilis	weatherfish
Neogobius fluviatilis	monkey goby
Perca fluviatilis	European perch
Platichthys flesus	European flounder
Pomatoschistus minutus	sand goby
Ponticola kessleri	bighead goby
Pseudorasbora parva	stone moroko
Pungitius pungitius	ninespine stickleback
Rhodeus amarus	European bitterling
Romanogobio belingi	northern whitefin gudgeon
Rutilus rutilus	roach

Latin name	common name
Salmo salar	atlantic salmon
Salmo trutta	brown trout
Sander lucioperca	zander
Scardinius erythrophthalmus	rudd
Silurus glanis	wells catfish
Solea solea	common sole
Squalius cephalus	common chub
Thymallus thymallus	grayling
Tinca tinca	tench
Umbra pygmaea	eastern mudminnow

www.gov.uk/natural-england