DOCTOR OF PHILOSOPHY

The Development of eDNA Metabarcoding for the Monitoring of Fishes in Estuaries

Gibson, Thomas

Award date:
2022

Awarding institution:
Bangor University

PRIFYSGOL
BANGOR
UNIVERSITY

The Development of eDNA Metabarcoding for the Monitoring of Fishes in Estuaries

A thesis submitted for the degree of Doctor of Philosophy

Thomas Ian Gibson

Molecular Ecology and Evolution at Bangor, School of Natural Sciences, Environmental Centre Wales, Bangor University, Bangor, Gwynedd, LL57 2UW

Ysgoloriaethau Sgiliau Economi Gwybodaeth
Knowledge Economy Skills Scholarships

Molecular Ecology and Evolution @ Bangor
Prifysgol Bangor University

Cronfa Gymdeithasol Ewrop
European Social Fund

General Information

Primary Supervisors: Professor Simon Creer, Dr Amy Ellison and Professor Gary Carvalho (Bangor University)

External Advisors: Professor Stefano Mariani (Liverpool John Moores University), Dr Lori Lawson-Handley (University of Hull) and Dr Tristan Hatton-Ellis (NRW)

KESS2 Partner Organisation: Natural Resources Wales

- Additional data was supplied by the Environment Agency (Chapter 4)

Funding Organisations: KESS2 Knowledge Exchange Scholarship (https://kess2.ac.uk) and the Fisheries Society of the British Isles (FSBI, Chapter 3; https://fsbi.org.uk)

Grant Number: BUK2147

- Chapter 3 was covered by an additional £5000 'Small Research Grant' from the FSBI

Authors Statement

I hereby declare that this thesis is the results of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards.

Ethics Statement

The fish surveys which generated the data used in this PhD were routine monitoring conducted by NRW and the EA and were not organised for this thesis. Therefore, any ethical considerations and licencing for fish sampling were already covered by these organisations. Where environmental samples were taken, and access to sites was required, the author contacted the relevant landowner and or the competent authority (e.g. NRW). This ensured all sampling was conducted in accordance with any specific local regulations. Therefore, I am confident that this project has been conducted in accordance with all relevant laws and regulations.

T.I. Gibson 31.06.22

Abstract

Estuaries are one of the most ecologically and economically valuable ecosystems on the planet, they are also among the most degraded. Currently, in the United Kingdom the ecological quality of estuaries ('transitional waters') is monitored by assessments of key ecological groups, such as fish, to meet the demands of legislation derived from the Water Framework Directive (WFD). At present within Wales and England, Natural Resources Wales (NRW) and the Environment Agency (EA), respectively, use several combined fishing methods to survey fish communities. Equivalent methodologies are used across Europe. However, multimethod fishing is expensive, difficult to implement consistently and potentially destructive. Environmental DNA (eDNA) analysis technology is an emerging survey method for fish. There are still relatively few studies using eDNA metabarcoding to assess fish biodiversity in estuaries. The aim of this project is to conduct the basic research which will enable the future development of an eDNA metabarcoding tool to assess the biodiversity of fishes in estuaries. Three studies were conducted in macrotidal estuaries in Wales and England. Comparable methodologies were used to generate data on fish eDNA. Replicate surface water samples were collected at each sampling event/station, DNA was extracted from samples and subjected to metabarcoding analysis using an assay targeting teleost fishes. Fish assemblage data was analysed using up to date statistical approaches. The first study aimed to compare the spatial fish assemblage composition detected via eDNA with conventional fishing surveys, and also investigate seasonal patterns, in the Dee (Wales). Sampling was conducted alongside three fishing gear types, in a spatially systematic design in October 2018. Seasonal changes in composition were investigated by re-sampling a subset of stations using eDNA only, in June 2018. In autumn, eDNA detected the majority of species caught by fishing, and detected a greater species richness, per a given sampling effort, than two gear types. Assemblage composition was also correlated with salinity, consistently across seasons. The aim of the second study was to investigate the short-term variability in the fish assemblage, in the Conwy Estuary (Wales). In Autumn 2020, samples were taken at a single station at high and low tide over 15 days, covering a spring to neap tidal cycle. Temporal variation in the assemblage composition of fish species were correlated with changes in salinity, which occurred at different tidal states and due to an episodic increase in river flow, and to a lesser extent tidal range. The third study aimed to compare the fish assemblage detected via eDNA to fishing gears in three estuaries in northeast England, over two seasons: early summer and autumn. Environmental DNA was sampled alongside fishing at multiple sampling stations in two estuaries in autumn 2016, and in all three estuaries in early summer and autumn 2017. The majority of species caught by fishing were detected by eDNA, including species of conservation interest and a none-native. Species richness estimates for each estuary were in some cases greater using eDNA compared to fishing. Numerous novel species were detected via eDNA and a different assemblage composition was detected relative to one netting type. Analyses of eDNA separately from fishing showed it could detect differences in assemblage composition between seasons and estuaries. In conclusion, eDNA may be an effective method to survey fishes in estuaries for biomonitoring purposes. Correlations in the eDNA assemblage with temporal and spatial variation in ecological variables, such as salinity, also have important implications for biomonitoring survey design. The implications of these studies for biomonitoring, and the requirements for further research are discussed throughout and summarised in the general discussion.

General Acknowledgements

I would like to thank my supervisors and collaborators, listed above, for their support during this project. Particularly mention should also be given to staff at NRW, EA, DAERA and Marland Marine Ltd. who facilitated the sampling or collected the fish data. I am also grateful to KESS2 (and the European Commission) for supporting this work. I am grateful to the FSBI and the Genetics Society for supporting my research. Although in the latter case I was unable to utilise the £1500 'Heredity Fieldwork Grant' (https://genetics.org.uk) I was awarded due to the pandemic.

Acknowledgement and credit specific to the project is provided with each chapter. However, I would also like to make the following thanks to people who have contributed to this project indirectly, or who allowed me to gain the background knowledge and skills that contributed to its completion. Firstly, I would like to thank my parents, Dr Ian Gibson and Grace Gibson for their encouragement and practical support at the end of the PhD project. I would also like to thank my sisters, Rosie and Mary, my grandfather Richard Bolton and my family more generally, for their encouragement and advice. Secondly, I would like to thank the scientists from which I have learned over the years either in education or in work. At Plymouth University: Dr Manuela Truebano-Garcia for supporting me during my first practical undertaking in molecular biology, Dr Oliver Tills for including me in his research projects and Dr Andy Foggo for his inspirational lectures on biodiversity science. At Aberdeen University: Dr Tara Marshall for her encouragement and getting me interested in fisheries science, Dr Phil Smith and Dr John Hartley (Hartley Anderson Ltd.) for all their insight on marine ecology in an applied context. At Marine Scotland Science: Dr Helen Dobby for teaching me how to properly write code in R, Jim Drewery and Adrian Weetman for showing how research surveys *should* be run. At Bangor University: my colleagues both within the immediate research group of which I was part (MEEB, formally MEFGL) and in the wider School and University. Specifically, I would like to thank the following here: Dr Georgina Brennan, Dr William Perry, Dr Alice Evans, Kirthana Pillay, Amy Gresham, Owain Barton, Tom Major, Bethan Pugh, Nicholas Welsby, Abdullah Munawar Rafiq, Dr Luke Hillary, Carlo Zampieri, Enrica Gargiulo and Sara Maggini. Thirdly, I would like to thank my friends for making life interesting and keeping me sane. In addition to those who are not my immediate colleagues, particularly Becky Irvin, Claire Brewster, Jamie Alison, Maya Gill-Taylor and Beth Ellsmore. Apologies if I have forgotten anyone. This work is dedicated to all the above, particularly my late grandmothers, Valerie Gibson and Claire Bolton.

Table of Contents

Forward and Aims

The thesis is structured around a literature review (Chapter 1), three data chapters (Chapter 2 to 4) which provide the core data for the project and finishes with a general discussion of the findings and suggestions for further work (Chapter 5). Chapters 2 to 3 were conducted by Bangor University in collaboration with NRW. The data for Chapter 4 was provided by the SeaDNA project (Liverpool John Moores University and University of Bristol) in collaboration with the EA. Supplementary material is provided with each data chapter. Relevant method development, as four appendices, follows Chapter 2. Throughout the work the emphasis is on considering eDNA metabarcoding as a complementary fish sampling technique compared to fishing, rather than considering the two approaches in an oppositional manner. Therefore, this research attempts to draw on the substantial research in estuarine and coastal science and on fishes in estuaries, in combination with molecular ecology, as much as possible.

Overall Aim

The aim of this project is to conduct the ecological research to enable the future development of an environmental DNA metabarcoding tool, which will be used to assess the biodiversity of fishes in estuaries.

Overview of Chapter Aims

- **Chapter 1.:** The Relevance of Environmental DNA to Biomonitoring of Fish in Estuaries: Present Knowledge and Future Directions
    - **Aim:** Provide a comprehensive literature review on estuarine ecology, the biomonitoring of fishes in estuaries, and how eDNA may be able to contribute to this.
- **Chapter 2.:** Environmental DNA Metabarcoding for the Biodiversity Assessment of Fishes in a Macrotidal Estuary: a Comparison with Established Fish Survey Methods.
    - **Aim:** Compare the fish assemblage detected via eDNA metabarcoding of surface water samples compared to conventional fishing gears in a macrotidal estuary, and determine if eDNA can detect ecological relevant spatial and seasonal patterns in assemblage composition.
- **Chapter 3.:** Environmental DNA Metabarcoding Detects Short-Term Temporal Variability in the Fish Assemblage within a Macrotidal Estuary.
    - **Aim:** Investigate the short-term tidal variability in the fish assemblage within a well-mixed macrotidal estuary detected by metabarcoding of eDNA from surface water samples.
- **Chapter 4.:** Environmental DNA Metabarcoding Reveals Ecologically Relevant Variation in Fish Assemblages Between Macrotidal Estuaries and Across Seasons.
    - **Aim:** To compare the fish assemblage detected via eDNA metabarcoding of surface water samples to sampling with conventional fishing gears in three macrotidal estuaries, over two seasons: autumn and early summer. In addition, to determine if eDNA can detect ecological relevant changes in assemblage composition between estuaries and seasons.
- **Chapter 5.:** The Implications of Fish eDNA Metabarcoding for Biomonitoring and the Requirement for Future Research.
    - **Aim:** To summarise the overall results, discuss how this data can be used in an applied monitoring context and outline future areas for research.

Chapter 1.

The Relevance of Environmental DNA to Biomonitoring of Fish in Estuaries: Present Knowledge and Future Directions

Authors: Thomas I. Gibson and Simon Creer

Author Contributions

Phase	Author
Research and Writing	TIG
Manuscript Review	TIG, SC

1. Introduction

At the interface between land, sea and river, estuaries are intermediate and dynamic habitats which show marked variation in their physical, chemical and topographic characteristics (McLusky and Elliott, 2004). Several definitions of estuaries exist, extensively reviewed in Wolanski (2007) and Elliott and McLusky (2002). However, an estuary can be defined as a "semi enclosed coastal body of water which is connected to the sea either permanently or periodically, has a salinity that is different from that of the adjacent open ocean due to freshwater inputs, and includes a characteristic biota". This definition includes tidal and closed non-tidal estuaries, hyperhaline systems and coastal systems without inflowing rivers where estuarine conditions are created by groundwater inputs (Whitfield and Elliott, 2011). Within certain European Union (EU) legislation, estuarine systems are included within the similarly broad definition of transitional waters. These are "bodies of surface water in the vicinity of river mouths which are partly saline in character as a result of their proximity to coastal waters but which are substantially influenced by freshwater flows" (EC, 2000). Transitional waters encompass estuaries and other brackish water systems such as lagoons, fjords, rias etc. (McLusky and Elliott, 2007). Most modern estuaries, particularly in Europe (Elliott and McLusky, 2002) were formed when valleys began to be flooded approximately 10, 000 to 12, 000 BP during Holocene sea level rises (Wolanski, 2007). One common feature of estuarine ecosystems is the gradient of environmental conditions from the open sea, through the estuary and into freshwater (Elliott and McLusky, 2002). Numerous physio-chemical variables show gradients with many operating as co-variables. Abiotic gradients occur both spatially, in the vertical, horizontal, cross-sectional and geographic planes and temporally, over tidal cycles, annual, seasonal and decadal time scales. However, the key spatial gradients are salinity, turbidity, sediment composition and oxygen, whereas the key prevailing temporal gradient in estuaries is temperature (McLusky, 1993). To give a spatial example, and as emphasised by the above definitions, salinity (practical salinity units, PSU) falls from 35 – 37 in the sea to < 0.5 up river, with intermediate (brackish) salinities within the estuary (McLusky and Elliott, 2004). To give a temporal example, temperature can vary substantially over tidal cycles, particularly in temperature northern estuaries (Elliott and Whitfield, 2011), and seasonally (Thiel *et al.*, 1995; Marshall and Elliott, 1998). Nevertheless this is a generalisation as spatial variables, e.g. salinity, can change substantially over tidal cycles (Elliott and Whitfield, 2011).

The environmental gradients within estuaries directly impact estuarine biodiversity. Estuaries are characterised by a low taxonomic diversity compared to adjacent marine and freshwater ecosystems, partly due to the physiological challenge of adapting to a highly

variable and consistently unpredictable environment. A key pressure is the osmotic stress caused by the highly variable salinity (Costanza *et al.*, 1993; Telesh and Khlebovich, 2010; Whitfield *et al.*, 2012). Species richness reaches its nadir at salinities of around 5 – 8 (McLusky, 1993; Whitfield *et al.*, 2012) the 'Artenminimum' (species minimum; Remane, 1934). The Artenminimum was initially described by Adolf Remane (Remane, 1934) for zoobenthic species in the Baltic sea (an open brackish sea). The salinity range from 5 – 8 is still considered an important ecological, physiological and evolutionary barrier, as only certain groups have evolved the required osmoregulatory traits to cross it (Telesh and Khlebovich, 2010). Although, the salinity zone from 5 – 8 is no-longer considered to mark the absolute downstream and upstream limits of freshwater and marine species respectively (Whitfield *et al.*, 2012) as Remane believed (McLusky, 1993). However, the overall trend in estuarine biodiversity still holds. Generally, species richness declines from marine through to brackish waters, as the number of marine species falls. Species richness then increases again, albeit to a lower level, as the number of freshwater species increases (figure 1; Whitfield *et al.*, 2012). Instead of showing absolute limits to the distribution of marine and freshwater communities, estuaries rather show ecoclines (Attrill and Rundle, 2002). Ecoclines are gradient zones between two communities which show gradual (generally spatial) divergence in one major environmental factor and contain relatively heterogeneous communities. A secondary environmental factor simultaneously influences the gradient by increasing total difference within it but maintaining all transitional states (van der Maarel, 1990). Estuaries show a two ecocline pattern with two overlapping gradients in the main stressor, salinity. One ecocline occurs from the river to mid-estuary for freshwater species and another from sea to mid-estuary for marine species, causing a continuum of relatively heterogeneous transitional assemblages in the zoobenthos. The secondary factor, freshwater input, influences the size and position of species ranges (Attrill and Rundle, 2002). An ecocline in a small-bodied nekton (largely fish) community has also been reported (Greenwood, 2007). Another factor driving low biodiversity may be the high mobility of taxa (as larvae and adults) in aquatic environments leading to few isolated populations. The connectivity of estuarine populations may cause functional replacements to be readily available causing limited selective advantage for specialism. Moreover, estuaries may contain an absence of physical structures and habitats created by organisms. Although, habitat forming species such as seagrasses can be important (Costanza *et al.*, 1993). Finally, the recent geological origin of brackish water ecosystems probably lies at the root of estuarine fauna's low species diversity (Cognetti and Maltagliati, 2000). Estuarine ecosystems may not have existed long enough for new species to evolve within them. For example, most European estuaries were formed at the end of the last ice age (Elliott and McLusky, 2002) and generally, estuaries have few endemic species with most species having a freshwater or marine origin (Costanza *et al.*, 1993).

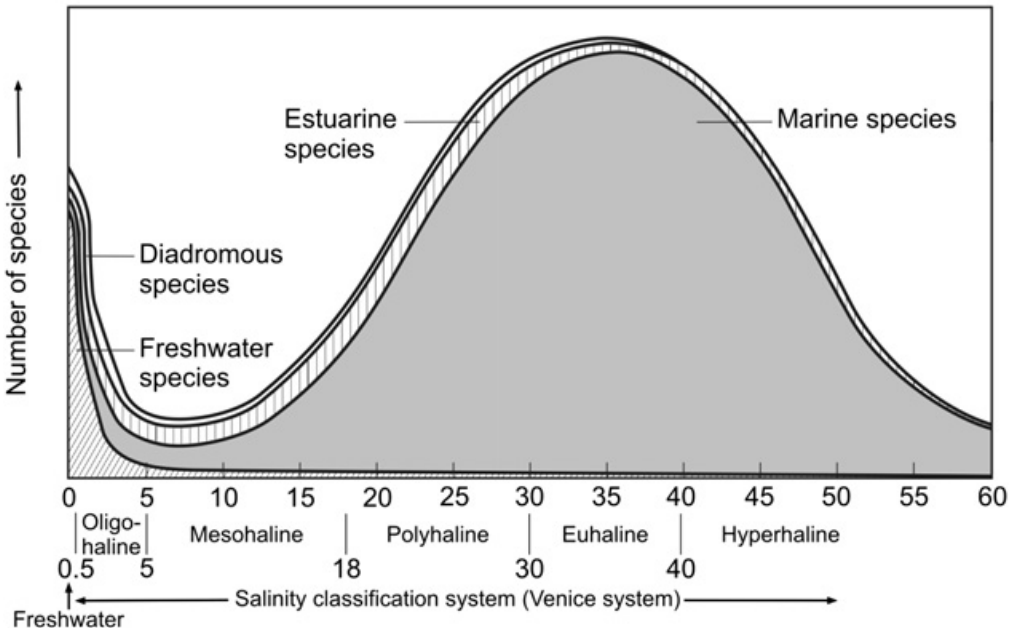


Figure 1: A conceptual model of species richness, for each group inhabiting estuaries, which may occur across the salinity gradient from freshwater to hyperhaline conditions (Whitfield *et al.*, 2012).

2. Ecological Importance of Estuaries

Estuaries contain numerous habitat types in addition to the pelagic component of the water column. In Europe, habitats include soft (fine silts to course sands or shingles) and hard (gravels to bed rock) substratum which occur intertidally and subtidally. Several habitats are also created by plants, including saltmarshes, intertidal and subtidal seagrass beds and reedbeds (Pihl *et al.*, 2002). Although, such habitats, e.g. seagrasses, are often considered ecosystems in their own right (Whitfield and Elliott, 2011). Globally, estuaries and seagrass ecosystems are among the most valuable ecosystems in terms of the services they provide. Estuaries provide numerous ecosystem services, chiefly nutrient cycling, but also food production and disturbance regulation among others (Costanza *et al.*, 1997). Similarly, seagrasses and salt marshes provide coastal protection, raw materials and food, carbon sequestration, erosion control and water purification amongst others. Estuarine ecosystem services are underpinned by ecosystem functions (Barbier *et al.*, 2011). Two examples of estuarine ecosystem functions are biological production and nursery provision. Estuaries have the highest primary production per unit area of all marine systems (Nixon, 1988) and estuaries and coastal marine systems have some of the highest levels of primary productivity in the world. High primary productivity is driven by high nutrient loading rates and efficient nutrient recycling caused by shallow depths and the proximity of sediments to the euphotic zone. In addition, estuaries have efficient nutrient retention due to physical circulation, particle trapping and density driven stratification (Costanza *et al.*, 1993). Similarly, estuaries show high secondary productivity. The production of fish communities in estuaries is often higher than other marine systems (Allen, 1982; Costa *et al.*, 2002) which is reflected in high fishery yields (Nixon, 1988). In fact, despite low prevailing levels of biodiversity, estuaries are generally associated with an increased abundance of many species (McLusky and Elliott, 2004). Estuaries, salt marshes and seagrass habitats may also act as nurseries for juvenile fish and thus contribute to the maintenance of fisheries (Beck *et al.*, 2001; Able, 2005; Barbier *et al.*, 2011). Nurseries are habitats that contribute, per unit area and on average, a greater production of recruits to the adult population than other such habitats (Beck *et al.*, 2001). Nursery function has also been shown to be important for the transport of nutrients, in the form of migrating fish (Gulf Menhaden,

Brevoortia patronus) from estuaries into the Gulf of Mexico (Deegan, 1993). However, it should be noted that the importance of estuaries as nurseries varies between species and also within them at geographic, annual and cohort specific scales (Able, 2005).

Estuarine ecosystems have long been of importance to the human race, as conduits for navigation and locations for settlement (McLusky and Elliott, 2004) and may even have influenced the development of human civilisation. Expansion of estuaries and other coastal margin habitats driven by sea level rise at the end of the last ice age may have been a factor in increasing primary and secondary production in coastal ecosystems. Subsequent increased access to high quality food, e.g. fish, may have facilitated the development of early human civilisations around 6000 BP following the stabilisation of sea levels around 7000 BP (Day *et al.*, 2007, 2012). In the modern context, several of the world's largest cities are located on estuaries e.g. Tokyo, New York and London. In the United Kingdom (UK), most main cities border estuaries into which urban, agricultural and industrial areas drain (McLusky and Elliott, 2004). Humans have, and will continue to impact estuarine systems. Impacts stem from human activities in the estuarine embayments and the broader drivers of overpopulation and uncontrolled developments in coastal watersheds (Kennish, 2002). Impacts can be defined as man-made changes to physical, chemical and biological components of an ecosystem. Comparably pressures are the cause of a change e.g. damage to the seabed caused by a particular anthropogenic driver e.g. fishing (McLusky and Elliott, 2004). These definitions will be loosely used here. Human impacts are expected to worsen in the future (Kennish, 2002). There are numerous anthropogenic impacts on estuarine systems and pressures driving these impacts, many of which interact (reviewed in Kennish, 2002; McLusky and Elliot, 2004; Wolanski, 2007). However, at present the main pressures are pollution and habitat degradation (summarised in table 2; Kennish, 2002). Almost all marine pollution is concentrated in coastal seas and estuaries (McLusky and Elliott, 2004). For an example of its consequences, the precipitous decline in European seagrasses, reduced in area by over 50% in most countries, is currently primarily associated with water quality degradation (Beck and Airoldi, 2007). Habitat degradation maybe more important in many estuaries than pollution (Kennish, 2002). For example, coastal developments and land claims have been associated with the decline of coastal wetlands, with a loss of over 50% of their area, in most European countries (Beck and Airoldi, 2007). In addition to the above, fishing is a pressure that may increase in importance with human population growth (Kennish, 2002). Most coastal and estuarine fish stocks are fully or overexploited. While fishing has numerous negative environmental impacts including killing of non-target organisms caught as bycatch, destruction of habitat and impairment of nursery function among others (Blaber *et al.*, 2000). Finally it should be noted that for estuaries and coastal seas in developed countries, most ecological degradation has already occurred over historical time scales e.g. the last 150 to 300 years (Lotze *et al.*, 2006).

Table 1: Main components of the overall pressures: pollution and habitat degradation (after Kennish, 2002)

		Individual Pressures	Impacts
Overall Pressure	Pollution	Excessive nutrient enrichment	Progressive enrichment and periodic eutrophication
		Organic carbon loading (e.g. sewage)	Promotion of anoxia and hypoxia in coastal waters
		Pathogen (e.g. bacteria, parasite and virus) input	Decline in water quality
		Toxic chemical release e.g. heavy metals, PAHs and halogenated hydrocarbons	Lethal and sub-lethal impacts on organisms endangering organismal health
		Oil spills	Habitat degradation
		Accumulation of floatable debris e.g. plastic	Harms organisms if ingested or entangled. Entangled animals are susceptible to predation, suffocation and drowning.
	Habitat Degradation	Wetland reclamation	Eliminate fish spawning, feeding and nursery grounds. Causes declines in recreational and commercial fisheries.
		Domestic and industrial construction	Destroys habitat in the water shed. May increase non-point source pollution to estuarine water. Large scale construction increases surface run-off aiding pollution input. Increases in sediment loading can cause changes to benthic communities.
		Human mediated structure changes to wetlands	Modification of hydrology and so water quality and salinity regimes in nearby estuaries.

3. Legislation

Given the severity of the pressures and impacts which threaten estuaries there is a requirement for effective governance. Governance is the policies, politics, administration and legislation developed to mitigate adverse pressures and impacts to a system and ensure sustainable management by involving stakeholder groups e.g. industry. Effective governance must ensure environmental function and services are maintained, while considering stakeholder needs (Lonsdale *et al.*, 2018). The EU has the most comprehensive and influential set of environmental policies in the world, acting as a model for other countries with comparable legislation to other developed nations e.g. the United States of America (Boyes and Elliott, 2016). The bulk of environmental legislation within EU member states originates from the EU. Although, member states may have legislation at the national level or have ratified global declarations and regional environmental conventions (McLusky and Elliott, 2004). Therefore, the focus here will be solely on EU policy. Key EU legislation for the environmental management of estuaries are the European Water Framework Directive (WFD; 2000/60/EC; EC, 2000), the Habitats Directive (92/43/EEC; EC, 1992) and the Wild Birds Directive (2009/147/EC; EC, 2010), although other legislation exists (Lonsdale *et al.*, 2015). Given the primacy of EU legislation for environmental management, the UK's departure from the EU has had an impact on environmental legislation, but the new UK legislation is derived from the original EU legislation. The Habitats Directive has been replaced by the equivalent 'The Conservation of Habitats and Species Regulations 2017' (UK Parliament, SI 2017/1012). Currently, the WFD has been replaced by several pieces of domestic legislation in the devolved regions of the UK (JNCC, 2021). For example, within England and Wales the WFD has been replaced by 'The Water Environment (Water Framework Directive) (England and Wales) Regulations 2017' (UK Parliament, SI 2017/407). At present this is equivelent to the WFD,

preserving the requirement for environmental montoring for example (UK Parliament, SI 2017/407). Although this does not preclude a shift in, and potential weakening of, the legislation overtime. However, it should be noted that many EU directives were implemented to respond to international commitments, e.g. the Habitats Directive implements the Bern Convention and the Convention on Biological Diversity (Boyes and Elliott, 2016). Finally, given the influence of EU legislation and its similarities to other countries it seems unlikely that domestic UK legislation will be radically different in tone and objectives from the existing EU directives.

4. Monitoring Ecosystem Health

Given the importance of EU directives for estuarine environmental management demonstrated above, it is worth briefly describing the aims of the three key directives. The Habitats Directive aims to conserve natural habitats (listed in Annex 1) of wild species (listed in Annex 2) to ensure biodiversity by maintaining or restoring listed habitats and species to favourable conservation status (McLeod *et al.*, 2005). Similarly, the Birds Directive provides protection, management and control of naturally occurring bird populations (Stroud *et al.*, 2001). The Habitats Directive Annex 1 habitats include estuaries and associated habitats e.g. subtidal sandbanks, salt marsh and sea grass. Annex 2 contains 12 estuarine and migratory fish species including Allis Shad (*Alosa alosa*), Twaite Shad *(A. fallax)* and Salmon (*Salmo salar*). Furthermore, several estuaries supporting wading bird populations are designated as Special Protection Areas under the Birds Directive (McLusky and Elliott, 2004). Comparably, the WFD provides a common framework for protecting surface waters, including transitional and coastal waters. The directive aims to prevent the deterioration of aquatic ecosystems, by protecting and enhancing their status and that of dependent terrestrial wetland systems. In addition, the directive aims to promote sustainable use based on long-term protection. The overall goal of the WFD was to achieve good ecological status for the majority of water bodies by 2015 (EC, 2000). Within the WFD, ecosystem status is derived with reference to 'ecological quality' rather than physicochemical thresholds alone. Ecological quality is determine by monitoring biological, hydromorphological and physio-chemical 'quality elements' specific to certain habitats (Hatton-Ellis, 2008). Traditionally, the concept of ecosystem health and associated monitoring for aquatic systems had limited scope, focusing on the physicochemical properties of water as indicators of an aquatic systems status e.g. dissolved oxygen content (Karr, 1981; Bain *et al.*, 2000; Hatton-Ellis, 2008). However, although useful, physicochemical measures do not provide broadly sensitive measures of ecosystem health (Bain *et al.*, 2000) because anthropogenic perturbations of ecosystems often interact in complex ways (Fausch *et al.*, 1990). Therefore, there is a requirement to directly use biological measures, e.g. community structure or function, in ecosystem monitoring (Bain *et al.*, 2000). However, biomonitoring in estuarine ecosystems is challenging as estuarine communities are resilient to high temporal and spatial variability in environmental factors. Environmental fluctuations have a similar character to anthropogenic stress, making it difficult to detect anthropogenic perturbations (the *Estuarine Quality Paradox*; Elliott and Quintino, 2007). Ecosystem health in water resource management is generally defined as an ecosystem's similarity to a least-impacted reference site or historical state by measuring its structure (biodiversity, temperature etc.) or function (productivity or thermal regime etc.; Palmer and Febria, 2012). This is essentially the definition used by the WFD. To measure ecological quality, quality elements are compared with a reference condition using an 'Ecological Quality Ratio' (Hatton-Ellis, 2008). Fish communities have been widely advocated as good indicators of the relative health of aquatic ecosystems in both freshwater and estuarine systems (Karr, 1981; Fausch *et al.*, 1990; Whitfield and Elliot, 2002). Monitoring fishes has numerous advantages over other taxonomic groups and any disadvantages (table 2)

often apply to other commonly used groups e.g. macroinvertebrates (Whitfield and Elliot, 2002). Although the usefulness of fish monitoring is greatest when it is combined with monitoring of physicochemical variables (Fausch *et al.*, 1990). Monitoring fish communities is one of the quality elements used to determine the ecological status of transitional waters under the WFD, along with benthic invertebrates, phytoplankton, angiosperms and macroalgae (EC, 2000), leading to the development of monitoring programs for fish across Europe (Coates *et al.*, 2007; Delpech *et al.*, 2010; Harrison and Kelly, 2013).

Table 2: Advantages and disadvantages of fish as environmental indictors (after Whitfield and Elliot, 2002)

Advantages	Disadvantages
Occur in all aquatic ecosystems except extremely polluted waters	Diverse fish assemblages can still be present in estuaries physically altered by humans
Most species have well characterised life history and environmental responses	Fish species may be more tolerant to substances which other species can succumb to
Acute stress and toxicity effects can be evaluated in the laboratory	Species may avoid localised exposure by swimming away from pollutants and disturbance
Relative ease of identification allows sample processing in the field and *can* be non-destructive	Sampling can be biased by seasonal and diel movement of fish
Good record of long-term environmental stress as comparatively long lived	Sampling gear is selective for certain habitats, sizes and species of fish
Diversity of trophic levels represented with species feeding on food originating from terrestrial and aquatic systems	
Reflect stressors in both the local and wider area due to sedentary and mobile species	
Many functional guilds and lifeforms that probably cover all components effected by anthropogenic stress	
General public are more likely to care about data from fish compared to invertebrates and aquatic plants	
Easy to evaluate the societal costs of degradation due to economic, conservation and aesthetic values attached to fish	

Generally, estuarine fish assemblages are sampled using nets to provide data for monitoring programmes (Harrison and Whitfield, 2004; Coates *et al.*, 2007; Delpech *et al.*, 2010; Harrison and Kelly, 2013). Although, other methods such as electrofishing (Warry *et al.*, 2013), hydroacoustics (Samedy *et al.*, 2015) and visual methods (Becker *et al.*, 2010) can be used to monitor fish assemblages in estuaries. However, some of the key disadvantages of using fish assemblages as ecological indicators (table 2) originate during sampling. It is well documented that specific gears (and not only netting methods) will be biased towards sampling a specific component of the assemblage present (Harmelin-Vivien and Francour, 1992; Guest *et al.*, 2003; Olin and Malinen, 2003; Rotherham *et al.*, 2012; Warry *et al.*, 2013). Firstly, a gear may not be deployable in every habitat and so may miss specific species due to the distribution of that species in specific habitats, e.g. the use of seine nets and trawls is generally constrained to firm unstructured strata (Warry *et al.*, 2013). Furthermore, a specific component of the habitat may not be sampled, e.g. beam trawls are designed to run along the seabed and so are poor at sampling pelagic roundfish (Hemingway and Elliot, 2002). A further complication is that diel patterns of fish movement through different habitats also influences their vulnerability to different gears at different times (Kubečka *et al.*, 2012). In addition, for netting methods the mesh size of the gear will influence the species for which a gear selects. Generally, finer meshes will catch more smaller species and hence influence the community composition of the catches

derived from different gears with different mesh sizes (Olin and Malinen, 2003; McKenna *et al.*, 2013). Another important component is variation in fish behaviour (Olin and Malinen, 2003; Rotherham *et al.*, 2012) which includes avoidance or escape behaviours (Kubečka *et al.*, 2012). For example, smaller species of fish are probably more susceptible to trawls (an 'active gear' which moves towards the fish), over gill nets (a 'passive gear' where the fish move to the gear; Hemingway and Elliot, 2002), due to their limited swimming ability (Rotherham *et al.*, 2012). A related factor is also the movement of active gears. Ineffective contact of a trawl with the bottom may lead to certain species being under sampled in comparison to a passive gear (Olin and Malinen, 2003). Given the above, it is generally advocated that multiple gear-types should be used concurrently to gain a more comprehensive understanding of the assemblage present (Hemingway and Elliot, 2002; Kubečka *et al.*, 2009). This is often the approach often taken for monitoring programmes of estuarine fish communities (Harrison and Whitfield, 2004; Coates *et al.*, 2007; Harrison and Kelly, 2013).

A key disadvantage of the use of multiple netting methods is that it makes monitoring fish in estuaries labour and time intensive (T. Hatton – Ellis, *pers. comm.*). High sampling costs therefore means less sampling occurs than befits the ecological importance of estuaries. For example, in Wales WFD fish surveys occur only once every three years in each estuary during autumn. Comparably from 2010 to 2016 sampling took place in each estuary annually, in both autumn and summer. Although, the individual sampling design and effort within each fish survey have improved substantially in recent years (T. I. Gibson, *pers. obs.*). Another disadvantage is that the most commonly used gears; trawls and seines, have low and variable catch efficiencies (Hemingway and Elliot, 2002). Catch efficiency is defined as the proportion of the target animals within the sample unit area which is enclosed by the gear (capture efficiency), and are later recovered and enumerated (recovery efficiency; Rozas and Minello, 1997). Large and unpredictable variations in capture efficiency, such as between habitats, can reduce the accuracy and precision of abundance estimates and hence ability to detect statistically significant changes in communities (Hemingway and Elliot, 2002). Catch efficiency can vary due to numerous factors including the movement of the gear and changes in fish behaviour (Hemingway and Elliot, 2002; Kubečka *et al.*, 2012). Finally, a more general disadvantage is many netting techniques can be destructive, e.g. bottom trawls can kill fish and damage the seabed. Although destructiveness varies between gears and with gears operation. However, there has been growing scientific interest in non-intrusive acoustic, visual and photographic techniques (Kubečka *et al.*, 2012). Another method which is attracting significant attention for monitoring species, including fish, within aquatic ecosystems is environmental DNA (eDNA; Rees *et al.*, 2014; Goldberg *et al.*, 2015). Environmental DNA refers to DNA isolated from an environmental sample, e.g. water or sediments, without capturing the organism (Taberlet *et al.*, 2012). In aquatic systems eDNA is derived from cells, faeces, saliva and urine of species occupying or visiting a water body (Rees *et al.*, 2014) or from biological material transported into that system e.g. by flowing water (Roussel *et al.*, 2015). Environmental DNA has been used to monitor individual fish species. For example, eDNA was used to map the distribution and biomass of Japanese Jack Mackerel (*Trachurus japonicus*) in a coastal bay, Japan (Yamamoto *et al.*, 2016; Jo *et al.*, 2017), and determine the distribution of Chinook Salmon (*Oncorhynchus tshawytscha*) in rivers in North America (Laramie *et al.*, 2015). Targeted assessments of single species may be of relevance to monitoring Annex 2 Habitats Directive species and are at present performed using PCR, quantitative PCR (qPCR) and digital droplet PCR (ddPCR; see section 5.). In addition, assemblage level studies can be performed, at present using a technique called metabarcoding (see section 5.). Environmental DNA has been used to assess the biodiversity of fish assemblages in lakes (Civade *et al.*, 2016; Hänfling *et al.*, 2016), rivers (Civade *et al.*, 2016; Shaw *et al.*, 2016), the marine environment

(Thomsen *et al.*, 2012; Yamamoto *et al.*, 2017) and estuaries (Stoeckle *et al.*, 2017; Hallam *et al.*, 2021). Environmental DNA has some advantages over capture-based methods. Environmental DNA is non–invasive, both for species and habitats (Evans and Lamberti, 2017). In addition, eDNA may have a higher detection probability for single (Hinlo *et al.*, 2017) and multiple species when compared to some established methods (Thomsen *et al.*, 2012). Sampling and analysis of eDNA also requires less effort in person hours than capture methods, but whether this makes it cheaper is currently ambiguous (Evans *et al.*, 2017). Finally, WFD assessment methods for lake fish communities have been developed using eDNA in the UK (Hänfling *et al.*, 2016) and is now being implemented for biomonitoring (L. Lawson-Handley, *pers. comm.*).

Despite the above, eDNA has not yet been implemented for biomonitoring fish communities in estuaries. Below, I will explore whether eDNA can be used in this context. Given that assemblage level assessments of fish biodiversity are primarily used as a proxy for ecosystem health, the focus will be primarily on assemblage level studies and will cover the following five broad areas. 1: The methods used in eDNA analysis will be briefly outlined. 2: The relevance of data provided by eDNA metabarcoding to biomonitoring will be discussed. 3: The ecological relevance of eDNA metabarcoding studies will be explored. Finally, although caveats will be discussed throughout, particular mention will be given to the effects of 4: eDNA transport and 5: The effects of other environmental variables given the ecological variance present in estuarine environments.

5. Current eDNA Methodologies

Before exploring the applicability of eDNA to biomonitoring fish in estuaries it is essential to briefly outline the current methods for sampling, lab analysis and data analysis. For more detailed reviews see Deiner *et al.* (2017a) and Evans and Lamberti (2017). Firstly, eDNA must be captured and isolated. Generally for water samples, replicated samples (generally 1 – 2 L) are collected from a waterbody and filtered through a media with a 0.45 – 3 μm pore size (Evans and Lamberti, 2017). However, empirical studies suggest that fish eDNA is most highly concentrated in particles between 1 – 10 μm in size (Turner *et al.*, 2014; Wilcox *et al.*, 2015; Jo *et al.*, 2019). Therefore, a wide range of pore sizes (0.2 – 10 μm) maybe used (Turner *et al.*, 2014). Samples can be kept cool (4^{o}C) until filtration in the lab (Hänfling *et al.*, 2016), or filtered in the field with filters preserved dry or in lysis buffer. Filtration in the field is advantageous as minimising the time between collection and filtration reduces eDNA degradation and eliminates the need for clean filtration facilities (Majaneva *et al.*, 2018). However, this is not always desirable if personnel are operating in poor weather conditions and laboratory facilities are only a few hours drive away. In addition, negative controls, e.g. sterile water samples, should be taken into the field and run through an identical workflow to true samples (Evans and Lamberti, 2017). For estuarine waters, filters will undoubtedly rapidly clog as estuarine waters are often highly turbid due to the low settling velocities of silt and clay particles. Many estuaries show a turbidity maxima in their middle and upper reaches and turbidity also varies with tidal cycles, being highest at low tide and lowest at high tide (McLusky and Elliott, 2004). Thus potential maximum sample volume may vary spatially and temporally. Because sample volume has been shown to affect species detection from metabarcoding (Sigsgaard *et al.*, 2017) this could bias results if the target volume cannot be reliably filtered. Several strategies have been used to consistently sample turbid water. Filtration of 2 L samples through large pore size, 10 – 20 μm, nylon filters is effective for targeting single species in turbid river waters (Robson *et al.*, 2016; Simpfendorfer *et al.*, 2016). Alternatively pre-filtration prior to final filtration have been performed with 10 – 20 μm, nylon filters (2 L samples; Robson *et al.*, 2016) and paper coffee filters (1 L samples; Stoeckle *et al.*,

2017). However, pre - filtration is costly and time intensive (Robson *et al.*, 2016) and costs will be particularly high if extraction is performed from both filters. Finally, smaller sample volumes, 500 ml, have been successfully filtered through 0.45 µm cellulose acetate filters from a glacial Fjord (Kelly *et al.*, 2018). However, none of these studies quantified the suspended particulate matter (SPM) concentration of samples making it difficult to generalise the results. Other studies should do this in future to improve development of filtering techniques across the research community. Regarding sediments, these can be sampled using sterile cores (Shaw *et al.*, 2016) or by careful operation of a grabs e.g. van Veen (Pochon *et al.*, 2015; Laroche *et al.*, 2016). The undisturbed surface layer, or first few cm, is then collected and sample tubes stored on ice, with or without a buffer such as Life Guard™, before freezing and later extraction (Pochon *et al.*, 2015; Laroche *et al.*, 2016; Shaw *et al.*, 2016).

Following sample collection DNA is extracted from filters or sediment. Extraction can be performed using a variety of methods including, phenol-chloroform-based extraction, physical lysis of cells and silica-based extractions. In addition, negative controls (sterile water) and positive controls (containing DNA) should also be extracted to indicate whether samples are contaminated and have been successfully extracted (Evans and Lamberti, 2017). During or following extraction, PCR inhibitors, substances which negatively affect the PCR reaction, must be removed if present. Inhibitors are common in environmental samples. Soils contain PCR inhibitors such as humic and fulminic acids and plant polysaccharides can also cause inhibition. In addition, filtration of water samples can cause the concurrent concentration of inhibitors (Schrader *et al.*, 2012). PCR inhibitors can have a large effect on eDNA studies. Inhibition of a qPCR assay targeting Brook Trout (*Salvelinus fontinalis*) in streams caused no amplification in the absence of procedures reducing inhibition for even high copy number samples. Inhibition also varied seasonally, increasing during autumn due to large depositions of leaf matter into streams (Jane *et al.*, 2015). PCR inhibition is likely to be an issue in estuarine environmental samples. Estuarine waters have high concentrations of humic compounds (around three times that of marine water) and potentially high levels of chemical pollutants which may inhibit PCR (Petit *et al.*, 1999). Therefore, studies should assess the presence of inhibition, which can be done by including internal positive controls within PCR reactions (Jane *et al.*, 2015). The effect of inhibitors can be reduced by selecting appropriate sample processing such as use of inhibitor removal kits (e.g. Mächler *et al.*, 2016) or dilution of the sample (e.g. Buxton *et al.*, 2017). Alternatively an appropriate nucleic acid extraction and a more robust DNA polymerase maybe selected (Schrader *et al.*, 2012).

After extraction and any required clean up, the current eDNA workflow bifurcates into two routes depending on whether the researcher is aiming to target a specific species or an entire community. Species specific assays use primers which only amplify specific sequences from the target species (Lawson Handley, 2015). Traditional PCR will allow determination of the presence or absence of a target species sequences. Comparably, qPCR and ddPCR are more sensitive and allow quantification of the target sequence (Evans and Lamberti, 2017). Sequencing may also be used to confirm the identify of a sub-set of positive PCR reactions (Mächler *et al.*, 2016). Comparably, metabarcoding can be used to describe whole assemblages by using conserved primers which amplify a variable region of a suitable taxonomy marker gene. Mitochondrial DNA is often used due to ease of amplification derived from the multi-copy nature of mitochondria in eukaryotic cells (Lawson Handley, 2015; Deiner *et al.*, 2017a). As with collection and extraction, negative and positive controls should be included during PCR to identify erroneous, proper and failed application of the target sequence, respectively (Evans and Lamberti, 2017). In 2018 there were five universal primer sets which have been developed for metabarcoding of fish eDNA (Nakagawa *et al.*, 2018). Following amplification, libraries are created for individual samples by the addition of sample specific indexes (unique

nucleotide sequences, incorporated using PCR or ligated onto different PCR products) and sequencing adapters to each sample which allows the pooling of samples into libraries ('multiplexing') making the sequencing process cost effective (Deiner *et al.*, 2017a). The amplified region in the pooled libraries is sequenced using High Throughput Sequencing' (HTS) e.g. on an Illumina MiSeq or HiSeq. The data produced is then analysed using bioinformatics 'pipelines' (Lawson Handley, 2015; Deiner *et al.*, 2017a). Bioinformatics generally proceeds with de-multiplexing of the samples based on the indices used and removal ('trimming') of the adapter sequences, the index sequences and the primer sequences (Deiner *et al.*, 2017a). Filtering data, among other steps is performed to remove erroneous sequences, that could arise from DNA degradation, PCR-generated errors and sequencing errors, which may be mistaken for rare taxa in later analysis (Coissac *et al.*, 2012). Clustering of multiple sequence reads into Molecular Operational Taxonomic Units (MOTUs) can also be performed before taxonomic assignment, but this is not always necessary (Deiner *et al.*, 2017a) if a full reference database is available e.g. Hänfling *et al.*, (2016). A more recent alternative approach to MOTUs is to generate Amplicon Sequence Variants (ASVs) via the *dada2* algorithm. This models and corrects for Illumina-sequenced amplicon errors, allowing exact inference of sample sequences rather than clustering into MOTUs (Callahan *et al.*, 2016). Reference databases are made up of sequences belonging to specific species of interest which have been identified by a taxonomist prior to sequencing. It is the final goal of DNA metabarcoding to link the unknown DNA sequences to a taxonomic name by comparing them to the known sequences in the reference database (Coissac *et al.*, 2012). Sequence matching is followed by appropriate data standardisation to account for variation in sequencing depth before final statistical analysis (Deiner *et al.*, 2017a).

6. Relevance of eDNA Data to Biomonitoring

An important question is: Does eDNA metabarcoding provide the data required by existing assessment methods for biomonitoring fish in estuaries? Existing assessment metrics for estuarine fish require information on species diversity and composition, species abundance and the ecological and feeding guilds of species. The latter give information on the nursey function and trophic integrity of an estuary (Harrison and Whitfield, 2004; Coates *et al.*, 2007; Harrison and Kelly, 2013). Guilds are important because information on ecosystem function is critical to determining estuarine ecosystem health. Estuarine ecosystems can maintain high levels of ecosystem function despite low levels of biological diversity. This makes structural measures of community health, e.g. species richness, less useful as indicators of ecosystem health (Elliott and Quintino, 2007). Given the above, reliable taxonomic information on the species present is essential. Clearly metabarcoding is able to do this by matching sequences in the sample to sequences associated with a known species in the reference database using bioinformatics. However, sequence matching can only occur if a species is present in a reference database. The reference database for 12S rRNA, commonly used in fish metabarcoding (see below) is far from complete. In August 2016, the NCBI nucleotide database only held sequence information on the 12S rRNA gene from ~ 30% of the known species of teleosts i.e. ~8,000 species out of ~27, 000 species (Andruszkiewicz *et al.*, 2017b). Therefore, it is likely a substantial amount of sequencing of missing species in the database will be required before biomonitoring may commence. In the UK there has recently been an increased effort to sequence the 12S rRNA gene in a larger proportion of the fishes found in freshwater and marine environments. The coverage of the 12S rRNA marker in 2021 was 75% (up from 70%) of the 530 total species which can be found in UK waters. While for the 176 most 'common' freshwater and coastal species (a somewhat arbitrary classification) coverages has increased from 76% to 93% (Collins *et al.*, 2021). In addition to reference databases, limited sequence variation in the short

barcoding regions currently used can prevent taxa being identified below the genus (Nakagawa *et al.*, 2018) or even the family level (Sigsgaard *et al.*, 2017). Comparably, most fish species are easy to identify morphologically (Whitfield and Elliot, 2002) and personnel can generally be brought up to speed, on the job, with a few hours or days of work (T. I. Gibson, *pers. obs.*). Another shortcoming of metabarcoding is that mismatches between the primer binding site and template, and purely stochastic effects, can prevent amplification of species which remain absent in the sequencing run (Piñol *et al.*, 2015). Such mismatches have been implicated in the failure to detect entire groups such as cartilaginous fish (Port *et al.*, 2016); challenges that are not insurmountable. Population of the reference databases requires a coordinated effort between research groups and government agencies. Whereas, use of multiple markers may improve detection rates due to reduced primer bias (Port *et al.*, 2016).

Further to the ability to accurately identify species, eDNA needs to be able to provide information on species abundance. Depending on the indices and metrics in question this could be absolute measures of abundance such as density (Delpech *et al.*, 2010), e.g. the number of individuals per unit area (Nicolas *et al.*, 2010b). Alternatively, it may be a relative abundance metric (Coates *et al.*, 2007; Harrison and Kelly, 2013), such as the relative numerical abundance (%) of estuarine species and marine migrant species for example (Harrison and Kelly, 2013). Generating quantitative assessments of abundance from eDNA relies on an assumed correlation between the eDNA particle concentration, or number of reads in the case of metabarcoding, and living biomass and, or species abundance (Hansen *et al.*, 2018). This itself is extremely challenging given that this relationship will be affected by variability in the eDNA production from fish, its degradation, transport and dilution of eDNA (Hansen *et al.*, 2018), alongside numerous other factors, such as methodological ones (Rourke *et al.*, 2022). For example, eDNA production per unit biomass varies at different life history stages. Larval fish produce more eDNA due to a larger surface area and metabolic rate per unit biomass. Size and age structures vary in fish communities naturally. Therefore, meaningful abundance/biomass estimates are likely only attainable if the size and age structure of a population is known (Hansen *et al.*, 2018). Despite these challenges several studies have found positive correlations between eDNA concentrations and abundance estimates and, or biomass for fish in rivers (Doi *et al.*, 2017; Tillotson *et al.*, 2018) and the sea (Jo *et al.*, 2017). Meta-analysis of all studies attempting to draw a quantitative link between eDNA concentrations or read counts and fish biomass, and or abundance, has shown that the vast majority of publications show a positive correlation. Overall, 46 studies (90% of those analysed) reported positive relationships with eDNA concentrations for species specific PCR methods, whereas 11 (92%) metabarcoding studies reported positive relationships with read counts (Rourke *et al.*, 2022). However, the overall strength of this relationship across studies was not investigated in this review. Therefore, this needs to be determined across multiple studies in future. More generally, across different taxonomic groups, meta-analysis has shown that a weak quantitative relationship may exist between the biomass and the read counts produced by metabarcoding (slope = 0.52 ± 0.34, $p < 0.01$), although with a high degree of uncertainty. Potential bias in this relationship can occur at the sample collection, DNA extraction, PCR, sequencing and bioinformatic stages (Lamb *et al.*, 2019). Overall, it seems unlikely that eDNA metabarcoding can be relied on, in an uncritical manner, to provide reliable abundance data for monitoring at present. However, studies such as Rourke *et al.* (2022) provide hope that this may be possible with further research and when the relevant context specific studies are carried out. Therefore, at present, eDNA should be combined with other quantitative methods, e.g. trawling, hydroacoustics etc., to carry out assessments of fish communities.

7. Ecological Relevance of eDNA

Despite the caveats discussed above, an increasing number of studies have demonstrated that eDNA metabarcoding can provide ecologically relevant information on fish communities in riverine, marine and estuarine ecosystems; both temporally and spatially. Such studies generally amplify short fragments of the mitochondrial 12S rRNA gene, the longest fragments being around ~182 bp in length (Miya *et al.*, 2015), but see Hallam *et al.* (2021). Although it is notable that entire mitogenomes (16 - 17 kb in length) and mitochondrial genes can be assembled (Deiner *et al.*, 2017b). Spatial patterns in the transition from typically lake to river dwelling fish species were detected in water samples from the Tier, France, by analysing presence/absence information from metabarcoding (Civade *et al.*, 2016). Similarly, metabarcoding of water samples from 51 rivers around Lake Biwa, Japan, showed that known distributional patterns in certain species could be detected using presence/absence and read data. This included longitudinal distributions, which occur due to interspecific competition and or differences in habitat preference, of two closely related species from the Sculpin genus *Cottus*. In addition, inter-river system patterns of the presence/absence of *Hemibarbus longirostris*, a Cyprinid, and the Loach *Niwaella delicate* were similar to known biogeographical patterns of the two species (Nakagawa *et al.*, 2018). Similarly, metabarcoding of water samples collected along a transect of habitats associated with Kelp forest ecosystems in Monterey Bay, north east Pacific, showed spatial trends in proportions of normalised eDNA read data for vertebrates, including fish. These trends were strongly consistent with expectations of species distributions e.g. Wrasse (*Oxyjulis spp.*) reads peaked in their core, kelp forest and rocky reef, habitats. In addition, individual habitats could be distinguished from each other by their community composition using presence/absence data (Port *et al.*, 2016). Metabarcoding of water samples from West Maizuru Bay, Sea of Japan, also showed that eDNA detection of some species e.g. the Eelpout, *Zoarchias major* occurred in their known habitat e.g. macroalgae beds, suggesting some of the eDNA signal was localised (Yamamoto *et al.*, 2017). Likewise, eDNA metabarcoding of surface water samples from the Thames, United Kingdom, showed differentiation in fish assemblage composition at the scale of 10s of kilometres between the estuary, the river and within the estuary. Changes in assemblage composition in the estuary was driven by the greater incidence of freshwater and anadromous species in the upper estuary, and the greater incidence of marine migrant and estuarine species in the lower estuary (Hallam *et al.*, 2021). Generally, the evidence leads to the conclusion that detection of known species distributions and community composition suggests the eDNA signal detected is largely of localised origin. Detection of generally similar species between eDNA and established methods, such as visual surveys and netting supports this conclusion (Civade *et al.*, 2016; Port *et al.*, 2016; Yamamoto *et al.*, 2016; Hallam *et al.*, 2021). Authors have suggested the spatial scale of the local community eDNA signal may be approximately 60 – 100 m in and around a Kelp forest (Port *et al.*, 2016), 2 – 3 km in a river (Civade *et al.*, 2016) and 800 m in a coastal bay. However, the accuracy of this local signal is invariably affected by eDNA transport, this transported eDNA should be considered as potential noise (Yamamoto *et al.*, 2017). This is examined in detail below. Generally estuaries are good test ecosystems for these kind of questions, as the fish fauna shows substantial longitudinal changes in taxonomic composition (Nicolas *et al.*, 2010b; Teichert *et al.*, 2018b) and species richness (Martino and Able, 2003).

Several studies on temporal patterns in fish assemblages using eDNA have also been carried out. Metabarcoding of water samples from Skovshoved Harbour, Denmark, an area with variable salinity (8 to 34) was performed over a year. This showed eDNA could detect the seasonal turnover of the marine fish community and recovered the majority of taxa detected by

snorkelling in that area (Sigsgaard *et al.*, 2017). Another time series of water samples was conducted at two sites on the Hudson estuary, an estuarine (lower flow and salinity) and more marine (higher salinity and flow) site. A strong seasonal increase in fish eDNA detection occurred from January to July, consistent with movement of fish populations into inshore waters and estuaries in the spring (Stoeckle *et al.*, 2017). Seasonal shifts in fish assemblage composition, detected via eDNA, have also been observed in an estuarine microtidal inlet between the North Sea and the Wadden Sea sampled weekly, over 18 weeks, from Spring to Autumn (Bleijswijk *et al.*, 2020). This ability to detect seasonal change in estuarine fish communities is important given that estuarine fish surveys are generally conducted during specific periods of the year in order to minimise the effect of seasonal variation on the results (Harrison and Whitfield, 2004) or to target a specific component of the fish community e.g. newly emergent fry in the Thames in May and June (Colclough *et al.*, 2002). However, few studies have currently assessed the temporal scale of eDNA signals from fish over short time scales, i.e. hrs to days, in estuaries. The signal from environmental DNA may vary over short periods with changes to fish distributions and activity (Sigsgaard *et al.*, 2017) and hydrographic effects. Regarding hydrographic effects, tides in a glacial fjord have been found to have a negligible effect on the community composition (covering 23 phylas and 8 kingdoms) as assayed using a 313 bp fragment of the mitochondrial COI gene. However, this study only sampled two outgoing and incoming tides over a 24 hr period (Kelly *et al.*, 2018). In addition, the authors failed to acknowledge the stratified nature of the Fjord (Paulson *et al.*, 1993), where the sill can restrict the inflow of tidal water (McLusky and Elliott, 2004). Comparably, preliminary evidence of the effect of tides on assemblage composition using a universal COI marker has been shown in the freshwater reaches of the Elbe (Germany); a well-mixed, meso/macrotidal estuary (Schwentner *et al.*, 2021). In a microtidal inlet between the North Sea and the Wadden Sea no substantial influence of tides was found on fish eDNA assemblage composition. However, the habitats either side of the inlet are similar and therefore, only small differences in assemblage composition are expected (Bleijswijk *et al.*, 2020). Further work should be conducted in an estuary with a clear differentiation in the assemblage composition between the river and the sea. This work should sample over two weeks to a month, to account for the variation in tidal range which occurs during spring to neap tidal cycles (McLusky and Elliott, 2004).

8. Influence of Transport on eDNA

Despite the evidence that eDNA metabarcoding can provide spatially and temporally representative data on fish communities in systems with flowing water, transport of eDNA will still affect results. An understanding of eDNA transport will be essential to relating detected eDNA to spatio–temporal inference of species presence (Barnes and Turner, 2016). Approximate calculations suggest that in marine systems fish eDNA could be transported between 40 km to 600 km before degradation. Assuming current speeds of 1 m/sec and a persistence time of 12 hrs to 7 days. Transport will depend on current, wind speeds and directions and factors effecting degradation such as water temperature (Thomsen *et al.*, 2012). In rivers, the eDNA of two species of lake dwelling macroinvertebrates could be detected 9 to 12 km down river from the source. While modelling suggested that distance could be 15 to 50 km before the detection rate decreased below a 5% threshold (Deiner and Altermatt, 2014). This means it is difficult to confirm from eDNA detection alone that a specific species is present in a system, even if most eDNA is produced locally. Examples of probable spurious detections include freshwater Common Carp (*C. carpio*) in Maizuru bay, Japan, (Yamamoto *et al.*, 2017) and the marine Mulloway (*Argyrosomus japonicas*) in the River Murray, South Australia (Shaw *et al.*, 2016). Identification of spurious species depends on background ecological

information, e.g. is the species freshwater or marine (Shaw *et al.*, 2016). This will be difficult in estuaries as the real fish fauna over the whole area sampled will almost certainly include both freshwater and marine species, e.g. Martino and Able (2003). Spurious detection of known human food species from sewage inflow is also an issue. Identification of species from sewage has been suspected in several studies, both in riverine (Nakagawa *et al.*, 2018), marine (Yamamoto *et al.*, 2017) and estuarine (Stoeckle *et al.*, 2017) systems. Spurious detection from pollution will likely be acute in estuarine systems as sewage input is a severe pollution problem in modern estuaries (Kennish, 2002). In the eDNA study of the Hudson estuary, around 10% of samples contained sequences matching fish that were locally rare or absent, but commonly consumed (Stoeckle *et al.*, 2017). This is potentially a serious concern as a fish eDNA biomonitoring survey may itself be confounded by one of the pressures expected to degrade ecosystem health. A way of accounting for this may be to remove commonly consumed species from an assessment of ecosystem health. This list could be static or generated each survey by sampling the known sewage inflows and factoring these species out of the assessment. Although the latter will invariably increase cost.

9. Influence of Environmental Variables on eDNA

The spatio – temporal inference of species presence from eDNA will also depend on the fate of eDNA, i.e. how it degrades and what factors influence degradation (Barnes and Turner, 2016). The persistence time of suspended eDNA particles is highly variable but suggests that eDNA can be used as a contemporary proxy of species presence spatially and temporal. In marine and brackish water environments persistence time from eDNA of several fish species is <1 to ~8 days. Whereas, in freshwater environments longer eDNA persistence times have been recorded for a range of organisms, from 8 to 44 days. Differences in environmental factors or fish osmoregulation between saline and freshwater systems may drive this. Although persistence is dependent on initial starting concentration and assay sensitivity. More comparable decay rates, in number of eDNA particles degraded per hour, show overlapping rates of 1.5 to 10.1% in brackish and marine environments and 5.1 to 15% in freshwater environments (reviewed in Hansen *et al.*, 2018). In contrast, fish eDNA may persist in sediments for substantial periods of time and is more concentrated than in water. Studies of eDNA released from Big Head Asian Carp (*Hypophthalmichthys spp.*) in artificial ponds and natural rivers showed a persistence of 132 days after carp removal. This may be due to particle settling, and or adsorption of eDNA to sediment particles which retards degradation. Hence, eDNA in sediments may not be an effective means of obtaining data with a high temporal resolution on fish species. In addition, resuspension of eDNA from the sediment could complicate spatio-temporal inferences from water sampling (Turner *et al.*, 2015). However, higher eDNA concentrations in sediments compared to water may not translate into higher numbers of species detected. This may be due to bias in sample weight from which eDNA is extracted from. It is technically easier, with commercially available extraction kits, to extract from large volumes of water compared to an equivalent weight of sediment (Shaw *et al.* 2016). Although such comparisons are probably better made between the concentrations of total DNA, or the metabarcoding marker, in the different sample types.

Both the characteristics of the DNA itself such as fragment length, conformation and association with cellular and organellar membranes and the characteristics of the abiotic and biotic environment will influence eDNA degradation (Barnes and Turner, 2016). Regarding fragment length, longer, 719 bp, fragments of the CytB gene in Japanese Jack Mackerel (*T. japonicus*) have a higher decay rate than shorter, 127 bp, fragments. The qPCR assays for *T. japonicus* also showed the concentration of longer fragments in surface sea water samples had a significant correlation with echo intensity (a proxy for biomass), while shorter fragments did

not (Jo *et al.*, 2017). A similar effect may influence the performance of 12S rRNA (~ 100 bp) and CytB (460 bp) fragments when metabarcoding lake fish communities. The shorter 12S rRNA fragments gave greater site occupancy and species detection than longer CytB fragments, potentially from greater persistence and transport through the lake (Hänfling *et al.*, 2016). Therefore, amplifying longer fragments may be a practical way to increase the spatial and temporal resolution of metabarcoding data in estuarine fish communities. Regarding abiotic characteristics of the environment, several factors may influence degradation. Increases in temperature will directly denature DNA and increase enzyme kinetics and microbial metabolism which will degrade DNA (Barnes *et al.*, 2014). Studies on various taxa including fish have shown a positive relationship between temperature and eDNA degradation in freshwater experimental settings (Pilliod *et al.*, 2014; Strickler *et al.*, 2015; Eichmiller *et al.*, 2016; Lance *et al.*, 2017) although see Merkes *et al.*, (2014). Estuaries in Northern Europe have higher water temperatures in summer than in winter (Thiel *et al.*, 1995; Marshall and Elliott, 1998; Selleslagh and Amara, 2008). These seasonal temperature differentials are comparable to differentials over which differences in eDNA decay rates have been observed. For example, temperatures in the Humber, UK varied from 2.5ºC in winter to 21.1ºC in summer (1992 to 1994; Marshall and Elliott, 1998). Comparably, incubation of Common Carp (*C. carpio*) eDNA at 5ºC showed a slower decay rate and a lower T_{90} value (time for initial concentration to decrease by 90%) of 6.6 days compared temperatures at or above 15ºC, T_{90} approximately 1 day (Eichmiller *et al.*, 2016). Therefore, the spatial and temporal resolution of the eDNA signal may increase during summer. Acidic conditions will also degrade eDNA by catalyse hydrolytic processes. Degradation rates of eDNA were faster in freshwater mesocosms of pH 4 than 10 in a lab setting (Strickler *et al.*, 2015). While eDNA have been shown to be higher in more acidic (pH of the most acidic mesocosm: 5.35 ± 0.05) than less acidic semi-natural stream mesocosms (pH of the least acidic mesocosm: 6.82 ± 0.04) over a smaller pH gradient (Seymour *et al.*, 2018). Although the pH of seawater is alkaline (typically 7.8 to 8.4), estuarine pH can be lower i.e. <7.5 pH and even be slightly acidic, <7.0 pH (Ringwood and Keppler, 2002). pH varies in estuaries temporally, seasonally (Selleslagh and Amara, 2008) and over tidal cycles (Ringwood and Keppler, 2002), and in some cases spatially (Martino and Able, 2003) and hence eDNA degradation may vary temporally and spatially. Although, more research on this area using the pH ranges regularly encountered in estuaries is probably required to draw robust conclusions. Finally, salinity may influence eDNA degradation. Higher salinities have been shown to slow the decay of bacterial DNA, for example (Eichmiller *et al.*, 2016). Although further research is required.

With regards to biotic environmental factors, numerous studies have shown an effect of microbial (biotic) communities on eDNA degradation. Microbes will both degrade eDNA directly and indirectly using extracellular enzymes. The effects of extracellular enzymes will be mediated by the abiotic conditions and their influence will continue in the absence of the microbes which produced them (Barnes *et al.*, 2014). For example, suppression of bacteria using antibiotics has been shown to reduce the overall loss of eDNA in the laboratory (Lance *et al.*, 2017). However, relationships between microbial activity and eDNA degradation are not always positive, as heterotrophic bacteria might preferentially feed on other nutrient sources (Barnes *et al.*, 2014). In addition, it has been suggested that the microbial communities associated with specific water sources can influence eDNA decay rates (Eichmiller *et al.*, 2016; Lance *et al.*, 2017). Another factor may be levels of dissolved organic carbon (DOC). DOC concentrations have been negatively correlated with decay rates of eDNA from Common Carp (*C. carpio*) in lake water. Potentially because humic substances contribute to the DOC pool in aquatic environments and binding of DNA to humic substances protects it from enzymatic degradation (Eichmiller *et al.*, 2016). As noted previously estuarine waters can have high

concentrations of humic compounds (Petit *et al.*, 1999), hence this may be an important factor in slowing eDNA decay in estuarine systems.

Not all environmental factors influence eDNA persistence, such as the direct physical effect of water turbulence (Lance *et al.*, 2017). Similarly, although ultraviolet radiation (UVA and B) is known to render DNA strands unamplifiable (Merkes et al., 2014), its influence as a component of visible light is ambiguous. Some studies have found no effect on degradation rate (Andruszkiewicz *et al.*, 2017a) or on species detection (Merkes *et al.*, 2014) while others have (Pilliod *et al.*, 2013; Strickler *et al.*, 2015). But it is unlikely to be important in estuaries given their often high turbidity (McLusky and Elliott, 2004). Finally, environmental factors which affect degradation may have no effect in detection of a species in the field. In a time-series analysis of Silver Carp (*Hypophthalmichthys molitrix*) eDNA detection in the Chicago area waterway system, only reverse river flow volume (back up into the system during dry periods) influenced detection out of a suite of environmental variables, including temperature and pH. This was probably due to reverse flow volume's negative correlation with the level of dilution of the eDNA source in the waterway system upstream (Song *et al.*, 2017). Given that estuaries often have dynamic hydrographic conditions (McLusky and Elliott, 2004), it seems likely that similar processes could occur within them. It is also notable that hydrographic and particle tracking models already exist for some well-studied estuaries such as the Conwy in the UK (Robins *et al.*, 2014, 2019). Therefore, these may be of use when combined with time series sampling and estimates of eDNA decay for assessing variation in the spatial and temporal resolution of eDNA detections. Overall, for estuarine systems a combined programme of work should be carried out, including controlled laboratory experiments examining the effect of the above variables over the range of environmental factors encountered in estuarine systems. In addition, short term (i.e. monthly) and longer (i.e. seasonal) time series should be conducted with the concurrent measurement of environmental variables to allow an understanding of how they relate to field measurements. Ideally, this data should be combined with hydrographic models and particle tracking to draw conclusions about the comparative effect of eDNA transport. However, it may be cheaper and simpler to discount large effects of environmental variables by attempting to recreate known spatial and temporal patterns which occur in estuarine fish communities in estuaries, as advocated in section 7.

10. Conclusion

In conclusion, fish communities are a key component of monitoring ecosystem health in estuarine systems. Metabarcoding of eDNA may be an effective new method of providing ecologically relevant assessments of fish biodiversity in estuaries and shows substantial promise in being able to do so, based on experience in other systems. However, this has yet to be empirically tested in a spatially extensive manner, with reference to physicochemical variables, or with comparison to established methods. In addition, limitations of the current metabarcoding method prevent it from *consistently* providing the abundance data required by current assessment methods. Therefore, it may have to be used alongside other methods until this is overcome. Finally, the environmental factors which influence eDNA are still poorly understood in estuarine systems. Therefore, the potential for environmental factors to affect assessments of ecosystem health needs to be fully established or their effects discounted by recreation of known spatial and temporal patterns in fish communities.

Chapter 2.

Environmental DNA Metabarcoding for Fish Diversity Assessment in a Macrotidal Estuary: a Comparison with Established Fish Survey Methods

Authors: Thomas I. Gibson, Gary Carvalho, Amy Ellison, Enrica Gargiulo, Tristan Hatton-Ellis, Lori Lawson-Handley, Stefano Mariani, Carlo Zampieri and Simon Creer

Author Contributions

Phase	Author
Grant Acquisition	SC, GC, TE-H, TIG
Study Design	TIG, SC, GC, SM, LL-H
Field Sampling	TIG, CZ, EG
Laboratory Analysis	TIG
Supporting Data Acquisition	TIG
Bioinformatic Analysis	TIG, AE
Statistical Analysis	TIG
Manuscript Writing and Review	TIG, SC, AE, SM, LL-H

Abstract

Fishes are a dominant component of the macrofaunal in estuaries and are important for assessing the health of these threatened ecosystems. Several studies have applied environmental DNA (eDNA) metabarcoding to assess the biodiversity of fishes in estuaries. However, none have combined measurement of physiochemical variables with a spatially extensive sampling design across the full salinity gradient. This study aimed to compare spatial fish assemblage composition detected via eDNA metabarcoding of surface water samples with conventional fishing gear surveys in a macrotidal estuary (the Dee, North Wales, UK). In addition, eDNA assemblage composition across seasons was investigated. In autumn 2018, triplicate eDNA samples were taken at 13 stations in a spatially systematic design alongside with seine, fyke and beam trawl sampling. In summer 2019, eDNA samples from eight of the 13 original stations were collected again in the upper and lower estuary. DNA was extracted from samples and subjected to metabarcoding analysis using an established assay targeting teleost fishes. The key findings were that in autumn, eDNA detected 17 of the 26 (71%) species caught by fishing gears, which included the most abundant species. Overall, eDNA detected a greater species richness, per 30 samples, than seine or fyke nets. Additionally, there was a clear correlation between salinity and assemblage composition, which was consistent across seasons. There were no significant changes in assemblage composition between seasons. Overall, the study indicates that eDNA metabarcoding could enhance existing fish sampling methods, by generating a more comprehensive picture of estuarine fish biodiversity and providing additional information for ecological inference and management actions.

Keywords: Estuaries, Fishes, environmental DNA, Water Framework Directive, metabarcoding, biomonitoring.

1. Introduction

Estuaries, which occur where rivers meet the sea, typically exhibit greater physicochemical variability than other aquatic systems, which in turn determines strong ecological gradients. Physicochemical variability is primarily associated with spatial variation in the salinity gradient (expressed in practical salinity units), from 0 in the river to 30-35 in coastal waters (reviewed in Whitfield & Elliott, 2011). Estuaries are of key ecological importance, possessing high primary (Nixon, 1988) and secondary productivity (Allen, 1982; Costa *et al.*, 2002). However, they are heavily impacted by anthropogenetic activities, mainly pollution and habitat degradation (Kennish, 2002). Therefore there are efforts to conserve and protect estuarine ecosystems and their associated species and habitats via environmental legislation, such as the European Water Framework Directive (WFD; 2000/60/EC; EC, 2000) and the Habitats Directive (92/43/EEC; EC, 1992), in Europe. Within the UK these have recently been replaced by equivalent domestic legislation following its departure from the European Union (JNCC, 2021; UK Parliament SI 2017/1012 and SI 2017/407).

Fishes are one of the dominant macrofaunal groups in estuaries (Martino & Able, 2003). Estuaries play a crucial role in the ecology of numerous fish species, providing an environment for truly estuarine fishes, nursey habitat for many marine species and a migratory route for diadromous fishes such as Atlantic Salmon (*Salmo salar*; McLusky & Elliott, 2004). Within estuaries, numerous abiotic and biotic environmental factors may influence the spatio-temporal distribution of fish directly and via interactions between environmental variables (reviewed in Martino & Able, 2003). Generally, salinity on a spatial scale (Selleslagh *et al.*, 2009; Whitfield *et al.*, 2012) and temperature on a temporal scale are the best predictors for the abundance and assemblage structure of fishes in estuaries (Selleslagh *et al.*, 2009). Overall, the species richness of the fish assemblage declines from its maximum in the euhaline (salinity > 30) marine environment through to the oligohaline (salinity 0.5 – 5.0) river (Martino and Able, 2003; Selleslagh and Amara, 2008; Whitfield *et al.*, 2012). Furthermore, fish assemblages in estuaries show strong spatial changes in taxonomic composition (Nicolas *et al.*, 2010b; Teichert *et al.*, 2018b). Freshwater and diadromous species dominate in the lower salinity reaches of the upper estuary, while marine and estuarine species dominate in the higher salinity zones towards the sea (Nicolas *et al.*, 2010b). This spatial turnover in species diversity (beta-diversity) underpins the overall biodiversity of fish assemblages at the estuary scale (gamma diversity; Teichert *et al.*, 2018b).

The study of fish assemblage structure in estuaries has an important practical applications as they are good indicators of the health of estuarine ecosystems (Whitfield and Elliot, 2002; Teichert *et al.*, 2016). Furthermore, data on fish assemblages, along with other taxonomic groups, is specifically required to assess the status of estuaries (and other transitional waters) under the WFD (EC, 2000). The WFD has led to the development of fish monitoring programs across Europe, which provide data to calculate multi-metric indices of fish assemblage health for each estuary (Coates *et al.*, 2007; Delpech *et al.*, 2010; Harrison and Kelly, 2013). At present the requirement for monitoring fish, equivelent to the WFD, is still required in England and Wales (UK Parliament, SI 2017/407). The metrics used to calculate these generally measure aspects of species diversity and composition, species abundance, and the ecological and feeding guilds of species (Coates *et al.*, 2007). Species guild classifications provide vital information on the nursery function and trophic integrity of an estuary (Harrison and Whitfield, 2004; Coates *et al.*, 2007; Harrison and Kelly, 2013). Currently, these metrics rely on data from capture-based methods using a variety of fishing gear types (e.g. beam and otter trawls, seine and fyke nets) at sampling stations, generally deployed longitudinally along the estuary in a spatially systematic manner (Colclough *et al.*, 2002; Hemingway and Elliot, 2002; Coates *et*

al., 2007; Harrison and Kelly, 2013). However, established captured based techniques have disadvantages, including known sampling biases, requiring the use of multiple fishing gear types to gain a comprehensive assessment of the assemblage (Hemingway and Elliot, 2002; Coates *et al.*, 2007). In addition, capture-based assessments can result in mortality of captured fish, damage to habitats (Kubečka *et al.*, 2012), and low detection probabilities (Evans and Lamberti, 2017). Therefore, there has been, and continues to be a need for the development of alternative methodologies. There has been substantial innovation in the application of hydroacoustics (Samedy *et al.*, 2015), electrofishing (Warry *et al.*, 2013) and visual methods (Becker *et al.*, 2010) to monitoring fishes in estuaries, as well as further development of existing fishing gear (Harrison *et al.*, 2017).

In addition to innovation in conventional fish sampling methodologies, in recent years there has been substantial advances in the analysis of environmental DNA (eDNA) for the detection of fish species. Environmental DNA is defined as DNA isolated from an environmental sample without capturing the organism (Taberlet *et al.*, 2012). The application of eDNA metabarcoding (see Lawson Handley, 2015; Deiner *et al.*, 2017) is potentially well suited to WFD assessment of fish, including those in estuaries (Hering *et al.*, 2018). The main conclusions of these studies are as follows. Firstly, in estuaries, fish eDNA metabarcoding generally detects greater species richness than other conventional methods, such as beam trawls (Zou *et al.*, 2020), baited remote underwater videos (BRUVs; Cole *et al.*, 2022) and combined multi-method netting techniques (Hallam *et al.*, 2021). However, while some species are detected by both eDNA and conventional methods, not all species are reliably detected by eDNA (Zou *et al.*, 2020; Hallam *et al.*, 2021; Cole *et al.*, 2022). In addition, eDNA may detect a different assemblage composition than conventional methods (Hallam *et al.*, 2021; Cole *et al.*, 2022). Secondly, eDNA metabarcoding can detect differences in the fish assemblage composition within estuaries at a number of spatial scales. At the scale of 100s of kilometres, fish assemblages can be distinguished between the St Lawrence river (Canada), its upper and middle estuary and the marine Gulf of St Lawrence (García-Machado *et al.*, 2021). At the scale of 10s of kilometres in the Thames (UK), differentiation in fish eDNA assemblage composition is detectable between the estuary and the river, and also between the upper and middle estuary. The dissimilarity within the estuary is caused by the greater incidence of freshwater and anadromous species in the upper estuary, and the greater incidence of marine migrant and estuarine species in the lower estuary (Hallam *et al.*, 2021). Even at scales of less than 1 km, differences in assemblage composition have been detected using eDNA between oyster reefs and sandy sediment habitats (Cole *et al.*, 2022). Thirdly, in addition to spatial shifts in biodiversity, eDNA is able to detect seasonal changes in the composition of fish assemblages in estuaries (Stoeckle *et al.*, 2017; Zou *et al.*, 2020), but not in all studies (Hallam *et al.*, 2021).

Despite the advances made in the study of fish eDNA in estuaries, further research is required. Few eDNA studies have collected concurrent data on physicochemical variables, chiefly salinity. Given that changes in assemblage composition in previous studies have not been associated with this key ecological parameter it is difficult to generalise the results between different estuarine ecosystems. A notable exception being (Ahn *et al.*, 2020), which correlated changes in proportions of freshwater and marine species at the mouths of five Japanese estuaries with changes in salinity. In addition, most studies have generally focused on specific sections of the estuary (Ahn *et al.*, 2020; Hallam *et al.*, 2021; Cole *et al.*, 2022) or amalgamated several surveys from different seasons (García-Machado *et al.*, 2021). Therefore, more spatially comprehensive studies, which not only compare eDNA metabarcoding with fishing gears, but directly correlate changes in fish assemblage composition with salinity are required. In addition, given that estuaries are generally open systems, species eDNA detections within estuaries are likely to be affected by eDNA transport from outside the ecosystem, which

is a widespread phenomenon in aquatic ecosystems (Deiner and Altermatt, 2014; Shaw *et al.*, 2016; Yamamoto *et al.*, 2017). Transport of eDNA could occur from the river, the sea and potentially wastewater outflows (Nakagawa *et al.*, 2018) and can be potential noise in the interpretation of fish eDNA data (Yamamoto *et al.*, 2017). In addition, the ability of eDNA to detect species from transported eDNA is of concern to environmental managers as spatially specific data is often required in ecological assessment (T. I. Gibson, *pers. obv.*). Therefore, there is a clear requirement to support eDNA detections using the wealth of previous fish survey data available for UK estuaries (Waugh *et al.*, 2019) to give environmental managers confidence that the results will be relevant for ecological assessment.

1.2 Aims and Objectives

The main aim of this study was to compare the fish assemblage detected via eDNA metabarcoding of surface water samples to conventional fishing gears in a macrotidal estuary (the Dee, Wales, UK), and determine if eDNA could detect ecologically relevant spatial and seasonal patterns in assemblage composition. The first objective was to compare eDNA data collected in autumn (October 2018) with data from fishing gears (seine, beams and trawls) collected concurrently. It was hypothesised that eDNA would detect more species in the estuary overall and would show a different assemblage composition (species presence/absence) to fishing gears. In addition, it was hypothesised that the assemblage composition would be correlated with salinity (and its correlates), across methods. The second objective was to determine if clear changes in eDNA assemblage composition could be detected longitudinally along the estuary in both autumn and summer (June 2019). The third objective was to determine if seasonal changes could be detected between autumn and summer using eDNA. It was hypothesised that there would be a greater species richness in the estuary overall in autumn than in summer, and that assemblage composition would differ between seasons. In addition, it was hypothesised that there would be a consistent correlation between eDNA assemblage composition and salinity between each season.

2. Methods

2.1 Study Location

The river Dee (on the border between North Wales and England, UK) ends in a macrotidal estuary, with a mean spring tidal range of ~ 10 m (at Hilbre Island), in addition the river can have high peak flows (300 m^3 s^{-1}). The physical and morphological characteristics of the estuary produce a complex circulation pattern and although the water column is generally well-mixed, stratification can occur in deeper channels at the mouth of the estuary (Bolaños *et al.*, 2013). There is substantial urban development and industrial activity around the estuary, the upper reaches of which are heavily canalised. Transitional and Coastal (TraC) fish surveys have been conducted from 2002 to 2017 in spring/summer (May to July) and, or autumn (September to November), for WFD assessment by Natural Resources Wales (NRW). The most consistent sampling occurred at eight stations along the estuary, from 2010 to 2016 using Beam trawls, seine and fyke nets and additional otter trawls at the estuary mouth (Supplementary Material - TraC Fish). In October 2018 a new WFD fish survey was implemented by NRW with assistance from DAERA (Department of Agriculture, Environment and Rural Affairs, Northern Ireland, UK) to provide data for the Estuarine Multi-metric Fish Index (EMFI; Harrison & Kelly, 2013).

2.2 Fish Survey

The EMFI fish survey was conducted from 15.10.2018 to 18.10.2018, approximately around low tide. The fishing gear used were seine nets (30m long × 2m deep, 14mm mesh body with a 5 m long × 6.5 mm central panel), double fyke nets (each trap was 0.5 m high, 2.5 m long with a 10 mm mesh cod end and was joined by a 6m long × 15mm mesh leader), and a beam trawl (1.5m wide × 0.5m high, the net body 3m long × 10/14 mm mesh with a 1 m long × 5/6.5 mm mesh cod end). Seine netting was conducted in shallow littoral areas, fyke nets were set for ~ 24 hrs in deeper waters and trawling was conducted in mid-channel areas at a speed of 1-2 knots for approximately 5 min or for a set distance of 100-200 m (Harrison & Kelly, 2013). Sampling stations were placed in a generally spatially systematic manner from the head to the mouth of the estuary (Rozas and Minello, 1997). Although, exact station placement locations had to account for the suitability of each location for the gear type used. Fifteen stations were sampled using fyke and beam trawls. A further 30 stations were sampled using seine nets, 15 of which were located near to the fyke and beam stations (figure 1). In general, a single sample per gear was taken at each station with a few exceptions at six beam trawl stations, where re-towing was required due to low initial sample yields.

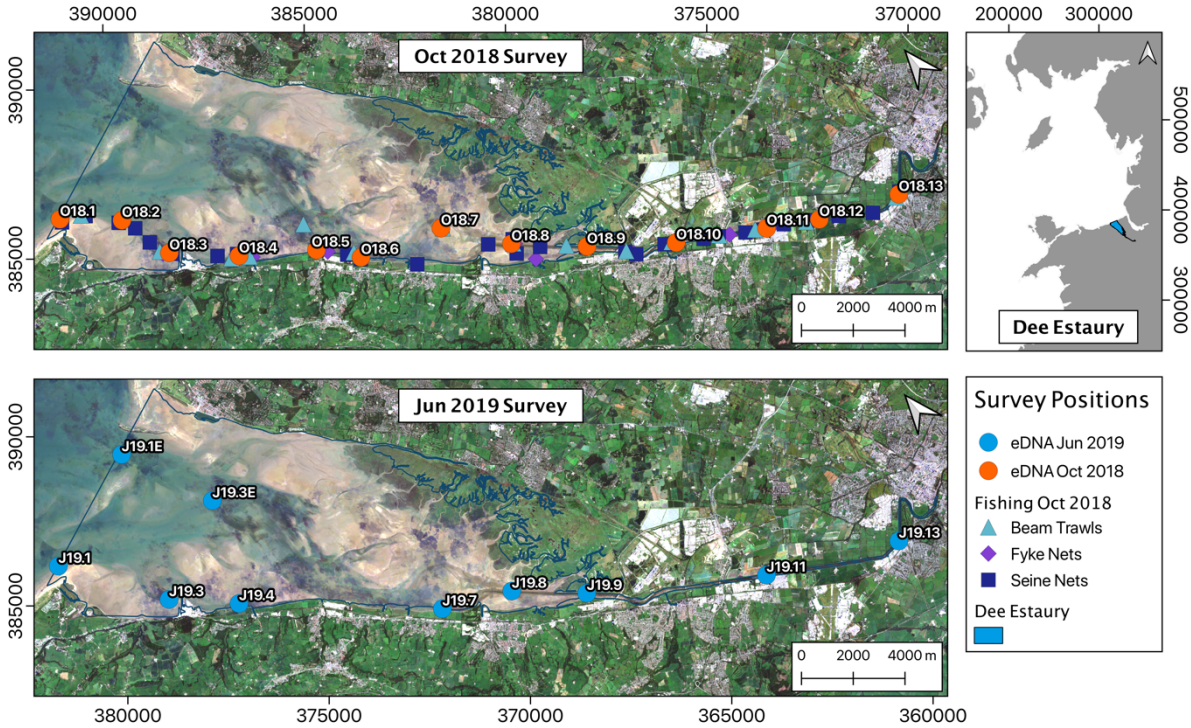


Figure 1: Map of the Dee Estuary, giving its geographic location within Britain and the distribution of eDNA sampling stations in October 2018 relative to fishing gear sampling stations. Coordinate System: British National Grid (EPSG:27700) axis in eastings/northings (m). British Coastline (Wessel & Smith, 1996 and 2017), Dee Satellite Photography (Copernicus, 2019), Dee Estuary Extent (Natural Resources Wales, 2019).

2.3 eDNA Sampling

Prior to the eDNA surveys and after each sampling day, equipment was prepared and decontaminated in a non-PCR laboratory. 1 L water Nalgene bottles (Nalgene Nunc International, Rochester, NY, USA) and silicone tubing (Cole-Parmer, Vernon Hills, IL, USA) were decontaminated in 5% commercial bleach solution (in tap water) for 4-5 hrs then rinsed

with tap water and air-dried. All other equipment, sample boxes etc., were decontaminated with a 10% commercial bleach solution, followed by rinsing in tap water.

In autumn 2018 (15-17.10.2018) eDNA sampling occurred in the vicinity of 13 of the 30 seine stations, eDNA sampling stations were sampled in a spatially systematic design (Rozas and Minello, 1997) around low tide. Sampling was conducted after fyke net deployment, but before beam trawling. At each station three 1 L water surface samples were taken (total samples: 39). The location of each station was recorded using a GPSMAP 64s (Garmin Ltd., Olathe, KS, USA). In summer (10-11.06.2019) eight of the original 13 eDNA stations were re-sampled in the upper and lower estuary at low tide, with two additional stations added at the request of NRW (total samples: 33). No fish sampling was conducted in summer 2019 and stations were as close to the 2018 stations as possible (figure 1). Before taking each water sample, the sampling bottle was filled once with water from the estuary, and this water discarded. The sampler wore gloves, which were changed between stations, and all care was taken to separate eDNA samples on the survey vessel from caught fish while it was sampled in October 2018. Samples were immediately sealed in plastic ziplock bags and placed on ice in a sealed box. At each station a Pro Plus Quatro (YSI Inc., Yellow Springs, OH, USA) multiprobe was used to measure temperature (°C), salinity (PSU), dissolved oxygen (DO) and pH at the surface (~ 1 m depth). Depth was also recorded using the vessel's eco-sounder or a weighted line. However, in October 2018 some variables were not recorded by NRW consistently, and so only temperature, salinity and DO were used in the analysis. At the end of each day a 1 L ddH_2O field blank was opened on board the vessel, sealed and placed inside the box with the other samples (3 field blanks per survey). Samples were then immediately transported back to the laboratory and stored on ice, or in a cold room (4°C) until filtration at the end of each day.

Before filtration, the outsides of sampling bottles were wiped with 5% commercial bleach solution to prevent contamination, then cleaned with tap water and dried with paper towels. Filtration was carried out in a non-PCR room on the day of collection. Silicon tubing was inserted into each sample and a small volume of water pumped through to pre-wash the tubing. Samples were filtered through an encapsulated 0.8 µm PES filter with an integrated 5.0 µm GF pre-filter (Nature Metrics Ltd., Egham, UK) using a Geopump Peristaltic Pump (Geotech Environmental Equipment, Inc., Denver, CO, USA). After filtration, filters were capped, bagged in sterile whirl packs and frozen at – 20°C in a non-PCR lab. The tubing and gloves were changed between filters. Field blanks were processed in an identical manner.

2.4 DNA Extraction

DNA extraction was carried out in a dedicated pre-PCR eDNA lab in a UV sterilised Biological Safety Cabinet (FASTER, SRL, Milan, Italy) to prevent contamination. The cabinet and equipment were sterilized with UV light for 20-30 mins before and after work. Batches of samples from different stations were extracted together in a random order. Total DNA was extracted from each filter capsule using DNeasy Blood and Tissue Kits (QIAGEN, Hilden, Germany) following a modification of the (Spens and Evans *et al.*, 2017) protocol. Briefly, 720 µl of Buffer ATL (QIAGEN) and 80 µl Proteinase K were added into each filter and incubated overnight at 56°C to allow sample lysis. To remove humic acids, 300 µl of flocculant solution (Sellers *et al.*, 2018) was added to ~ 1 ml of sample lysate, vortexed and incubated for ~ 1 hr at 4°C in the refrigerator (G. Sellers and R. Donnely *pers. comm.*). Each sample was then centrifuged at 10,000 x g for 2 min and 1200 µl of the supernatant removed (Sellers *et al.*, 2018). The rest of the DNA extraction followed Spens and Evans *et al.* (2017) and 70 µl of AE buffer (QIAGEN) was used for the final elution. Extraction blanks, consisting only of Buffer ATL and Proteinase K, were added at the sample lysis step and treated identically to samples.

All plastic tubes used for handling eDNA were DNA LoBind (Eppendorf, Hamburg, Germany). DNA extracts were stored at -20°C in DNA LoBind tubes (Eppendorf) in the pre-PCR lab. Total DNA concentrations were measured using 1 µl of each sample on a Nanodrop ND-1000 Spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, USA). In total 35 samples were successfully extracted for October-2018; four samples from four separate stations dropped out of the analysis due to a laboratory error. In June 2019 all samples were successfully extracted.

2.5 Library Preparation & Sequencing

The 12S rRNA fragment was amplified with PCR using Tele02 primers (Miya *et al.*, 2015; Taberlet *et al.*, 2018). Set-up was conducted in a biological safety cabinet, in a pre-PCR eDNA lab and all equipment was sterilized with UV as above. For each survey: samples, field and extraction blanks were randomly ordered on 96-well plates with three PCR negative controls (RNase Free-H_2O; QIAGEN) and three PCR positive controls (~ 0.2 ng/µl DNA zebra mbuna cichlid, *Metriaclima zebra*). All round one PCR reactions were conducted in triplicate. Samples and blanks from each survey were processed on separate plates to reduce cross contamination between seasons. Total reaction volume for each PCR reaction was 25 µl containing 12.5 µl 2x QIAGEN Multiplex PCR Master Mix (QIAGEN), 1.25 µl of template (or control), 0.5 µl of forward and reverse Tele02 primer at 10 µM concentration and 10.25 µl of RNase-Free Water (QIAGEN). The thermal cycle profile was: *Taq* activation at 95°C for 15 min; 35 cycles of 94°C for 30 sec, 60°C for 1 min 30 sec, 72°C for 20 sec and finally 72°C for 10 min. All steps following the initial PCR were carried out on the bench in a post PCR lab. For each triplicate, 20 µl from each reaction was pooled to account for PCR bias between plates. From each pool, 15 µl of product was cleaned of DNA fragments under 200 bp (bead ratio: 1.75X) using ProNex® Size-Selective paramagnetic beads (Promega Corporation, Madison, WI, USA). Amplification in pooled PCR products was checked by gel electrophoresis. Each cleaned PCR product was indexed and with its own unique i5/i7 dual index combination (Integrated DNA Technologies, Inc. Coralville, IA, USA) via a second-round of PCR. PCR master mix and indexes were combined inside a dead air cabinet before template addition to prevent contamination. Total reaction volumes for each PCR were 25 µl, containing 12.5 µl 2x QIAGEN Multiplex PCR Master Mix (QIAGEN), 3.0 µl of template, 1 µl of premixed i5/i7 indexes at 10 µM concentration (Integrated DNA Technologies, Inc.) and 8.5 µl of RNase-Free Water (QIAGEN). The second step thermal cycling profile was: 95°C for 15 min; 15 cycles of 94°C for 30 sec, 60°C for 1 min 30 sec, 72°C for 20 sec; 72°C for 10 min. Second-round PCRs were checked for amplification using gel electrophoresis. Following PCR, reaction concentrations were quantified using a Qubit™ dsDNA Broad Range Assay Kit (ThermoFisher Scientific, Waltham, MA, USA). All PCR products were then pooled in equimolar quantities. This final pool was loaded into a 1.5% agarose gel, run for 30 mins at 90V, imaged and the target amplicon manually excised to remove primer dimer and high molecular weight material from PCR. The gel slice was purified using a QIAEX II Gel Extraction Kit (QIAGEN) and quantified using a Qubit™ dsDNA HS Assay Kit (Invitrogen). Finally, the purified pool was diluted to a concentration of 12 pM, with a 3% PhiX spike and sequenced with the Illumina MiSeq Reagent v2 (500 cycles; Illumina, San Diego, CA, USA).

2.6 Bioinformatics

The quality of the demultiplexed files from the MiSeq run was assessed by generating a report for each sample and controls using *FastQC* (Andrews, 2010) and then assessing quality over all files using *MultiQC* (Ewels *et al.*, 2016). The Tele02 primers were then removed using *Cutadapt* (Martin, 2011). Quality trimming of the files then proceeded using *fastp*, using a

sliding window method to drop low quality bases from the head and tail of each read (in both 5' to 3' and 3' to 5' direction, mean quality = phred 30; Chen *et al.*, 2018). The high phred score was used to minimise the potential for the misidentification of a species due to sequencing error. The quality of the files was then reassessed with *FastQC* and *MultiQC* as before. The reads were denoised, dereplicated, merged and cleaned of chimaeras in *dada2* (Callahan *et al.*, 2016). Taxonomic assignment broadly followed Collins *et al.* (2019). The exported Amplicon Sequence Variants (ASVs) were assigned an approximate taxonomy using a "global" BLAST search (evalue: 1×10^{-4}; Camacho *et al.*, 2009) against the 'nt' BLAST database (NCBI, 2021) to provide an approximate classification. Following this, all ASVs were searched using a "local" BLAST against the Meta-Fish-Lib, a curated UK fish database (Collins *et al.*, 2021; Collins, 2021a) to identify UK fish species (evalue: 1×10^{-5}). All the ASVs identified by both searches as fish were then assigned a taxonomy using an Evolutionary Placement Algorithm to verify the result of the local BLAST search (Barbera *et al.*, 2019) and GAPPA (Czech and Stamatakis, 2019; Czech *et al.*, 2020). ASVs were then assigned to species based on the following rules (see Collins *et al.*, 2019 for rationale). Rule 1: Species level taxonomy was assigned if both the best scoring BLAST hit and species level EPA result were identical and sequence identity was ≥ 97%. Rule 2: Species level taxonomy was assigned if both the BLAST hit and the species level EPA results were identical, BLAST identity was ≥ 95% and EPA probability ≥ 90%. Rule 3: Species level taxonomy was assigned if the BLAST hit alone was 100% and no species had tied on their BLAST score, even if EPA had not given a species level result. However, for rule 3 for European Flounder (*Platichthys flesus*; 3 ASVs) and European Plaice (*Pleuronectes platessa*; 4 ASVs) the threshold for assigning a species assignment based on BLAST alone was dropped to a 99% BLAST match identity. This was because it was apparent from other 100% matches for these species that the EPA algorithm struggled to place these species and the reference library was complete for close relatives. Following this, the BLAST and EPA results for any query sequences which had passed through these filters were manually appraised. Higher-level phylogenetic levels as determined from EPA, up to order level, were assigned to any remaining query sequences. Most assignments were to species level. For brevity, all assignments are referred to as species when discussed collectively.

In addition to taxonomic assignment, species were assigned to estuarine use functional guilds using the guild classification system for European estuaries (Franco *et al.*, 2008). Guild assignments were based on this system initially, but the classifications in Elliott & Hemingway (2002) and Elliott & Dewailly (1995) were used when multiple guild assignments were present for individual species. Briefly, guilds are Marine Stragglers (MS), Marine Migrants (MM), Estuarine Species (ES), Anadromous Species (A), Catadromous Species (C), Freshwater Species (F; see Franco *et al.*, 2008). Where no guild assignment was found or could be assigned, species were counted as Unassigned (UA) and were retained in the analysis. However, the ambiguity between *Melanogrammus aeglefinus*/*Merlangius merlangus* was resolved by assigning the sequences to whiting (*M. merlangus*) due to haddock being absent from all the WFD catch data, both from October 2018 and from 2002 to 2017.

To account for contamination, a conservative, per species read threshold cut-off was calculated based off the approach in adapted from Yamamoto et al., (2017). Species that contributed equal to or less than 0.55 % of the total target fish reads in a sample were considered absent. This cut-off was generated by taking the reads of the most abundant fish species in negative controls (*Anguilla anguilla*: 1464 reads) and dividing it by the number of negative controls (21) to give a per species, per control, contamination value (70). This value was multiplied by the number of water samples (65) to give the total potential per taxa contamination (4531). Total potential per species contamination was then divided by the total target fish reads in the water samples (823183) to give the threshold cut-off (0.005505 ≈

0.55%). Following this, sample level rarefaction curves of species richness vs. reads were inspected (supplementary figure 1 and 2). Samples where species accumulation curves did not show signs of plateauing and, or that had an extremely low species richness, were considered likely sequencing failures and rejected. The assemblage composition of these samples was also checked using nMDS (presence/absence; Bray-Curtis distance) to ensure they were definite outliers (supplementary figure 3 and 4).

2.7 Statistical Analysis

All statistical analysis and data manipulation was conducted using the R programming language (R Core Team, 2021). Firstly, species (and higher level taxa) detected by eDNA were checked against the TraC Fish data from 2002 to 2017 to ensure they had been detected previously in the estuary. In addition, species were checked to see if they had been detected by fishing gears in October 2018. For comparisons of eDNA in October, the closest station for each gear type to the eDNA stations was confirmed using nearest neighbour analysis using the Distance Matrix tool in QGIS (coordinate reference system: OSGB 1936; QGIS Development Team, 2019). In addition, this was used to assign physiochemical data collected at Fyke Stations to the nearest Seine and Beam trawl stations.

2.7.1 Species Richness

Sample-size-based rarefaction and extrapolation (R/E) sampling curves were used to compare species richness at the level of the estuary using the 'iNEXT' package (Chao *et al.*, 2014; Hsieh *et al.*, 2016). Curves were calculated separately for each gear type and the eDNA data for October 2018, using the presence/absence data per sample. For the eDNA data three curves were generated, one using all species, a second using species previously detected by TraC Fish and a third using species only detected by the fishing gears in October 2018. R/E curves were generated for twice the sample size of each gear type (Hsieh *et al.*, 2016) and 95% confidence intervals (CI) and standard errors calculated using 1000 bootstrap replicates. In addition, asymptotic species richness and a richness estimate per 30 samples was calculated and compared. Finally, separate R/E curves and asymptotic species richness were compared for the eDNA data and species detected by TraC Fish between autumn 2018 and summer 2019.

2.7.2 Spatial Variation in Fish Assemblage Composition

For any spatial comparisons of assemblage composition (presence/absence) the reads of samples at each eDNA sampling station were summed together, to treat each station as an independent sample. This was to account for non-independence of samples within each station (Hurlbert, 1984). Only species previously detected in the TraC Fish surveys (2002 - 2017) were used for the following analyses. This was to reduce the influence of species which were more likely to have been transported into the estuary, given they had never once been detected in the estuary previously across a wide range of gear types, sites and years.

Spatial variation in eDNA assemblage composition was determined separately for Autumn and Summer and analysed using ordinations with generalised linear latent variable models (GLLVM; R Package: gllvm; Niku *et al.*, 2019). The binomial distribution (probit link) was used, with 50 initial runs including random variation (jitter = 0.01) to minimise sensitivity to the chosen starting values (Niku *et al.*, 2019). GLLVMs were fitted with 2 to 5 latent variables and model fit assessed using the AICc to check 2-D ordinations were appropriate. In addition, model selection was performed using AICc to check if including total per sample read counts as a fixed factor influenced the assemblage composition. A similarity profile routine (SIMPROF) test was conducted using the 'clustsig' package (Whitaker and Christman, 2015)

based on Bray-Curtis distances from species presence/absences to test if there were no statistically significant differences within ordination groupings. A conservative p-value < 0.001 threshold was used to account for multiple-testing (Clarke *et al.*, 2008). Finally, an indicator species analysis was conducted using the 'indicspecies' package (Cáceres and Legendre, 2009) to identify species associated with the SIMPROF groupings, and combinations of groupings. The association between species and assemblage groupings, was calculated using the Indicator (IndVal) index (Dufrene and Legendre, 1997; Cáceres and Legendre, 2009; De Cáceres *et al.*, 2010). The statistical significance of associations was calculated using 10000 bootstrap iterations. The *p*-values were adjusted using the "fdr" method to account for multiple testing (see Benjamini & Hochberg, 1995). In addition to the IndVal index, its two constituents A_{pa} and B_{pa} were calculated. A_{pa} is the probability a station belongs to the grouping given the fact the species has been found. Whereas B_{pa} is the probability of finding the species when the station belongs to the target station group (Dufrene and Legendre, 1997).

2.7.3 Comparisons of Assemblage Composition with Fishing Gear

A comparison of assemblage composition (presence/absence) at all 13 eDNA stations against nearby seine nets was conducted, then a comparison of composition at 9 eDNA stations with a full complement of nearby fishing gear stations (Fyke, Beam and Seine) was made. For each analysis all explanatory environmental variables, temperature (°C), salinity (PSU), DO (%) were checked for collinearity visually and using the Pearson correlation coefficient. Both temperature and dissolved oxygen had correlations > 0.8 with salinity and so were considered colinear (supplementary figure 5; Zuur *et al.*, 2007). Therefore, only salinity (PSU) was retained as it is the primary environmental driver of fish assemblage structure (Whitfield *et al.*, 2012). Graphical comparisons of assemblage composition, in both analyses, were made using GLLVM ordinations as above. Direct comparisons of the fixed effects of sampling method (factor), salinity (continuous) and the interaction between method and salinity, on fish assemblage composition and the presence/absence of individual species was made for both datasets using multivariate GLMs in the mvabund R package (Wang *et al.*, 2012). Analysis of Variance tests were conducted using the Wald test statistic and assuming correlation between species response variables with ridge regularisation. *P*-values were calculated using the PIT-trap resampling method (5000 bootstrap iterations), while *p*-values for univariate tests for each species response variable were adjusted for multiple testing. Model selection was then applied using backwards selection and assessing the AIC (Zuur *et al.*, 2007).

2.7.4 Comparisons of Assemblage Composition between Seasons

For comparisons of seasonal changes in assemblage composition, the subset of stations that were sampled in both autumn and summer were retained. For GLLVM ordinations all the stations in summer were used, but for formal analysis using multivariate GLMs the additional two stations in the lower estuary were removed (i.e. 8 stations per season). Temperature, salinity and DO were checked for collinearity, and salinity and DO were retained in the analysis as temperature was colinear with salinity (supplementary figure 6). Direct comparisons of the fixed effects of season, salinity, DO, and the interaction between season and salinity, on fish assemblage composition and the presence/absence of individual species, was made for both datasets using multivariate GLMs as above. Model selection was performed as above.

Following each analysis, model validation of GLLVM and multivariate GLMs was conducted by plotting the residuals. In addition, experimental directional variograms were calculated for the residuals of each species from multivariate GLMs in a South-Eastly (135º) direction to determine if stations were statistically independent (Zuur *et al.*, 2007, 2009).

3. Results

3.1 Sequencing Results

A total of 13,409,393 paired reads were generated from the MiSeq sequencing run. Of the total, 5,321,787 (39.7%) reads passed quality filters and returned a BLAST hit to chordates against the nt database and 828,499 reads (6.2 % of total; omitting the positive controls) were assigned to fish present in the Meta-Fish-Lib database (supplementary table 1 and supplementary figure 7). A total of 6415 reads, from 12 fish species, were detected in negative controls, 98% of these reads were from one field and one extraction blank from October 2018. No contamination was detected in positive controls. Contamination was accounted for with the 0.055 % read contribution, per sample cut-off threshold. Of the contaminating species, only the Guppy (*Poecilia reticulata*; table 1), a common aquarium species, was unlikely to have occurred in the estuary although its eDNA may have been. However, where the 11 target fish species found in the negative controls passed this threshold, they were retained in the analysis due to their ecological relevance.

Two samples from October 2018 (10C and 2B) and June 2019 (J19.1B and J19.4A) were removed due to extremely low read depth. Prior to the cut-off threshold the total number of species detected in October 2018 was 42; following cleaning this fell to 39. Specifically, a lamprey species (*Lampetra sp.*), a possible flatfish (pleuronectiform) and Nilsson's pipefish (*Syngnathus rostellatus*) were lost from the samples. The median species richness of the samples also dropped from 16 (interquartile range, IQR: 11) to 14 (IQR: 8) after threshold application. For June 2019, prior to filtering, the total number of species was 48, that was reduced to 38 taxa after filtering. The 10 filtered species included *Lampetra sp.*, a pleuronectiform and two dragonet taxa: *Callionymus lyra*, *Callionymus sp.* taxa among others. The median species richness of the samples dropped from 19 (IQR: 10.5) to 17 (IQR: 4.25) after threshold application. Investigation of species accumulation curves showed that the number of taxa in the cleaned datasets in both October 2018 and in June 2019 were approaching saturation (supplementary figures 1 and 2). Only the cleaned eDNA data was used in the following analyses (table 1).

Table 1: Taxonomic and Guild Assignments for each species and comparison to contemporary and previous TraC fishing

Taxa	Common Name	Guild	Oct-18				Jun-19		Previously Detected
			eDNA		Fishing		eDNA		
			Total Reads	Incidence	Total Catch	Incidence	Total Reads	Incidence	
Platichthys flesus †	European flounder	MM	78091	31	201	41	77903	26	●
Merlangius merlangus †	Whiting	MM	38163	13	31	2	673	2	●
Phoxinus phoxinus †	Eurasian minnow	F	34113	27	-	-	79402	20	○
Cottus gobio †	Bullhead	F	24414	28	-	-	36344	22	○
Salmo trutta	Brown trout	A	18426	25	-	-	16633	19	●
Anguilla anguilla †	European eel	C	13107	23	24	5	18644	19	●
Pleuronectes platessa †	European plaice	MM	11947	11	106	19	14941	14	●
Salmo salar †	Atlantic salmon	A	9667	20	-	-	6050	16	●
Oncorhynchus mykiss †	Rainbow trout	F	8185	24	-	-	6960	18	○
Barbatula barbatula	Stone loach	F	7020	22	-	-	12310	19	○
Pomatoschistus microps	Common goby	ES	6145	13	1707	29	4318	11	●
Leuciscus leuciscus †	Common dace	F	5853	17	16	8	9842	16	●
Rutilus rutilus	Roach	F	5456	19	-	-	10503	18	●
Gasterosteus aculeatus †	Three-spined stickleback	A	4633	15	8	5	31095	21	●
Thymallus thymallus	Grayling	F	4068	18	-	-	8233	14	○
Dicentrarchus labrax †	European seabass	MM	3461	12	-	-	11734	11	●
Squalius cephalus	Chub	F	2914	13	-	-	7733	17	●
Gymnocephalus cernua	Ruffe	F	2708	12	-	-	10227	16	○
Perca fluviatilis	European perch	F	2539	15	1	1	7679	17	●
Pomatoschistus minutus	Sand goby	ES	2473	10	61	15	8560	9	●
Gobio gobio	Gudgeon	F	2431	11	1	1	5373	16	●
Clupea harengus	Atlantic herring	MM	2028	10	4	3	4401	6	●
Sprattus sprattus	European sprat	MM	2002	5	-	-	57473	13	●

Table 1 Continued

Taxa	Common Name	Guild	Oct-18				Jun-19		Previously Detected
			eDNA		Fishing		eDNA		
			Total Reads	Incidence	Total Catch	Incidence	Total Reads	Incidence	
Ammodytidae	Sand lances	UA	1702	6	-	-	45673	10	■
Solea solea	Common sole	MM	1401	2	1	1	4173	9	●
Barbus barbus	Barbel	F	1351	5	-	-	186	1	○
Limanda limanda	Dab	MM	1324	3	21	1	5554	6	●
Gadidae	Cods	UA	1009	1	-	-	-	-	-
Esox lucius	Northern pike	F	976	9	-	-	948	5	●
Poecilia reticulata †	Guppy	UA	944	2	-	-	-	-	○
Osmerus eperlanus	European smelt	A	920	5	2	2	3072	8	●
*Chelon sp.**	Mullet genus	UA	794	2	6	4	1011	3	■
Cottidae	Sculpins	UA	759	1	-	-	-	-	-
Gadus morhua	Atlantic cod	MM	636	4	9	2	-	-	●
Coregonus sp.	Whitefish	F	394	5	-	-	-	-	○
Abramis brama	Freshwater bream	F	266	2	-	-	1116	5	○
Pungitius pungitius	Ninespine stickleback	F	161	2	-	-	213	2	○
Salvelinus fontinalis	Brook trout	UA	83	1	-	-	-	-	○
Scophthalmus rhombus	Brill	MS	64	1	2	1	-	-	●
Buglossidium luteum	Solenette	MS	-	-	-	-	1795	5	●
Carassius auratus	Goldfish	F	-	-	-	-	101	1	○
Cyprinus carpio	Common carp	F	-	-	-	-	368	1	○
Echiichthys vipera	Lesser weever	MS	-	-	1	1	144	3	●
Myoxocephalus scorpius	Shorthorn sculpin	ES	-	-	-	-	1147	2	●
Scardinius erythrophthalmus	Rudd	F	-	-	-	-	126	1	●

Incidence: Total presence/absence in samples. **Previously Detected Cat.:** Exact Species Present: ● No Species or Clade Present: ○ For Identified Undifferentiated "Species" - Species Present: ■ Taxa where comparison not possible indicated with a dash (-). **Chelon sp*. refers to *Chelon ramada* in contemporary fishing data. † Indicates species detected in blanks.

3.2 Physicochemical Environment

In October the survey covered the full spatial transition in salinities around low water from 0.06 in the upper estuary to 30.04 in the lower estuary. Salinity increased little in the upper estuary and then increased rapidly in the middle and lower estuary. Surface water temperature showed limited spatial variation, generally increasing from 12.6°C in the upper estuary to 13.4°C in the lower estuary. DO saturation was relatively high and increased from ≥ 84.5 to ≤ 92.1 % in the upper estuary to 99.5 % in the lower estuary. There was little variation in water depth where measured, the median depth was 1.5 m with some outliers (maximum depth of 8.2 m). pH was not measured consistently but was 8.0 to 8.4 at two stations in the lower estuary. In June 2019, at low water, the survey captured the full spatial transition in salinities around low water from 0.07 in the upper estuary to 30.23 - 27.65 in the lower estuary. Similarly, to October 2018 salinities increased little in the upper estuary and then rapidly increased in the middle and lower estuary. Surface water temperature showed limited spatial variation ranging from 12.2 °C in the upper estuary and increasing to 14.1 to 14.3 °C in the lower estuary. Dissolved oxygen was high, increasing from 95% in the lower estuary to 101.3 - 101.2 % in the upper estuary. Whereas pH varied from 7.55 in the upper estuary, and 8.01 in the lower estuary. In June 2019, depth generally increased from 1.15 m to 2.1 – 4.5 m in the lower estuary, with some outliers. The volume of samples filtered varied between surveys due to variation in turbidity, but was consistent within surveys. In October 2018 mean sample flowthrough was 513 ml (SD: 23 ml), in June 2019 mean sample flowthrough was 907 ml (SD: 67 ml).

3.3 Comparison with Fishing Survey

A detailed comparison of intersecting species from eDNA and gear types in October, and TraC data, is given in figure 2 and table 1. Overall, in the TraC Fish data from 2002 to 2017 a total of 63 fish species were detected (figure 2). In October 2018, a total of 24 species were detected by a combination of fyke nets, seine nets and beam trawls. All these species had been detected at least once in the Dee estuary by TraC Fish surveys, in autumn from 2002 to 2017 (figure 2). No fish were caught in two beam trawls (station 4 and 7) and three fykes (station 11, 12 and 13). These stations were omitted from any further analysis.

Of the 24 species detected during the fishing survey, 16 had direct matches in the eDNA data in October 2018. In addition, detection of *Chelon sp*. by eDNA was probably the *Chelon ramada* detected in the fishing survey. Overall, these 17 species (71% of the species detected by fishing) made up 57.0 % of the total reads in the cleaned October 2018 eDNA dataset. This included the 11 most abundant species in the fishing survey which accounted for 98.8 % of the total catch abundance. In October 2018, eDNA detected an additional 22 species which were not detected by the fishing survey. Of these species, 8 had been detected on the Dee in autumn historically, accounting for 14.7 % of the reads in October. Therefore, there is evidence that 71.8 % of the reads in October data came from fish that *could* have been present in the estuary. Of the 14 taxa not accounted for, two represented higher-level taxa: Gadidae and Cottidae, potentially unsequenced species, local variants, or partially degraded sequences. These could not be compared with species level records for fishing data. Of the remaining 12 species, 11 belong to the Freshwater guild and the detection of *Coregonus sp.* was likely from a freshwater upstream population. These freshwater species represented 27.7% of the reads in the October. It is probable these never previously detected freshwater species represent eDNA transported into the estuary from further upstream in the river.

A comparison of intersecting species between the 2018 and 2019 eDNA datasets and the TraC data is shown in figure 3 and table 1. Of the species detected in June 2019 with eDNA,

22 had been detected in TraC Fish data (2002 to 2017) in summer and a further 5 detected in autumn. Overall, this accounted for 69.7 % of the reads in June 2019. This might suggest that most species data was of local origin within the estuary. In addition, this proportion seems relatively consistent across the two seasons despite the differences in survey design. A further 11 Freshwater guild species had never previously detected by TraC Fish. Similarly to autumn, these species accounted for 30.3 % of the total reads. In addition, 9 of these new species had been detected in October by eDNA. This may suggest eDNA transport was relatively consistent between seasons.

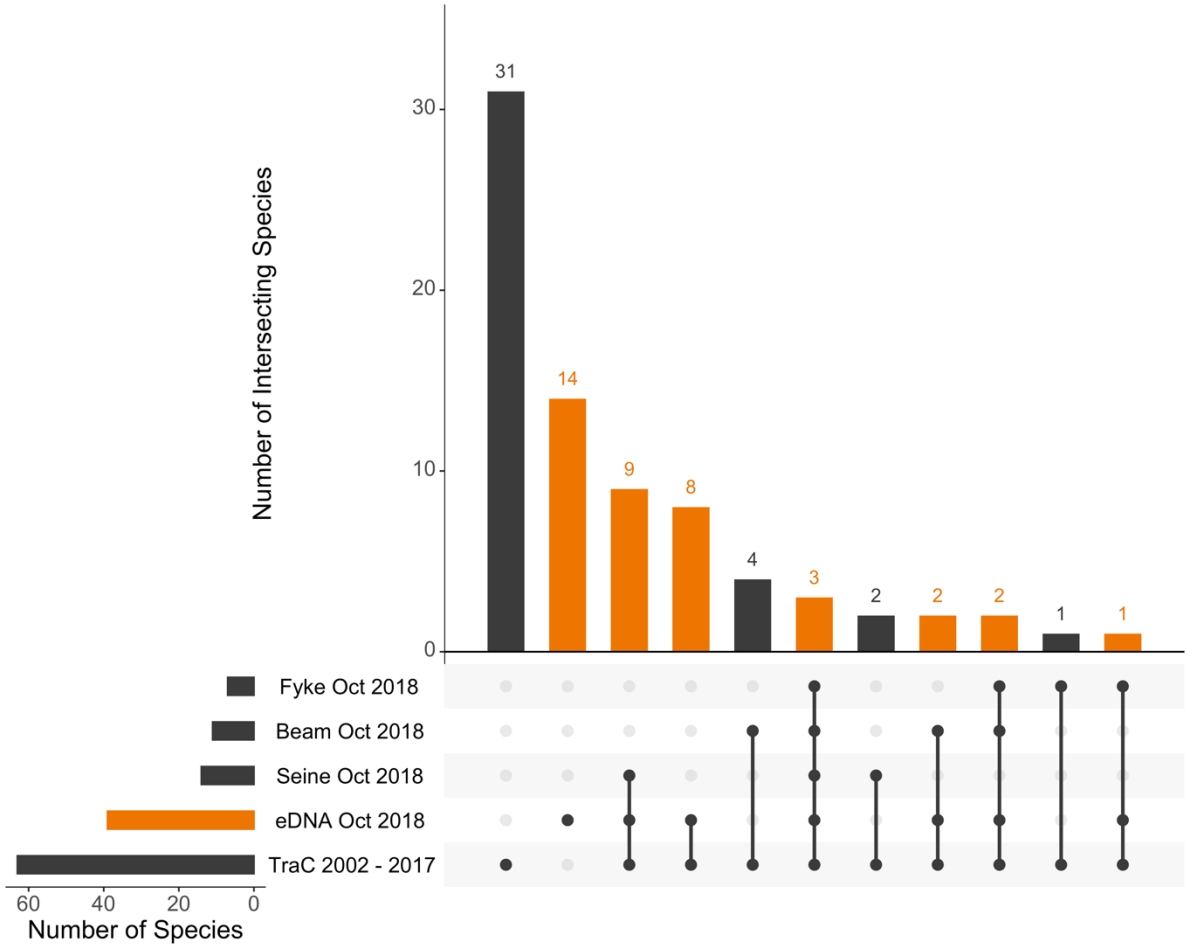


Figure 2: UpSet plot (Conway *et al.*, 2017) showing the number of intersecting species between each species list generated for each dataset. Species lists were generated for October 2018 fyke, beam and seine catch datasets, the cleaned eDNA October 2018 dataset and the TraC Fish data (2002 - 2017). The bottom table shows each dataset and which intersections between species lists they contributed to. Single black dots indicate no intersections between species lists. Black dot connected by lines indicate which lists shared species. The top bar graph gives the number of species shared between each species list in each intersection, intersections including species detected in the eDNA data are indicated in orange. The left-hand bar chat shows the total number of species within each species list (eDNA species list indicated in orange).

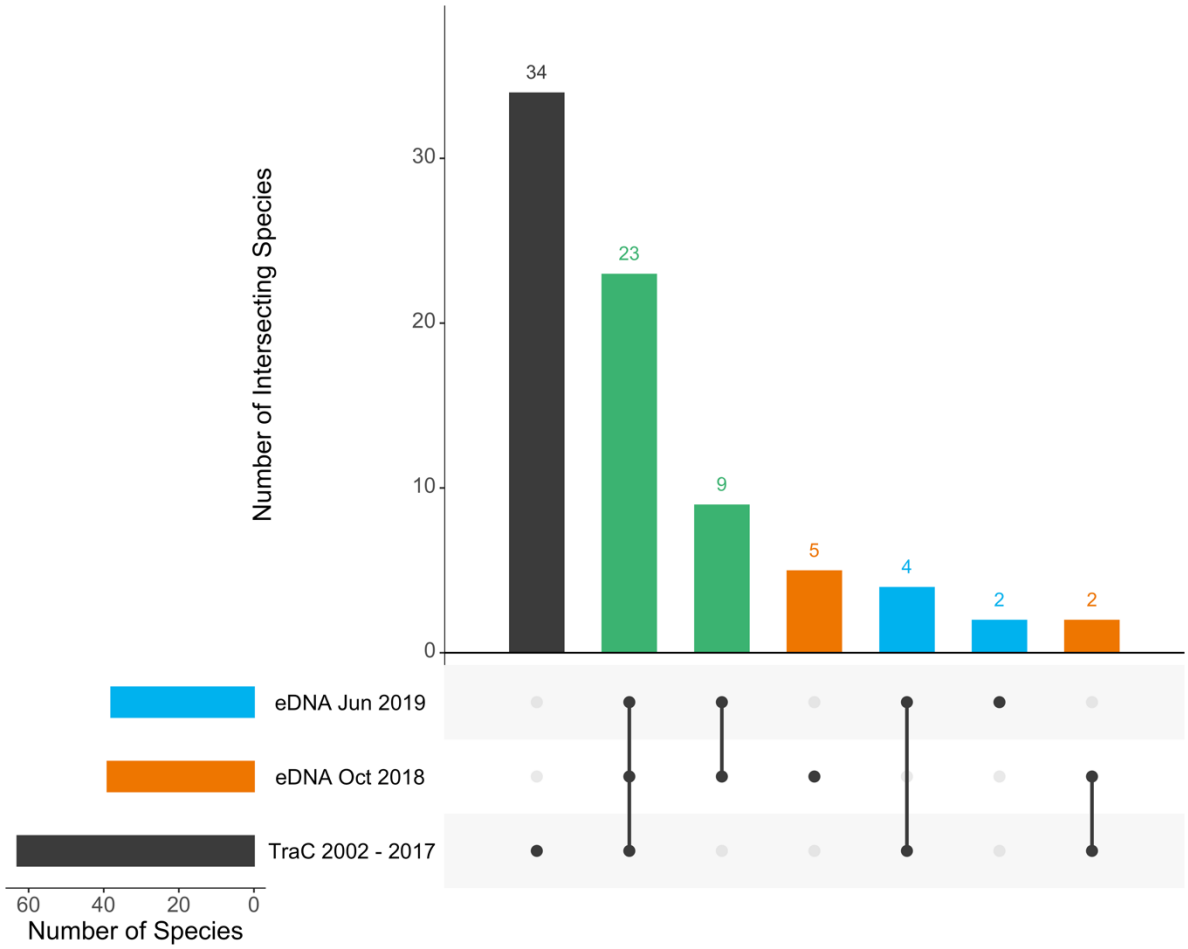


Figure 3: UpSet plot (Conway *et al.*, 2017) showing the number of intersecting species between each species list generated for each dataset. Species lists were generated for the cleaned eDNA October 2018 and June 2019 datasets and the TraC fish data (2002 - 2017). The species lists for June are shown in blue and the species lists for October are shown in orange in the table. Intersections which only overlapped with October are shown in orange, intersections which only overlapped with June are shown in blue. Intersections which overlapped with June and October are shown in green.

3.4 Species Richness

In October 2018 the full eDNA dataset had the highest asymptotic species richness. After filtering the eDNA by prior TraC Fish detections from 2002 to 2017 ('eDNA TraC Fish') and filtering by species caught in fishing gear in October 2018 ('eDNA Catch') asymptotic species richness progressively fell (CI none-overlapping; table 2; figure 4). The two filtered datasets showed a more rapid plateau in the species accumulation curve than the full eDNA data (figure 4). Comparing between the two filtered eDNA datasets and different gear types showed similar patterns of species accumulation to seine and fyke nets. Beam trawls showed a steeper R/E curve with higher variance. However, the two filtered eDNA datasets showed more rapid increases in species richness at lower sampling levels than the gear types (figure 4). Comparisons of asymptotic species richness between the filtered datasets and gear types showed that 'eDNA TraC Fish' data, only had a higher estimated asymptotic species richness than fyke nets (CL none-overlapping, $p = < 0.05$; table 2). All other comparisons were not statistically significant (CL overlapping, $p = \geq 0.05$; table 2). However, this result may have been due to extrapolating from relatively low initial starting sample sizes. Therefore, to account for this, comparisons were also made for the estimated species richness for 30 sample (the number of seine net samples). When species richness estimates were given for 30 samples it

was shown that in addition to the statistically significant difference with fyke nets, 'eDNA TraC Fish' detections had a significantly higher species richness than the seine nets (CL none-overlapping; $p = < 0.05$; table 3). In addition, the 'eDNA Catch' data also had a statistically significantly higher species richness than the fyke nets (CL none-overlapping; $p = < 0.05$; table 3).

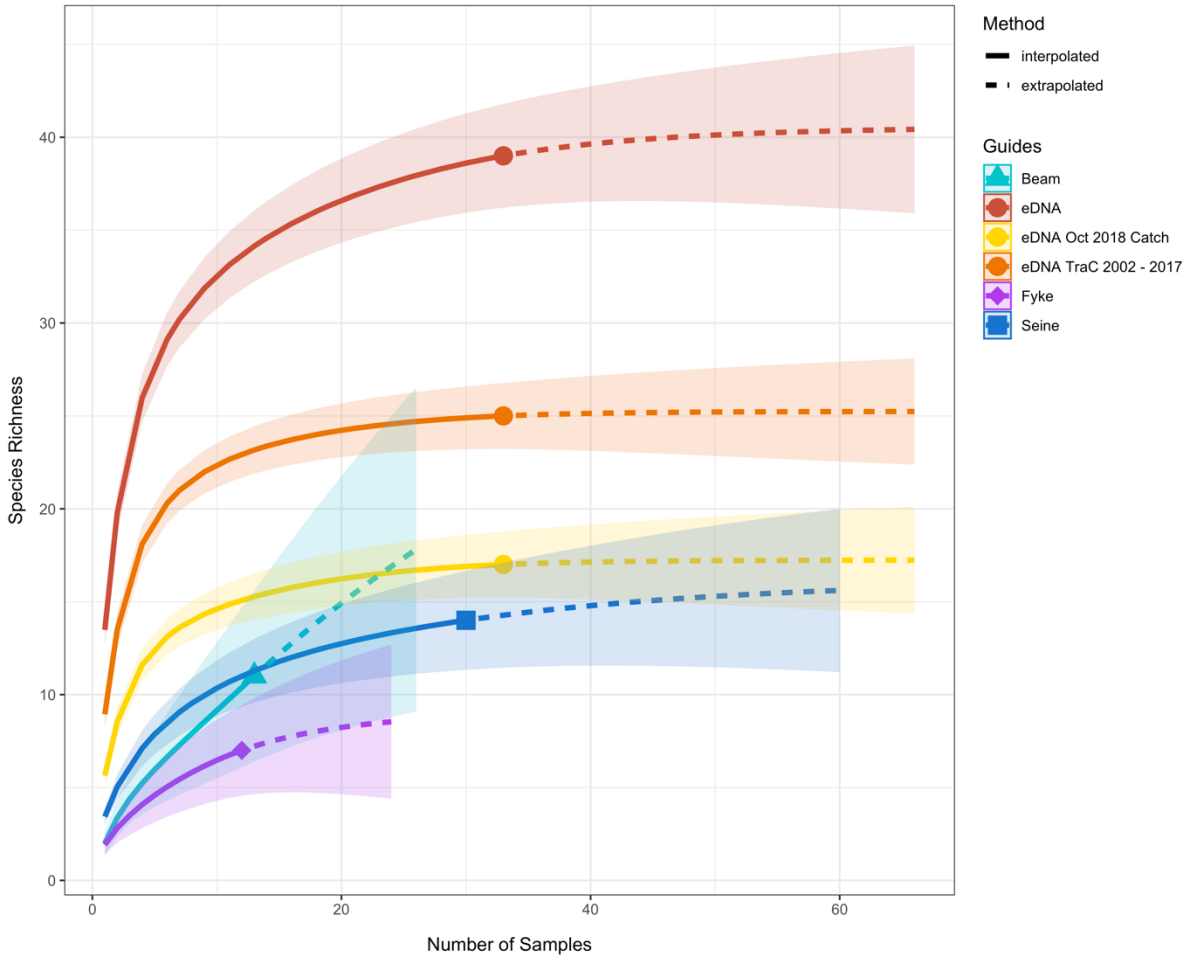


Figure 4: Species Rarefaction and Estimation (R/E) curves and 95% confidence intervals calculated on species (including some higher-level taxa) presence/absence data for eDNA and catches from October 2018. R/E curves were calculated individually for each gear type (beam, fyke and seine). For the 2018 eDNA data, R/E curves were calculated including all species (eDNA), species only previously detected in TraC surveys (eDNA TraC 2002 - 2017) and species only detected in the October 2018 catch data (eDNA Oct 2018 Catch). R/E curves were calculated, to double the observed sample size, using the iNEXT software (Hsieh *et al.*, 2016), confidence intervals were calculated using 1000 bootstrap iterations.

Table 2: Estimated Asymptotic Species Richness and Confidence Intervals for fishing data, eDNA data and eDNA data filtered by species detected in TRaC surveys (2002 - 2017) and filtered by the October 2018 catch.

Sampling Method	Species Richness		SE	95% CI	
	Observed	Estimated		Lower CI	Upper CI
eDNA – All Taxa	39	40.55	2.11	39.21	50.51
eDNA – TraC Fish	25	25.24	0.71	25.01	29.63
eDNA – Oct 2018 Catch	17	17.24	0.71	17.01	21.63
Beam Trawls	11	36.85	19.64	17.88	108.05
Seine Nets	14	16.18	3.29	14.26	32.48
Fyke Nets	7	9.06	3.14	7.24	24.65

CI = Confidence Interval. SE = Standard Error.

Table 3: Estimated Species Richness and Confidence Intervals for 30 samples fishing data, eDNA data and eDNA data filtered by species detected from TraC surveys (2002 - 2017) and filtered by the October 2018 catch.

Sampling Method	Method	SC	Species Richness	95% CI	
				Lower CI	Upper CI
eDNA – All Taxa	Interpolated	0.99	38.61	36.04	41.18
eDNA – TraC Fish	Interpolated	1.00	24.90	23.32	26.47
eDNA – Oct 2018 Catch	Interpolated	0.99	16.90	15.22	18.57
Beam Trawls	Extrapolated	0.80	19.52	10.04	29.00
Seine Nets	Interpolated	0.97	14.00	11.54	16.46
Fyke Nets	Extrapolated	0.99	8.80	4.10	13.50

CI = Confidence Interval. SC = Sample Coverage.

3.5 Spatial Variation in eDNA Assemblage

After removing species that had not previously been detected in TraC Fish data, GLLVM ordinations (binomial distribution; 2 latent variables) and SIMPROF analysis (Bray-Curtis distances) of species presence/absence data in October 2018 and June 2019 showed clear changes in assemblage composition across stations (figure 5 and 6). In October 2018 three groupings of stations in assemblage composition (groups: A - C) along the estuary were obvious from the ordination (figure 5 and 6). The AICc fell when station read depth was removed as a covariate (483 to 427), indicating read depth did not affect the composition. SIMPROF analysis confirmed that there were no substantial differences in assemblage structure within the three groups (p = <0.001; supplementary figure 8). However, station 10 was a relative outlier (group D, Oct 2018; figure 5). An analysis of the strength of the association of each species with each grouping, and combinations of groups, using the IndVal index showed the Marine Migrant species whiting (*M. merlangus*) and sea bass (*Dicentrachus labrax*) were both associated with group A and B. Whereas the Freshwater common roach (*Rutilus rutilus*) was associated with group B and C (table 4). Overall, the analyses suggest the presence of two, overlapping assemblages characterised by Marine Migrant and Freshwater species in the lower and upper estuary, respectively.

In June 2019 a comparable split was seen in assemblage composition in the ordination between stations in the lower estuary (A) and those stations in the upper estuary (B), with station 1E identified as a relative outlier (figure 5 and 6). The AICc fell when station read depth was removed as a covariate (from 473 to 374) as with October 2018. SIMPROF showed two main groupings of stations displaying no substantial differences in assemblage composition within them (p = <0.001; supplementary figure 8), whilst station 1E was clustered within group

A. Indicator species analysis showed that six Marine Migrant, one Estuarine Species and the Unassigned Ammodytidae were associated with the A grouping. The B grouping contained no statistically significant indicator species. It is notable that the June 2019 A group was in the same area of the estuary as the A + B grouping in October 2018 and shared an indicator species, *D. labrax* (table 4).

Table 4: Indicator Species (fdr $p = < 0.05$) for SIMPROF groupings, October 2018 A-C and June 2019 A-B.

Year	SIMPROF Group	Indicator Species (Guild)	A	B	Index	fdr *p-value*
October 2018	**A + B**	*Merlangius merlangus* (MM)	1.00	1.00	1.00	0.0144 *
		Dicentrachus labrax (MM)	1.00	1.00	1.00	0.0144 *
	B + C	*Rutilus rutilus* (F)	1.00	1.00	1.00	0.0256 *
June 2019	**A**	*Ammodytidae* (UA)	1.00	1.00	1	0.0382 *
		Clupea harengus (MM)	1.00	1.00	1	0.0382 *
		Dicentrachus labrax (MM)	1.00	1.00	1	0.0382 *
		Pomatoschistus minutus (ES)	1.00	1.00	1	0.0382 *
		Solea solea (MM)	1.00	1.00	1	0.0382 *
		Sprattus sprattus (MM)	1.00	1.00	1	0.0382 *

A: Positive Predictive Value. B: Species Fidelity. Index: square-root of IndVal index.

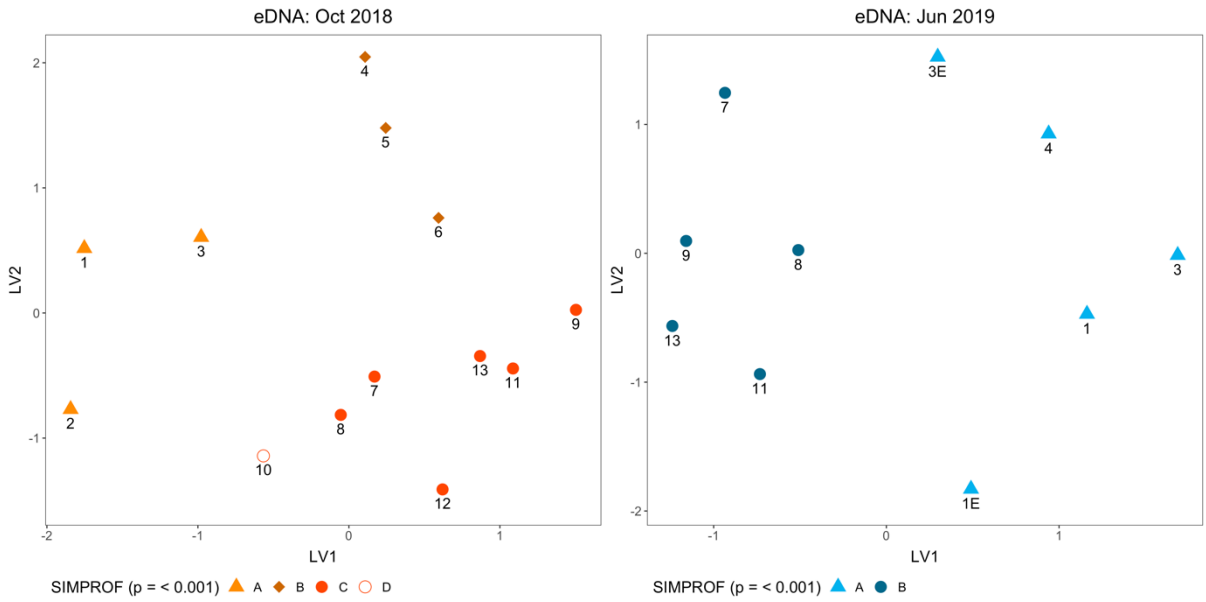


Figure 5: Presence/absence of fish species per station in October 2018 (left) and June 2019 (right) modelled using a binomial GLLVM (log link; two latent variables; 50 iterations) for eDNA (filtered by TraC Fish detections 2002-2017). In both panels colours and symbols indicate the SIMPROF groupings for each season ($p = < 0.001$) calculated using Bray-Curtis dissimilarities calculated using species presence/absences.

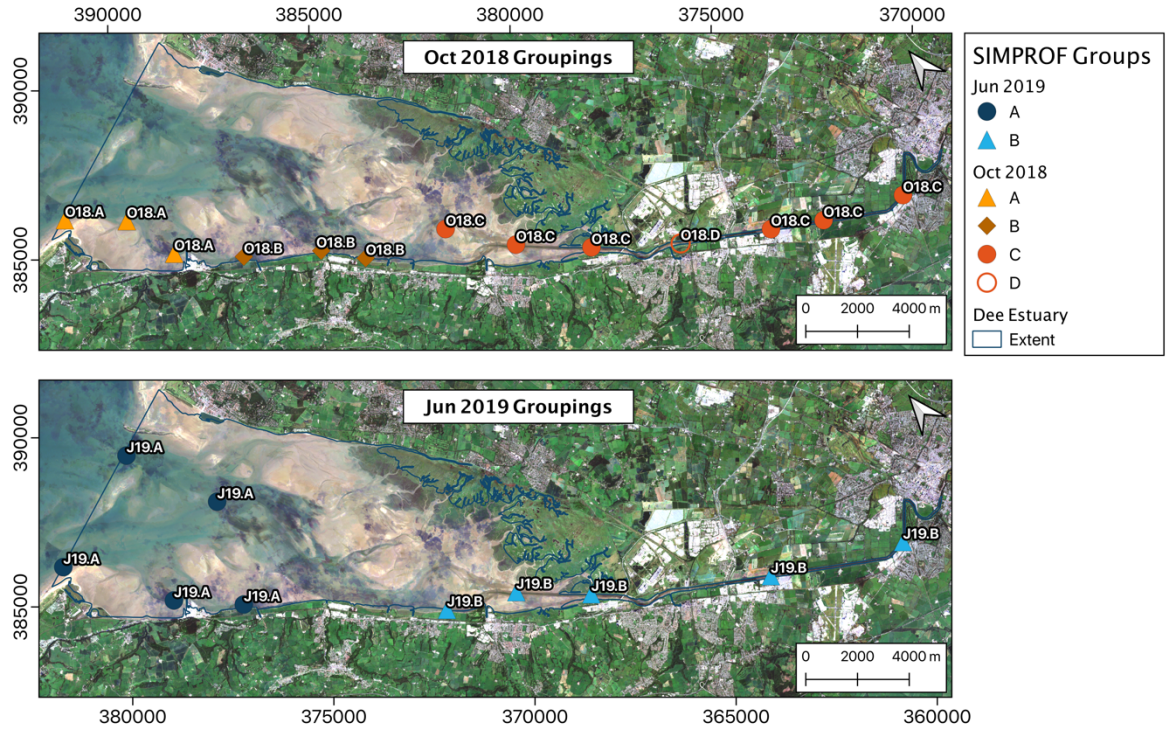


Figure 6: The distribution of SIMPROF (p = < 0.001) fish assemblage groupings (presence/absence; Bray-Curtis Distances) from eDNA data in October 2018 and June 2019 in the Dee Estuary. Coordinate System: British National Grid (EPSG:27700) axis in eastings/northings (m). British Coastline (Wessel & Smith, 1996 and 2017), Dee Satellite Photography (Copernicus, 2019), Dee Estuary Extent (Natural Resources Wales, 2019).

3.6 Fish Assemblage Composition

3.6.1 eDNA vs. Seine

After removing species in the eDNA not detected by TraC Fish (2002 to 2017), the ordination of 13 seine net and 13 eDNA stations showed an overall separation in species presence/absences between the two methods, although the eDNA data showed a greater variation in assemblage composition (figure 7). The spatial trend in species presence/absence for the October 2018 eDNA data previously identified was still apparent (figure 7). Further to this, the seine net data showed a comparable spatial shift in species presence/absences over the same set of stations. There was no obvious pattern in the model residuals (supplementary figure 9). Multivariate GLM analysis showed comparable ecological patterns to the ordination. The AIC of the final GLM was improved (from 511 to 477) by omitting the interaction term between Method and Salinity. The final model showed marginally statistically significant differences in the species presence/absence between seine nets and eDNA and a clearly statistically significant effect of salinity on species incidence (table 5). This suggests that the effect of salinity on species presence/absence was consistent between methods. Regarding species level patterns, two freshwater species, dace (*Leuciscus leuciscus*), perch (*Perca fluviatilis*) and the anadromous three-spined stickleback (*Gasterosteus aculeatus*) were detected less frequently in seine net stations than the eDNA data. In addition, four Marine Migrant and one Estuarine species showed statistically significant positive relationships with salinity (table 6). Model validation showed a somewhat non-random distribution (supplementary figure 10), suggesting that additional explanatory variables or mixed modelling could have improved the fitted model (Zuur *et al.*, 2007), if such approaches were available. There was no obvious trend in the semi-variance of the residuals suggesting spatial non-independence was not a factor.

Table 5: ANOVA for multivariate GLM – species presence/absence

AIC: 477			
Explanatory Variable	**Residual DF**	**Wald-Test**	**P-Value**
Intercept	25		
Salinity	24	6.883	$< 2 \times 10^{-16}$ ***
Method (Seine vs. eDNA)	23	4.944	0.035 *

Significance codes: *** 0.001 ** 0.01 * 0.05

Table 6: Species level ANOVA results (*p-adjusted* for multiple testing) and model coefficients for species with a statistically significant association ($p > 0.05$) for either salinity of method.

Species (Estuarine Use Guild)	**Salinity**			**Method (Seine)**		
	Wald Stat.	*p*-value	Coefficient	Wald Stat.	*p-value*	Coefficient
Gasterosteus aculeatus (A)				2.551	0.017*	-3.748
Leuciscus leuciscus (F)				2.310	0.041*	-2.821
Perca fluviatilis (F)				2.663	0.007**	-3.809
Clupea harengus (MM)	2.607	0.013*	0.122			
Pleuronectes platessa (MM)	2.623	0.012*	0.197			
Pomatoschistus minutus (ES)	2.597	0.013*	0.117			
Merlangius merlangus (MM)	2.321	0.042*	1.680			
Dicentrarchus labrax (MM)	2.321	0.042*	1.680			

Significance codes: *** 0.001 ** 0.01 * 0.05. Tests and coefficients for Method were calculated relative to eDNA

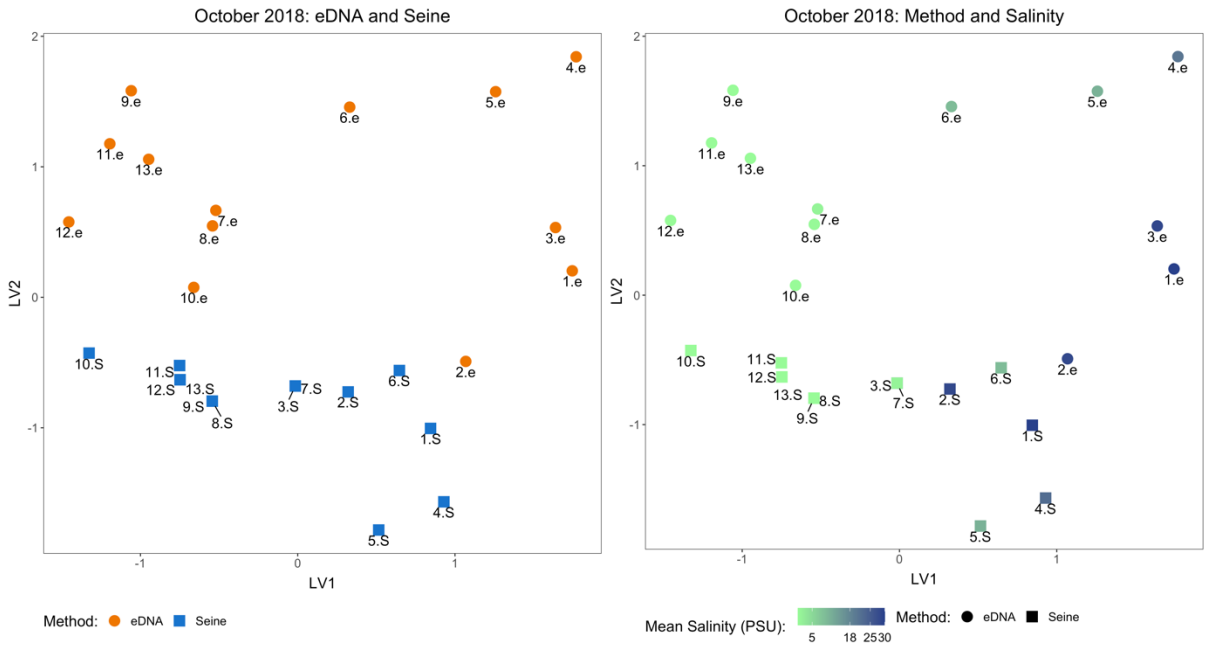


Figure 7: Presence/absence of fish species per station in October 2018, modelled using a binomial GLLVM (log link; two latent variables; 50 iterations) for eDNA (filtered by TraC Fish detections 2002-2017) and seine net stations. Left panel shows method at each station, and the right panel shows salinity at each station (PSU).

3.6.2 eDNA vs. All Fishing Gears

For comparisons between the October 2018 eDNA data (filtered for TraC Fish detections) and all gear types only 9 stations had a full complement of fishing gears at a close distance to the eDNA station. The GLLVM ordination of 9 sites sampled with eDNA, seine, fyke and beam trawl showed some separation between species incidences in eDNA and the fishing gears. Spatial variation between eDNA stations was still apparent. More limited spatial variation was obvious in the assemblage composition of the selected gear types, with a trend in the seine and beam trawl stations (figure 8). For the multivariate GLM, following model selection, the interaction term between salinity and gear was dropped to improve AIC (665 to 523). In the final model there was no statistically significant difference in assemblage composition between the methods but there was a highly statistically significant effect of salinity on assemblage composition (table 7). At the species level, two Marine Migrant species whiting (*Merlangius merlangus*) and plaice (*Pleuronectes platessa*) showed positive associations with salinity, as they had in the previous eDNA vs. seine comparison (table 8). The distributions of model residuals were comparable to the previous GLLVM and multivariate GLM (supplementary figure 11 and 12).

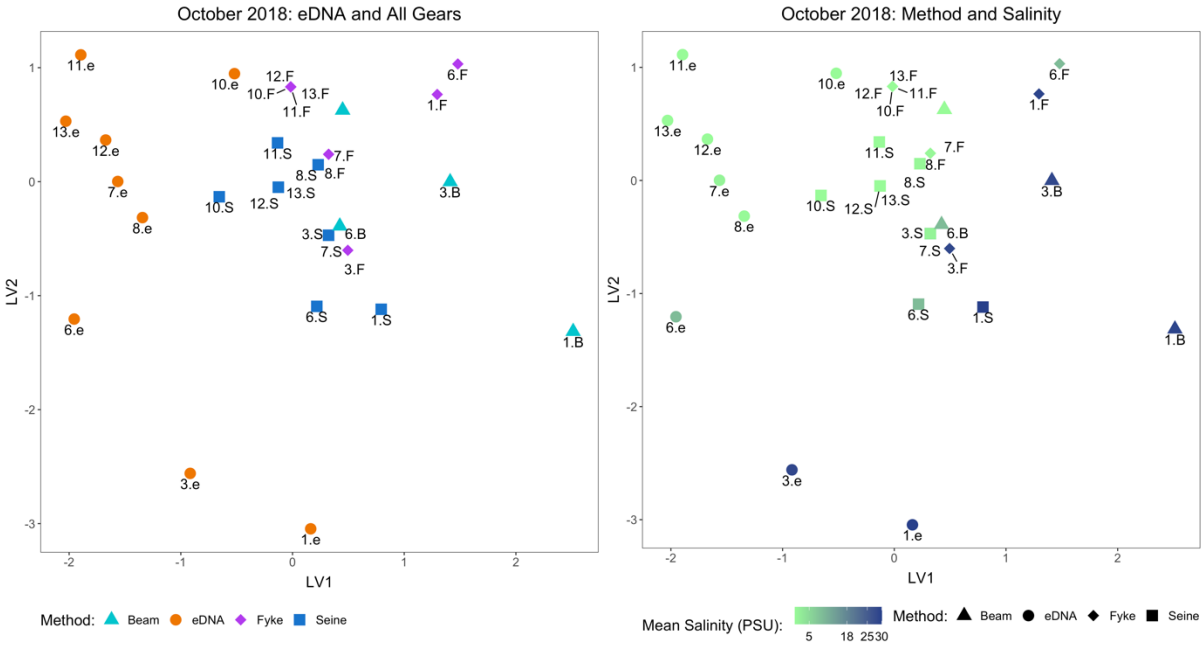


Figure 8: Presence/absence of fish species per station in October 2018, modelled using a binomial GLLVM (log link; two latent variables; 50 iterations) for eDNA (filtered by TraC Fish detections 2002 - 2017), seine and fyke nets and beam trawls. Left panel shows method at each station, and the right panel shows salinity at each station.

Table 7: ANOVA for multivariate GLM – species presence/absence

AIC: 523			
Model: Species Presence/Absence ~ Salinity + Method			
Explanatory Variable	**Residual DF**	**Wald-Test**	**P-Value**
Intercept	35		
Salinity	34	6.515	$< 2 \times 10^{-16}$ ***
Method (Seine vs. Fishing)	31	4.755	0.487

Significance codes: *** < 0.001 ** < 0.01 * < 0.05

Table 8: Species level ANOVA results (*p-adjusted* for multiple testing) and model coefficients for species with a statistically significant association (p > 0.05) with salinity.

Species (Estuarine Use Guild)	Salinity		
	Wald Stat.	*p*-value	Coefficient
Merlangius merlangus (MM)	2.653	0.012	1.163
Pleuronectes platessa (MM)	3.091	0.001	2.436

Significance codes: *** < 0.001 ** < 0.01 * < 0.05

3.7 Seasonal Comparisons

3.7.1 Seasonal Species Richness

The R/E curves calculated for October 2018 and June 2019, using all samples collected in each survey, showed that both surveys were approaching an asymptote in both the full dataset and the data filtered by species previously detected by TraC surveys (figure 9). Comparisons of asymptotic species richness between seasons for both the full eDNA dataset, and datasets reduced by TraC detections showed that there was no statistically significant difference in asymptotic species richness between the seasons regardless of data subset (figure 9; table 9).

Table 9: Estimated Asymptotic Species Richness and Confidence Intervals for eDNA data, and eDNA data filtered by species detected from TraC Fish (2002 - 2017), in October 2018 and June 2019.

Data Set	Species Richness		SE	95% CI	
	Observed	Estimator		Lower CI	Upper CI
eDNA Oct 2018	39	40.55	2.11	39.21	50.51
eDNA Jun 2019	38	40.57	3.38	38.36	56.28
eDNA Oct 2018 TraC	25	25.24	0.71	25.01	29.63
eDNA Jun 19 TraC	27	27.24	0.71	27.01	31.61

CI = Confidence Interval. SE = Standard Error.

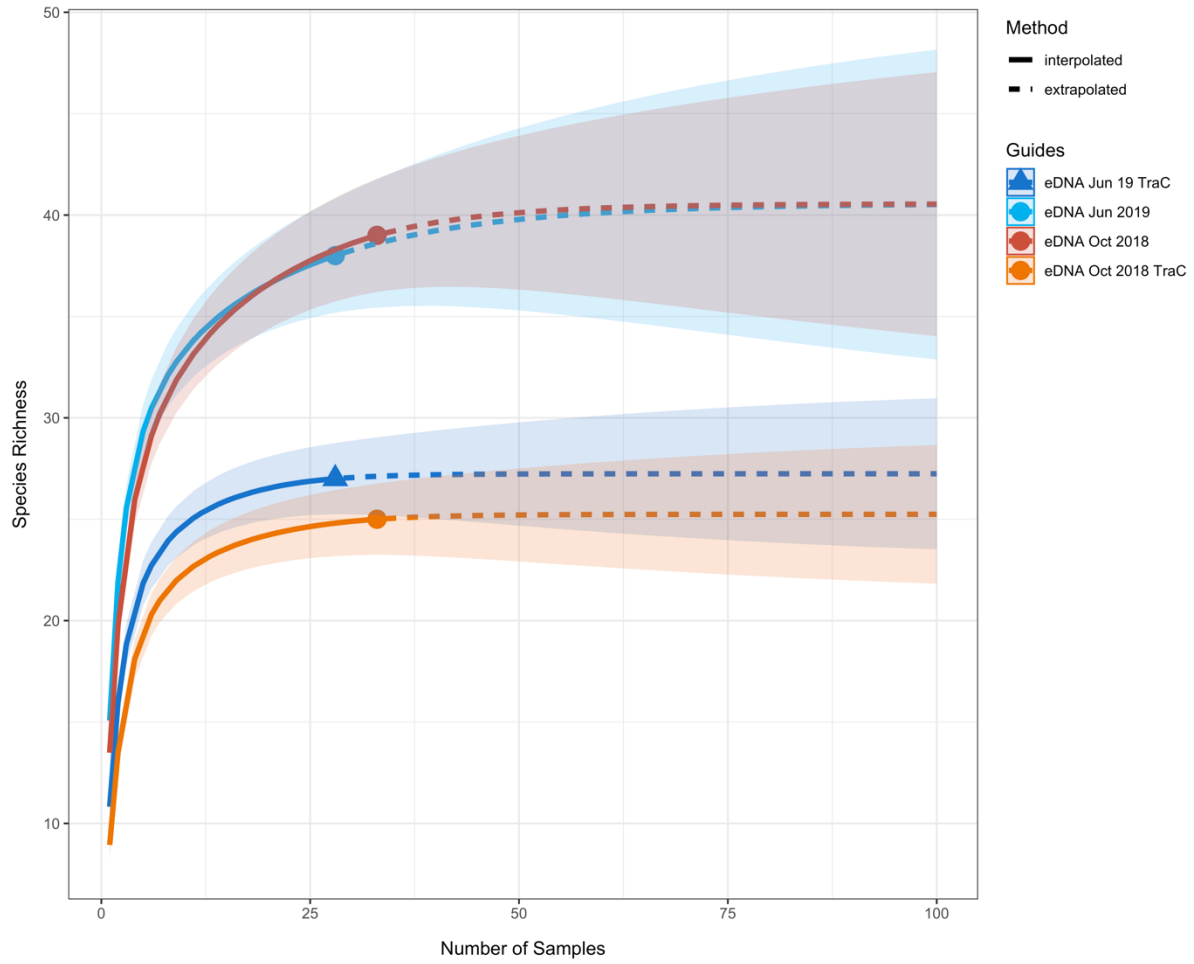


Figure 9: Species Rarefaction and Estimation (R/E) curves and 95% confidence intervals calculated on species (including some higher-level taxa) presence/absence data for eDNA from October 2018 and June 2019 (all samples and sites). For the 2018 and 2019 eDNA data, curves were calculated including all species and species only previously detected in TraC surveys from 2002 to 2017 (suffix - TraC). R/E curves were calculated, to double the observed sample size, using the iNEXT software (Hsieh *et al.*, 2016), confidence intervals were calculated using 1000 bootstrap iterations.

3.7.2 Seasonal Assemblage Composition

For seasonal comparisons between the eDNA data in autumn (October 2018) and summer (June 2019) a subset of 8 stations in the upper and lower estuary were compared. However, for ordination, the two additional stations sampled in June 2019 were retained in the analysis. GLLVM ordination (binomial distribution; 2 latent variables) showed a general overlap in assemblage composition between the two seasons with both seasons showing a shift in assemblage composition across the estuary (figure 10). AICc fell when read depth was removed from the analysis (AICc: 519 to 581). The results of the multivariate GLM confirm the ordination. Firstly, model selection showed a model with only salinity as an explanatory variable had the lowest AIC (297). All other explanatory variables and interactions were dropped by model selection and had non-significant effects in the initial model ($p > 0.05$; initial model AIC: 379). Therefore, season and dissolved oxygen had no statistically significant effect on species presence/absence and there was no interaction between salinity and season. In the final model, salinity had a statistically significant effect on species presence/absence (Table 10). Therefore, the effect of salinity on species presence/absence between the two seasons was consistent. In addition, salinity had a positive effect on the incidence of the Estuarine Species

sand goby (*Pomatoschistus minutus*) and a negative effect on the incidence of the Freshwater species chub (*Squalius cephalus*; Table 11). Model residuals for GLLVM and multivariate GLM were comparable to the previous models (supplementary figure 13 and 14).

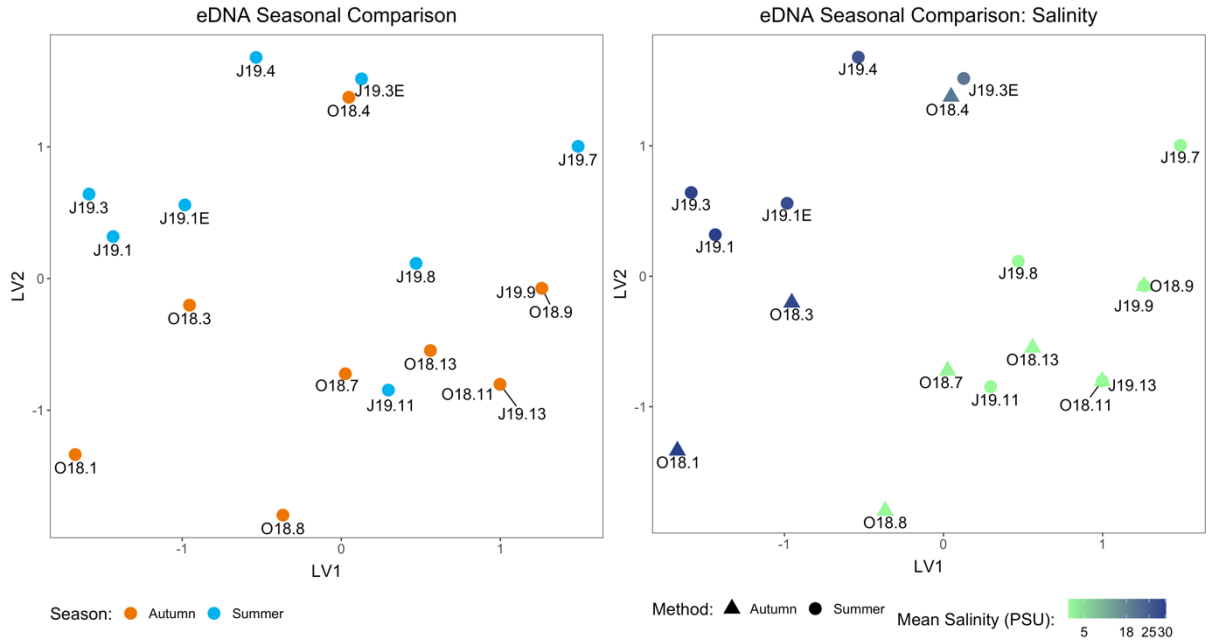


Figure 10: Presence/absence of fish species per station as modelled using a binomial GLLVM (log link; two latent variables; 50 iteration) for eDNA (filtered by TraC Fish detections 2002-2017) in October 2018 and June 2019. Left panel shows season, and the right panel shows salinity at each station (PSU).

Table 10: ANOVA for multivariate GLM – species presence/absence

AIC: 297			
Final Model: Species Presence/Absence ~ Salinity			
Explanatory Variable	**Residual DF**	**Wald-Test**	**P-Value**
Intercept	15		
Salinity	14	6.683	$< 2 \times 10^{-16}$ ***

Significance codes: *** < 0.001 ** < 0.01 * < 0.05

Table 11: Species level ANOVA results and model coefficients for species with a statistically significant association (p > 0.05; adjusted for multiple testing) for salinity.

	Salinity		
Species (Guild)	Wald Stat.	*p*-value	Coefficient
Pomatoschistus minutus (ES)	2.425	0.007	0.132
Squalius cephalus (F)	2.306	0.016	-0.130

Significance codes: *** < 0.001 ** < 0.01 * < 0.05

4. Discussion

4.1 Overview

This study provides a comprehensive comparison of the Teleost fish assemblage detected via eDNA metabarcoding of surface water samples and conventional fishing gears in a macrotidal estuary. The study primarily focused on species which had previously been detected by prior fish surveys (from 2002 to 2017). This gave confidence that the species detected with eDNA had the *potential* to occur in the estuary, rather than being the result of eDNA transport alone. Regardless of this conservative approach, it was possible to resolve clear ecologically relevant spatial patterns in assemblage composition and determine the differences between the assemblage composition detected by eDNA and fishing gears. Overall, eDNA detected a greater species richness, per 30 samples, than seine or fyke nets and eDNA showed a different assemblage composition to seine nets. Additionally, there was a clear correlation between salinity and assemblage composition, which was consistent between eDNA and fishing gears, and was consistent over seasons. Comparably there were no significant changes in assemblage composition between seasons. These results are comparable to those found in previous comparative studies of eDNA and conventional methods in estuaries (Zou *et al.*, 2020; Cole *et al.*, 2022), particularly work conducted in the tidal Thames (UK, Hallam *et al.*, 2021)

4.2 Detection and Composition between eDNA and Fishing Gears

Overall, 17 of the 26 species (71%) detected in fishing gears in autumn were detected by eDNA, including the 11 most abundant. This was comparable to coverage in the tidal Thames, where 13 out of 18 species (72%) detected by fishing gears were detected using a 12S rRNA and a CO1 marker (Hallam *et al.*, 2021). At present, the greatest coverage of fish species would be provided by a combination of eDNA and sampling using fishing gears, as with previous studies in estuaries (Zou *et al.*, 2020; Hallam *et al.*, 2021; Cole *et al.*, 2022). Species coverage in the present study would have risen without the specific contamination threshold used, e.g. Nilsson's pipefish (*Syngnathus rostellatu*s) was removed. In addition, two species detected by fishing gears not detected in October 2018 were detected in June 2019: lesser weever (*Echiichthys vipera*) and common dragonet (*Callionymus lyra*; prior to data cleaning). Failure to detect these species in October maybe because samples were stored at - 20°C for 12 months prior to DNA extraction, compared to 6 months in June, due to challenges associated with method development, or stochastic factors. In addition, a further two species not detected by eDNA in October 2018: pogge (*Agonus cataphractus*) and sand smelt (*Atherina presbyter*) were absent from the reference database at the time of analysis. Therefore, it is possible that species detections would have risen with further, ongoing, development of the refence database.

Metabarcoding of eDNA detected more species (previously detected in TraC fish data) in the estuary overall, than seine and fyke nets, per 30 samples. This was also the case for eDNA compared to fyke nets when only species detected in October were retained in the eDNA data. Comparisons between asymptotic species richness showed more limited statistically significant differences, probably due to the high bias associated with extrapolating to the asymptote (Hsieh *et al.*, 2016). However, eDNA metabarcoding is clearly a more sensitive method for detecting differences in species richness at the estuary level than seine, and particularly fyke nets. Therefore, there is broad support for the hypothesis that eDNA will detect more species in the estuary overall than each gear type, except for beam trawls. In addition, eDNA richness estimates approached an asymptote. Therefore, eDNA may be particularly useful for calculating metrics for fishes in estuaries that rely on species richness estimates and are

particularly sensitive to sampling effort (Gamito *et al.*, 2012). The demonstrated sensitivity to species richness is comparable to other studies in estuaries where eDNA generally detects greater species richness, per site, than conventional methods (Zou *et al.*, 2020; Hallam *et al.*, 2021; Cole *et al.*, 2022). Although, in the tidal Thames no statistically significant differences in eDNA and fishing gears in species richness estimates at the level of the estuary (CI overlapping) were detected. This was probably because data from the multimethod technique was aggregated rather than each gear type being compared individually (Hallam *et al.*, 2021). It is arguable that in the present study, eDNA samples should also have been aggregated for each site when compared to the fishing gears. However, this was not required as the aim of this particular analysis was to compare different methodologies at the estuary scale, on a sample per sample basis, rather than explicitly test hypotheses with a spatial component.

Comparing assemblage spatial composition with fishing gears, metabarcoding of eDNA showed a different assemblage composition to that detected with seine nets. No difference was detected in the less well replicated comparison with all gear types, probably due to a reduced sample size. Therefore, there is partial support for the hypothesis that eDNA would show a different assemblage composition compared to fishing gears. Differences in assemblage composition between eDNA and seine nets is comparable to other studies that have shown different compositions between a multimethod netting technique (Hallam *et al.*, 2021) and BRUVs (Cole *et al.*, 2022) in estuaries. The species which were detected less frequently with the seine netting method: three-spined stickleback (*Gasterosteus aculeatus*), common dace (*Leuciscus leuciscus*) and European perch (*Perca fluviatilis*), were not completely absent from seine nets. The latter suggests that the greater detection of these species by eDNA was being driven by the greater detection probability of eDNA and potential transport within the estuary, rather than potentially spurious detections in the eDNA alone from transport into the estuary.

4.3 Assemblage Composition and Salinity

Exploration of spatial changes in assemblage composition showed that in both seasons, regardless of differences in design, it was possible to detect a clear shift in assemblage composition along the estuary. This is remarkable given the river was in flood during both surveys (supplementary figure 15 and 16). In October the three sub-assemblages that showed a change in assemblage composition over space, also exhibited overlapping indicator species, with a similar pattern in June. Two Marine Migrant species, *M. merlangus* and *D. labrax,* were associated with the assemblages in the lower estuary. Whereas common roach (*R. rutilus*) was associated with the assemblage in the upper estuary and one of the assemblages in the lower estuary. This supports the growing body of evidence that despite the impact of eDNA persistence and transport, localised patterns in fish assemblages can be detected within estuaries using eDNA (García-Machado *et al.*, 2021; Hallam *et al.*, 2021; Cole *et al.*, 2022) as well as more generally in marine assemblages (Port *et al.*, 2016; Yamamoto *et al.*, 2017; Jeunen *et al.*, 2019). In this study it is likely that eDNA transport influenced the results to some degree given its prevalence in aquatic systems (Deiner and Altermatt, 2014; Shaw *et al.*, 2016; Yamamoto *et al.*, 2017). This study assumed that eDNA of species not detected previously in TraC Fish data from 2002 to 2017 had been transported into the estuary. This appears to be a valid, if simplistic, assumption, given that all the previously undetected taxa identified by eDNA to species level were freshwater species rather than being a random assortment across different estuarine use guilds. That being said, these detections may be useful in other contexts, such as the study of changes in species distributions due to global climate change.

In addition, comparisons between the assemblage composition of eDNA and fishing gears showed a consistent correlation with salinity, regardless of the methodology, therefore

supporting the initial hypothesis relating to this. Focusing on the comparison between seine nets and eDNA, four Marine Migrant and one Estuarine Species were detected more frequently at higher salinities. This included *M. merlangus* and *D. labrax*, previously identified as the indicator species for the eDNA assemblages in the lower estuary. All five species were absent in the upper estuary, where salinity was below 1. These results support those reported from Japanese estuaries where the proportion of marine species detected increased, compared to freshwater and brackish species, with increasing salinity (Ahn *et al.*, 2020). In addition, salinity was a consistent correlate with the assemblage composition in both seasons. As stated previously, salinity is a key environmental variable influencing the structure of fish assemblages (Selleslagh *et al.*, 2009; Whitfield *et al.*, 2012). Salinity influences the distribution of fish through their physiological salinity tolerance, although the range of salinities in which fish are habitually found is often narrower than their tolerance range (Marshall and Elliott, 1998). In the context of eDNA detections, the correlation with salinity could also be due in part to changes in the relative quantities of transported eDNA from the marine environment and the river. It is notable that no freshwater species showed a species level correlation with salinity, potentially due to downstream transport obscuring these associations.

4.4 Comparison between Seasons

Compared to spatial changes in assemblage composition, a direct comparison between the assemblage composition of eight stations in autumn and summer showed no difference in assemblage composition, as was also found by Hallam *et al.* (2021). This contrasts with other eDNA studies which have been able to detect seasonal changes in the assemblage composition within estuaries (Stoeckle *et al.*, 2017; Zou *et al.*, 2020) and in coastal fish assemblages more generally (Sigsgaard *et al.*, 2017). In addition, there was no difference in the estimated asymptotic species richness at the estuary level between either season, comparable to the observation that there were no statistically significant differences in species richness between winter and summer in the tidal Thames (Hallam *et al.*, 2021). Seasonal changes in the structure of the fish fauna in temperature estuaries are a well-established phenomenon (Maes *et al.*, 2005; Henderson and Bird, 2010; Selleslagh *et al.*, 2012). Consistent seasonal changes in the fish fauna are caused by sequential immigration and emigration of marine, freshwater, estuarine and diadromous species, probably controlled by the spawning times and the time needed for larval and juvenile stages to recruit into the estuary (Maes *et al.*, 2005). However, given that much of these seasonal changes are related to changes in abundance it maybe that a presence/absence analysis using a relatively sensitive survey method is not the best way to assess seasonal variation. The design employed in the current study was also not optimal for assessing seasonal changes in assemblage composition. Given that there were effectively only two temporal replicates in the analysis (two seasons) a design with higher levels of temporal replication as in Sigsgaard *et al.* (2017) or Stoeckle *et al.* (2017) would have been more appropriate. In addition, the threshold cut-off to remove contamination applied across species may have smoothed out any differences in assemblage composition between the two seasons. Sampled water volume also varied between the two surveys, ~ 0.5 L in Autumn and ~ 1 L in June which could also have affected the results. Further study and method development is required in this area.

4.5 Methodological Caveats and Considerations

Despite the clear results and aside from those already discussed, there are a number of methodological caveats and considerations. Overall, the total yield of fish reads in the study from target species was low, at 6.2 % of the total reads, when compared to other such studies using Tele02 primers (Aglieri *et al.*, 2021; Zhang *et al.*, 2020). However, it was shown that

read depth had a limited effect on assemblage composition via species rarefaction curves and model selection of GLLVM ordinations. In addition, several samples dropped out of the analysis due to poor sequencing and other operational constraints, but this does not appear to have influenced the overall results. Finally, contamination was detected in this study, field contamination in October 2018 was probably due to working alongside a fish survey whereas lab contamination was likely operator error. Contamination was addressed using a threshold cut off, a standard approach in eDNA metabarcoding (Sepulveda *et al.*, 2020) using an adaption of the methodology in Yamamoto *et al.* (2017). The cut off did have some influence on the results, particularly in the detection of rare species. Whereas, most of the species found in the blanks were not omitted as they were of ecological importance and the results were ecologically coherent. Regarding sampling design, samples were taken only from the surface water, however, given the shallow depths of the sampled area and the generally well-mixed character of the estuary in the areas sampled (Bolaños *et al.*, 2013) it seems unlikely substantial gradients in eDNA would be present. Finally, regarding the statistical analysis using multivariate GLMs, some structure in the residuals was present. Given that the models were correctly specified for presence/absence data it have been that unmeasured covariates or mixed effects should have been added (Zuur *et al.*, 2007). Clearly, using site as a random variable in a mixed effects model would have been more appropriate. However, multivariate GLMs incorporating random effects (Wang *et al.*, 2012) have not been developed yet to the authors' knowledge. Finally, this study has not assessed the ability of metabarcoding to provide semi-quantitative data on species abundance via relative quantification.

4.6 Implications for Management

This study has focused on providing direct comparisons of species richness and assemblage composition between eDNA and fishing gears. Further research should build on these findings to devise ways to calculate eDNA-based metrics for the EMFI (Harrison & Kelly, 2013) or other indices of interest (Coates *et al.*, 2007; Delpech *et al.*, 2010). For example, eight of the 14 metrics in the EMFI use varying measures of species richness (Harrison & Kelly, 2013) and therefore should be straightforward to calculate using eDNA data. In addition, further surveys should be conducted on other estuaries which are known to show variation in anthropogenetic impacts, and the health of the fish assemblage, to assess if eDNA reflects this variance. However, the current research does have some important implications for survey design in an applied setting. Firstly, eDNA sampling in estuaries should have a spatially replicated design to take account of spatial variation in the fish assemblage. Secondly, future surveys should record salinity, and other physicochemical parameters, at each sampling station to contextualise the results and provide valid comparisons, this is already best practice in fish surveys (Elliott & Hemingway, 2002). Thirdly, any eDNA detections should be scrutinised to determine if they are likely to occur in the ecosystem before species are included in bioassessments of an estuary's health. The present study used a liberal approach for species filtering, given only a single detection was required in the TraC fish data from 2002 to 2017 to include the species in further analysis. A more rigorous approach should be used focusing on species over a standardised time-series of baseline data for assessment, particularly when comparing between ecosystems. Expert judgment would then be required to incorporate any additional species detected by the eDNA analysis, which were not detected in TraC fish surveys. In addition, careful consideration should be given to the inclusion of species which are widely eaten, e.g. Atlantic salmon (*S. salar*), that could result from erroneous detections from wastewater effluent.

5. Conclusion

Overall, it can be concluded that eDNA metabarcoding is an effective way to assess the biodiversity of fishes in large, macrotidal estuaries. Although care needs to be taken when interpreting the data yielded, it is likely with further study that this technology will become a complementary method to more conventional approaches to assess the health of fish assemblages in transitional waters.

Acknowledgements

T. I Gibson would like to thank and acknowledge the following individuals for their invaluable help on this project. Firstly, Mark Kyriacou, Adam Cooper, Adam Lyshon and the rest of the NRW staff and survey teams. Dr Trevor Harrison and the rest of the DAERA survey team. Ian Marland of Marland Marine Services Ltd. All these individuals contributed to or facilitated fielded work. Dr Rupert Collins for his excellent bioinformatic advice and providing clear examples of complete bioinformatic pipelines via Github and Dr Molly Czachur for providing equivalent data to practice on. Finally, Dr Graham Sellers, Dr Robert Donnelly, Dr Charles Baillie, Dr Georgina Brennan and Kirthana Pillay for their excellent advice on molecular biological matters at an early stage of this project. In addition, Dr Marco Distaso for carrying out a professional job on the sequencing. This work would not have been possible without the support of the KESS2 staff.

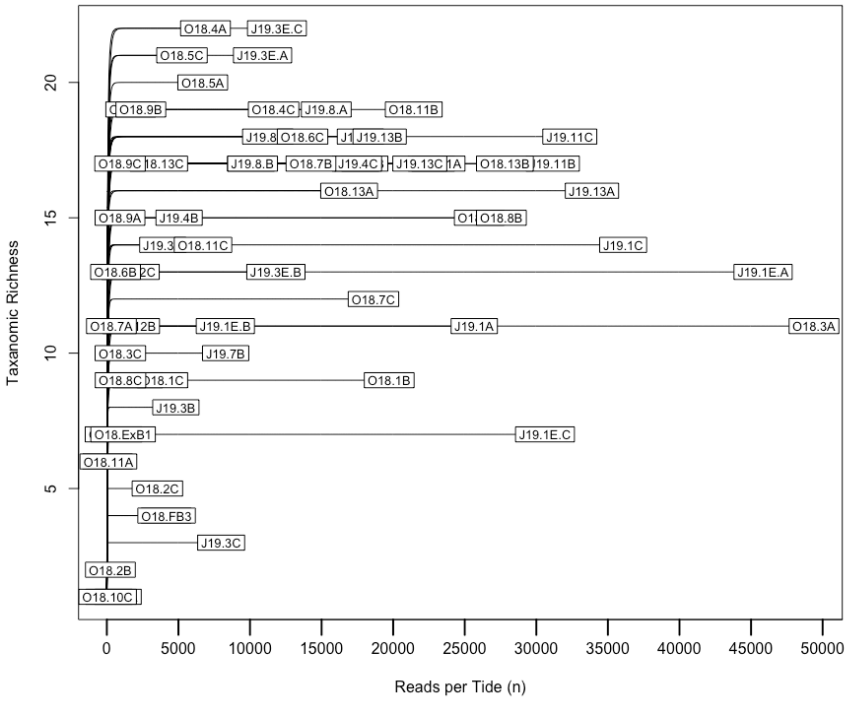
Chapter 2: Supplementary Material

TraC Fish Surveys

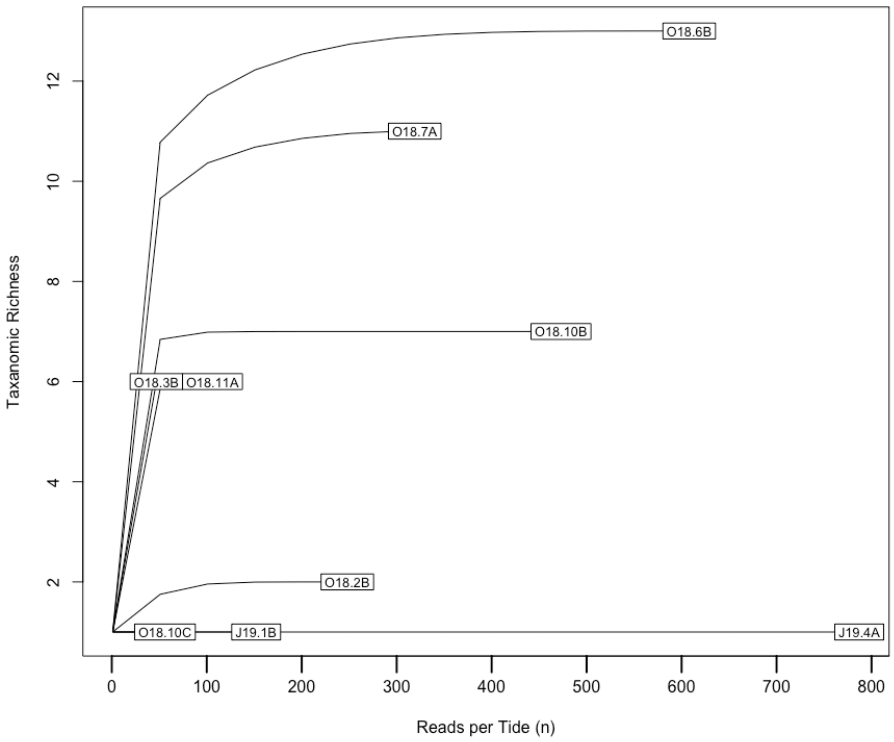
The majority of fish sampling from 2002 to 2017 was conducted by Natural Resources Wales and the Welsh Environment Agency to provide data for the Transitional Fish Classification Index (TFCI; Coates *et al*., 2007). The most consistent fish sampling occurred in spring/summer and autumn in 2010 to 2015, and in 2016 when only sampling in autumn was conducted. This is broadly based on the methodology from Colclough *et al.* (2002). The fishing gear used were seine nets (43 m long by 4 m deep, with a 6.5 mm knotless mesh centre panel and 14 mm knotless wings), double fyke nets (each trap 0.5 m high, 2.5 m long with a 10 mm mesh cod-end, and joined by a 6 m long x 15 mm mesh ladder) and 1.5 m wide beam trawls (net is 4 m long, 1.2 m cod end, with a main knotless mesh size of 20 mm, and a cod-end mesh size of 8 mm knotless mesh; M. Kyriacou and T. Gray, *pers. comm.*). From 2010 to 2016 sampling was conducted at, generally, eight stations in a rough transect from the bottom of the estuary at Talacre to the head of the estuary at Saltney. Seine nets were deployed at all eight stations with two hauls per sampling station. In addition, less regular beam trawling and fyke netting occurred at four and one station(s) respectively, with one sample per station. Additional otter trawling (gear specification not known) was conducted in the outer estuary at two stations.

Supplementary Table 1: Number of paired-end reads at each step of bioinformatic pipeline.

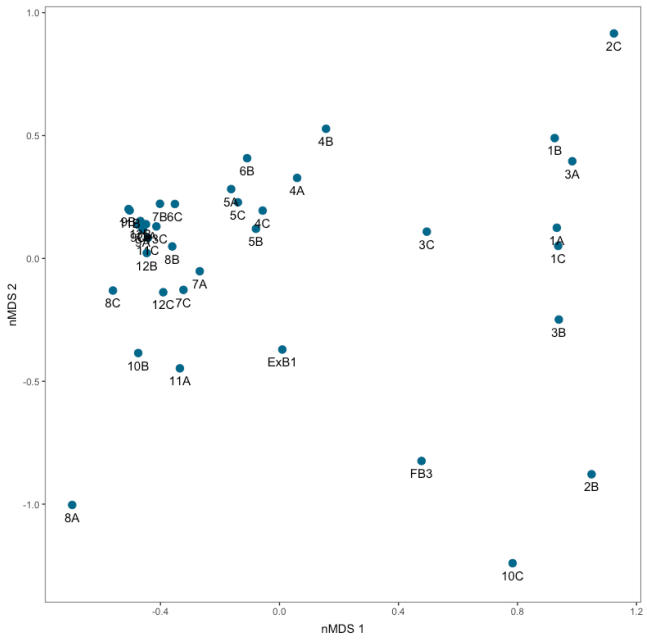
Step Number	Number of Paired-End Reads	Pipeline Step	Proportion of Total Reads (%)
1	13409393	MiSeq Sequencing Run	100.0
2	12635073	Primers Removal	94.2
3	6344589	Quality Filtering	47.3
4	6321528	Denoising	47.1
5	5742819	Merging Paired-End Reads	42.8
6	5468324	Chimera Removal	40.8
7	5338905	Reads Matched to nt BLAST (All ASVs)	39.8
8	5321787	Reads Matched to nt BLAST (All Chordates)	39.7
9	1595220	Reads Matched to Meta-Fish-Lib (All Data)	11.9
10	828499	Reads Matched to Meta-Fish-Lib (No Positive Controls)	6.2



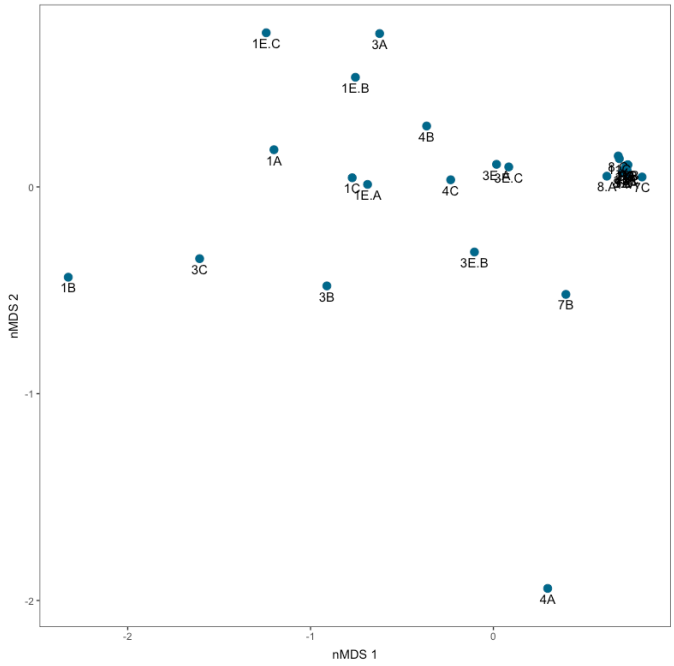
Supplementary Figure 1: Rarefaction curve of fish taxonomic richness (≈ species richness) against read depth for fish reads for each individual sample, following the 0.055% read contribution cut-off. This creates a flat profile relative to the raw data (not shown) due to the removal of rare species. Calculated using rarecurve(step = 50) in *Vegan* (Oksanen *et al.*, 2015).



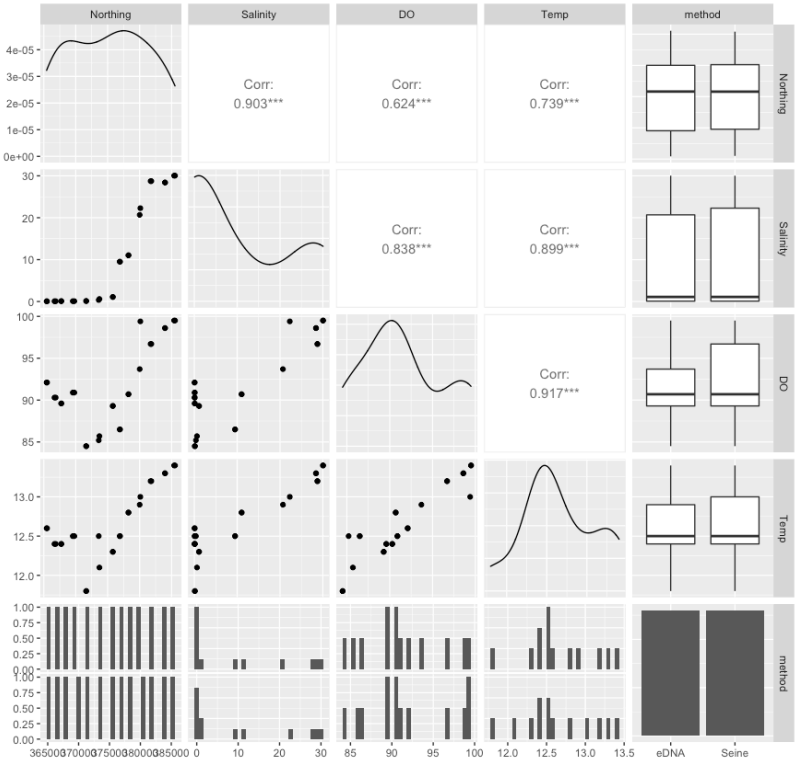
Supplementary Figure 2: Rarefaction curve of fish taxonomic richness (≈ species richness) against read depth, for samples following the 0.055% cut-off, showing only samples with below 800. Calculated using rarecurve (step = 50) in *Vegan* (Oksanen *et al.*, 2015).



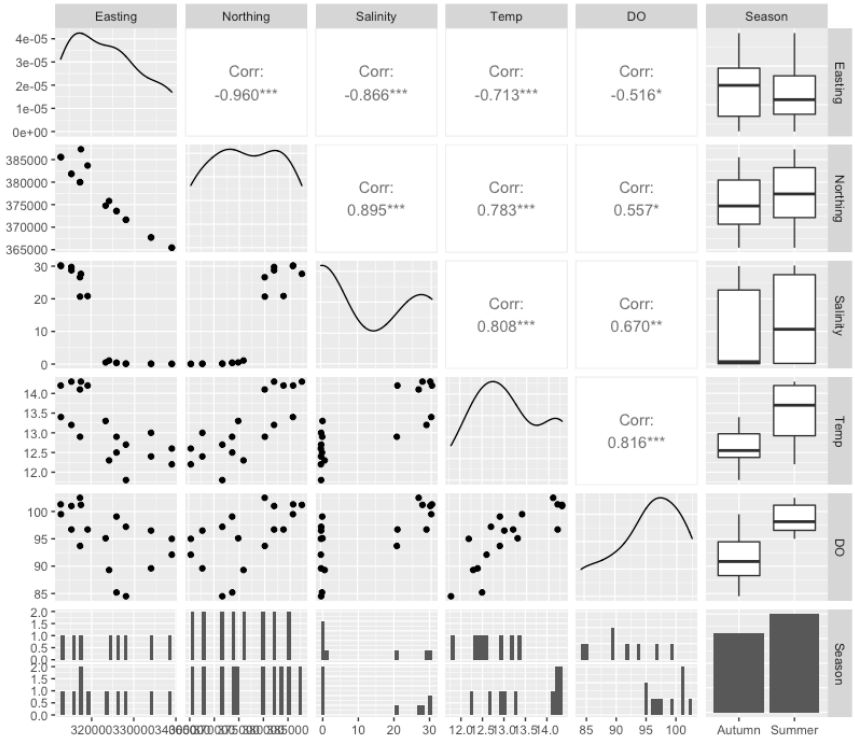
Supplementary Figure 3: nMDS ordination for the October 2018 data generated using Bray-Curtis distances calculated on taxonomic (≈ species) presence/absence, following the 0.055% read contribution cut-off. This shows the position of the contaminated blanks: ExB1 and FB3 (centre and bottom left) and the position of the samples 10C and 2B (bottom left) sequenced poorly. All of these were omitted from the final analysis.



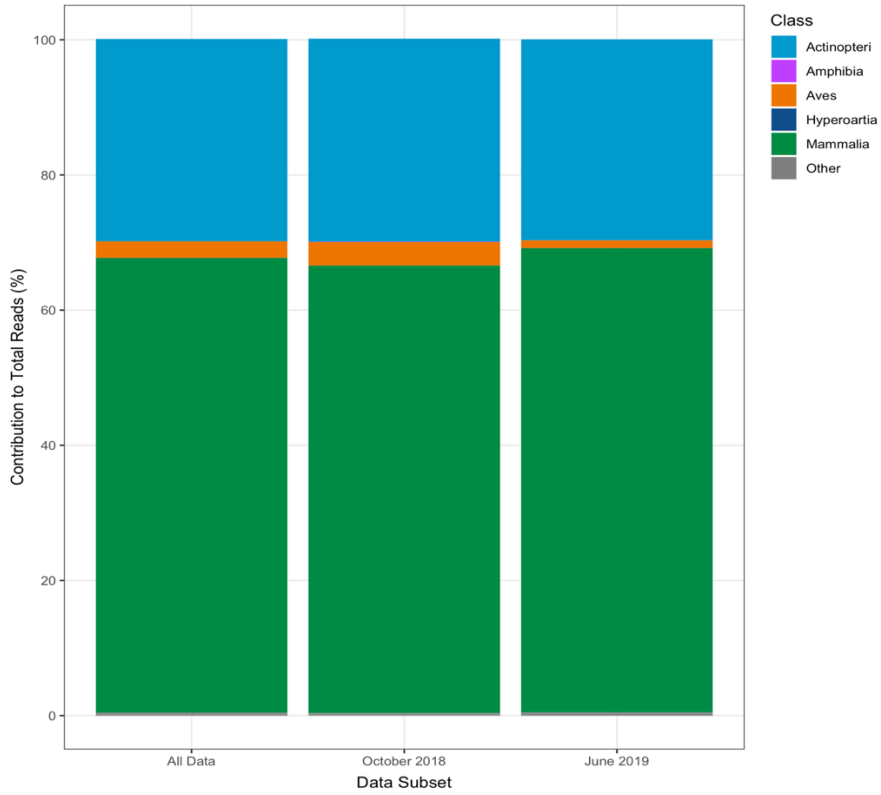
Supplementary Figure 4: nMDS ordination for the June 2019 data generated using Bray-Curtis distances calculated on taxonomic (≈ species) presence/absence, following the 0.055% read contribution cut-off. This shows position of the samples which sequenced poorly, J19.1B (far left) and J19.4A (bottom left). Both these were omitted from the final data analysis.



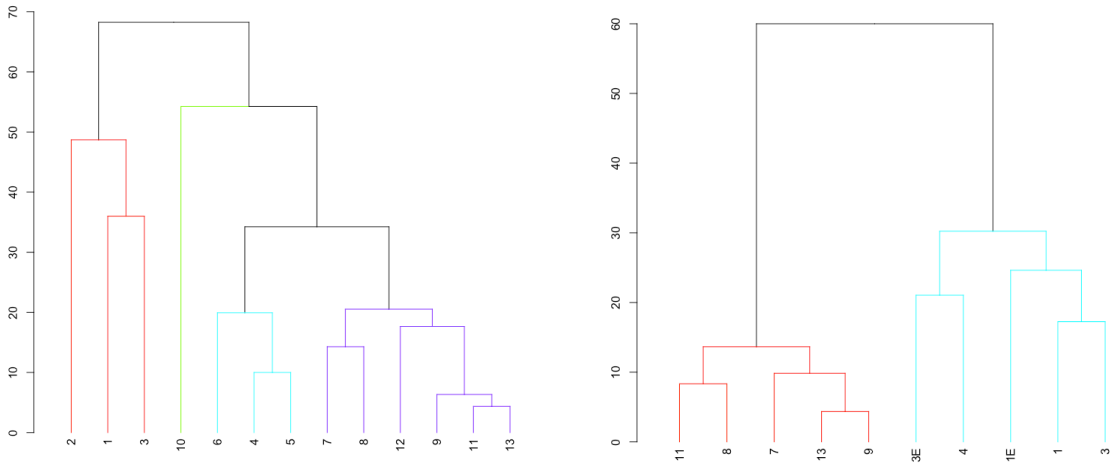
Supplementary Figure 5: Correlations between each combination of environmental variables for the comparison between eDNA and seine net stations in October 2018, with Pearson correlation coefficients calculated between continuous variables.



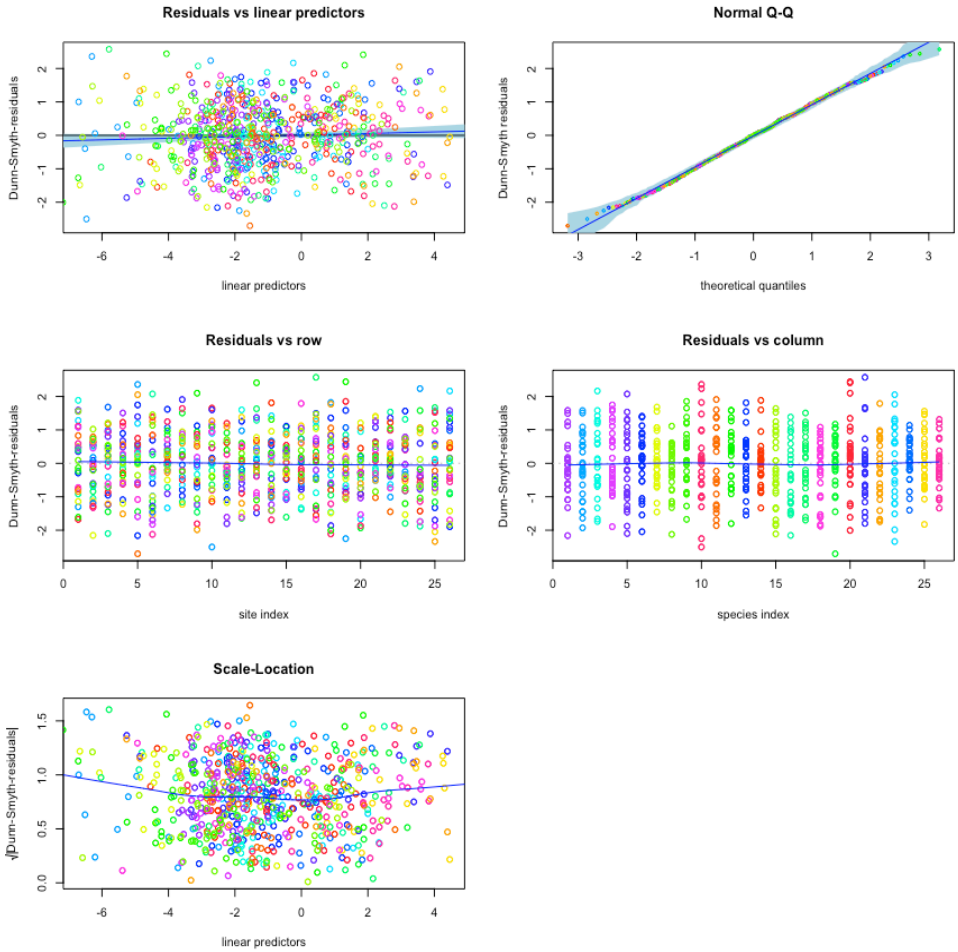
Supplementary Figure 6: Correlations between each combination of environmental variables for the comparison between eDNA stations in October 2018 and June 2019, with Pearson correlation coefficients calculated between continuous variables.



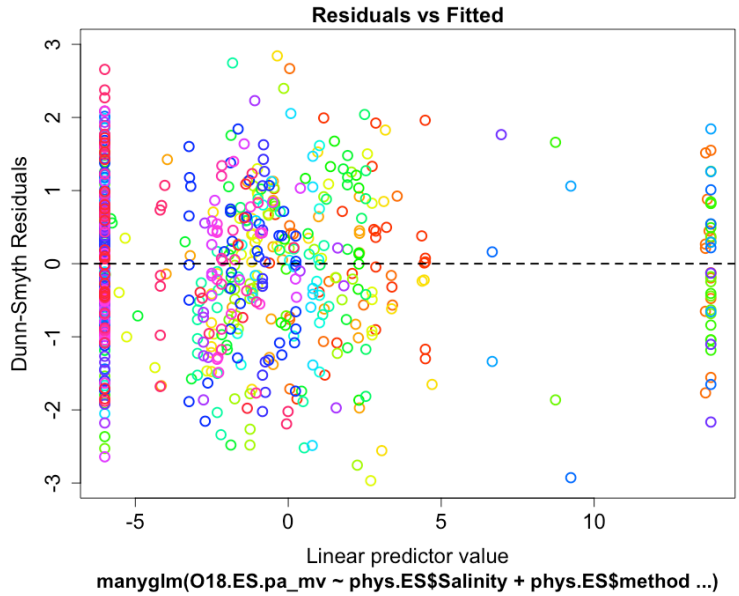
Supplementary Figure 7: Overall percentages of reads belonging to each Chordate Phyla (all other Phyla shown as Other) present in samples after a global BLAST search against the "nt" BLAST database (Step 7. in supplementary table 1.). Taxonomy assigned according to highest scoring BLAST hit.



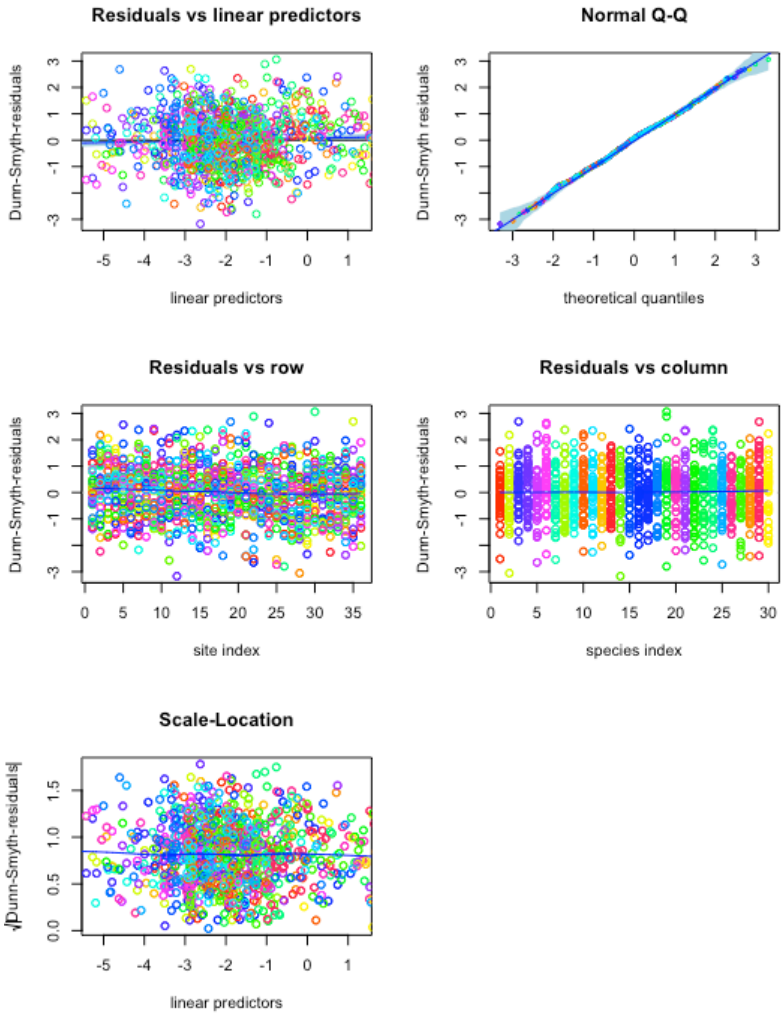
Supplementary Figure 8: Separate SIMPROF and HAC (hierarchical agglomerative clustering) analyses for the cleaned October 2018 and June 2019 dataset, including only species previously detected in TraC Fish data. X-axis indicates percentage dissimilarity, colours indicate groupings of stations which contain no internal group structure in species incidences. Analysis was conducted on species presence/absence, converted to Bray-Curtis dissimilarities. HAC was conducted with the group average linkage. A $p = < 0.001$ threshold was used due to the multiple testing inherent in this method (Clarke *et al.*, 2008). The analysis was conducted using the 'clustig' package in R (Whitaker and Christman, 2015).



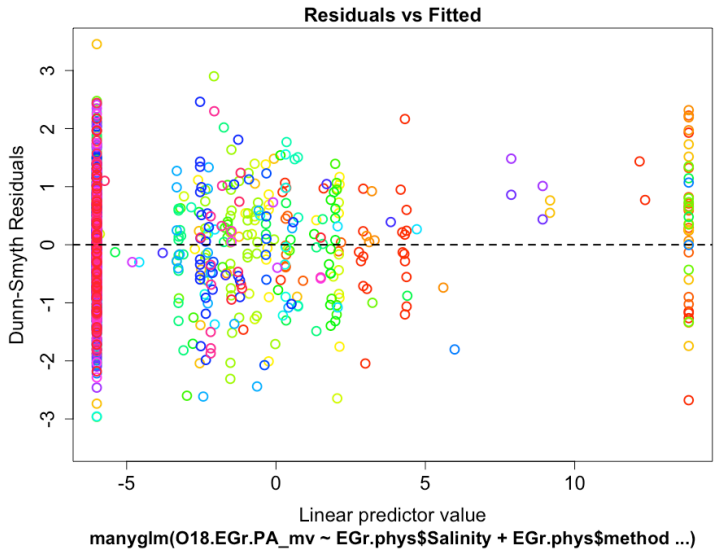
Supplementary Figure 9: Residuals plots for the binomial GLLVM of presence/absences (probit link; two latent variables; 50 iterations) for the comparison between eDNA and seine nets.



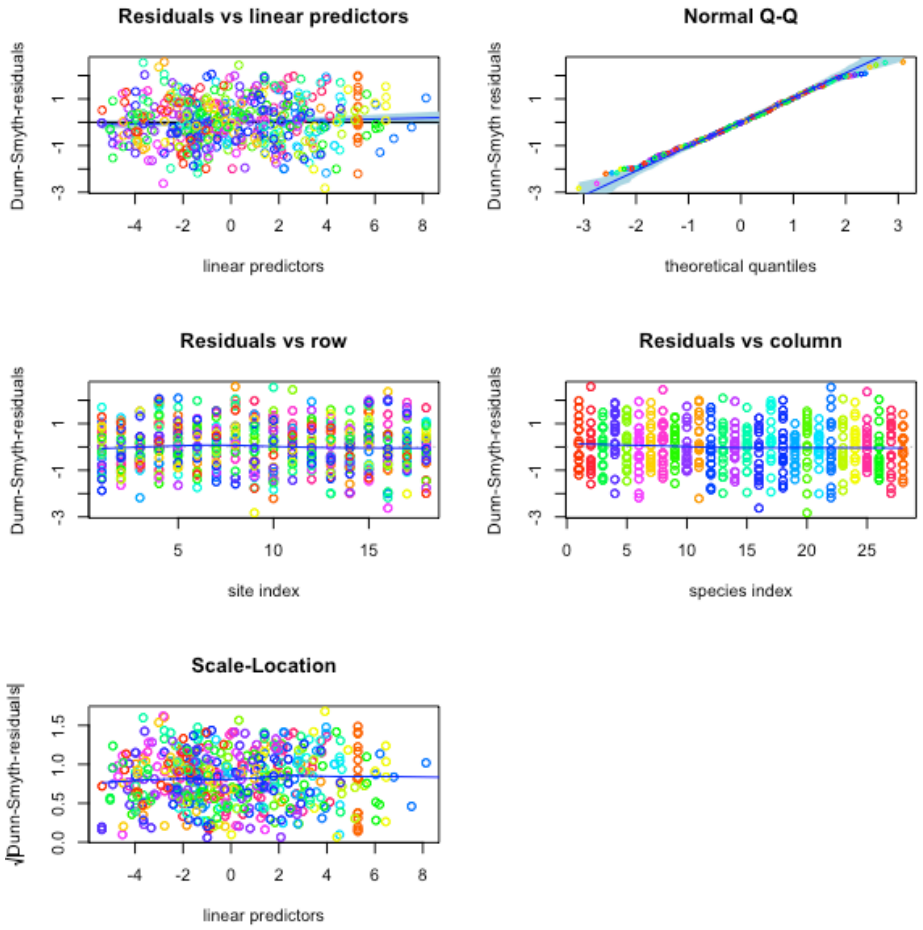
Supplementary Figure 10: Residuals plots for the binomial multivariate GLM of presence/absences (logit link) for the comparison between eDNA and seine nets.



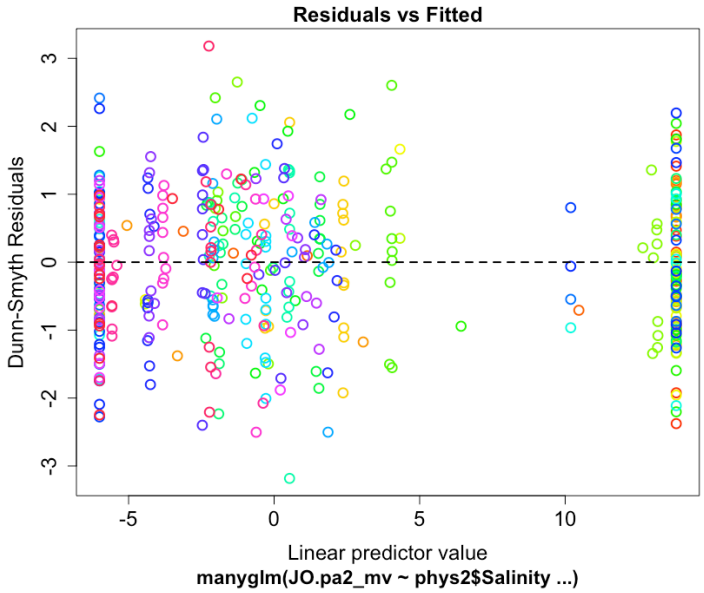
Supplementary Figure 11: Residuals plots for the binomial GLLVM of presence/absences (probit link; two latent variables; 50 iterations) for the comparison between eDNA and all gears.



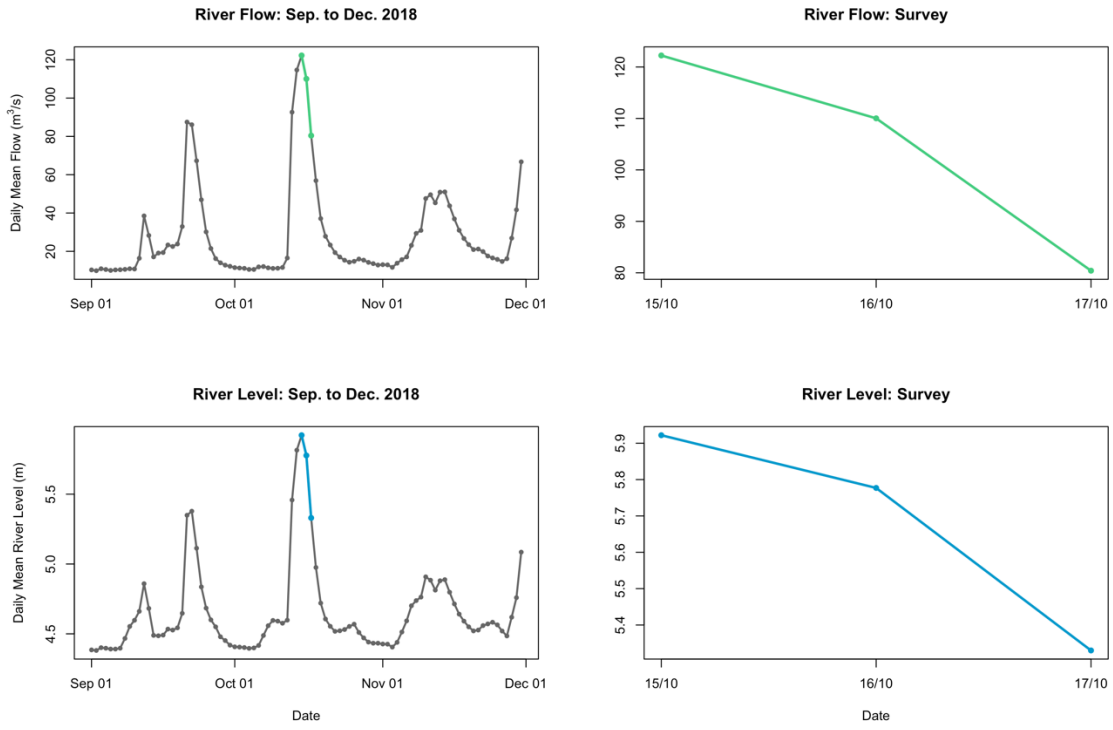
Supplementary Figure 12: Residuals plots for the binomial GLLVM of presence/absences (probit link; two latent variables; 50 iterations) for the comparison between eDNA and all gears.



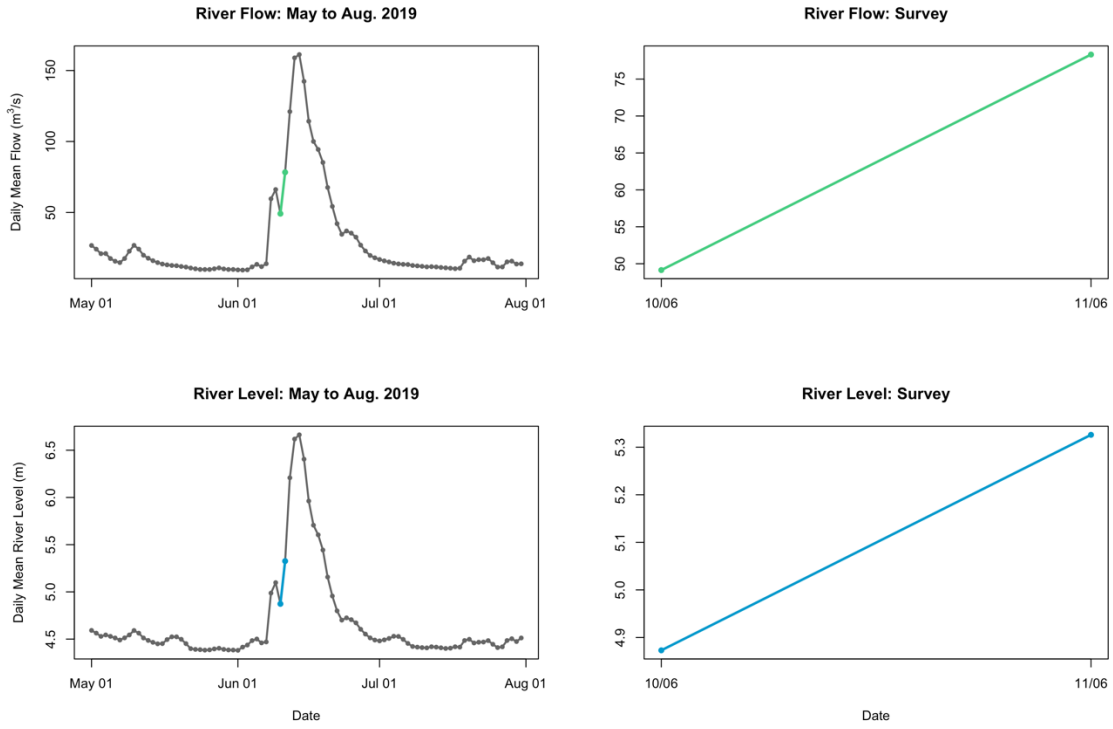
Supplementary Figure 13: Residuals plots for the binomial GLLVM of presence/absences (probit link; two latent variables; 50 iterations) for the seasonal comparison.



Supplementary Figure 14: Residuals plots for the binomial GLLVM of presence/absences (probit link; two latent variables; 50 iterations) for the seasonal comparison.



Supplementary Figure 15: Daily mean river flow (m^3/s) and daily mean river level (m) at Iron Bridge (Dee River; Lat: 53.133887; Long: -2.870553), from Sept to Dec 2018 and from over the survey period (15 to 17 Oct 2018, Flow: Green; River Level: Blue). Data source: NRW, data request.



Supplementary Figure 16: Daily mean river flow (m^3/s) and daily mean river level (m) at Iron Bridge (Dee River; Lat: 53.133887; Long: -2.870553) from May to Aug 2018 and from over the survey period (10 to 11 Jun 2019, Flow: Green; River Level: Blue). Data source: NRW, data request.

Chapter 2: Appendix A
Refining the Spens and Evans Methodology to Extract Amplifiable Fish eDNA from Estuarine Water

Author Contributions

Project Phase	Author
Study Design	TG, SC
Lab Work	TG
Write-Up	TG
Review	AE

1. Introduction

Extraction of DNA from environmental samples is an essential component of any environmental DNA (eDNA) analysis workflow. Silica-based commercial kits are commonly used in fish eDNA studies (Hänfling *et al.*, 2016; Port *et al.*, 2016; Shaw *et al.*, 2016). Following sample lysis DNA is bound onto a silica matrix under chaotrophic conditions. DNA remains within the matrix while other biopolymers, e.g. RNA, are removed using solvents. DNA is then recovered by rehydration using an aqueous buffer (Green and Sambrook, 2012). Specifically, the DNeasy Blood and Tissue kit (QIAGEN, Hilden, Germany) has been co-opted to extract eDNA from filters in several studies of marine (Thomsen *et al.*, 2012; Sigsgaard *et al.*, 2017; Yamamoto *et al.*, 2017), freshwater (Nakagawa *et al.*, 2018) and estuarine fish (Hallam *et al.*, 2021). The initial Proteinase K and ATL buffer lysis from this kit can be combined with encapsulated filters. Lysis occurs within the sealed filter capsule, reducing handling of the filter and potential contamination (Spens and Evans *et al.*, 2017). Encapsulated filters can also allow filtration and preservation in the field (Majaneva *et al.*, 2018). In this study industry standard encapsulated filters (Nature Metrics Ltd., Guildford, UK) were used. These incorporate a 5.0 µm Glass Fibre (GF) membrane on top of a 0.8 µm Polyethersulfone (PES) membrane, allowing consistent filtration of water samples with high suspended particulate matter (SPM) concentrations. However, Blood and Tissue Kits are not designed for environmental samples and contains no steps to remove environmental PCR inhibitors (QIAGEN, 2020). These are likely to be in high concentrations in estuarine water samples (Chapter 1). Therefore, the aim of the following experiments was to modify the Spens and Evans *et al.* (2017) method to reliably extract amplifiable fish eDNA filters used to process estuarine water samples. The objective was to determine a DNA extraction method and *Taq* polymerase combination which yielded consistent amplification of a fragment of the 12S rRNA gene (mean length: 172 bp), used for fish metabarcoding, amplified using the MiFish U primers (Miya *et al.*, 2015).

2.1 Experiment 1. DNA Extraction Technique

Three replicate surface water samples were collected on 15.10.2018 at a station in the middle Dee estuary ('The Grinds', NGR: SJ2907570566), alongside the eDNA survey described in Chapter 2. Samples were processed and stored as detailed in Chapter 2, 525 – 475 ml of water was filtered per sample. All filtering and extractions were conducted in the same facilities, and to the same standard, as Chapter 2. One water sample was used for each of the following treatments. One sample (TG_C) was extracted following an adaption of the Spens and Evans *et al.* (2017) protocol (Spens-Evans). Briefly, 720 µl of Buffer ATL (QIAGEN) and 80 µl Proteinase K were added into each filter and incubated overnight at 56°C to allow sample

lysis. The rest of the extraction proceeded as normal. A second sample (TG_B) was extracted identically to the first, and the extract cleaned following extraction using a OneStep PCR Inhibitor Removal kit (Zymo Research, Irvine, CA, USA). A third sample (TG_A) was extracted using Spens-Evans with the inclusion of a particle flocculation step following sample lysis. This flocculant solution was a mixture of 180 mM aluminium ammonium sulphate dodecahydrate (aq), 3% calcium chloride (aq), and 5 M Ammonium acetate (aq) in a 0.5:0.25:0.25 volume ratio (Sellers *et al.* 2018). Consumables were molecular grade where possible. The final solution was poured into a 50 ml falcon tube and sterilised using UV light in a biological safety cabinet for 2 hrs. This has been shown to render DNA unamplifiable (G. Sellers, *pers. comm*. 2018). To remove PCR inhibitors, 300 µl of flocculant solution (Sellers *et al.*, 2018) was added to ~ 1 ml of sample lysate, vortexed and incubated for ~ 1 hr at 4°C in the refrigerator (G. Sellers and R. Donnelly, *pers. comm.*). Each sample was then centrifuged at 10,000 x g for 2 min and 1200 µl of the supernatant removed (Sellers *et al.*, 2018). The rest of the DNA extraction followed Spens and Evans *et al*. (2017). See Chapter 2 Appendix B for details. An extraction blank (ExNC) was added for the Spens-Evans + Flocculent extraction as this treatment was at highest risk of contamination, given the flocculant solution was manufactured in the eDNA lab. Following DNA extraction total DNA concentration was measured using spectrophotometry with a Nanodrop ND-1000 (Thermo Fisher Scientific Inc., Waltham, MA, USA) to confirm successful extraction of DNA (table 1).

Table 1: Sample codes and Nanodrop results for DNA extracts

Sample Code	Treatment	ng/µl	260/280	260/230
TG_A	Spens-Evans + Flocculent	35.62	1.58	0.41
TG_B	Spens-Evans + OneStep	205.20	1.88	1.64
TG_C	Spens-Evans	328.31	1.62	1.14
ExNC	Spens-Evans + Flocculent	0.43	2.66	0.08

TG_B values after OneStep cleaning.

PCRs were set up in the main post-PCR lab for ease. Contamination precautions: filter tips, benches and equipment were cleaned with 10% bleach, equipment was sterilised using 20 min irradiation with UV light, and PCR reactions were set up in individual caped tubes. In addition, the positive control was never handled alongside filter extracts, and added as the very last step. PCR assays used MiFish-U forward and reverse primers (Miya *et al.*, 2015) with the addition of a Universal Tail and Overhang (IDT, Newark, NJ, USA), which gives the amplicon a final fragment size of ~ 300 bp. Reaction set up was as follows, total reaction volume was 25 µL containing 2.5 µL of DNA extract, 12.5 µL of Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA, USA), 0.5 µL of MiFish-U-F, 0.5 µL of MiFish-U-R and 9.0 µL of PCR water was added. PCR amplifications were performed as singletons for each sample, along with a negative (2.5 µL of PCR water) and positive control (2.5 µL of PCR water of ~ 2 ng/µL *Salmo salar* DNA extract in water). The volume of sample added to each reaction was comparable to Hänfling *et al.* (2016) and Yamamoto *et al.* (2017). The thermal cycle profile was an initial denaturation for 98.0°C for 30 sec, followed by 40 cycles of denaturation at 98°C for 10 sec, annealing at 60°C for 30 sec and extension at 72.0°C at 10 sec. This was followed by a final extension at 72.0°C for 2 min. The annealing temperature, cycle number and extension time had been selected using previous experiments with two filters collected from Mostyn (in October 2018, NGR: SJ166803, as in Chapter 2) and extracted using Spens-Evans + Flocculant. These had tested annealing temperature (45, 50, 55 and 60°C, cycles: 35, extension time: 7s), cycle number (35, 40, 42, 44, annealing temp: 60°C, extension time: 7s) and extension time (7, 8, 10 and 15 sec., annealing temp.: 60°C, cycle number: 40), sequentially. The final conditions listed above were optimal for ensuring PCR amplification, analysed using gel electrophoresis, as described later (data not shown).

Due to the result of the initial round of PCRs, samples were spiked into reactions to determine if they could inhibit PCR in a positive control, and therefore were inhibited. Positive PCR reactions were set up as described above, but 2.5 µl of the PCR water volume was replaced with 2.5 µl of an eDNA sample for each reaction. Following this spiking experiment, all samples were again cleaned using OneStep kits (Zymo). Standard PCRs were then repeated with all samples, and a final round of spiking tests was to check if the final PCRs caused inhibition. PCR products were analysed by electrophoresis for 40 – 50 min at 74V on 2% agarose gels stained with SafeView Nucleic Acid Stain (NBS Biologicals, Huntingdon, UK). For each sample 5 µl of 5:3 of product to 6X Blue/Orange Loading Dye (Promega Corporation, Madison, WI, USA) was run with two 100bp DNA Ladders per gel (Promega). Success/failure of each test was scored by presence/absence of a clearly identifiable band.

2.2 Experiment 1: Results

No amplification was detected in the PCR reactions for samples of any treatment (figure 1), despite the presence of total DNA in sample extracts (table 1). The positive control amplified and the PCR and extraction blank were negative. The DNA extracts had a brownish colouration, 260/280 and 260/230 ratios also indicated contamination from proteins and other contaminates. In addition, when PCR reactions were spiked using the sample extracts, no amplification occurred (figure 2). This indicates the presence of inhibitors in the samples. Following all sample extracts being cleaned using a OneStep PCR Inhibitor Removal kit (Zymo) a band at ~ 300bp in the Spens-Evans + Flocculent treatment (TG A) was shown, both in the standard PCR reaction and the positive control spiked with that sample (figure 2). This indicated that the use of the Spens-Evans + Flocculent treatment, followed by a OneStep clean-up is the most robust option. Therefore, the consistency of this effect was investigated over multiple sites and samples in Experiment 2.

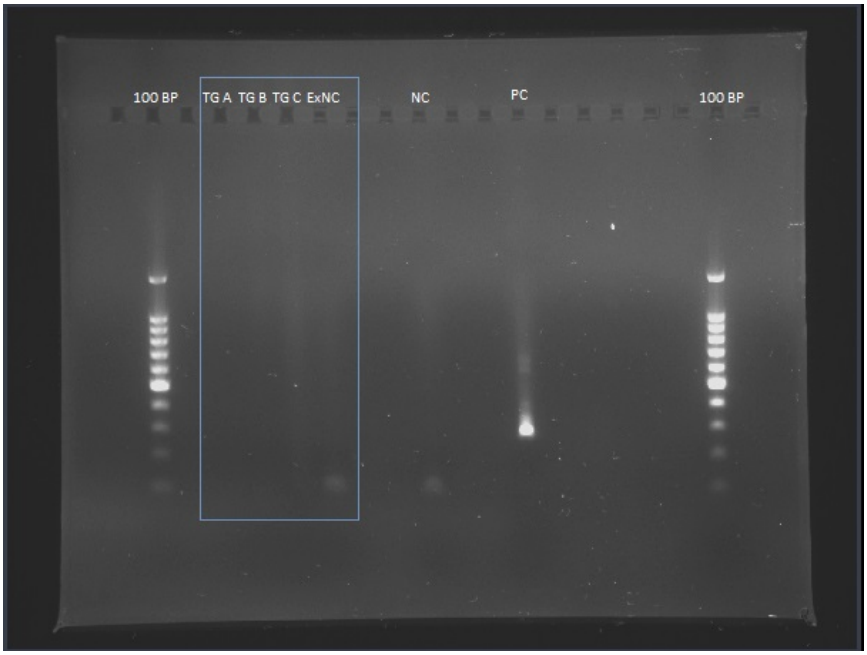


Figure 1: 2% agarose gel showing PCR products after amplification with MiFish-U primers. Products were run for 40-50 mins at 74V alongside 100 BP ladder (Promega). MiFish-U amplicon: ~ 300 bp. TG_A: Spens-Evans + Flocculent, TG_B: Spens-Evans + OneStep, TG_C: Spens-Evans, ExNC: Spens-Evans + Flocculent Extraction Blank, NC: PCR Blank, PC: Positive Control.

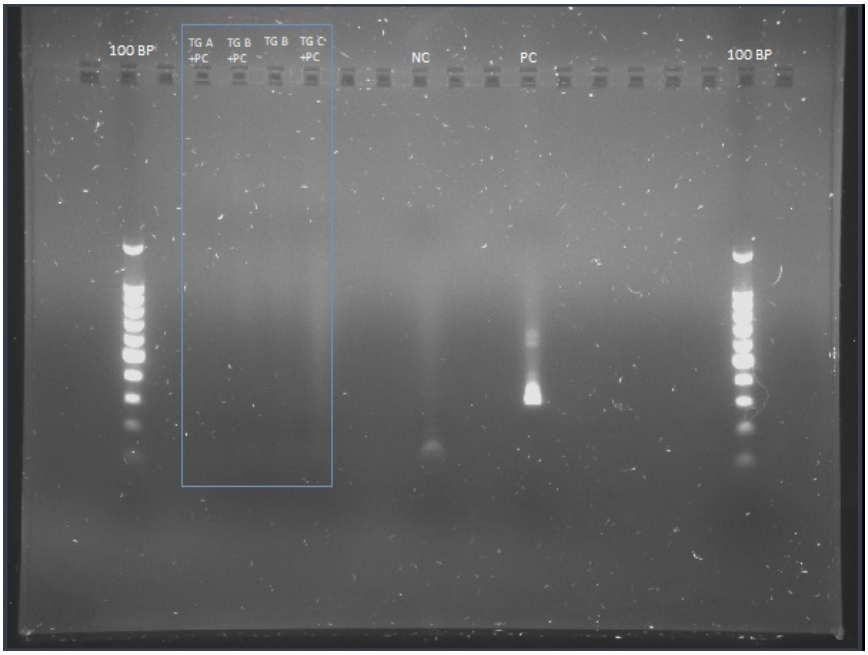


Figure 2: 2% agarose gel showing PCR positive controls, spiked with 2.5 µl of PCR extract, after amplification with MiFish-U primers (target amplicon: ~ 300 bp). Products were run for 40-50 mins at 74V alongside 100 BP ladder (Promega). +PC suffixed denotes a positive control spiked with a sample. The TG_B sample was also re-run independently to check amplification was negative in the first gel. Sample Codes, TG_A: Spens-Evans + Flocculent, TG_B: Spens-Evans + OneStep, TG_C: Spens-Evans, NC: PCR Blank, PC: Positive Control.

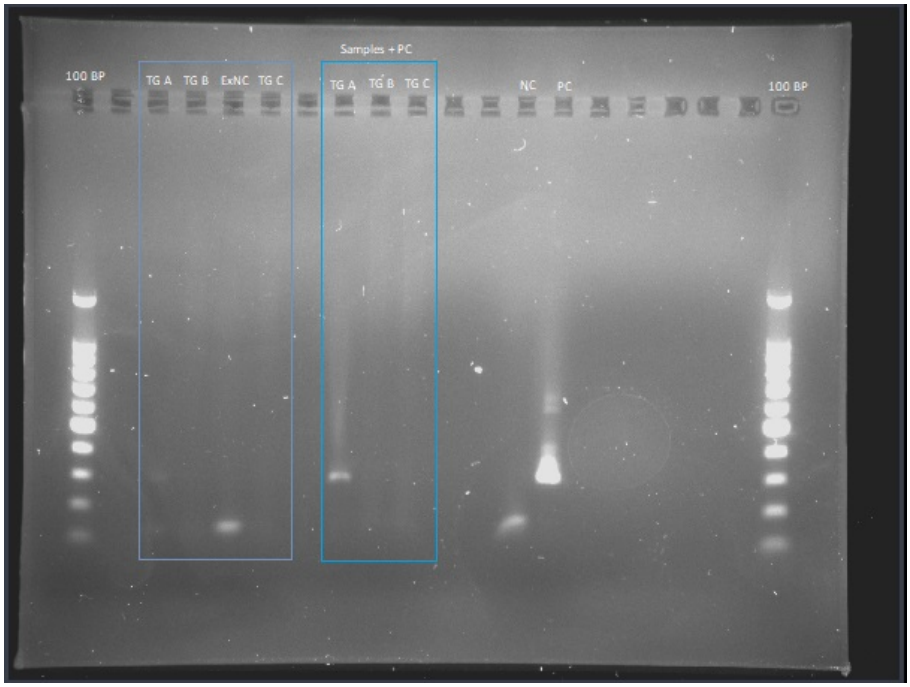


Figure 3: 2% agarose gel showing PCR products, for extracts cleaned using OneStep PCR Inhibitor Removal kit (Zymo) and positive controls spiked with those extracts, after amplification with MiFish-U primers. Products were run for 40-50 mins at 74V alongside 100 BP ladder (Promega). +PC heading denotes positive controls spiked with a sample. MiFish-U amplicon: ~ 300 bp. TG_A: Spens-Evans + Flocculent, TG_B: Spens-Evans + OneStep, TG_C: Spens-Evans, ExNC: Spens-Evans + Flocculent, NC: PCR Blank, PC: Positive Control. Samples + PC indicates positive controls where samples were spiked into them.

3.1 Experiment 2: DNA Extraction and *Taq* Polymerase Choice

To test the consistency of DNA extraction using Spens-Evans + Flocculant + OneStep clean up. Duplicate surface water samples were collected at two stations on the banks of the Dee estuary at Mostyn Docks (NGR: SJ166803, Samples: Ma + Mb), in the lower estuary and Sandy Croft (NGR: SJ336677, Samples: SCa + SCb), in the upper estuary at low tide on 23/08/19. Samples were processed as described in Chapter 2. The samples collected at Mostyn had a volume of ~ 1020 and ~ 970 mL, the samples at Sandy Croft both had a volume of ~ 980 mL. DNA Extraction was carried out using Spens-Evans + Flocculent as described above. A blank extraction was not included as it was negative in the previous experiment. 50 µl of DNA extract was cleaned using a OneStep kit (Zymo), and an aliquot of each extract was left uncleaned to provide comparison. PCR reactions were conducted as above for cleaned and uncleaned aliquots. Samples were then spiked into positive controls, as above. Following this, PCRs were repeated using a different master mix, Multiplex PCR Master Mix (QIAGEN), known to be effective with inhibited samples (A. Ellison and K. Pillay, *pers. comm.*). PCR reaction set up was 25 µL containing 2.5 µL of DNA extract, 12.5 µL of 2x Multiplex PCR Master Mix (QIAGEN), 0.5 µL of MiFish-U-F, 0.5 µL of MiFish-U-R and 9.0 µL of PCR water. Negative and positive PCR reactions were added as above. All PCR experiments were conducted in the main laboratory and analysed using gel electrophoresis, as above. The thermal cycle profile was an *Taq* activation for 95.0°C for 15 min, followed by 40 cycles of denaturation at 94°C for 30 sec, annealing at 60°C for 90 sec and extension at 72.0°C at 20 sec. This was followed by a final extension at 72.0°C for 10 min.

Table 2: Sample codes and Nanodrop results for DNA extracts

Sample Code	Treatment	ng/µl	260/280	260/230
Ma	Spens-Evans + Flocculant + OneStep	6.13	1.42	0.50
Mb	Spens-Evans + Flocculant + OneStep	4.14	1.33	1.46
SCa	Spens-Evans + Flocculant + OneStep	18.39	1.73	0.95
SCb	Spens-Evans + Flocculant + OneStep	34.00	1.98	1.72

All values after Zymo clean up

3.2 Experiment 2: Results

DNA was successfully extracted from each sample (Table 2). However, initial PCR using Q5 Master Mix found no amplification, regardless of cleaning method. Except for extremely weak amplification between 300 and 400 bp in one sample collected at Mostyn (Mb; figure 4). Surprisingly two samples from Mostyn and one from Sandy Croft did not inhibit PCR when spiked into positive controls and were therefore not inhibited (figure 5). However, after rerunning PCRs for each sample using Multiplex PCR Master Mix (QIAGEN), clear amplification was detected in both samples at Sandy Croft, regardless of cleaning method. Whereas at least one sample showed faint amplification at ~ 300 bp at Mostyn for each cleaning method (figure 6). None-specific amplification was seen at different fragment sizes other than ~ 300 bp, particularly for the two samples collected at Sandy Croft. Multiplex PCR Master Mix was able to amplify the fragment of interest from DNA extracted using the Spens-Evans + Flocculent solution method, regardless of whether additional cleaning was carried out using the OneStep kit.

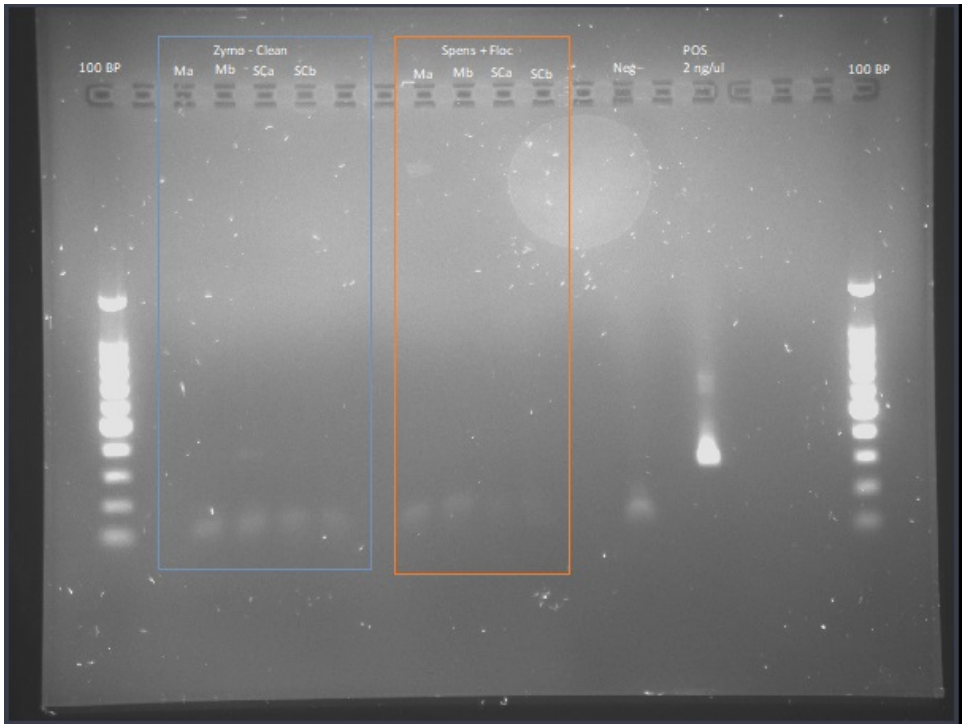


Figure 4: 2% agarose gel showing PCR products, for raw extracts and extracts cleaned using OneStep PCR Inhibitor Removal kit (Zymo), after amplification with MiFish-U primers. Products were run for 40-50 mins at 74V alongside 100 BP ladder (Promega). MiFish-U amplicon: ~ 300 bp. Samples: Ma, Mc, SCa, SCb. Treatments: 'Spens + Floc' indicates samples before Zymo cleaning, 'Zymo-Clean' indicates samples after inhibitor removal cleaning. Neg: PCR negative, POS: PCR positive.

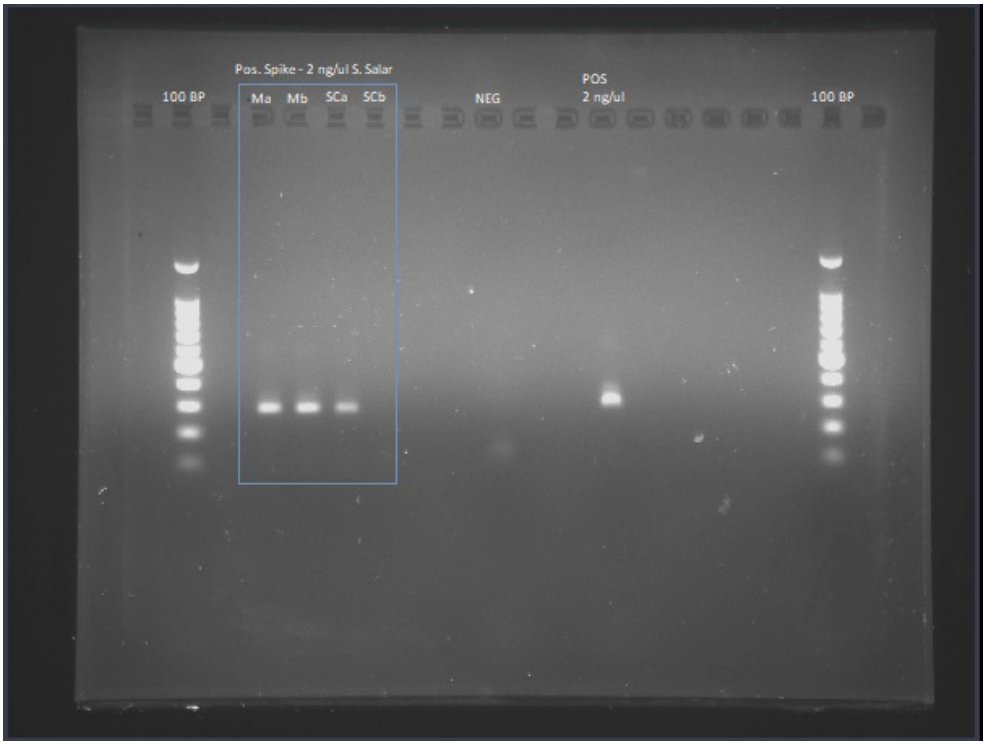


Figure 5: 2% agarose gel showing PCR products of positive controls, spiked with 2.5 µl of PCR extract from samples cleaned with OneStep PCR Inhibitor Removal kit (Zymo), after amplification with MiFish-U primers. Products were run for 40-50 mins at 74V alongside 100 BP ladder (Promega). MiFish-U amplicon: ~ 300 bp. Samples spiked into positive controls: Ma, Mc, SCa, SCb. NEG: PCR negative, POS: PCR positive.

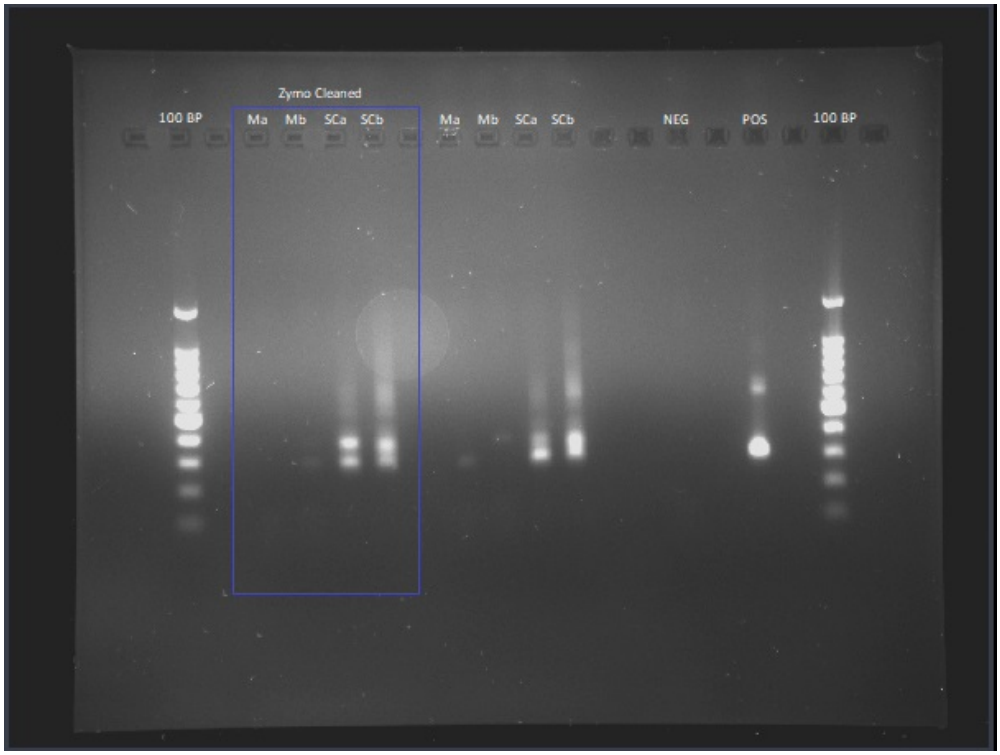


Figure 6: 2% agarose gel showing PCR products, for raw extracts and extracts cleaned using OneStep PCR Inhibitor Removal kit (Zymo), after amplification with MiFish-U primers using Multiplex Master Mix (QIAGEN). Products were run for 40-50 mins at 74V alongside 100 BP ladder (Promega). MiFish-U amplicon: ~ 300 bp. Samples: Ma, Mc, SCa, SCb. Treatments: 'Zymo-Cleaned' indicates samples after inhibitor removal cleaning, whereas the unlabelled samples are before Zymo cleaning, Neg: PCR negative, POS: PCR positive.

4.1 Experiment 3. Confirmation of Amplification Across Samples

The consistency of the final extraction and amplification method was investigated at a subset of four stations from the Dee October 2018 survey. The stations were: eDNA 2, 6, 10 and 12, in the upper and lower estuary (Chapter 2). Two replicates from each site were collected, filtered and stored as described in Chapter 2. DNA extraction was carried out following the Spens-Evans + Flocculent methodology (Chapter 2 Appendix B). A single DNA extraction blank was included (Chapter 2). PCR trials were performed in an identical manner using Multipex PCR Master Mix as in the previous experiment.

4.2 Experiment 3: Results

Amplification at ~ 300 bp was detected in at least one sample collected at every station and 6 out of the 8 samples extracted in total. The extraction blank and the PCR blank were negative and the positive control amplified (figure 7). Therefore, this indicates that the use of Spens-Evans + Flocculent solution method combined with the use of Multiplex Master Mix is an acceptable method for extracting amplifiable DNA from estuarine water samples collected at low tide, across the salininty gradient.

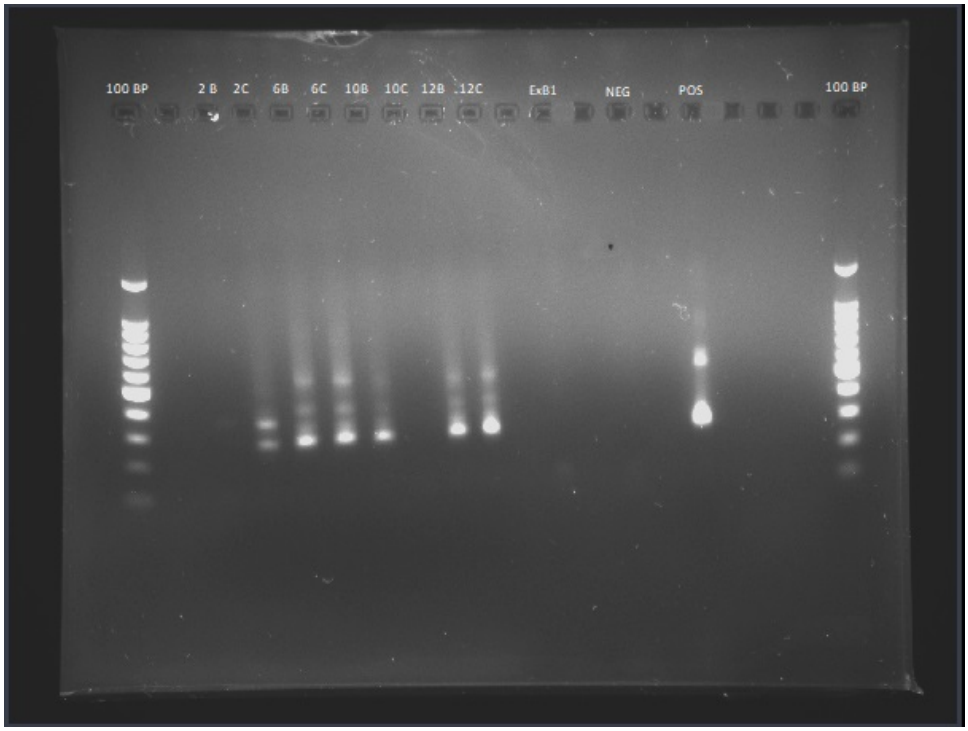


Figure 7: 2% agarose gel showing PCR products for samples extracted with Spens-Evans + Flocculant after amplification with MiFish-U primers using Multiplex Master Mix (QIAGEN). Products were run for 40-50 mins at 74V alongside 100 BP ladder (Promega). MiFish-U amplicon: ~ 300 bp. Samples: 2B, 2C, 6B, 6C, 10B, 10C, 12B and 12C, ExB1: extraction blank, NEG: PCR negative, POS: PCR positive.

5. Discussion

The aim of these experiments was to develop a modification to the Spens and Evans *et al.* (2017) DNA extraction to allow reliable amplification of eDNA from filtered estuarine water samples. There are numerous methods of reducing PCR inhibition (reviewed in Chapter 1). Several *ad hoc* trials on the use of dilution and Bovine Serum Albumin on PCR amplification, from DNA extracted using the Spens-Evans method, were conducted prior to this work. But the results were inconclusive (data not shown). The most appropriate method for extracting amplifiable DNA was the Spens and Evans *et al.* (2017) methodology with a particle flocculation step adapted from Sellers *et al.* (2018), following sample lysis. However, this only gave consistent amplification when combined with Multiplex PCR Master Mix (QIAGEN). This effect was consistently found across sites in the upper and lower estuary in the final trial. Therefore, *Taq* polymerase master mix choice should be a key consideration in eDNA studies, although removing inhibitors during extraction is still important. Ideally the Multiplex *Taq* should have been retested against the other DNA extraction and clean up methods previously discounted. However, it did make sense to remove sediment particles early in DNA extraction using flocculation to prevent spin column blocking. The flocculant solution contains ammonium acetate which precipitates proteins, along with aluminium ammonium sulphate and calcium chloride which remove inhibitors. Chemical flocculation is pH sensitive (Sellers *et al.*, 2018). The efficacy of Al^{3+} (released from the aluminium ammonium sulphate) in removing humic acid increases with decreasing pH (from 6.0 to 8.0) although the removal of DNA also increases (from pH 8.1 to 6.0, Dong *et al.*, 2006). The pH of the ATL lysis buffer (QIAGEN) is 8.3 (QIAGEN, 2020), lower than that of the original protocol, pH 9.0 (Sellers *et al.* 2018). However, this is still higher than those tested by Dong *et al.* (2006) and is unlikely to substantially influence DNA loss (G. Sellers, *pers. comm.*).

It was not surprising the DNA extracts contained PCR inhibitors (reviewed in Chapter 1). The water samples and initial stages of DNA extraction clearly contained fine suspended particulate matter (T I. Gibson, *pers. obs.*). Estuarine waters are known to have high concentrations of PCR inhibitors (Petit *et al.*, 1999). In addition, the capacity of the encapsulated filters used to filter a large volume of water probably enhanced the concentrations of PCR inhibitors, as water filtration concentrates PCR inhibitors (Schrader *et al.*, 2012). An additional factor was that sampling was around low tide when SPM concentrations are generally highest (McLusky and Elliott, 2004). Fundamentally the final methodology was an *ad hoc* fix. It was initially assumed that the Spens and Evans extraction and Q5 *Taq* would provide consistent amplification based on previous work in rivers and lakes (S. Creer, *pers. comm.*). Future studies should use a proprietary DNA extraction method (e.g. DNeasy PowerWater, QIAGEN) to give greater confidence in results, with less method development. These should be selected on the assumption that samples *are* inhibited to avoid delays.

Acknowledgements

I would like to acknowledge Dr Graham Sellers, Dr Robert Donnelly, Dr Amy Ellison, Kirthana Pillay and Professor Simon Creer for their technical advice, and Dr Charles Baillie and Dr Amy Ellison for their contribution of molecular consumables to this endeavour.

Chapter 2: Appendix B
Protocol: Spens and Evans Method with Flocculant Solution

Intro

This method is an adaption of the capsule method from Spens and Evans *et al.* (2017) with the addition of a particular flocculation step from Sellers *et al.* (2018) for DNA extraction from encapsulated disk filters (GF 5.0 µm; PES 0.8 µm, Nature Metrics Ltd.).

- Checked once, on 28.05.18, against Spens and Evans *et al.* (2017) Appendix S1.
- Modifications by Seymour, M. (*pers. comm.*) added 30.05.18.
- Re-assessed steps against original on 05.03.2019, checked once.
- Addition of flocculent solution steps (Sellers *et al.* 2018) and final modifications 22.10.2019.

Filter Preparation and Lysis

1. **Before extraction:**

   Carefully wipe the outer surfaces of all the collection tubes and filter capsules with 5% bleach using clean tissue paper. Dry and wipe with 70% Ethanol using tissue paper.

   - *Work quickly and keep the samples on ice or in the fridge, before being cleaned and while awaiting addition of lysis buffer.*

2. **Addition of lysis buffer:**

   Keep the outlet end (luer-lock - male) closed with the outlet cap. Remove the cap from the inlet (luer-lock - female) and add **800 µL** of pre-mixed **lysis master mix** (80 µL of **Proteinase K** to 720 µL of **Buffer ATL**). Close with an inlet cap.

   - *Be careful when adding in lysis master mix. Add 400 – 600 µL initially then smaller quantities until 800 µL is added. If filter overtops, record volume added.*

3. **Handshake** filter capsules vigorously 5 times and **vortex** each filter once.

4. **Incubate** at **56ºC.** Initially incubate all filters with inlet facing up. After 30 mins, vortex each filter and turn the filters so the outlet faces up, re-incubate. After another 30 mins, vortex the filters and incubate them with inlets facing all facing up overnight (~ 16 hrs).

   - *This turning of the filters is to aid penetration of the lysis buffer through the filter.*

5. **Transfer:**

   Take a sterile 5 ml luer-lock syringe and suck 4 to 5 ml of air into the syringe. Attach this to the inlet end of the filter. Remove the cap from the filter outlet and carefully expel all of the liquid from the filter into a 5 mL LoBind tube. Do not expel the final white foam as it can easily spill and contaminate the tube exterior. Spin the 5 ml tubes down.

Optional: Lysed samples can now be stored at ambient temperature. QIAGEN states lysed tissue samples can be stored for 6 months without a reduction in DNA quality. However, given the already degraded nature of eDNA samples we suggest storage at room temperature (RT) for, an arbitrary, 7 days maximum.

Particulate Flocculation (Sellers *et al.* 2018)

6. **Transfer:** Assuming **1000 µL** of transferred lysate, add all of this to a 2 ml LoBind tube containing **300 µl** (X 0.3) of **flocculent solution**. Vortex this briefly and incubate at 4°C or on ice for a minimum of 1 hr.

- *1000 µL appears to be the typical volume in my filters.*

7. **Centrifuge** at 10,000 * g for 2 min at RT (use a "soft stop setting").
8. **Transfer 1200 µL** of the sample, without disturbing the pellet, to a 5 ml tube containing **1200 µL** of **Buffer AL** and **1200 µL** of molecular grade 99% **ethanol** (1:1:1 ratio).
    - *Sample volume transfer standardised to maintain ratios for column binding (pers. comm. G. Sellers, 2019)*

DNA Binding, Washes and Final Elution

9. **Vortex** each sample vigorously.
10. **Pipet** the mixture (max 600 µL at a time) into a DNeasy MiniSpin column in a 2 mL collection tube provided in the kit.
11. **Spin** in micro-centrifuge preferably at 6000 * g 1 min.
12. **Discard** flow through.
13. **Repeat** steps 10-12 until all sample is filtered through DNA Mini spin column.
14. Place the **DNeasy Mini Spin Column** in a new 2 ml collection tube (provided), add **500 µl Buffer AW1**, and centrifuge for 1 min at 6000 * g. Discard flow-through and collection tube.
15. Place the **DNeasy Mini Spin Column** in a new 2 ml collection tube (provided), add **500 µl Buffer AW2**, and centrifuge for 3 min at 20,000 * g to dry the DNeasy membrane. Discard flow-through and collection tube. Place spin column in a new collection tube, centrifuge at 1 min at 17,000 * g.
16. **Transfer** spin column to a new 1.5 or 2 mL DNA LoBind tube with caps removed.
17. **Add 70 µl** of **Buffer AE** (Mat Seymour's Suggestion)
18. **Incubate** at RT for 10 min.
19. **Centrifuge** for 1 min at 6,000 * g.
20. **Re-elute DNA** from DNA LoBind tube.
21. **Incubate** at RT for 10 min.
22. **Centrifuge** for 1 min 6,000 * g.
23. **Discard** the spin column.
24. **Transfer** DNA to pre-marked DNA LoBind tube with lid intact.
25. **Optional:** aliquot 2 µL in a separate tube for DNA measurement.
    - *Note: > 20 µL of DNA required for PCR. 1 µL for Nano drop, 1 µL for Qbit and 1 µL for gels (as standard). So 10 µL?*
26. **Store** at -20°C or at -80°C.

Chapter 2: Appendix C
Optimisation of Sequence Library Preparation

Project Phase	Author
Study Design	TG, SC, AE
Lab Work	TG
Write-Up	TG
Review	AE

1. Introduction

Prior to sequence library preparation for Chapter 2 and 3, PCR reactions had to be optimised. A 2-step PCR approach was used to reduce PCR bias (see Chapter 2). All experiments were conducted using the Tele02 primers (Taberlet *et al.*, 2018). These primers replaced the MiFish-U Primers (Miya *et al.*, 2015), used in previous experiments (Chapter 2 Appendix A). Tele02 primers are an *in silico* optimisation of the MiFish-U Primers (Taberlet *et al.*, 2018) and cover the same region of the 12S rRNA sequence (R. Collins, *pers. comm.*). 12S rRNA primers are considered to outperform COI primers for studies of low concentration environmental DNA, as they show less non-specific amplification of prokaryotic and non-target eukaryotic DNA (Collins *et al.*, 2019). The coverage of UK fish species in the relevant reference database has also expanded recently for both these primer sets (Collins *et al.*, 2021). Tele02 primers were used because *in silico*, for UK fishes, they show the same level of taxonomic discrimination compared to MiFish-U, i.e. the proportion of amplified DNA sequences unambiguously associated with a single species. However, they have a higher primer universality *in silico* than the MiFish primers, i.e. the consistency with which primers will bind to species of the target taxonomic group (Collins *et al.*, 2019). In addition, use of Tele02 primers increased the consistency between work conducted at Bangor (Chapter 2 and 3), and the SeaDNA project (Chapter 4). The aim of the following experiments was to optimise PCR reactions for sequencing library preparation. Specifically, PCR conditions and bead clean up methods were tested to reduce none-specific amplification.

2. Methods

All samples were collected from the Dee estuary and extracted as in Chapter 2. A single sample was randomly selected (*sample* function, R Core Team, 2021) from three stations in the upper and lower estuary from the October 2018 and June 2019 (table 1). All procedures were conducted following contamination control standards and appropriate category of laboratory as stated in Chapter 2.

Table 1: Samples used for optimisation

Survey	Samples	Volume (ml)	Salinity Classification*
Oct 2018	Oct-18 1C	520	Euhaline
	Oct-18 7B	n/a	Oligohaline
	Oct-18 13C	460	Limnetic
Jun 2019	Jun-19 1A	896	Euhaline
	Jun-19 6B	838	Oligohaline
	Jun-19 9C	1008	Limnetic

*Salinity Classification: Venice System (1959)

2.1 Experiment 1: Amplification using Tele02 at Reduced Volume.

The first test determined if Tele02 primers (Integrated DNA Technologies, Inc. Coralville, IA, USA) amplified the target amplicon at a reduced sample input volume, to increase the possible PCR reactions per sample. A single PCR reaction was set up for each sample. PCR reaction set up was 25 µL containing 1.25 µl of DNA extract, 12.5 µl of 2x Multiplex PCR Master Mix (QIAGEN), 0.5 µl of Tele02-F, 0.5 µl of Tele02-R and 10.25 µl of PCR water. A PCR negative was included (sample replaced with 1.25 µl of water). For the positive control, 1.25 µl of an eDNA sample from the Dee, which had previously amplified using MiFish-U primers was used (Appendix A: Experiment 2). The thermal cycle profile was *Taq* activation for 95.0°C for 15 min, followed by 40 cycles of denaturation at 94°C for 30 sec, annealing at 60°C for 90 sec and extension at 72.0°C at 20 sec. This was followed by a final extension at 72.0°C for 10 min. PCR products were analysed by electrophoresis for 45 – 60 min at 74V on 1.5% agarose gels stained with SafeView Nucleic Acid Stain (NBS Biologicals, Huntingdon, UK). For each sample 5 µl of product was mixed 5:1 to pre-diluted 6X Purple Loading Dye (NEB, Ipswich, MA, USA) and was run alongside one to two 100bp DNA Ladders per gel (NEB). Success/failure of each test was scored by presence/absence of a clearly identifiable band on the gel. The mean length of the 12S rRNA fragment amplified by Tele02 is 167 bp (min: 129, max: 209, Taberlet *et al.*, 2018). This translates to a mean amplicon size of 277 bp (min: 239, max: 319) following addition of the primers and index tag.

Amplification at ~300 bp occurred across all samples, from across the environmental gradient, with some non-specific amplification at higher molecular weights. The negative control was negative, whereas the positive control showed weak amplification (figure 6). Therefore, all further trials were conducted with 1.25 µl of DNA extract added to each PCR.

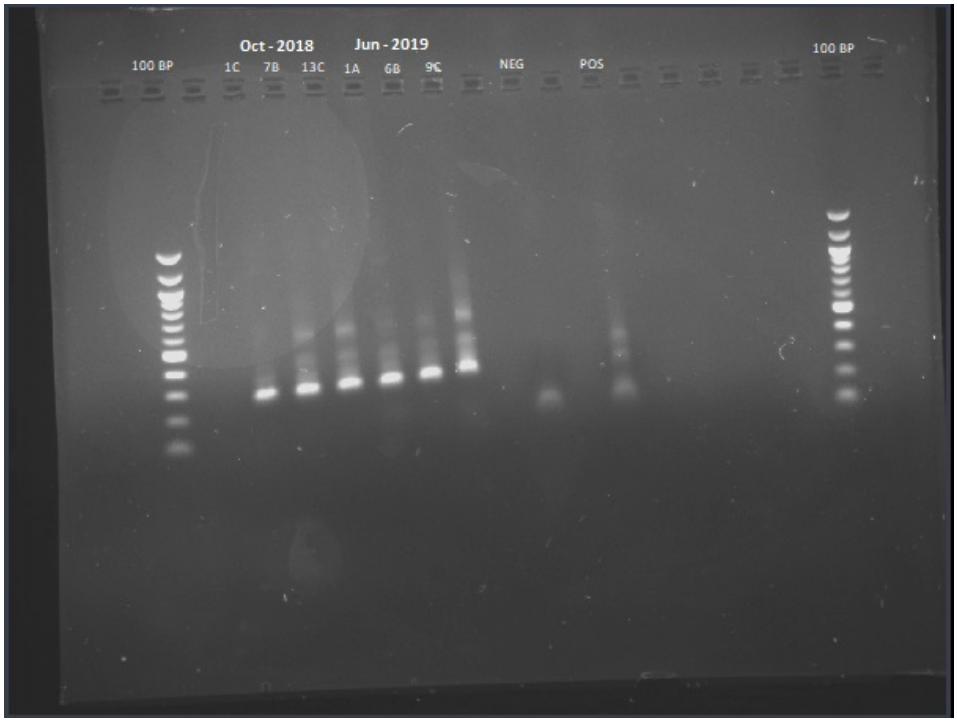


Figure 1: 1.5% agarose gel showing PCR products after amplification with Tele02 primers for round 1 PCR at 40 cycles, with 1.25 µl of DNA extract. Products were run for 45-60 mins at 74V alongside 100 BP ladder (NEB). Tele02 amplicon: ~ 300 bp. Survey and samples are indicated as in table 1. NEG: PCR Blank, POS: Positive Control.

2.2 Experiment 2: Round 1 PCR, 30 and 35 Cycles

Given the presence of none-specific amplification at higher molecular weights than 300 bp in Experiment 1. An additional PCR trial was conducted at 30 and 35 cycles to determine if none-specific amplification could be reduced. All setup conditions were identical to the previous experiment. PCR set-up was identical with 1.25 µl of sample added to each PCR. Two PCRs set-ups were conducted at 30 and 35 cycles, compared to the previous 40 cycles, all other thermocycling conditions were identical. Analysis was identical to above.

At 30 cycles amplification was poor, with bands barely visible on the gel. Comparably at 35 cycles, bands were present and none-specific amplification was reduced in comparison to the previous experiment at 40 cycles. Although some none-specific amplification was still seen at samples from the upper estuary in June 2019 (9C). Therefore, 35 cycles with 1.25 µl of template was adopted as the standard methodology for round one PCR (Chapter 2 and Chapter 3).

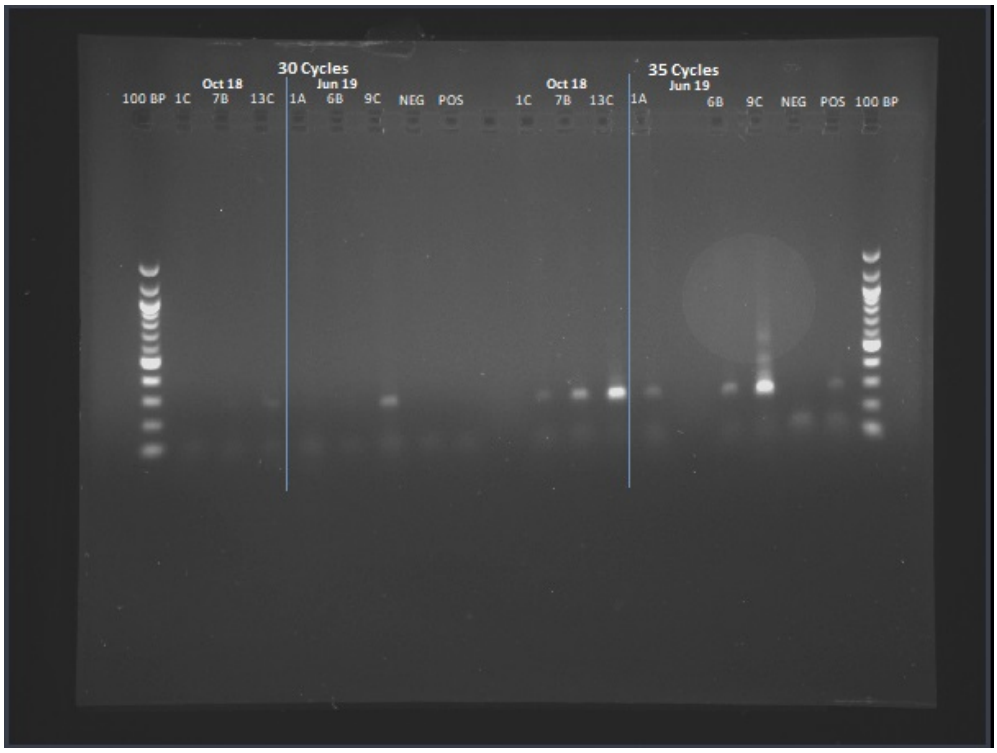


Figure 2: 1.5% agarose gel showing PCR products after amplification with Tele02 primers for round 1 PCR at 30 and 35 cycles, with 1.25 µl of DNA extract. Products were run for 45-60 mins at 74V alongside 100 BP ladder (NEB). Tele02 amplicon: ~ 300 bp. Survey and samples are indicated as in Table 1. The blue line separates samples from different surveys. NEG: PCR Blank, POS: Positive Control.

2.3 Experiment 3: Round 1 Beads Clean Up

All gel images prior to this point showed low molecular weight DNA around 100 bp, likely primer dimer. Therefore, to remove this low molecular weight DNA, 10 µl of the PCR product from the 35 cycles trial (above) was treated with ProNex® size-selective paramagnetic beads (Promega Corporation, Madison, WI, USA). A bead ratio of 1.75X compared to the sample was used. This ratio was intermediate between 1.5X (size cut-off: ~250 bp) and 2.0X (size cut-off: ~150 bp) and was assumed to have a size cut-off of ~200 bp (Promega, 2018). It was therefore unlikely to remove any shorter PCR products produced by the Tele02 primers (minimum size: 239 bp). Initially double-sided bead clean ups were investigated to see if they

also eliminated any higher molecular weight material. However, these trials were abandoned as the results were inconsistent and it was technically challenging (data not shown).

A bead ratio of 1.75X clearly removed the low molecular weight material below ~300 bp from the samples (figure 3, compare with 35 cycles in figure 2). Although some limited primer dimer was still present in the PCR negative and positive. Therefore, all bead clean ups were conducted using a 1.75X bead ratio (Chapter 2 and 3).

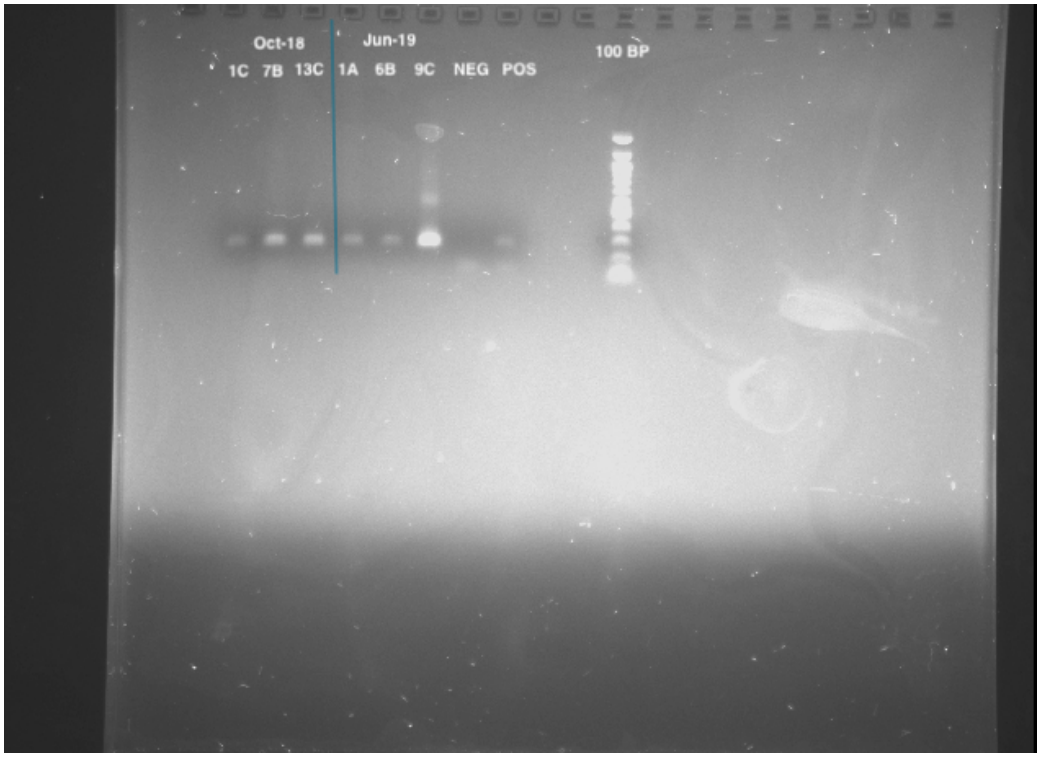


Figure 3: 1.5% agarose gel showing PCR products from round 1 PCR with Tele02 primers (35 cycles, 1.25 µl of DNA extract) after cleaning with Pronex beads (Promega) at a 1.75X ratio. Products were run for 45-60 mins at 74V alongside 100 BP ladder (NEB). Tele02 amplicon: ~ 300 bp. Survey and samples are indicated as in Table 1. A blue line separates the samples from different surveys. NEG: PCR Blank, POS: Positive Control.

2.4 Experiment 3: PCR Round 2

To confirm the cleaned round 1 PCR products from Experiment 2 could be indexed by PCR, PCR products were amplified using i5/i7 indexes (Integrated DNA Technologies, Inc). A single PCR reaction was set up for each sample and control. PCR reaction set up was 25 µl containing 3.0 µl of PCR product, 12.5 µl of 2x Multiplex PCR Master Mix (QIAGEN), 1 µl of premixed i5/i7 indexes at 10 µM concentration and 8.5 µl of RNase-Free Water (QIAGEN). 3.0 µl of PCR product was used to maximise the potential for re-running PCR reactions. The thermal cycle profile was *Taq* activation for 95.0°C for 15 min, followed by 15 cycles of denaturation at 94°C for 30 sec, annealing at 60°C for 90 sec and extension at 72.0°C at 20 sec. This was followed by a final extension at 72.0°C for 10 min. 15 cycles were used because this was the lower range used by other scientists (W. Perry and A. Ellison, *pers. comm*).

Clear amplification was seen across all samples, the negative control was negative whereas the positive control amplified (figure 4). Therefore, this method was used as the protocol for round two PCR (Chapter 2 and 3). The final fragment size, after addition of indexes, had a mean size of 343 bp (min = 304 bp, max = 384 bp). However, substantial none-specific amplification was found at higher molecular weights (figure 4). 1:10 and 1:100 dilution of the

round two PCR product in water did not show a substantial reduction in the larger bands relative to the target fragment (data not shown). This indicated the none-specific amplification was real rather than being a product of too much DNA being loaded onto the gel. Reducing the number of cycles in round two PCR to 10 cycles as an alternative method, showed substantially weaker amplification (data not shown). Therefore, a method to remove both higher and lower molecular weight DNA from PCR product, was required.

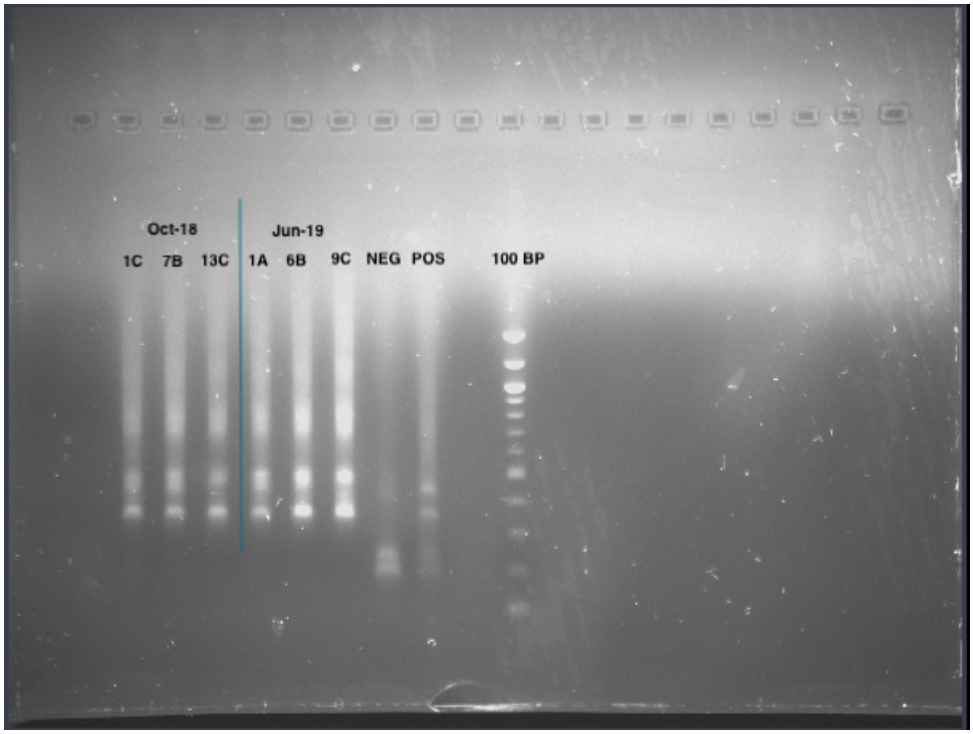


Figure 4: 1.5% agarose gel showing PCR products from round 2 PCR (15 cycles, 3.0 µl of PCR product) with i5/i7 indexes (Integrated DNA Technologies, Ltd.), PCR products used had been amplified for 35 cycles (figure 3). Products were run for 45-60 mins at 74V alongside 100 BP ladder (NEB). Indexed amplicon: ~ 350 bp. Survey and samples are indicated as in Table 1. A blue line separates the samples from different surveys. NEG: PCR Blank, POS: Positive Control.

2.5 Experiment 4: Round 2 Beads Clean Up

An experiment using a double-sided bead clean-up was conducted to attempt to remove both smaller and larger fragments than the target amplicon. Briefly, the first bead binding step removes high molecular weight fragments, the second removes the low molecular weight fragments (Promega, 2018). 10 µl of PCR product was cleaned and the bead ratio used was 1.2X in the first step, and 0.35X in the second step, to ensure preservation of the amplicon (Promega, 2018).

Following cleaning, lower molecular weight material was removed. Some of the higher molecular weight material was also removed, although it was difficult to tell if this was just an artifact of dilution (figure 5). Overall, it may have been possible to clean the PCR products using this method in Chapter 2 and 3.

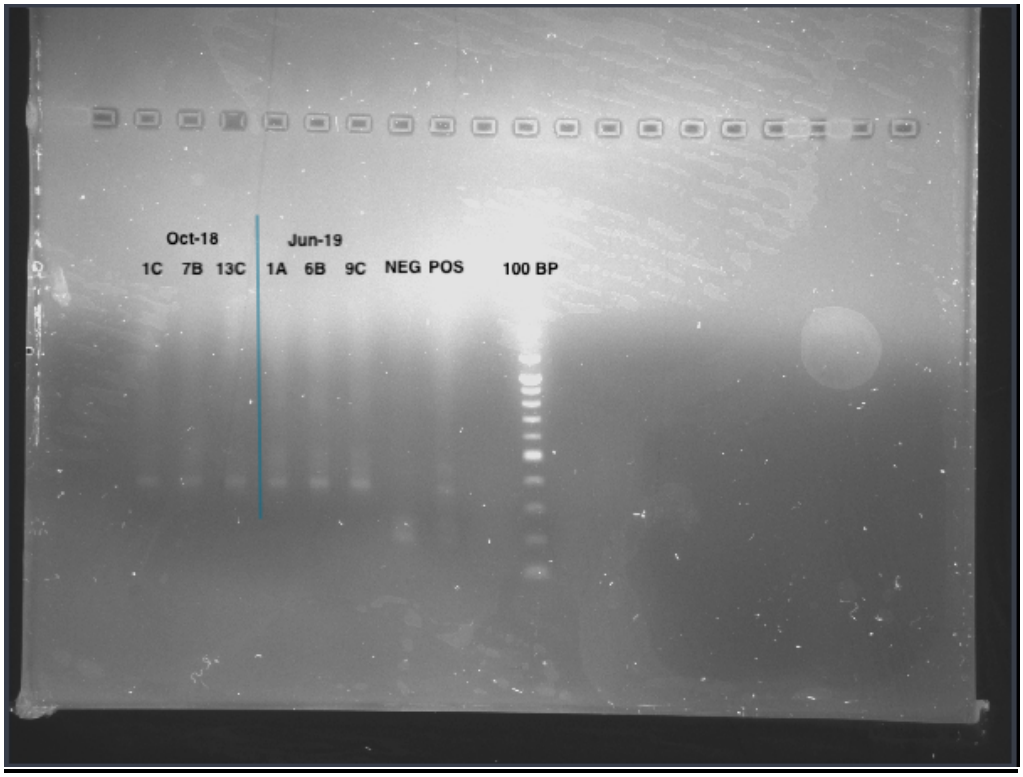


Figure 5: 1.5% agarose gel showing PCR products from round 2 PCR (15 cycles, 3.0 µl of PCR product) with i5/i7 indexes (Integrated DNA Technologies, Ltd.) after cleaning using Pronex beads (Promega) using a 1.2X/0.35X ratio. Products were run for 45-60 mins at 74V alongside 100 BP ladder (NEB). Indexed amplicon: ~ 350 bp. Survey and samples are indicated as in table 1. A blue line separates the samples from different surveys. NEG: PCR Blank, POS: Positive Control.

3. Discussion

The aim of the experiments was to optimise PCR reactions for sequencing library preparation. The optimal protocol for both PCRs, with a bead clean up step between these, was included in Chapter 2 and Chapter 3. In the final protocol, the double-sided bead clean-up of the round 2 PCR product was replaced with a final gel slice after pooling all samples (the stage shown in figure 4, Chapter 2 and 3). This was because it was difficult to replicate the success of this initial work, due to the very small volumes of liquid being used. In retrospect, greater time should have been spent investigating other aspects of PCR optimisation e.g. extension time, annealing temperature, lower initial cycle numbers in PCR 1, normalisation of DNA extracts prior to PCR etc. Perhaps this may have increased the number of fish reads returned by the final sequencing reaction (Chapter 2 and 3), however this is unknown. At this point in the project time was very short. Therefore, a quick method using a final gel slice was considered appropriate for practical reasons. More generally it was perhaps not wise to use the Tele02 primers, given they had only been tested *in silico*, at the beginning of this project (Taberlet *et al.*, 2018; Collins *et al.*, 2019). Whereas the MiFish-U primers had undergone substantial *in vitro* testing (Miya *et al.*, 2015; Collins *et al.*, 2019). Although it is unknown if this would have increased fish sequence yields for Chapter 2 and 3 and the Tele02 primers appeared to perform well in Chapter 4.

Acknowledgements

I would like to acknowledge Dr Amy Ellison, Dr Graham Sellers, Dr Lori Lawson-Handley and Professor Simon Creer for their technical advice throughout these experiments. I would

like to thank Dr Rupert Collins and Professor Stefano Mariani for their advice on the choice of primers.

Chapter 2: Appendix D
Bioinformatics Protocol for Chapter 2 and 3

Author Contributions

Project Phase	Author
Bioinformatic Analysis	TIG and AE
Protocol Writing	TIG
Review of Protocol	SC, AE

Introduction

Command line bioinformatic programmes were accessed via a Unix shell on a MacBook Pro 2019 (Apple Inc., California, CA, US). Command line scripts were organised via Jupyter Notebook (Project Jupyter, 2020). Reads were tracked throughout the pipeline. Any trivial helper functions e.g. *grep* and shell scripts are not detailed here but were written using the Unix Shell, the R language (R Core Team, 2021) via RStudio (RStudioTeam, 2019), or used functions from *seqkit* (Shen *et al.*, 2016).

Prior to **Step 1.** Raw sequencing data were converted into demultiplexed fastqc files using *bcl2fastq v2.20* (Illimina, Inc., San Diego, CA, USA). Raw sequencing files from different sequencing runs were merged (using *cat*), if required.

Step 1. Initial Quality Assessment

The initial quality of the demultiplexed fastq files was assessed by generating a quality report for each file using *FastQC,* a quality control tool for high throughput sequence data (Andrews, 2010). The quality reports generated by FastQC were then summarised using *MultiQC* (Ewels *et al.*, 2016) to create a single report and visually appraised.

Step 2. Primer Trimming

The position of primers in the sequences was double checked on paired sequences first using *seqkit locate* (Shen *et al.*, 2016). The sequences of the forward and reverse Tele02 primers (Miya *et al.*, 2015; Taberlet *et al.*, 2018) were removed from the foreword and reverse reads using the command line adaptor trimming tool *Cutadapt* (Martin, 2011). Primers were removed following trimming of the foreword hexamer (-- *cut* 5) and an error tolerance of 0.1 in the primer sequence was allowed (-- *error-rate* 0.1). Any untrimmed sequences were discarded (-- *trimmed only*). The quality of output reads was assessed as in step 1.

Step 3. Quality Filtering

Following primer trimming, quality filtering of the files then proceeded using *fastp* to take advantage of its sliding window capability. *Fastp* is an ultra-fast fastq pre-processor developed in C++ (Chen *et al.*, 2018). Sequences were trimmed by moving a sliding window (--cut_window_size: 4) from the front (5') to trail (3'; --cut right) and dropping the window if it fell below the quality threshold of phred score 30 (--cut_mean_quality: 30) and also the proceeding bases to the right. Reads were also omitted if they contained any N bases (--n_base_limit: 0), and/or if more than 40% of their bases were below phred score 15 (--unqualified_percent_limit: 40) and/or if their length was below 50 bp (--length_required: 50). The quality of output reads was assessed as in step 1.

Step 4. DADA2 Workflow

Following quality filtering, data was inputted into *dada2* (Callahan et al., 2016). *Dada2* models and corrects for Illumina-sequenced amplicon errors and allows inference of exact sample sequences as Amplicon Sequence Variants (ASV), rather than clustering into Operational Taxonomic Units (OTUs). The pipeline followed DADA2 Pipeline Tutorial 1.16 (Callahan, n.d.) and used the R language (R Core Team, 2021) via RStudio (RStudioTeam, 2019). Although the default settings were generally used, the functions and settings used are detailed here for clarity.

4.1 Learn the Error Rates

Firstly, the error model was generated for foreword and reverse reads using *learnErrors()*. The observed and estimated error rates for each possible transition were generated and visually appraised with reference to the error rates expected under the nominal definition of the Q-score using *plotErrors()*.

4.2 Dereplication

Amplicon sequences within the foreword and reverse fastq files were then dereplicated to unique sequences and quality information retained using *derepFastq()*.

4.3 Sample Inference

The sample inference algorithm was then used to remove all sequencing errors and infer ASVs for foreword and reverse reads, using *dada(pool = T)*. During this step samples were pooled prior to sample inference, to increase the sensitivity to low frequency sequencing variants.

4.4 Merge Paired Reads

Each pair of denoised forwards and reverse reads were then aligned and merged to produce a contiguous sequence, using *mergePairs*(). Paired reads were only merged if they overlapped by at least 12 bases and had no mismatches in the overlap region.

4.5 Construct sequence table

An ASV table was then constructed using *makeSequenceTable()*.

4.6 Remove Chimeras

Following this chimeric sequences were identified if they could be reconstructed from combining a left and right segment for two more abundant "parent" sequences (bimeras). Specifically, each samples was independently checked for bimeras and a consensus decision on each sequence variant was made. This used, *removeBimeraDenovo (method = "consensus")*. Named unique sequences were then exported from DADA2 for taxonomic assignment and the ASV table for further analysis.

Step 5. Taxonomic Assignment

Taxonomic assignments for the ASVs (query sequences) were generated using a multi-step approach broadly following the steps outlined in Collins *et al*., (2019). This used a combination of BLAST against a local database, the global nt database and verification of results using an evolutionary placement method.

5.1 Local BLAST Search

The query sequences were subjected to BLAST searches against a curated local 12S rRNA database, Meta-Fish-Lib, using BLAST+ (Camacho *et al.*, 2009). Meta-Fish-Lib is a dynamic, reference library for fish species in the UK, derived from the NCBI GenBank and BOLD and is updated with each GenBank release (Collins, 2021a; Collins *et al.*, 2021). The most up to date Meta-Fish-Lib version, at the time, was used (21.08.21 search; GenBank version 245). Meta-Fish-Lib was downloaded and the 12S rRNA component corresponding to the Tele02 primers extracted using scripts available with Meta-Fish-Lib (Collins, 2021a). A local BLAST database was generated from the Meta-Fish-Lib 12S rRNA library using *makeblastdb*. Following this, a nucleotide BLAST search was performed for each unique ASV (evalue: 1 x 10^{-5}) against the local database, using *blastn*. Further data manipulation of blast results were conducted using R (R Core Team, 2021).

5.2 Global BLAST Search

Section 5.2 was conducted by Dr Amy Ellison. The query sequences were subjected to BLAST searches (evalue: 1 x 10^{-4}) against a recent version of the preformatted 'nt' BLAST database (NCBI, 2021). Searches were conducted using Super Computing Wales. Taxonomic classification for query sequences was added to the highest scoring hit using Taxallnomy, an extension of NCBI Taxonomy (Sakamoto and Ortega, 2021). This allowed the general identification of taxa which were not present in the reference database. Given the noncurated nature of this database any fish species identified only from the 'nt' BLAST database were not included in downstream statistical analyses.

5.3 Evolutionary Placement Algorithm

In order to verify the results of the BLAST searches and assign higher level taxonomic ranks to sequences where it was required, we used a phylogenetic (or evolutionary) placement method. Phylogenetic placement methods identify query sequences by placing them on a phylogenetic tree inferred from the reference sequences. Therefore, incorporating information about the evolutionary history of the studied species. Comparably, BLAST does not use or provide phylogenetic information about the query sequence which can decrease identification accuracy (Barbera *et al.*, 2019; Czech and Stamatakis, 2019). This method was implemented by adapting the scripts and approach in the bioinformatics pipeline Meta-Fish-Pipe (Collins, 2021b).

In order to ensure clear results (R. Collins, *pers. comm.*), only those query sequences which had either returned a BLAST hit vs the local reference database, and, or had returned a BLAST hit vs the nt database which was assigned to a species which was part of one of the fish Classes (e.g. Actinopterygii, Chondrichthyes etc.). In addition, the reference sequences in Meta-Fish-Lib were dereplicated to remove any duplicate sequences within each species to allowing building the phylogenetic tree (R. Collins, *pers. comm.*).

Firstly, multiple sequence alignment was constructed for both the query and the dereplicated reference sequences, together, using MAFFT v7 (Katoh *et al.*, 2002; Katoh and Standley, 2013). Using a maximum of 1000 iterations *(-- maxiterate 1000*) and the E-INS-i algorithm (--*genafpair*), which is an iterative refinement method using both weighted sum of pairs and consistency scores and is highly accurate (Katoh, 2013). The query and reference sequences in the alignment were then separated from each other following alignment.

Secondly, a maximum likelihood phylogenetic tree was then generated using the aligned reference sequence using Randomised Axelerated Maximum Likelihood (RAXML) implemented using RAXML-ng (Kozlov *et al.*, 2019). The model of sequence evolution used the substitution matrix define by Tamura and Nei (1993) and the GAMMA model of among site rate heterogeneity (*--model TN93+G*). Following tree inference, the best tree was evaluated by computing the likelihood of the tree topology by optimise the model parameters (*--evaluate*).

Thirdly, an Evolutionary Placement Algorithm (implemented with EPA-ng) (Barbera *et al.*, 2019) was used to place the query sequences onto the maximum likelihood reference tree using the optimised model parameters from the tree evaluation. Following placement of the sequences, taxonomic assignment of the placed query sequences was carried out using GAPPA (subcommand *gappa examine assign*, Czech et al., 2020; Czech & Stamatakis, 2019) and the best hits extracted for each query sequence.

5.4 Final Taxonomic Assignment.

Following each of these steps we assigned taxonomy based on the following criteria. Rule 1 and 2 are adapted from Collins *et al.* (2019).

Firstly, all BLAST hits with an alignment length of less than 90% of the total length for each ASV they were assigned to were omitted. This is an arbitrary cut off selected after appraising the data (median length in the reference database: 166 bp; min: 83 bp; max: 183 bp).

Rule 1: Species level taxonomy was assigned if both the best scoring BLAST hit and species level EPA result were identical and sequence identity was ≥ 97%.

This rule finds assignments that are congruent between both the BLAST search and the EPA, but rejects assignment with low similarity and very short match lengths (Collins *et al.*, 2019). The ≥ 97% identity BLAST cut off was obtained via a simulation study of the threshold at which erroneous assignments appeared in a comparable UK marine fish dataset (R. Collins *pers. comm.*).

Rule 2: Species level taxonomy was assigned if both the BLAST hit and the species level EPA results were identical, BLAST identity was ≥ 95% and EPA probability was ≥ 90%.

This rule allows for dissimilar hits, but only those with a high level of phylogenetic probability (Collins *et al.*, 2019).

Rule 3: Species level taxonomy was assigned if only the BLAST hit only was 100% and no species had tied on their BLAST score.

This final rule allows for identical hits, most likely a positive detection of the species, not being rejected due to any potential errors in the EPA assignment tree.

Rule 4: Following this, the BLAST and EPA results for any query sequences which had passed through these filters were manually appraised. Higher-level phylogenetic levels as determined from EPA (e.g. genus, order, family) were assigned to any remaining query sequences. Any sequences assigned above family level e.g. class (e.g. Actinopterygii) or subphylum (Vertebrata) were rejected as they would not be ecologically meaningful.

Acknowledgments

I would like to acknowledgment the advice of Dr Rupert Collins, Dr Amy Ellison, Professor Simon Creer and Dr Thomas Brekke on various bioinformatic issues, both practical and theoretical. I would also like to thank Dr Molly Czachur for access to initial data used to develop this pipeline.

Notes

I am a co-author on the publication Collins *et al.* (2021) 'Meta-Fish-Lib: A generalised, dynamic DNA reference library pipeline for metabarcoding of fishes' which describes the development of the Meta-Fish-Lib reference database. I contributed to discussions on the development of the database as an early user, obtained fish samples via my contacts to expand the libraries taxonomic coverage (although these were not used in the end) and proofread and commented on the final manuscript.

Chapter 3.

Environmental DNA Metabarcoding Detects Short-Term Temporal Variability in the Fish Assemblage within a Macrotidal Estuary

Authors: Thomas I. Gibson, Georgina Brenna, Gary Carvalho, Amy Ellison, Tristan Hatton-Ellis, Lori Lawson-Handley, Stefano Mariani and Simon Creer.

Author Contributions

Phase	Author
Grant Acquisition	TIG, SC, GC, TH-E
Study Design	TIG, SC, GC, SM, GB, AE, LL-H
Field Sampling and Laboratory Analysis	TIG
Supporting Data Acquisition	TIG
Bioinformatic Analysis	TIG, AE
Statistical Analysis	TIG
Manuscript Writing and Review	TIG, SC, AE, SM, LL-H

Abstract

Estuaries often show a wide temporal variability in their physicochemical characteristics over short timescales (hours to days), which in turn influences the structure of the fish assemblage within them. The analysis of environmental DNA (eDNA) via metabarcoding is quickly developing as an effective method for the biodiversity assessment of fish assemblages in various ecosystems, including estuaries. However, limited research has been conducted on the temporal variability in estuarine fish eDNA composition and the implications for biomonitoring. Therefore, the short-term tidal variability in the fish assemblage, detected by eDNA metabarcoding, within a well-mixed macrotidal estuary was investigated. In autumn at the mouth of the Conwy Estuary (Wales, UK), duplicate ~1L surface water samples were collected at high and low tide over 15 consecutive days, covering before and after a full spring to neap tidal cycle. DNA was extracted from samples and subjected to metabarcoding analysis using an established assay targeting teleost fish. Multivariate statistical analysis showed temporal variation in the relative reads of fish species were correlated with changes in salinity, which occurred at different tidal states and due to an episodic increase in river water entering the estuary. Temporal variation in species presence/absence was also associated with changes in salinity and in addition tidal range. The overall conclusion is that short-term variability in fish eDNA assemblages can be substantial in macro-tidal estuaries and that future biomonitoring must account for this.

Keywords: Estuaries, Fishes, Environmental DNA, Metabarcoding, Biomonitoring, Time-series.

1. Introduction

The analysis of environmental DNA (eDNA), isolated from an environmental sample without capturing the organism (Taberlet *et al.*, 2012), is sufficiently well advanced for biodiversity assessment. At the same time the required techniques for eDNA analysis are continually being refined and developed (Hering *et al.*, 2018). More specifically, metabarcoding of eDNA (see Lawson Handley, 2015; Deiner *et al.*, 2017a) extracted from

water samples is an effective method for the assessment of fish assemblages in freshwater (Hänfling *et al.*, 2016; Hallam *et al.*, 2021), estuarine (Bleijswijk *et al.*, 2020; García-Machado *et al.*, 2021) and marine (Yamamoto *et al.*, 2017; Valdivia-Carrillo *et al.*, 2021) environments. This is encouraging given that fish assemblages are good indicators of the relative health of aquatic ecosystems and therefore are an important component of biomonitoring (Fausch *et al*., 1990; Whitfield & Elliot, 2002). In addition, monitoring of fish assemblages is explicitly required in the EU and UK, under the Water Framework Directive (EC, 2000) and equivalent legislation (JNCC, 2021; UK Parliament SI 2017/1012 and SI 2017/407), respectively. One such group of ecosystems of which fish are an important indicator group for, are estuaries (Whitfield & Elliot, 2002). This is not a homogenous category and estuaries show substantial variation in their physical characteristics, such as water circulation (Whitfield & Elliott, 2011). Estuaries are of key ecological importance, possessing high primary (Nixon, 1988) and secondary productivity (Allen, 1982; Costa *et al.*, 2002). They rank among the most valuable ecosystems in terms of their ecosystems service provision (Costanza *et al.*, 1997). However, estuaries are heavily impacted by human activities (Kennish, 2002) and most ecological degradation in developed countries has occurred historically (Lotze *et al.*, 2006). Therefore, it is critical that the biodiversity of estuarine ecosystems is monitored, including the fish assemblage.

Estuaries are characterised as highly dynamic environments with steep spatial gradients in physicochemical parameters such as salinity, that rapidly change temporally in response to tides, and also seasonally in response to freshwater inflow, which in turn influences the biota (Whitfield & Elliott, 2011). Therefore, a key challenge in the application of any ecological sampling technique to estuaries, including eDNA metabarcoding of fish, is to understand and account for the inherent variability which occurs at short-time scales (hours to days), associated with tidal cycles. More generally, documenting the natural variation in fish assemblage structure and function, against which degradation can be compared, is a key research goal for biomonitoring (Whitfield & Elliot, 2002). Although less studied than the variation documented over seasonal and inter-decadal time scales in estuaries (Yeoh *et al.*, 2017), short-term changes in the distributions of fishes in estuaries occur due to a variety of factors. These include including diel changes (Yeoh *et al.*, 2017), episodic changes in river flow (Gillson, 2011) and movements with the tides. Regarding the influence on tides, there is good evidence for tidal influence on the distributions of individual species and assemblages. Telemetry studies of individual juveniles have shown fish may, “ride the tide”, up and down the channel of an estuary with the flood and ebb tides (Childs *et al.*, 2008; Næsje *et al.*, 2012). At the assemblage level, greater numbers of fish may enter an estuary from the sea during the flood than the ebb tide, and then swim out again on the ebb (Becker *et al.*, 2016a). Further to this, more fish have also been shown to move in and out of an estuary on a neap than a spring tide (Becker *et al.*, 2016b). Ebb tides can also cause estuarine resident species to become concentrated into subtidal channels at low water as intertidal areas dry and the salinity upstream falls. While marine species can increase in abundance at high water, as the net movement of water into the estuary on the flood tide facilitates immigration, and then emigrate on the ebb tide (Greenwood and Hill, 2003). The spring to neap cycle may influence the structure of fish assemblages by changing the amount of the intertidal zone that can be accessed on spring compared to neap tides. On spring tides, the upper parts of the intertidal zone can be accessed at high tide, although fish are forced completely out of the intertidal zone at low tide. During neap tides the limited variation in range means the lower intertidal area is available for longer periods (Wilson and Sheaves, 2001). Tidal influences can be seen across life history stages, e.g. on larvae and early juveniles (Beckley, 1985). Overall, changes in tidal currents can lead to shifts in the functional organisation of fish assemblages (Ribeiro *et al.*, 2006). Despite the evidence for the

influence of tides, fishes are able to move in and out of estuaries independently of them (Becker *et al.*, 2016b, 2016a), and even larvae and early juveniles may migrate against the outgoing ebb tide (Pattrick and Strydom, 2014). In addition, tidal changes may not always cause overall changes in the assemblage organisation (Ribeiro *et al.*, 2006). Nevertheless, short-term variability in fish assemblages is an important consideration when sampling and trying to understand their structure and function in estuaries.

Given the evidence for the inherent short-term variability of fish assemblages in estuaries, this raises the question does the eDNA of these fishes also shows variation in response to tides? If the eDNA signal accurately reflects the short-term tidal changes in fish communities, we would predict a higher proportion of sequence reads from species associated with marine conditions on higher tides at higher salinities, and conversely a higher proportion of sequence reads from freshwater and potentially diadromous species when freshwater inputs are more dominant, on lower tides and at lower salinities. Such a prediction would be driven by both the distribution of fishes and eDNA transport from the river and the sea. Several studies have assessed the influence of tides in estuarine ecosystems with varied results. For example, Kelly, Gallego and Jacobs-Palmer (2018) sampled surface waters on two incoming and outgoing tides over a 24-hr period and assessed assemblage level change using a universal eukaryotic COI marker in a fjord (Hood Canal, USA). The effects of tidal direction or individual tides did not consistently drive differences in assemblage composition, compared to site and physicochemical parameters. Instead, changes in the physicochemical parameters associated with different water masses were more important (Kelly *et al.*, 2018). However, this particular fjord has sluggish circulation with a stratified water column, little vertical mixing and two deep sills similar to a "classic fjord" (Paulson *et al.*, 1993). Where the sill reduces the inflow of tidal water (McLusky and Elliott, 2004). Therefore, these results are not generalisable to well mixed estuaries. Comparably, no substantial effect of tides was found on the eDNA fish assemblage composition in surface water samples from a microtidal inlet between the North Sea and Wadden Sea (Bleijswijk *et al.*, 2020). Either for an individual tidal cycle, or for high and low tide samples taken weekly over 18 weeks. It should be noted that there were probably limited differences in fish assemblage composition either side of the tidal inlet (Bleijswijk *et al.*, 2020). Comparably, an assessment of assemblage composition at high and low tide at two stations using universal COI primers in the freshwater reaches of the Elbe (Germany), a well-mixed, meso/macrotidal estuary, showed evidence of strong tidal influence (Schwentner *et al.*, 2021). At low tide, downstream stations had a highly similar assemblage compositions to the upstream station. Whereas, at high tide the assemblage composition changed at the downstream stations, but not at the upstream stations. However, fish species were poorly represented and the sampling effort makes the results preliminary (Schwentner *et al.*, 2021). Overall, these studies suggest that the short-term variation in assemblage composition driven by hydrographic factors, such as tides, is likely to be highly dynamic, context-dependent, and certainly requires more empirical research.

1.2 Study Aims, Objectives and Hypothesis

The overall aim of this study was to investigate the short-term tidal variability in the fish assemblage within a well-mixed macrotidal estuary detected by metabarcoding of eDNA from surface water samples, using a Teleost specific assay. Assemblage composition was inferred from the species presence/absence, relative contribution of species reads to each sample, and estuarine use guild assignments. Estuarine use guilds describe the overall ecological use of an estuary by a species and its links between the estuary and marine and freshwater areas (Elliott *et al.*, 2007). The first objective was to determine the temporal variation in assemblage composition at high and low tide over 15 days, covering before and after a full spring to neap

cycle. The eDNA assemblage was expected to be characterised by marine species at high tide and transition to one characterised by freshwater species at low tides, and the difference between these two assemblages would decrease as tidal range fell. The second objective was to formally test to what extent any changes in assemblage composition were associated with changes in key environmental variables: salinity, tidal range, pH and turbidity. It was hypothesised that changes in the fish eDNA assemblage composition are correlated with salinity and tidal range. In addition, it was expected that changes in the fish eDNA assemblage, focusing on relative read counts, would be driven by the differential correlations of estuarine use guilds with the key environmental variable(s). For example, Marine Migrants fishes may have a positive relationship with salinity, whereas Freshwater species may not.

2. Methods

2.1 Sampling Location

The Conwy estuary (N. Wales) is a vertically well-mixed macrotidal estuary with relatively shallow water depths, typical for UK estuaries. It is tidally dominated (typical spring tidal range: 6 m), with tidal volume exchange exceeding mean river input. The Conwy river accounts for the majority of the estuary's freshwater input (Robins *et al.*, 2014). The estuary has been monitored in spring/early summer and autumn by a catch-based fish survey which uses multiple net types and collects physicochemical data (Colclough *et al.*, 2002). The survey was conducted each year from 2004 to 2016, although sampling in each season was not consistent over this time. This data is for WFD assessment and was collected at low tide, slack water. The eDNA sampling site was located at Beacons Jetty (Lat: 53.29404, Lon: -3.83908; figure 1). This site is easily accessible across the tidal cycle and is close to the estuary mouth where the tidal influence is relatively strong (P. Robins *pers. comm.*). From 2007 to 2015 the median salinity of a nearby downstream WFD sampling site at low water (Figure 1) was 23.30 (IQR: 7.32) characterising the sampling site as polyhaline (McLusky, 1993) and so representing intermediate environmental conditions between the sea and the river.

2.2 Field Sampling and Filtration

Equipment was prepared and decontaminated prior to each individual sampling event in a non-PCR laboratory. 1 L water Nalgene bottles (Nalge Nunc International, Rochester, NY, USA) and silicone tubing (Cole-Parmer, Vernon Hills, IL, USA) were decontaminated in <5% commercial bleach solution (in tap water) for 4-5 hrs respectively. The equipment was washed with tap water and dried. All other equipment, sample boxes etc., were decontaminated with a 10% commercial bleach solution, followed by rinsing in tap water.

On each sampling day, duplicate 1 L surface water samples were collected at high and low tide for 15 days from the 29.09.20 to the 13.10.20 (total samples: 60). Water samples were taken from a depth of <1 m using a sampling pole. Sampling for each tide started approximately 5 min before and finished within approximately 22 min of slack water. Before collecting replicates on each tide, the sampling pole and bottles were immersed in the estuary and this water discarded. Field blanks (1 L of ddH_2O) were taken into the field approximately once every two days and treated identically to samples (total field blanks: 7). Following collection, bottles were placed on ice in individual zip lock bags, in an insulated box, until filtration. Immediately following eDNA sampling, physicochemical parameters were measured at a depth of <1 m for 20-30 second (2-3 readings per tide) using an AquaTROLL600 (In-Situ, Inc., Fort-Collins, CO, USA). Specifically, Temperature (°C), Dissolved Oxygen (% saturation), pH, Turbidity (Nephelometric Turbidity Units; NTU) and Salinity (PSU) were recorded. Water samples were then transported to the laboratory and filtered within 2-3 hrs of collection in a

non-PCR laboratory. Benches were cleaned with 10% commercial bleach solution before and after filtering. Samples were mixed by inversion, silicon tubing inserted and a small volume of water pumped through to pre-wash the tubing. Samples were filtered through an encapsulated 0.8 µm PES filter with an integrated 5.0 µm GF pre-filter (Nature Metrics Ltd., Gilford, UK) using a Geopump Peristaltic Pump (Geotech Environmental Equipment, Inc., Denver, CO, USA). The mean volume of sample flowthrough was 980 ml (sd: 140 ml). Following filtration, filters were capped, bagged in sterile whirl packs and frozen at – 20°C in a non-PCR lab. The tubing and gloves were changed between filters. Field blanks were processed in an identical manner, mean volume: 960 ml (sd: 40 ml).

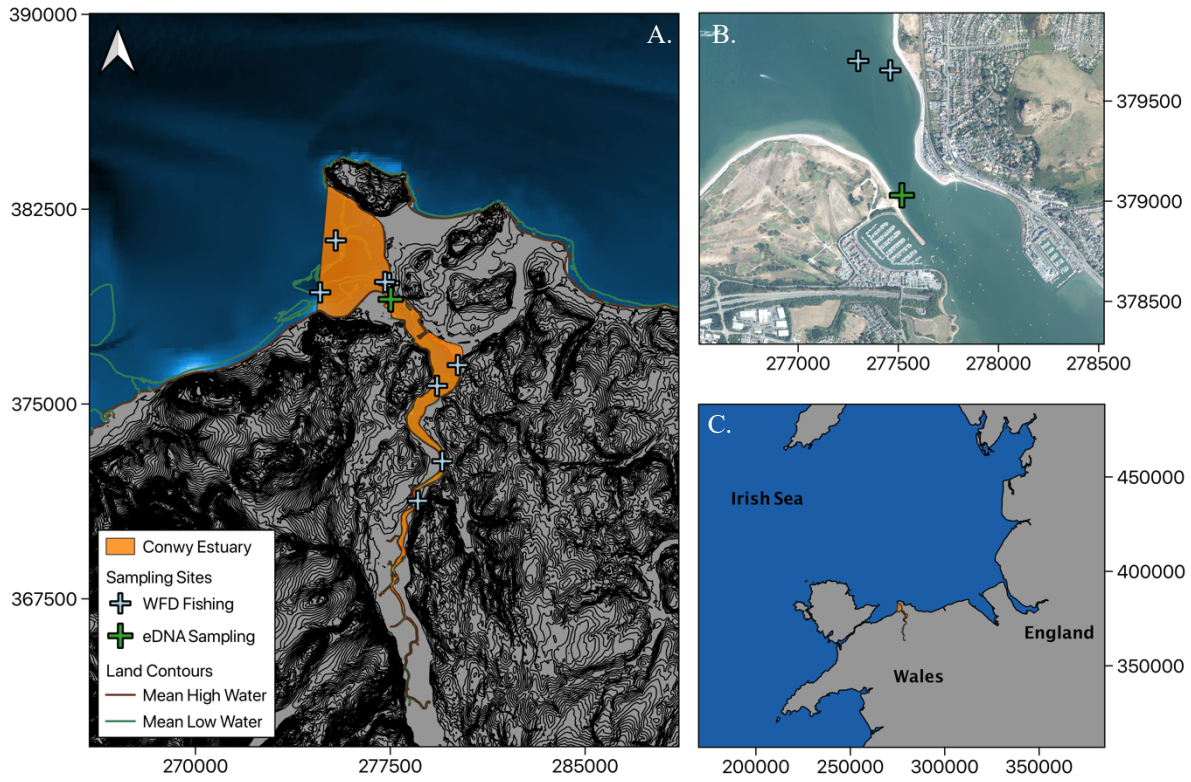


Figure 1: Map of the Conwy Estuary (A), the mouth showing the eDNA sampling site (B) and the geographic location within Britain (C). Coordinate System: British National Grid (EPSG:27700) axis in eastings/northings (m). Map data Sources: EMODnet Digital Bathymetry (DTM; EMODnet Bathymetry Consortium, 2018), land terrain contours (OS Terrain® 50; Ordnance Survey, 2021), British Coastline (Wessel & Smith, 1996 and 2017), Conwy Aerial Photography (Getmapping PLC, 2018) and Conwy Estuary Extent (Natural Resources Wales, 2019).

2.3 Laboratory - DNA Extraction

DNA extraction was carried out in a pre-PCR lab in an enclosed hood to reduce contamination. The workspace and equipment were sterilized with UV light for 20-30 mins to eliminate DNA contamination prior and after work. Batches of samples from different tides were extracted in a random order. Total DNA was extracted from each filter capsule using DNeasy Blood and Tissue Kits (QIAGEN, Hilden, Germany) following a modification of the Spens and Evans *et al.* (2017) protocol. Briefly, 720 µl of Buffer ATL (QIAGEN) and 80 µl Proteinase K was added into each filter and incubated overnight at 56°C to allow sample lysis. A flocculant solution was then used to remove humic acids (Sellers *et al.*, 2018). Briefly, a 880 µl subsample of lysate from each sample was added to 264 µl of flocculant solution, vortexed

and incubated for 30 mins at ~ 4°C. Each sample was then centrifuged at 10,000 x g for 2 min and 1000 µl of the supernatant removed (Sellers *et al.*, 2018). The rest of the DNA extraction followed Spens and Evans *et al.* (2017) and 70 µl of buffer AE (QIAGEN) was used for the final elution. Extraction blanks, consisting only of Buffer ATL and Proteinase K, were added at the sample lysis step and treated identically to samples. All plastic tubes used for handling eDNA were DNA LoBind (Eppendorf, Hamburg, Germany). DNA extracts were stored at -20°C in DNA LoBind tubes (Eppendorf, Hamburg, Germany) in the pre-PCR lab. Total DNA concentrations were measured using 1 µl of each sample on a Nanodrop ND-1000 Spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, USA).

2.4 Laboratory – PCR

Firstly, the 12S rRNA fragment was amplified with PCR using Tele02 primers (Miya *et al.*, 2015; Taberlet *et al.*, 2018). Set-up was conducted in a pre-PCR lab, any equipment and the enclosed PCR hood were irradiated with UV light for 20-30 mins to eliminate contamination. Samples, field, and extraction blanks were randomly ordered on a 96-well plate with five PCR negative controls (RNase Free-H_2O; QIAGEN) and three PCR positive controls (~ 0.2 ng/µl DNA Zebra mbuna cichlid, *Metriaclima zebra*) were included. Each PCR plate was repeated in triplicate. Total reaction volumes for each PCR was 25 µl containing 12.5 µl 2x QIAGEN Multiplex PCR Master Mix (QIAGEN), 1.25 µl of template (or control), 0.5 µl of forward and reverse Tele02 primer at 10 µM concentration and 10.25 µl of RNase-Free Water (QIAGEN). The thermal cycle profile was: *Taq* activation at 95°C for 15 min; 35 cycles of 94°C for 30 sec, 60°C for 1 min 30 sec, 72°C for 20 sec and finally 72°C for 10 min. All steps following the initial PCR were carried out on the bench in a post PCR lab. Amplification in PCR products was checked by gel electrophoresis. For each triplicate, ~15 µl from each reaction was pooled to account for PCR bias between plates. From each pool, 15 µl of product was cleaned of DNA fragments under ~200 bp (bead ratio: 1.75X) using ProNex® Size-Selective paramagnetic beads (Promega Corporation, Madison, WI, USA). Each cleaned PCR product was indexed and with its own unique i5/i7 dual index combination (Integrated DNA Technologies, Inc. Coralville, IA, USA) via a second-round of PCR. PCR master mix and indexes were combined inside a dead air cabinet before template addition to prevent contamination. Total reaction volumes for each PCR were 25 µl, containing 12.5 µl 2x QIAGEN Multiplex PCR Master Mix (QIAGEN), 3.0 µl of template, 1 µl of premixed i5/i7 indexes at 10 µM concentration (Integrated DNA Technologies, Inc.) and 8.5 µl of RNase-Free Water (QIAGEN). The second step thermal cycling profile was: 95°C for 15 min; 15 cycles of 94°C for 30 sec, 60°C for 1 min 30 sec, 72°C for 20 sec; 72°C for 10 min. Second-round PCRs were checked for amplification using gel electrophoresis. Following PCR, reaction concentrations were quantified using a Qubit™ dsDNA Broad Range Assay Kit (ThermoFisher Scientific, Waltham, MA, USA). Variation in successful PCR amplification across samples created challenges in pooling samples prior to sequencing. Therefore, following a failed sequencing run, three sequencing pools were created. Pool 1: reactions with identifiable bands in equimolar quantities. Pool 2: reactions without identifiable bands in equimolar quantities. Pool 3: control reactions pooled in equal volumes.

Each pool was loaded onto a 1.5% agarose gel, run for 50 mins at 74 V, imaged and the target amplicon excised and purified using a QIAEX II Gel Extraction Kit (QIAGEN) and quantified using a Qubit™ dsDNA HS Assay Kit (ThermoFisher Scientific). A final pool was created in the following arbitrary ratios: 70% Pool 1, 25% Pool 2 and 5% Pool 3 (referring to contribution to the final quantity of DNA). Finally, the purified pool, with a 10% PhiX spike, was sequenced using an Illumina MiSeq Reagent v2 kit (500 cycles; Illumina, San Diego, CA, USA). The library was sequenced initially using a nano kit, and then using a standard kit.

2.5 Bioinformatics

The data from the MiSeq and nano MiSeq run was combined. The quality of the demultiplexed files was assessed by generating a report for each sample and controls using *FastQC* (Andrews, 2010) and then assessing quality over all files using *MultiQC* (Ewels *et al.*, 2016). The Tele02 primers were then removed using *Cutadapt* (Martin, 2011). Quality trimming of the files then proceeded using *fastp*, this uses a sliding window method to drop low quality bases from each reads head and tail (in both 5' to 3' and 3' to 5' direction, mean quality = phred 30; Chen *et al.*, 2018). High phred scores were used to minimise the potential for the misidentification of a species due to sequencing error. The quality of the files was then reassessed with *FastQC* and *MultiQC* as before. The reads were denoised, dereplicated, merged and cleaned of chimaeras in *dada2* (Callahan *et al.*, 2016). For taxonomic assignment we broadly followed the methodology outlined by Collins *et al.* (2019). The exported Amplicon Sequence Variants (ASVs) were assigned an approximate taxonomy using a "global" BLAST search (evalue: 1×10^{-4}; Camacho *et al.*, 2009) against the 'nt' BLAST database (NCBI, 2021) to provide an approximate classification. Following this, all ASVs were searched using a "local" BLAST against the Meta-Fish-Lib, a curated UK fish database (Collins, 2021 and Collins *et al.,* 2021) to identify UK fish species (evalue: 1×10^{-5}). All the ASVs identified by both searches as fish were then assigned a taxonomy using an Evolutionary Placement Algorithm to verify the result of the local BLAST search (Barbera *et al.*, 2019) and GAPPA (Czech and Stamatakis, 2019; Czech *et al.*, 2020). ASVs were then assigned to species based on the following rules, see Collins *et al.* (2019) for rational, to ensure high confidence taxonomic assignments. Rule 1: Species level taxonomy was assigned if both the best scoring BLAST hit and species level EPA result were identical and sequence identity was ≥ 97%. Rule 2: Species level taxonomy was assigned if both the BLAST hit and the species level EPA results were identical, BLAST identity was ≥ 95% and EPA probability ≥ 90%. Rule 3: Species level taxonomy was assigned if only the BLAST hit only was 100% and no species had tied on their BLAST score. Following this, the BLAST and EPA results for any query sequences which had passed through these filters were manually appraised. Higher-level phylogenetic levels as determined from EPA, up to order level, were assigned to any remaining query sequences. Any assignments higher than this were rejected. Taxonomic assignments were checked to see if they had been detected in the WFD catch data described previously.

In addition to taxonomic assignment, species were assigned to estuarine use functional guilds. This followed the guild classification system for European estuaries (Franco *et al.*, 2008). Species were assigned to guilds based on this system initially, but where multiple guild assignments were present for individual species, classification from Elliott & Hemingway (2002) and Elliott & Dewailly (1995) was used. The only amendment was to assign Artic Charr (*Salvelinus alpinus*) to the Freshwater guild. Species not positively identified to species level could not be assigned to guilds and were marked as Unassigned but retained in the analysis. However, the *Melanogrammus aeglefinus*/*Merlangius merlangus* assignment could clearly be identified as whiting (*M. merlangus*), due to haddock being absent from the WFD catch data and whiting being commonly found from Autumn 2007 to 2015 (omitting 2013).

2.6 Hydrographic and Physicochemical Data

Prior to statistical analysis tidal height data for high and low tide was obtained for the survey period, from a local UK National Tide Gauge Network monitoring station, Llandudno (Lat: 53.33167; Lon: -3.82522). Sea level data is relative to Admiralty Chart Datum (British Oceaographic Data Centre *et al.*, 2020). Tidal range was calculated for each tide by taking the difference in sea level from the proceeding high or low tide and assigning it to the sampled

tide. To ensure quality control, physicochemical data from the AquaTROLL600 was plotted for each 20-30 sec reading and the first three measurements in each reading removed, as they were consistently erroneous. A mean and standard deviation for each tide was then calculated using data collected from all readings on that tide. All explanatory environmental variables were checked for collinearity, visually and using the Pearson correlation coefficient (Zuur *et al.*, 2007). Any variable with a correlation of 0.8 with another variable was considered colinear and dropped from the analysis (supplementary figure 1; Zuur, Ieno and Smith, 2007) and the most ecologically relevant variable retained in its place. After this process Salinity (PSU), pH, Turbidity (NTU) and Tidal Range (m) were used as the final explanatory variables. Salinity was correlated with temperature and sea level. Tidal range was correlated with date and time (supplementary figure 1). Although the selection of one colinear variable over another is arbitrary, all four variables are known to have an influence on fish communities and, or their eDNA (see Discussion).

2.7 Statistical Analysis

The fish reads from each replicate water sample collected on an individual tide were summed together to give a single sample per tide, this accounted for statistical none-independence. Each tide was treated as a single independent sample (Hurlbert, 1984).

Changes in assemblage composition between tides were compared by plotting the percentage contribution of each taxa to a sample and using ordinations implemented with generalised linear latent variable models (GLLVM; R Package: gllvm; Niku *et al.*, 2019). These are extensions of the generalised linear model to multivariate data using latent variables, which capture the main axes of covariation of abundance (or presence/absence) between samples (Niku *et al.*, 2019). Read count data used a negative binomial distribution (log link) and the log of the total number of fish reads per sample was used as a model offset (Zuur *et al.*, 2009), therefore, modelling relative read counts. Presence/absence data used the binomial distribution (logit link) and read depth was incorporated as a fixed factor. GLLVM's were implemented with 50 initial runs with random variation (jitter = 0.01) to minimise sensitivity to the chosen starting values (Niku *et al.*, 2019). The GLLVMs were fitted with 2 to 5 latent variables and model fit assessed using the AICc to check it was appropriate to display the ordination in two dimensions.

To directly test the influence of the four explanatory environmental variables on assemblage composition, and the response of individual species, multivariate GLMs were implemented using the mvabund R package (Wang *et al.*, 2012). Relative read counts and presence/absence data was modelled as with the GLLVM, with the exception that presence/absence models used the logit link function. Environmental variables were included as fixed factors. Analysis of Variance tests were conducted using the Wald test statistic and assuming correlation between species response variables with ridge regularisation. *P*-values were calculated using the PIT-trap resampling method (999 bootstrap iterations). *P*-values for univariate tests for each species level response were adjusted for multiple testing. Model selection was then applied using backwards selection and assessing the AIC (Zuur *et al.*, 2007). Model validation was conducted by plotting the residuals and plotting them against time and calculating the autocorrelation function for each taxa's residuals (Zuur *et al.*, 2007). Any explanatory variables with a statistically significant effect were visualised by plotting them on the ordinations.

To study the differential response of species in different estuarine use guilds to environmental variables, a trait-environment interaction model (or "Fourth Corner Model") using GLLVM was fitted to the relative read count data. Only the variable identified by

multivariate GLM as having a statistically significant effect on relative read counts was included (i.e. salinity, see Results), as these models are computationally demanding. This model allows us to test the interaction between environmental variables and species traits which may explain species responses (Niku *et al.*, 2019). The GLLVM was fitted with all four environmental variables as fixed factors, and an interaction between salinity and estuarine use guild (50 initial runs; jitter. = 0.01; 2 latent variables). Random slopes were added for each environmental variable to capture variation in environmental response not explained by the trait model to improve robustness to any potential missing predictors.

3. Results

3.1 Data Overview

A total of 14,997,384 paired reads were generated from the MiSeq and MiSeq nano sequence runs. A total of 2,983,122 (19.9 % of total raw reads) reads passed quality filters and returned a BLAST hit to chordates against the nt database and 385,428 reads (2.6 % of total; omitting the positive control) were assigned to fish present in the Meta-Fish-Lib database (supplementary table 1 and supplementary figure 2). Unfortunately, 16 samples (12 high and 4 low tide) contained no fish reads and were dropped out of the analysis. This resulted in zero reads being recovered in both replicates at high tide on day 5, 6, 7, 8 and 12 and at low tide on day 6. The remaining 44 samples contained median 3689.5 fish reads, but with substantial variation (IQR: 13113.3). After pooling of reads between replicates to form a single sample for each tide, high tide day 9 was identified as an outlier with 26 reads and dropped from the analysis.

In total, 1189 ASVs were identified by *dada2* and returned a BLAST hit against the nt database. 269 were assigned to chordates and 80 (omitting the positive control) passed the taxonomic assignment filters using BLAST and EPA and were identified as fish taxa by Meta-Fish-Lib. These were assigned to 55 taxa, 44 were identified to species level. 11 taxa were assigned to a higher taxonomic level (up to Order), 5 had no variation between closely related species (referred to as undifferentiated) and another 6 had no species level matches and all belonged to Families or Orders with unsequenced species. Assignments to higher-level taxa were retained in the analysis. No contaminating fish reads were detected in any field, extraction, or PCR negative or positive controls. In addition, ASVs associated with the positive control were not detected in any of the samples. No abundance filtering was conducted on any of the taxa as no quantitative threshold was available to conduct filtering. Rarefaction curves of species richness against read depth generated for each sample showed that most samples either had, or were beginning to, plateau in their species richness for the number of reads sequenced. This was the case even for samples with less than 5000 reads (supplementary figure 3 and 4). Therefore, the combination of samples in differing quantities during library preparation appears not to have biased the analysis. Overall, the species accumulation curve for all samples showed that the total expected mean species richness was beginning to saturate over the number of samples collected (supplementary figure 5).

3.2 Variation in Environmental Parameters

There was a clear shift in environmental conditions across the time series (figure 2). Tidal range increased from day 1 to day 5 (spring tide: day 5 – 6), before falling to day 13 (neap tide), then increasing again. Salinity was consistently higher at high compared low tide, but fell across both tidal states from day 6, before recovering slightly at high tide from day 10. There was a clear increase in mean river level upstream of the Conwy estuary, until day 10, which probably accounted for this decline in salinity (supplementary figure 6). Turbidity was

generally higher at low tide than high tide and showed a general increase from day 5 to day 10, with a substantial spike at low tide on day 6. Finally, pH showed a clear drop at low tide from day 9 to day 15.

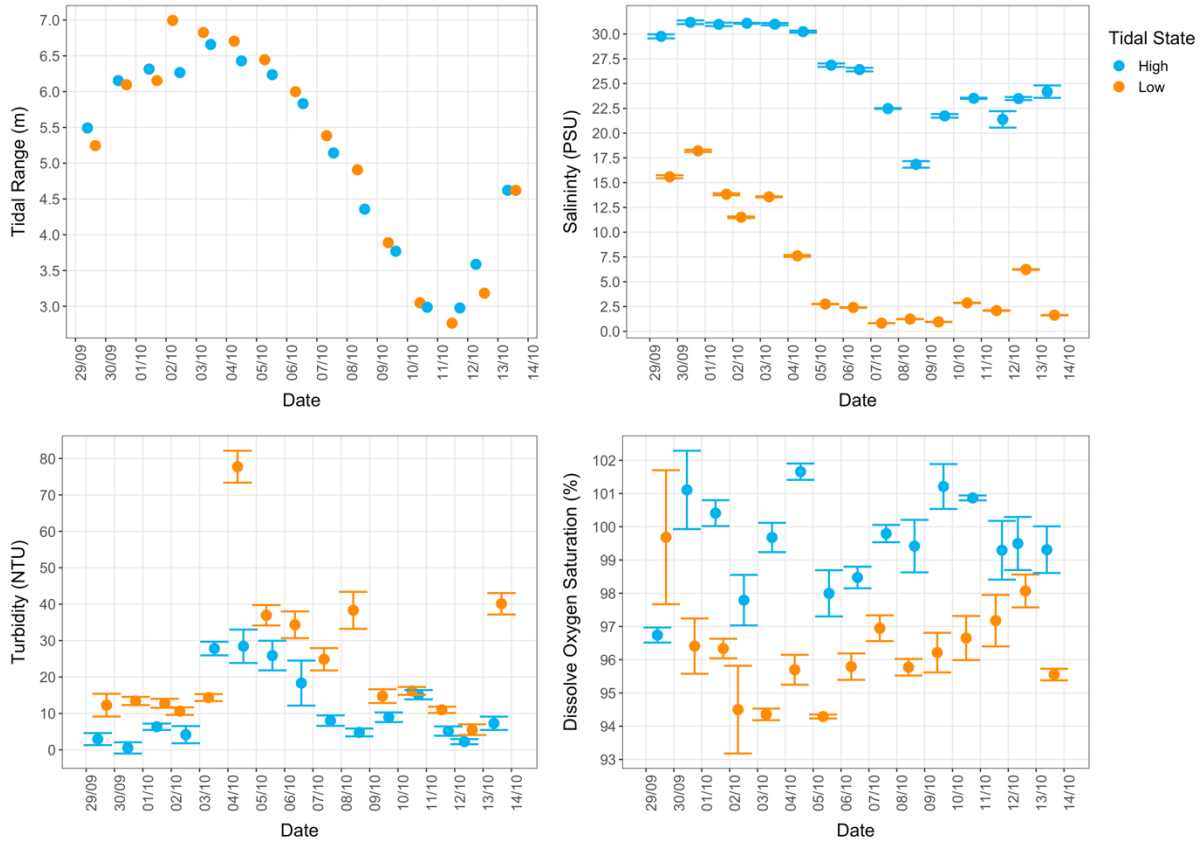


Figure 2: Environmental parameters: Salinity (PSU), Turbidity (NTU) and pH as mean and standard deviation and Tidal Range between the current and previous tide (m; relative to Admiralty Chart Datum) at high and low tide from 29/09/20 (Day 1) to 13/10/20 (Day 15).

3.3 Assemblage Structure and Historical Data

Of the 44 positively identified species, 28 had been detected in the estuary previously by catch based methods in early summer and autumn between 2004 and 2016. These species represented 88% of the fish reads in the analysis and included 8 of the 9 most abundant species (table 1). Further to this, it is likely the assignment for big-scale sand smelt (*Atherina boyeri*) was sand smelt (*A. presbyter*). Big-scale sand smelt are generally rare in British estuaries and this species was present in the historical data, but absent from the reference database. For the undifferentiated taxa, four of these five taxa contained representatives in the historical data. As mentioned before, *M. merlangus/M. aeglefinus* was identified as whiting. Only shad (*Alosa sp.*) was not present from historical data. For the six potentially unsequenced taxa, the Gobiidae assignment and the *Pomatoschistus sp.* assignment belong to a clade where an unsequenced species, *Pomatoschistus lozanoi*, had previously been detected in the estuary by WFD data. A further three taxa belonged to orders (or a lower-level taxonomic group) to which species previously detected in WFD data belonged to. So, it is possible these might represent rare local sequence variants or potentially degraded sequences rather than unsequenced species.

Table 1: Taxonomic and guild assignments for each species and comparison to WFD data

Scientific Name	Common Name	Total Reads (n)	Presence in Tide (n)	Contribution to Tides (mean %)	Estuarine Use Guild	Presence in WFD Data
Sprattus sprattus	European sprat	97807	19	25.15	MM	●
Phoxinus phoxinus	Eurasian minnow	62713	20	17.93	F	●
Salmo trutta	Brown trout	41813	19	14.02	A	●
Platichthys flesus	European flounder	34702	23	9.59	MM	●
Dicentrarchus labrax	European seabass	24596	21	9.68	MM	●
Gasterosteus aculeatus	Three-spined stickleback	18539	18	6.81	A	●
Salmo salar	Atlantic salmon	14452	16	5.99	A	○
Anguilla anguilla	European eel	13405	18	4.36	C	●
Pomatoschistus microps	Common goby	12448	20	4.04	ES	●
Ammodytidae	Sand lances	10804	21	3.34	UA	■
Merlangius merlangus	Whiting	8936	15	3.60	MM	●
Pomatoschistus minutus	Sand goby	7139	16	2.79	ES	●
Barbatula barbatula	Stone loach	5147	13	2.20	F	○
Trisopterus luscus	Pouting	4763	3	5.77	MM	●
Pholis gunnellus	Rock gunnel	3178	12	3.02	ES	●
Taurulus bubalis	Longspined bullhead	3131	12	0.91	MS	●
Clupea harengus	Atlantic herring	2959	14	0.84	MM	●
Ciliata mustela	Fivebeard rockling	2633	8	1.66	MM	●
Lipophrys pholis	Shanny	2499	8	2.46	MS	●
Pleuronectes platessa	European plaice	1951	19	1.06	MM	●
Cottidae	Sculpins	1485	3	2.45	UA	◇
Sardina pilchardus	European pilchard	1150	8	1.33	MM	○
Pomatoschistus pictus	Painted goby	967	4	0.90	MS	●
Limanda limanda	Dab	835	7	0.69	MM	●
Alosa sp.	River herring	755	1	1.01	A	○
Chelon sp.	Mullet	679	7	0.42	UA	■
Syngnathus rostellatus	Nilsson's pipefish	632	1	1.16	ES	●
Myoxocephalus scorpius	Shorthorn sculpin	575	7	0.52	ES	●
Belone belone	Garfish	567	1	1.04	MM	●
Pungitius pungitius	Ninespine stickleback	539	6	0.65	F	○
Gobiidae	True gobies	488	4	0.23	UA	◆
Scophthalmus rhombus	Brill	396	2	1.84	MS	○
Scomber scombrus	Atlantic mackerel	371	5	0.40	MS	●
Gadus morhua	Atlantic cod	312	3	0.14	MM	●

Table 1. Continued

Scientific Name	Common Name	Total Reads (n)	Presence in Tide (n)	Contribution to Tides (mean %)	Estuarine Use Guild	Presence in WFD Data
Buglossidium luteum	Solenette	281	3	0.41	MS	●
Oncorhynchus mykiss	Rainbow trout	261	2	3.30	F	○
Pollachius pollachius	Pollack	240	1	0.32	MM	●
Pleuronectiforme	Flatfishes	201	12	0.03	UA	◇
Symphodus melops	Corkwing wrasse	151	2	1.46	ES	○
Gadidae	Cods	144	1	0.26	UA	◇
Salvelinus alpinus	Arctic char	113	2	0.43	F	○
Echiichthys vipera	Lesser weever	85	1	0.83	MS	●
Liparidae	Snailfishes	75	1	0.92	UA	○
Spinachia spinachia	Sea stickleback	73	2	0.37	ES	●
Solea solea	Common sole	72	2	0.10	MM	●
Ciliata septentrionalis	Northern rockling	69	1	0.67	MS	○
Triglidae	Sea robins	65	1	0.63	UA	■
Pomatoschistus sp.	Goby genus	47	4	0.02	UA	◆
Raja clavata	Thornback ray	43	4	0.08	MS	●
Scyliorhinus canicula	Lesser spotted dogfish	34	6	0.06	MS	●
Scardinius erythrophthalmus	Rudd	28	1	0.44	F	○
Gobius paganellus	Rock goby	27	1	1.15	ES	○
Atherina boyeri	Big-Scale sand smelt	12	1	2.64	ES	○
Mustelus asterias	Starry smooth-hound	9	1	0.01	MS	○
Conger conger	European conger eel	6	1	0.01	MS	○

Presence in WFD Categories

Exact Species Present: ● For Identified Undifferentiated "Species" - Species Present: ■

No Species or Clade Present: ○ For Unidentified Species - Clade Present: ◇

For Unidentified Species - Unsequenced Species in Clade Present: ◆

3.4 Assemblage Composition

The negative binomial GLLVM (log link; two latent variables; 50 iterations) showed a clear difference in the relative read counts of fishes between high and low tides early in the time series (day 1 to 4), before and around the spring tide (figure 3). In contrast for day 10 to 15, slightly before and following the neap tide, high and low tides *generally* had similar assemblage compositions to low tides. There was also a clear shift in assemblage composition between high tide on day 1 to 4 and high tide on day 10 to 15. There was a comparable but more variable trend for low tides, on days 1 – 3 and 5 assemblage composition differed compared to most low tides later in the time series (figure 3). There was no substantial trends in model residuals (supplementary figures 8 and 9). The change in assemblage composition was clearly visible in graphs of percentage contribution to tides (figure 4 and 5). Prior to day 6, across both tides, reads were generally dominated by those of Marine Migrant species, particularly European spratt (*Sprattus sprattus*). However, low tides had a greater proportion of reads from Freshwater and diadromous categories. Following day 6 there was a substantial increase in the proportion of reads from the Freshwater and Anadromous guild across the samples. The freshwater species, Eurasian minnow (*Phoxinus phoxinus*) and the (potentially) anadromous brown trout (*Salmo trutta*) made up a large proportion of the reads. Although the proportion of Marine Migrant and Marine Stragglers species in high tide samples remained somewhat higher than low tide samples. Species presence/absences modelled with a binomial GLLVM (probit link, two latent variables, read depth as a fixed effect; 50 iterations), showed a general shift in fish presence/absences between high and low tides. Although with substantial overlap in composition between most high and low tides between day 1 and day 7. In addition, there was a shift in assemblage composition between high tides during the time series, which was also the case between the assemblage compositions of low tides (figure 6 and 7). There was no substantial trend in model residuals (supplementary figures 10 and 11).

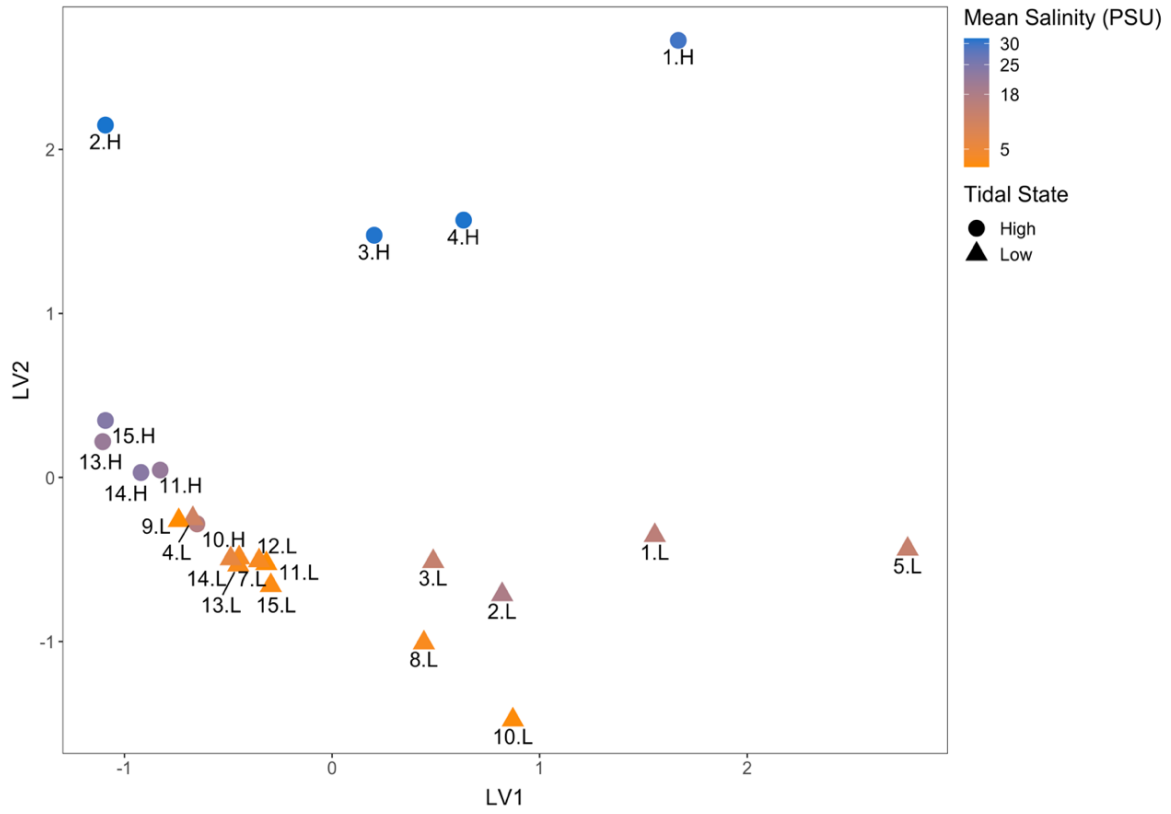


Figure 3: Relative read counts of fish taxa per tide modelled using GLLVM (log link; two latent variables; 50 iterations; offset: log total fish reads) with salinity (PSU; mean per tide) superimposed.

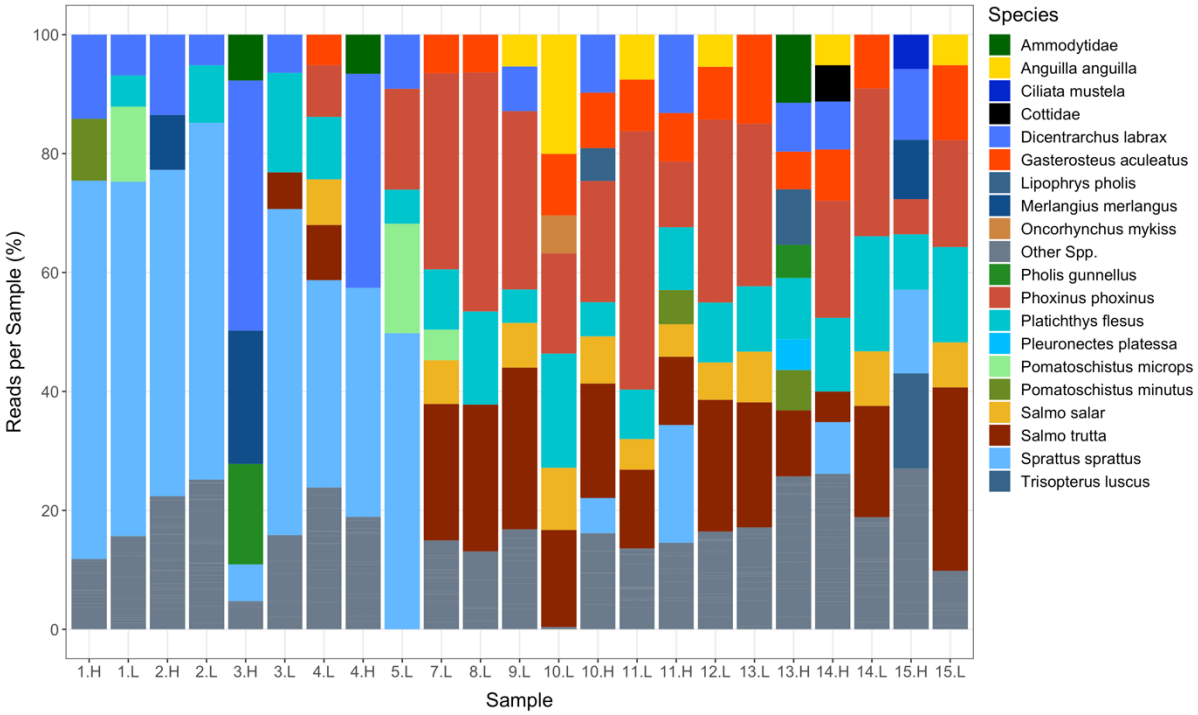


Figure 4: Percentage contribution of each species to total sample reads for each tide from day 1 (29/09/20) to day 15 (13/10/20), ordered by time. Species contributing less than 5% of total sample reads labelled as 'Other Spp.'.

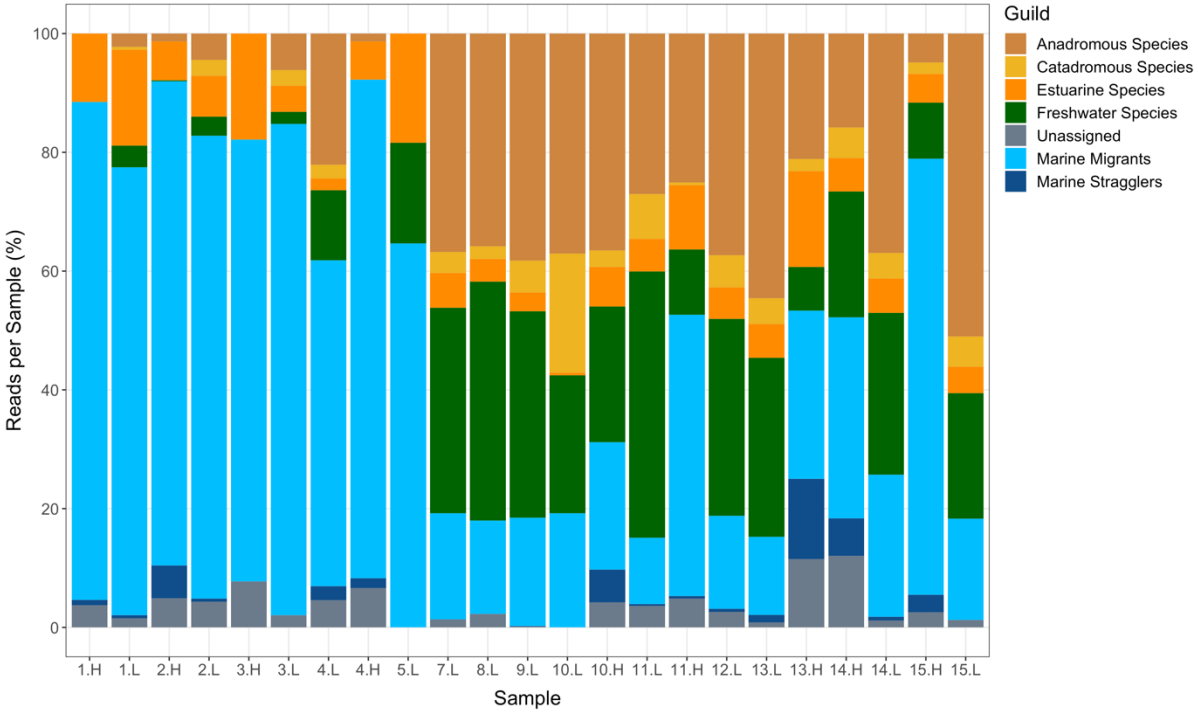


Figure 5: Percentage contribution of each estuarine use guild to total sample reads for each tide from day 1 (29/09/20) to day 15 (13/10/20), ordered by time. Species which could not be assigned a guild are marked as Unassigned.

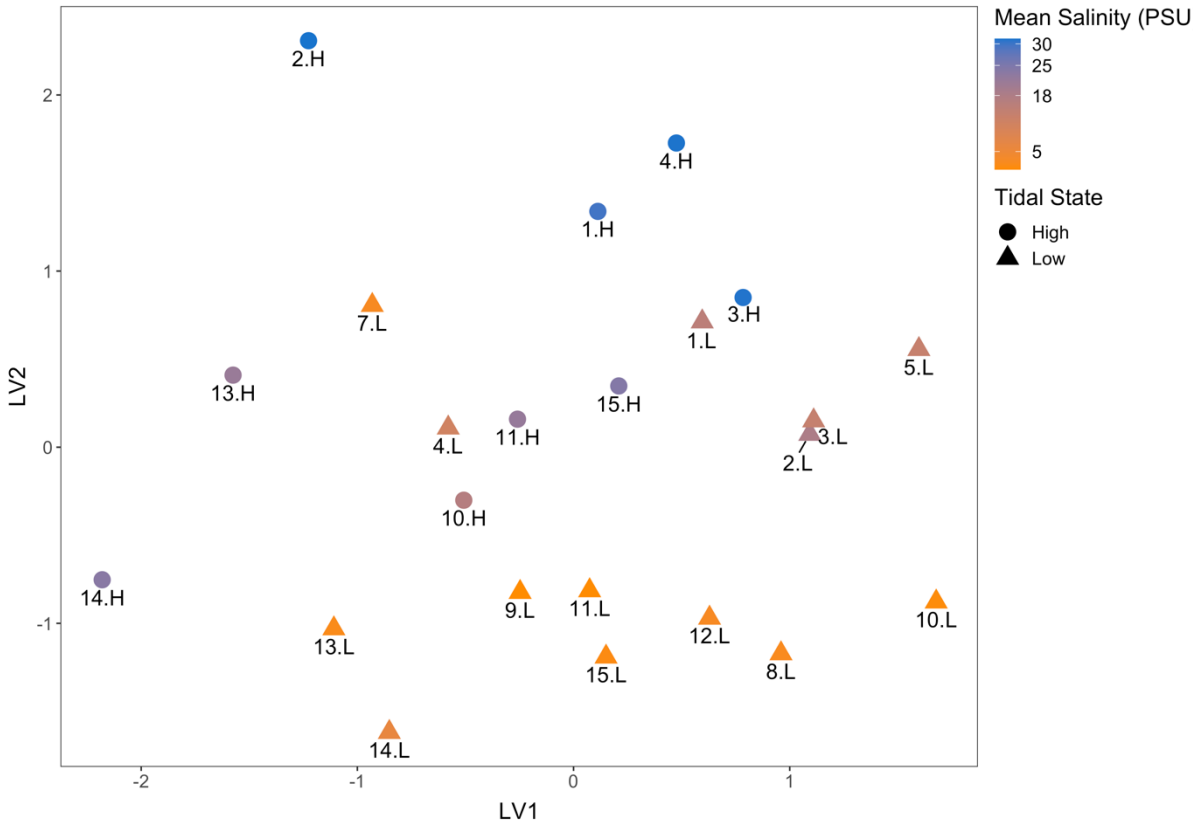


Figure 6: Presence/absence of fish taxa per tide modelled using GLLVM (log link; two latent variables; 50 iterations; total fish reads as a fixed factor) with salinity (PSU; mean per tide) superimposed.

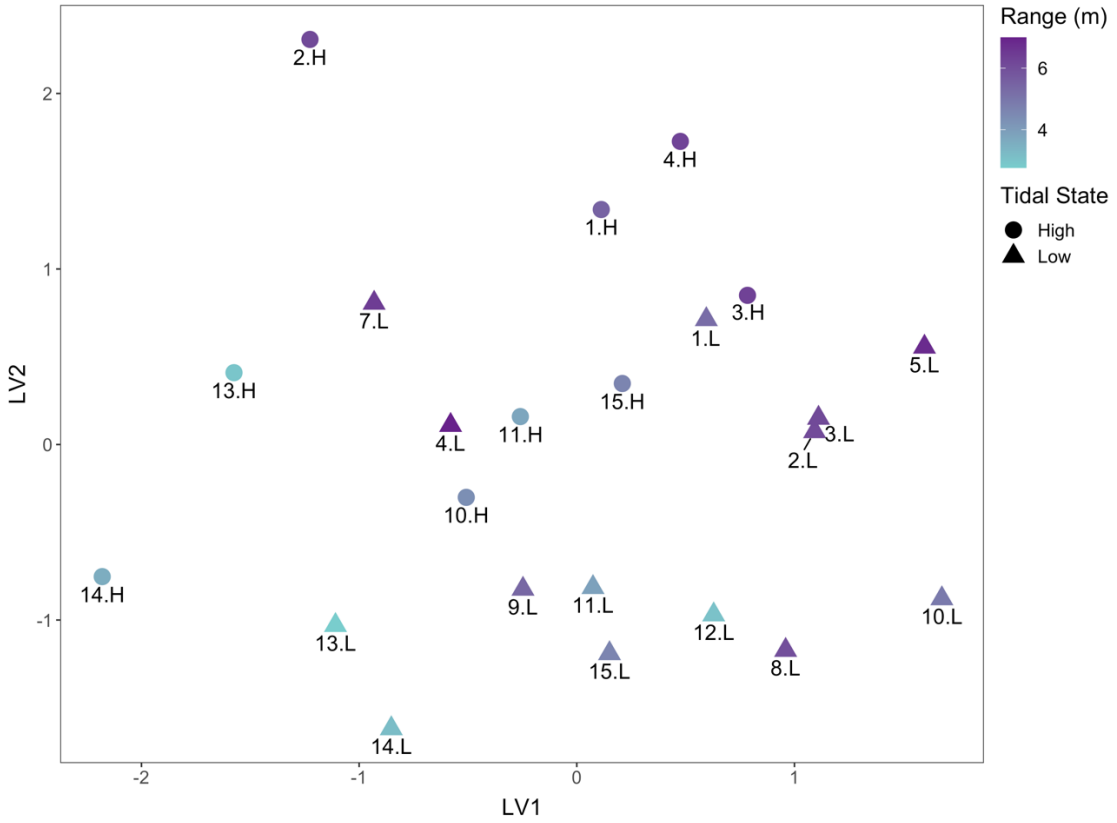


Figure 7: Presence/absence of fish taxa per tide modelled using GLLVM (log link; two latent variables; 50 iterations; total fish reads as a fixed factor) with tidal range (m) superimposed.

3.5 Environmental Variables and Assemblage Structure

The multivariate GLM of species relative read counts retained all four explanatory variables (salinity, tidal range, turbidity, and pH) following model selection. However, only salinity had a marginally statistically significant effect on the assemblage composition (table 2). Superimposing salinity onto the ordination of species counts showed a clear trend that tides with similar salinities had similar assemblage compositions (figure 3). However, on an individual level, no species showed a statistically significant response to any of the explanatory variables (*p*-values: >0.05). When considering only species presence/absences and including read depth as a fixed effect: salinity, tidal range, and read depth were retained in the optimal model and had a clear statically significant influence on the assemblage composition (table 2). Superimposing salinity and tidal range onto the ordination of presence/absences showed a shift in the community composition related to salinity and tidal range (figure 6 and 7). However, only the shanny (*Lipophrys pholis*) showed a statistically significant species level response, and only to tidal range (Wald Statistic: 2.601; *p*-value: 0.002). The proportion of *L. pholis* presences compared to absences on each tide declined as tidal range increased from ~ 3 to ~ 5.5 m (supplementary figure 7). All other environmental responses, and for all other species, were not statistically significant (*p*-value: > 0.05). It is likely that although the sample size was sufficient to detect an overall response of the fish eDNA assemblage composition, it was not large enough to consistently detect responses at the species level. There were no substantial trends in model residuals (supplementary figures 12 and 13).

Table 2: ANOVA for multivariate GLM – species relative read counts

Explanatory Variable	Residual DF	Wald-Test	*p*-value
Intercept	22		
Salinity	21	16.09	0.042 *
pH	20	10.89	0.308
Turbidity (NTU)	19	12.24	0.338
Range (m)	18	15.05	0.424

Significance codes: *** 0.001 ** 0.01 * 0.05

Table 3: ANOVA for multivariate GLM – presence/absence

Explanatory Variable	Residual DF	Wald-Test	*p*-value
Intercept	22		
Salinity	21	10.26	0.001 ***
Range (m)	20	7.96	0.002 **
Read Depth (n)	19	7.34	0.001 ***

Significance codes: *** 0.001 ** 0.01 * 0.05

3.6 The Interaction between Estuarine Use Guild and Salinity

Given the clear correlation between salinity and relative read counts a trait-environment interaction model was fitted using GLLVM (binomial distribution; log-link; 2 latent variables; 50 iterations) to study the differential response of estuarine use guilds to salinity, with tidal range, pH and turbidity retained in the model as fixed effects. There was a clear interaction between guild and salinity on relative reads counts. The inclusion of the interaction between guild and salinity had a statistically significant effect vs. a null model containing the environmental effects only (deviance = 19.44806, difference in degrees of freedom: 6, *p*-value: 0.0035). Plotting the model coefficients for each guild interaction with salinity (calculated relative to the interaction with the Anadromous guild) showed increased salinity had a positive effect on the relative read counts of Marine Stragglers, Estuarine Species, Marine Migrant guilds, and the Unassigned Species when compared to Anadromous species (figure 9).

Comparably there was no difference between the response of Anadromous species and Freshwater and Catadromous species. After re-levelling and re-running the model using Unassigned species as the comparison for calculating coefficients, it was shown that the relative reads of Freshwater, Anadromous and Catadromous guilds had a negative relationship with salinity. Whereas there was no difference between the responses of Unassigned species, Marine Stragglers, Estuarine Species and Marine Migrant guilds (figure 9). There were no substantial overall trends in model residuals, but one outlier was present (supplementary figure 14 and 15).

The results from all multivariate GLM and GLLVM models are robust to the inherent temporal non-independence between tides, as there was minimal observed cyclical temporal-autocorrelation. There were no strong patterns in the residuals vs. time plots overall for GLLVMs (supplementary figure 9, 11 and 15). In addition, the autocorrelation function calculated for each species residuals, for all models, showed correlations mainly below 0.4.

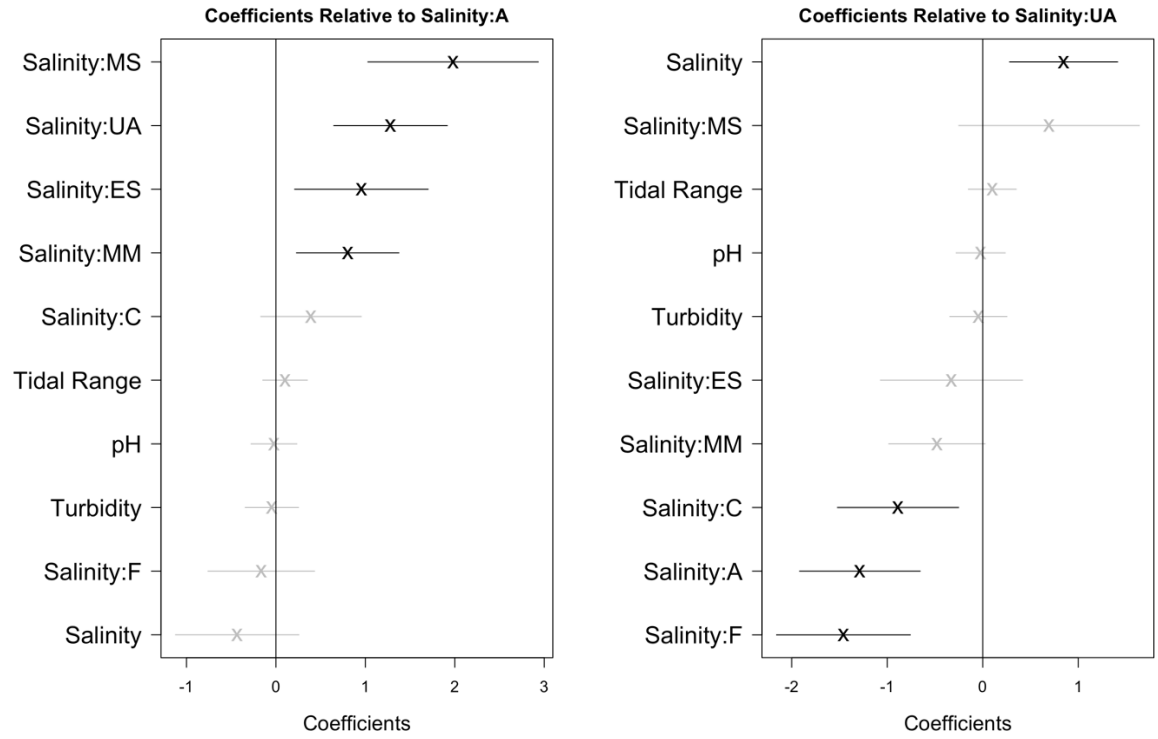


Figure 9: Estimated coefficients (crosses) and their 95% confidence intervals (lines) for all variables and interactions in the trait-environment interaction model between estuarine use guild and salinity, describing relative reads per species (binomial distribution; log-link; 2 latent variables; 50 iterations). Guilds are as follows, A: Anadromous species, C: Catadromous species, F: Freshwater species, MM: Marine Migrants, MS: Marine Stragglers, ES: Estuarine Species, UA: Unassigned. The main effects of salinity, tidal range, pH and turbidity on fish relative reads are also included. The left-hand diagram shows the coefficients calculated relative to the coefficient for Anadromous Species, the right-hand diagram shows the coefficients calculated relative to the Unassigned Species (hence their respective guild interactions are not shown).

4.Discussion

4.1 Overview

Overall, this study has revealed the short-term variability that can occur in the fish eDNA assemblage composition detected with metabarcoding using a Teleost specific assay, within a well-mixed, macro-tidal estuary. This study is also one of the first fish eDNA studies in estuaries to classify species according to their functional use of estuaries, using estuarine use guilds. A clear differentiation in assemblage composition was detected between tidal states, particularly early in the time series prior to and around the spring tide. Comparably for the time points just before and following the neap tide, assemblage composition was more similar as was initially predicted. Strikingly, the most substantial change in composition, particularly when taking the relative read counts of species into account, occurred after day 6 in the study as the tidal range began to fall. Given the correlation between the assemblage composition and salinity (colinear with sea-level and temperature) it seems likely this shift occurred in part due to an increase in the amount of freshwater entering the estuary. This complicates inferences regarding the influence of the spring-neap cycle. Interestingly, different estuarine use guilds showed different relationships with salinity and their relative read counts. There was also an effect of tidal range on species presence/absence. It was not possible to detect an effect of pH or turbidity on the assemblage composition.

4.1 Methodological Considerations

Firstly, there are inherent constraints in the inferences which can be made from this study due to the experimental design and analytical method. The results are only representative of eDNA in surface water, there is no data on fish eDNA gradients with depth. However, the Conwy is vertically well-mixed (Robins *et al.*, 2014), so substantial gradients in eDNA seems unlikely. Similarly, the design was not spatially replicated and results are representative only of the estuary mouth. More generally, no contemporary fish data was collected, partially due to COVID-19 restrictions. Therefore, it cannot be definitively determined if any patterns in the eDNA assemblage composition also occurred in the real fish assemblage. Despite this, 88% of the total fish reads in the dataset came from 28 species known to occur in the estuary from historical fish data which suggests a substantial portion of the eDNA could be of local origin. Clearly, this remains an area for further study. In addition, it should be clearly stated the data on relative fish reads are interpreted only as changes in the relative proportion of fish eDNA. Although most fish metabarcoding studies report positive correlations between read counts and abundance and/or biomass numerous factors influence this relationship. Read counts are not precise indicators of fish biomass and abundance (Rourke *et al.*, 2022).

Secondly, only 2.6% of total paired end reads were assigned to fish despite 19.9 % of reads being assigned to Chordates (mostly mammals). Other studies using Tele02 primers have shown higher fish read yields (Aglieri *et al.*, 2021; Zhang *et al*., 2020). However, Tele02 primers amplify a wide variety of vertebrate groups *in silico* (Zhang et al., 2020) and this has been shown in estuarine water samples (Mariani *et al.*, 2021). Further to this, water samples analysed with Tele02 from the coastal Mediterranean yielded only 8.7% of total reads from Mediterranean fish (Aglieri *et al.*, 2021). It is likely that high yields of none-fish chordate reads were due to high concentrations of eDNA from non-target groups being washed into the estuary from the surrounding watershed. This is evidenced by an increase in the percentage contribution of fish reads at high tide compared to low tide (supplementary figure 2). This is expected if transport of non-fish eDNA into the estuary from the river is maximal at low tide. Despite the low recovery of fish reads, sample level rarefaction curves of species richness vs.

depth showed this has not substantially impacted inferences of the eDNA fish assemblage (supplementary figure 3 and 4). An initial technical fix would be to investigate the use of the MiFish U primers (Miya *et al.*, 2015), which amplify a lower percentage of non-target vertebrate groups compared to fish *in silico* (Zhang *et al.*, 2020b).

Thirdly, due to poor amplification, no fish reads were retrieved for high tides on day 5 to 8 and 12, or either tide on day 6, therefore limiting the statistical analysis. High tides probably had a lower concentrations of fish eDNA than low tides, which may have been exacerbated by the low specificity of the primers. In a tidal inlet of the estuarine Wadden Sea, the total vertebrate eDNA concentration was slightly higher, although more variable, at low tide than high (Bleijswijk *et al.*, 2020). In addition, the dilution of fish eDNA may reduce species detectability after high rainfall in a neotropical river (Sales *et al.*, 2021). It seems likely therefore that the increased volume of the estuary at high tide led to an analogous process. The concentration of suspended particulate matter (SPM) in estuaries generally falls as the tide rises and the water volume of the estuary increases (McLusky and Elliott, 2004). Fish eDNA is generally concentrated within SPM particles between 1 – 10 µm in size (Turner *et al.*, 2014; Wilcox *et al.*, 2015; Jo *et al.*, 2019) and SPM transport in rivers is analogous to fine particulate organic matter (Wilcox *et al.*, 2016; Pont *et al.*, 2018). Although nephelometric turbidity, a relative index and difficult to generalise (Davies-Colley and Smith, 2001) was measured here instead of SPM, turbidity does have a generally strong positive relationship with SPM (Davies-Colley and Smith, 2001; Jafar-Sidik *et al.*, 2017). Over the study, turbidity was generally higher at low tide than at high tide, supporting the supposition that fish eDNA was also diluted. The potential aggregation of fish into subtidal areas at low tide (Greenwood and Hill, 2003), may have also contributed to greater detection. A simple solution to low yields of reads at high tide would be to filter more water. It is unlikely that inhibition played a role in the poor amplification of most of the high tides, given their low turbidity and that the assay had been optimised to reduce inhibition (Chapter 2 Appendix A). However, on day 6 (close to the spring tide) turbidity increased across both tides, but particularly at low tide, so inhibition seems more likely in this context. Turbidity potentially increased due to river water flushing the position of the Estuarine Turbidity Maximum downstream as observed in other estuaries (Grabemann and Krause, 2001; McLusky and Elliott, 2004). In addition, SPM concentrations have been shown to increase in other estuaries around a spring tide (Lindsay *et al.*, 1996; Grabemann and Krause, 2001). Therefore, these two events probably drove the increase in turbidity and potentially inhibition around day 6.

4.2 Fish Assemblage Composition

Despite methodological constraints, a convincing eDNA fish assemblage was detected. Of the 55 taxa detected 44 species were clearly detected, 28 of which had been detected by WFD fish surveys previously and accounted for the majority of reads in our dataset. In addition, another 4 of the identified but undifferentiated taxa had representative species detected in the estuary by past WFD surveys. Therefore, this suggests the eDNA of these 32 taxa could have been produced locally within the estuary. These 32 detected taxa represented the full complement of estuarine use guilds present in European estuaries (Franco *et al.*, 2008). In addition, the total numbers of taxa identified was comparable to the maximum species richness estimated for the Conwy, 42 (95% Confidence Limits: 38 to 60) from seine netting data (2006 to 2015; Waugh et al., 2019). However, a proportion of the eDNA must have been transported into the estuary from the sea and the river. The transport of eDNA is a well-established phenomenon and can occur in rivers over scales of a few 100 meters to over 100 km, depending on the size and hydrography of the river (reviewed in Pont *et al.*, 2018). Only one species, Artic Charr (*Salvelinus alpinus*) was an undoubtedly clear example of downstream transport. It

should not occur in the estuary, but does exist within the catchment. All the other species detected have the *potential* to be present in the estuary (Franco *et al.*, 2008). It is also possible that many of the edible fish species detected, e.g. European Pilchard (*Sardina pilchardus*), Atlantic Salmon (*Salmo salar*) etc., could be inputted into the estuary from wastewater outflows from local settlements.

4.3 Fish Assemblage Composition and Environmental Effects

The analysis showed that changes in assemblage composition occurred due to changes in salinity (colinear with temperature and sea-level) and to a lesser extent tidal range. The effect of salinity on the assemblage structure was clearly seen in the relative reads of species/taxa and their presence/absences which showed a similar overall trend. Although the analysis could not detect any individual species level responses, salinity was positively associated with the relative read abundance from Marine Migrant, Marine Straggler and Estuarine Species guilds as shown by the trait-environment interaction model. Comparably the relative reads of species in Freshwater and diadromous guilds increased as salinity fell. Clearly such an approach has caveats given species may be assigned to multiple guilds categories (Franco *et al.*, 2008), but classification of fishes into single categories is a standard approach (Harrison & Kelly, 2013). The sources of variation in salinity in this study were the regular rise and fall of the tides and an episodic increase in freshwater inputs from the river, which occurred as tidal ranges fell, causing an overall decline in salinity across tidal states in the estuary. Therefore, as salinity cyclically increased and decreased in the early part of the time series, so the relative composition of marine species increased and decreased. These data support preliminary results suggesting eDNA assemblage composition can change with tides in the macro/meso tidal Elbe (Schwentner *et al.*, 2021). However, the overall decline in salinities around day 6 precipitated an overall shift to an eDNA assemblage dominated by freshwater and diadromous fishes. The concurrent shift in assemblage and salinity is comparable to observations that changes in eDNA assemblages are correlated with changes in physicochemical parameters, e.g. salinity and temperature, associated with specific water masses in a Fjord (Kelly *et al.*, 2018). These changes probably occurred as combination of changes in eDNA transport, eDNA resuspension from sediment caused by the movement of water (Bleijswijk *et al.*, 2020) and shifts in fish distributions. Salinity is a key driver of the spatial distribution of fishes in estuaries (Thiel *et al.*, 1995; Marshall and Elliott, 1998) and its influence may contribute to changes in the distribution of fishes as the tides rise and fall along with the net movement of water (see intro; Greenwood & Hill, 2003). Whereas it is also intuitive that changes in the relative composition of eDNA transported into the ecosystem from the river and sea accounted for a proportion of the correlation between salinity and community composition at different states of the tide, along the time-series. High rainfall events and flooding can impact the discharge, dilution and source of eDNA in ecosystems (Harrison *et al.*, 2019). If large enough, flood events can reduce the abundance of marine taxa by forcing them to emigrate from the estuary to escape rapid declines in salinity (Gillson, 2011). Therefore, it is unsurprisingly that the eDNA assemblage changed as freshwater inputs increased. Clearly further research with concurrent measurements of fish abundance and composition are required to disentangle these relative effects.

Secondly, compared to the effect of salinity, tidal range only influenced the assemblage composition as measured by presence/absences. Its effect was probably too subtle to be detected in the analysis incorporating relative read abundance. This may suggest its direct influence was greater on species with lower read abundances, which will have more influence on a presence/absence analysis than an quantitative one (Clarke and Warwick, 2001). As noted in the introduction, changes in tidal range can lead to changes in the available habitat for species (Wilson and Sheaves, 2001). It is notable that the only species level correlation which could be

detected was the intertidal Shanny (*L. pholis*), which showed a correlation with low tidal ranges. This is interesting given this fish species shows activity rhythms in response to tides (Northcott, 1991). Changes in the spring to neap cycle are also likely to also have influenced transport into the estuary.

Compared to salinity and tidal range, pH and turbidity had no overall effect on assemblage composition. Measured pH did fall overall from generally above 8, to generally between 8 to 7.25 from day 8 at low tide onwards, probably due to increased freshwater inputs. Low pH has been associated with the degradation of eDNA. But the values at which pH has been observed to have had an effect (Strickler *et al.*, 2015; Seymour *et al.*, 2018), were below the minimum pH value in the present time-series. In addition, given these declines in pH were cyclical and temporary it seems unlikely any change in pH would have much time to influence community composition compared to changes in predominant drivers such as eDNA transport and fish distributions. For example, time-series analysis of Silver Carp (*Hypophthalmichthys molitrix*) eDNA detection showed only reverse river flow volume influenced detection out of a suite of environmental variables, including pH and temperature (Song *et al.*, 2017). Turbidity had no effect on the fish eDNA assemblage, the patterns in turbidity and caveats with this parameter have already been discussed. Turbidity has been shown in influence fish assemblage structure in inshore areas and estuaries (Blaber and Blaber, 1980; Cyrus and Blaber, 1992), probably due to its effect on feeding and predation (Blaber and Blaber, 1980). Although it does not have an effect in every estuary (Marshall and Elliott, 1998). Variation in turbidity has been associated with changes in fish assemblage eDNA composition between, and within water masses, in St. Lawrence River (Quebec, Canada) at small spatial scales (Berger *et al.*, 2020). The most parsimonious explanation for the result in the present study is the loss of data during the middle of the time series which covered a large fluctuation in turbidity. This probably reduced the power of the analysis to the point where not effect could be detected.

4.4 Implications for Management

This study is focused on characterising variation in the eDNA assemblage rather than providing data of immediate use to environmental managers. Further research and the development of specific criteria is required to filter the data to eliminate species which are unlikely to be truly present in the estuary. However, the vast majority of reads were from species which had a direct match in the WFD data, therefore the conclusions relating to relative read abundance should be robust. Aside from any methodological points made earlier, point sampling of eDNA should be standardised to time points with similar salinities i.e., the same tidal state, avoiding episodic high flow events and ideally at similar time points in the spring-neap cycle to avoid large fluctuations in the eDNA assemblage composition. Accounting for temporal variation in environmental factors is already best practice in surveys of fishes in estuaries using direct capture methods (Hemingway and Elliot, 2002), although clearly episodic changes in river flow are harder to plan for. Comparably, to detect the full breadth of the eDNA assemblage using the present methodology, sampling would have to cover a range of different salinities. A related point is that any fish eDNA assemblage data needs to be interpreted in the context of the hydrographic and physicochemical conditions in the estuary at the time of sampling. Collection of this data is critical and comparisons between ecosystems without such background data are likely to provide limited, or potentially erroneous, ecological insight.

5. Conclusion

This study has clearly described the variability in a fish eDNA assemblage within a macro-tidal estuary at high and low tide over a period of 15 days and covering a spring to neap cycle.

Changes in this assemblage were largely due to changes in marine influence, due to the tide and the effect of an episodic river event, with demonstrable links to salinity. Overall, this study suggests that the influence of short-term variation in assemblage composition driven by hydrographic factors such as tides are likely to be substantial, context-dependent, and certainly require more empirical research. Further research into the relative response of fish assemblages in estuaries as measured with eDNA and capture-based methods will allow scientists and manager to better interpret the biodiversity data which is generated.

Funding Organisations

Aside from the core KESS2 funding, the majority of the funding for the practical component of this project was provided by the Fisheries Society of the British Isles (FSBI) via a Small Research Grant. SC would also like to acknowledge the support of NERC Highlight Topic Grant NE/N006216/1.

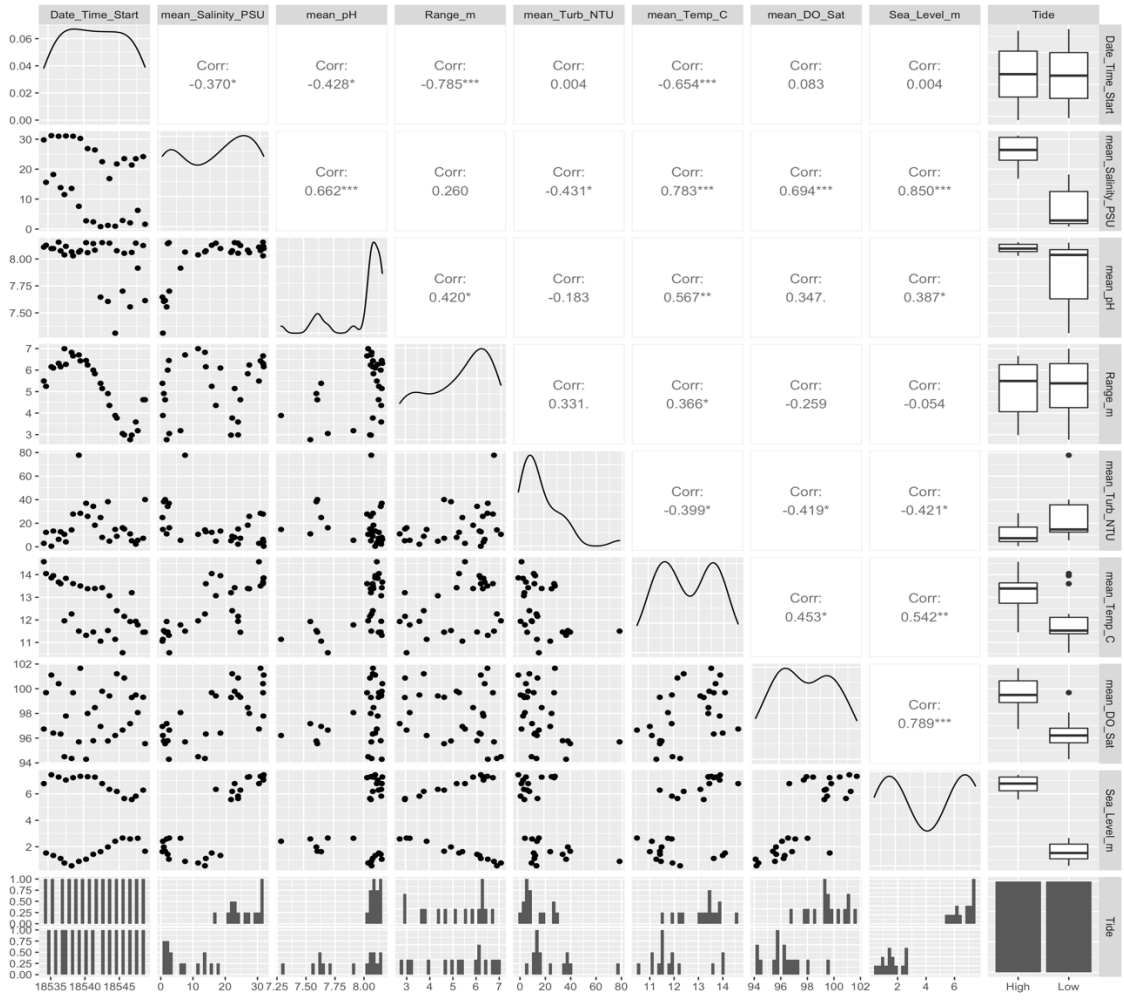
Acknowledgements

T. I Gibson would like to thank and acknowledge the following individuals for their invaluable help on this project. Tom Major, Bethan Pugh, Rebecca Irvine, and Peter Hughes for their excellent support in, and relating to, the field sampling. Dr Rupert Collins for his excellent bioinformatic advice and providing clear examples of complete bioinformatic pipelines via Github and Dr Molly Czachur for providing equivalent data to practice on. Dr Suzie Jackson for insightful discussions on the Conwy estuary and SPM sampling, regrettably T.I. Gibson did not have the time to conduct this work. Stefanie Krafft and Ian Pritchard for allowing me to prepare SPM filters in their laboratory. Finally, Dr Graham Sellers, Dr Robert Donnelly, and Dr Charles Baillie for their excellent advice on molecular biological matters at an early stage of this project. This work would not have been possible without the support of the KESS2 staff and the FSBI's exemplary attitude to supporting research into fish ecology.

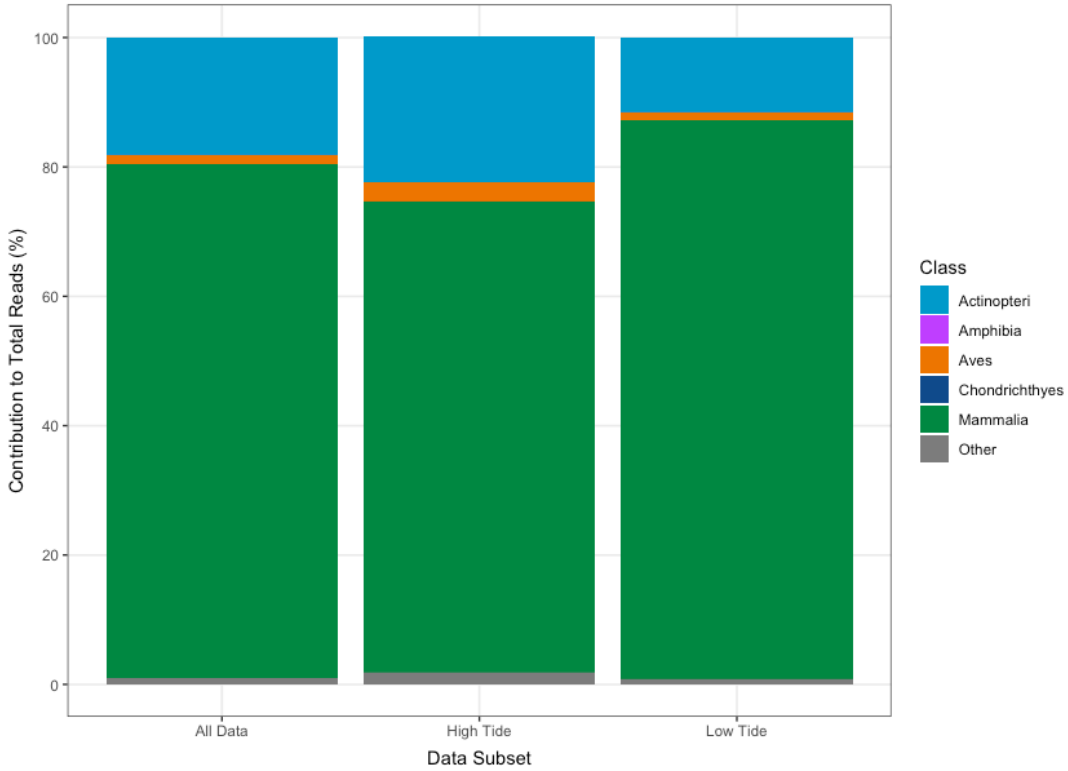
Chapter 3: Supplementary Material

Supplementary Table 1: Number of Paired-end Reads at each step of bioinformatic pipeline.

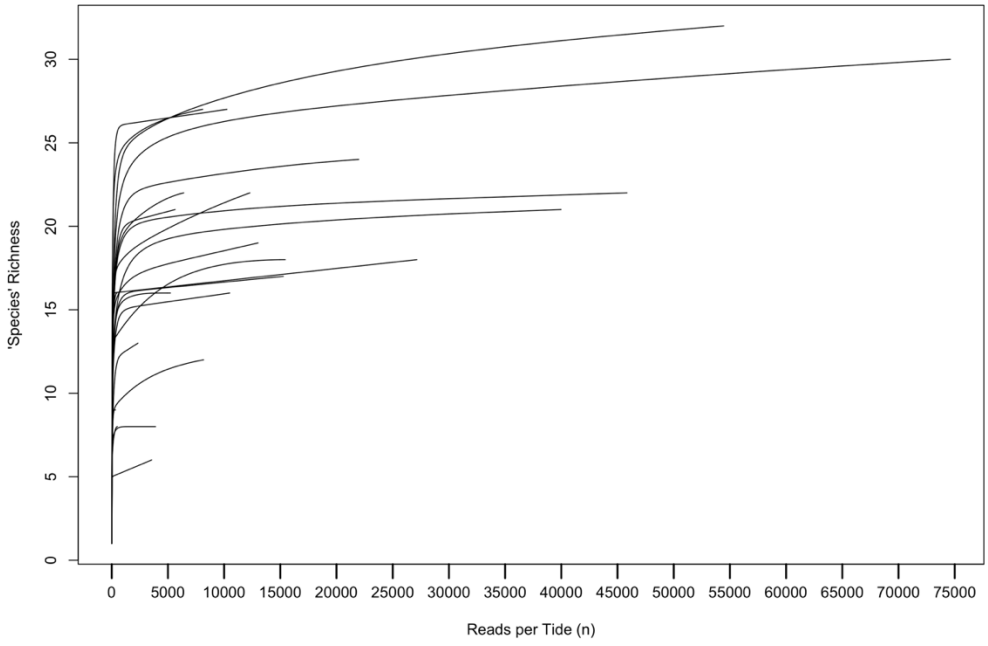
Step Number	Pipeline Step	Number of Paired-end Reads	Percentage of Total Paired End Reads Sequences (%)
-	Standard MiSeq Sequencing Run	14708556	98.1
-	Nano MiSeq Sequencing Run	288828	1.9
1	Combined Reads	14997384	100.0
2	Primers Removal	13699852	91.3
3	Quality Filtering	3946097	26.3
4	Denoising	3857518	25.7
5	Merging Paired-end Reads	3603840	24.0
6	Chimeras Removal	3399836	22.7
7	Reads matched to nt BLAST (All ASVs)	3008877	20.1
8	Reads matched to nt BLAST (All Chordates)	2983122	19.9
9	Reads matched to Meta-Fish-Lib (All Data)	540890	3.6
10	Reads matched to Meta-Fish-Lib (No Positive Controls)	385428	2.6



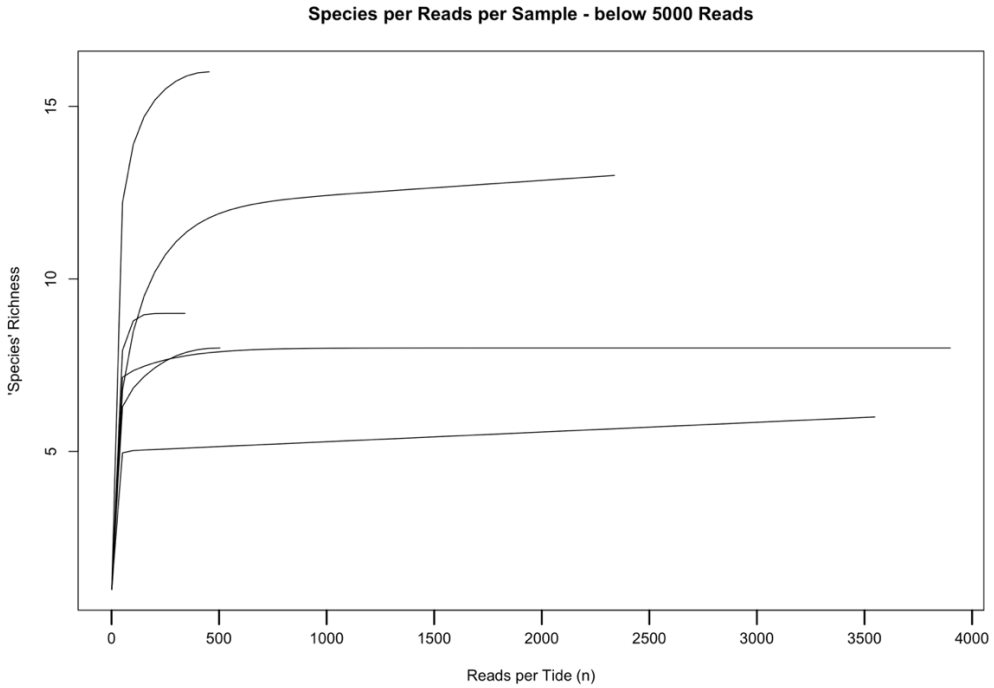
Supplementary Figure 1: Correlation between each combination of environmental variables, Pearson correlation coefficient calculated between continuous variables.



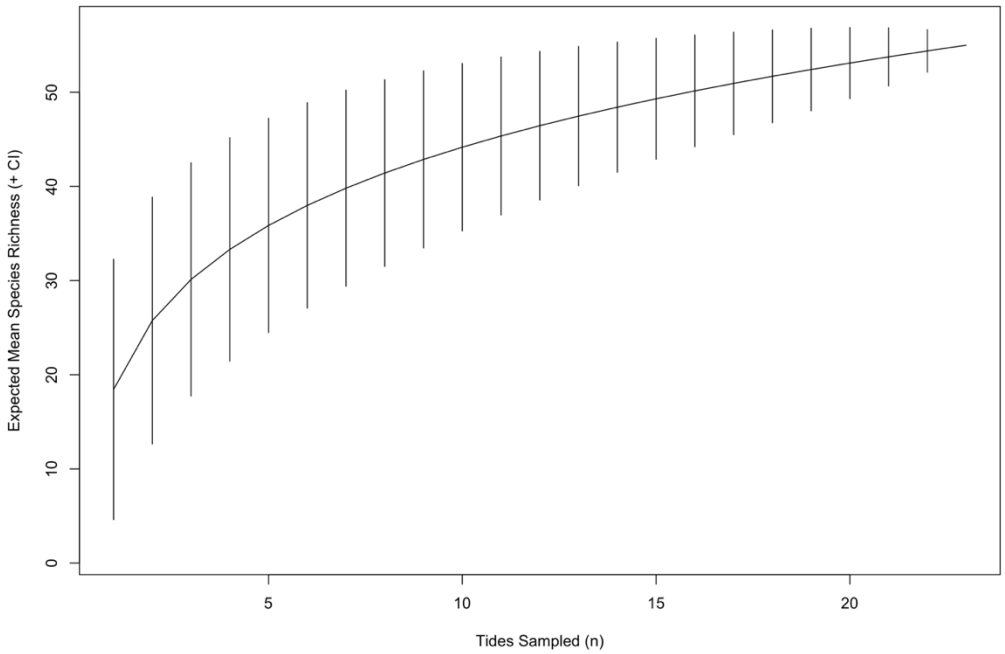
Supplementary Figure 2: Overall percentages of reads belonging to each Chordate Phyla (all other Phyla shown as Other) present in samples after global BLAST search against the "nt" BLAST database (Step 7. in Table 1.). Taxonomy assigned according to highest scoring BLAST hit.



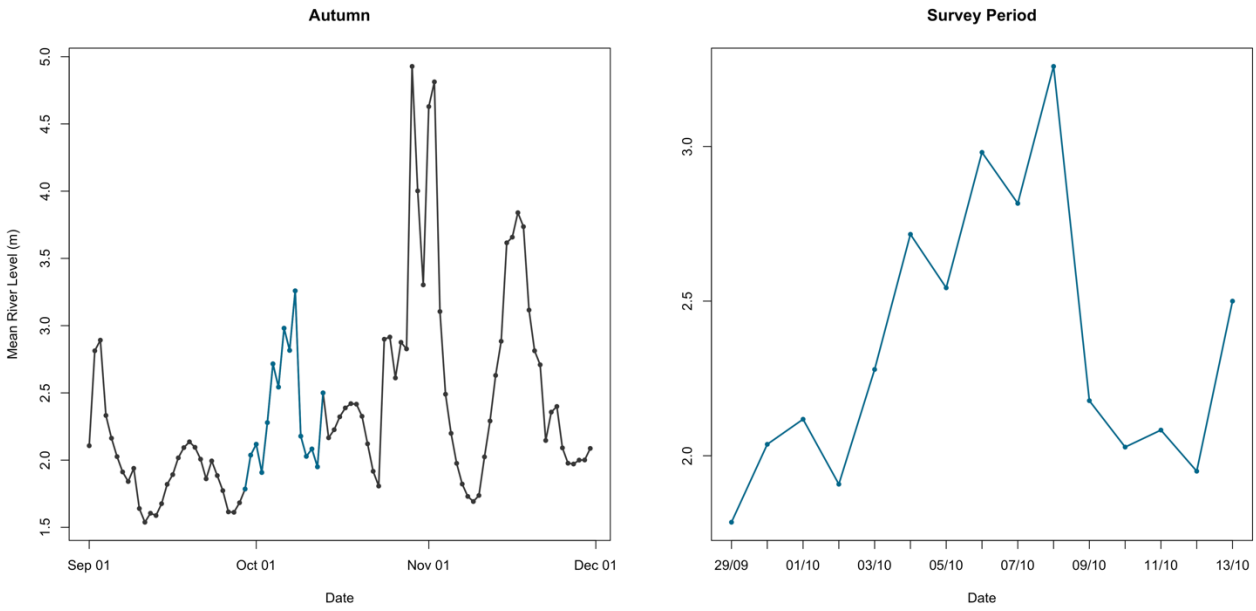
Supplementary Figure 3: Rarefaction curve of fish 'species' richness against fish read depth summed over each tide. Calculated using rarecurve(step = 50) in Vegan (Oksanen *et al.*, 2015).



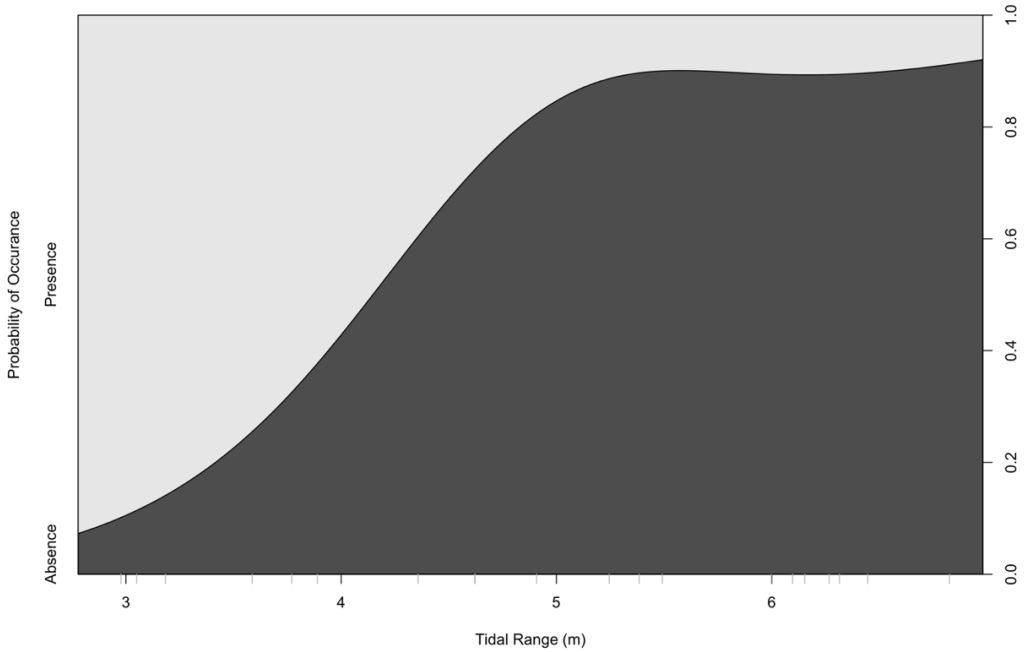
Supplementary Figure 4: Rarefaction curve of fish 'species' richness against fish depth summed over each tide, for tides with < 5000 total fish reads. Calculated using rarecurve (step = 50) in Vegan (Oksanen *et al.*, 2015).



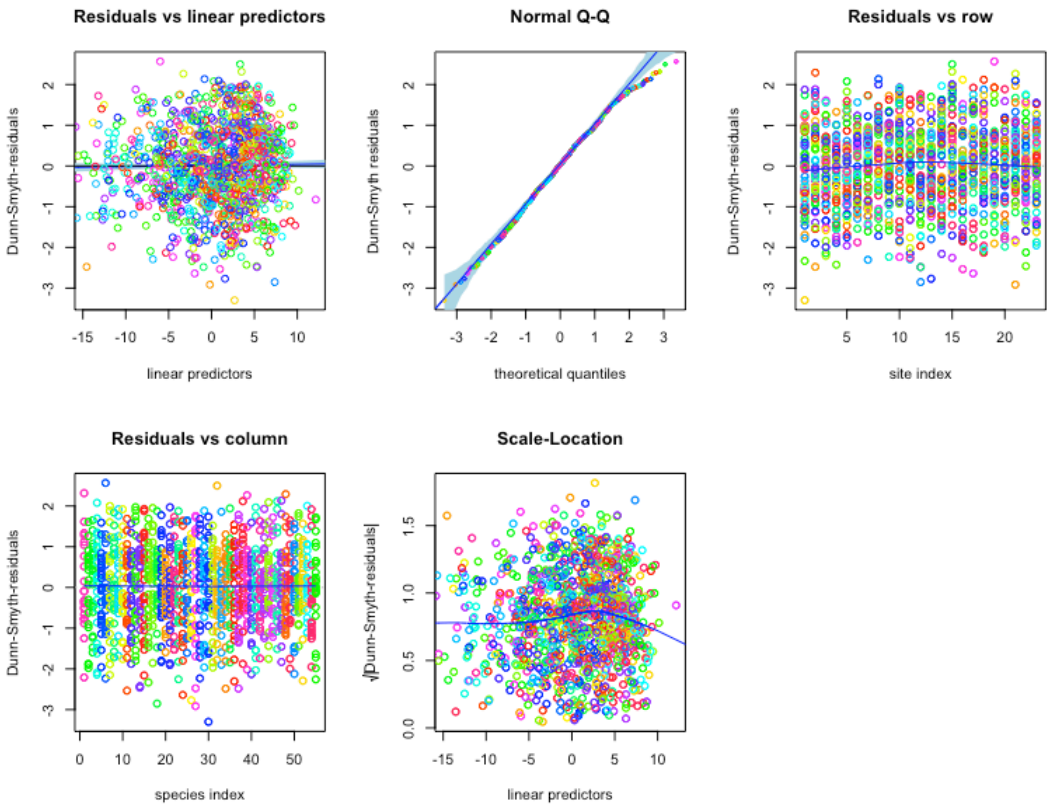
Supplementary Figure 5: "Species" accumulation curve of expected mean "species" richness with confidence intervals (CI; standard error of the estimate) for all fish taxa (step. 10) in increasing numbers of tides sampled. Calculated in 'Vegan' (Oksanen *et al.*, 2015) using specaccum(method = "exact", conditioned = T, permutations = 10,000).



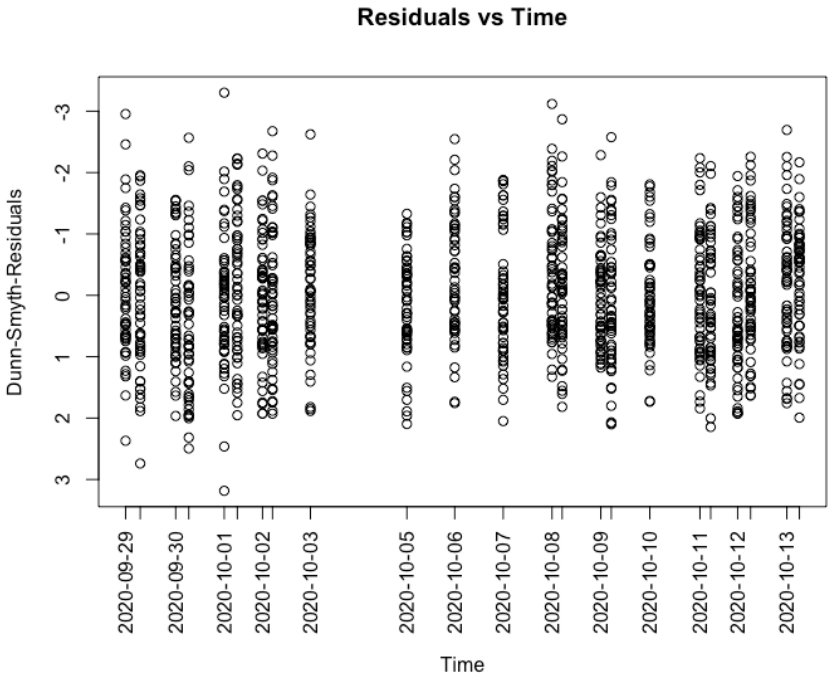
Supplementary Figure 6: Mean River Level (m) per day at Trefriw, Conwy (Lat: 53.160391; Lon: -3.824468) for the autumn period (Grey) and the survey period (Blue). Data source: NRW data request.



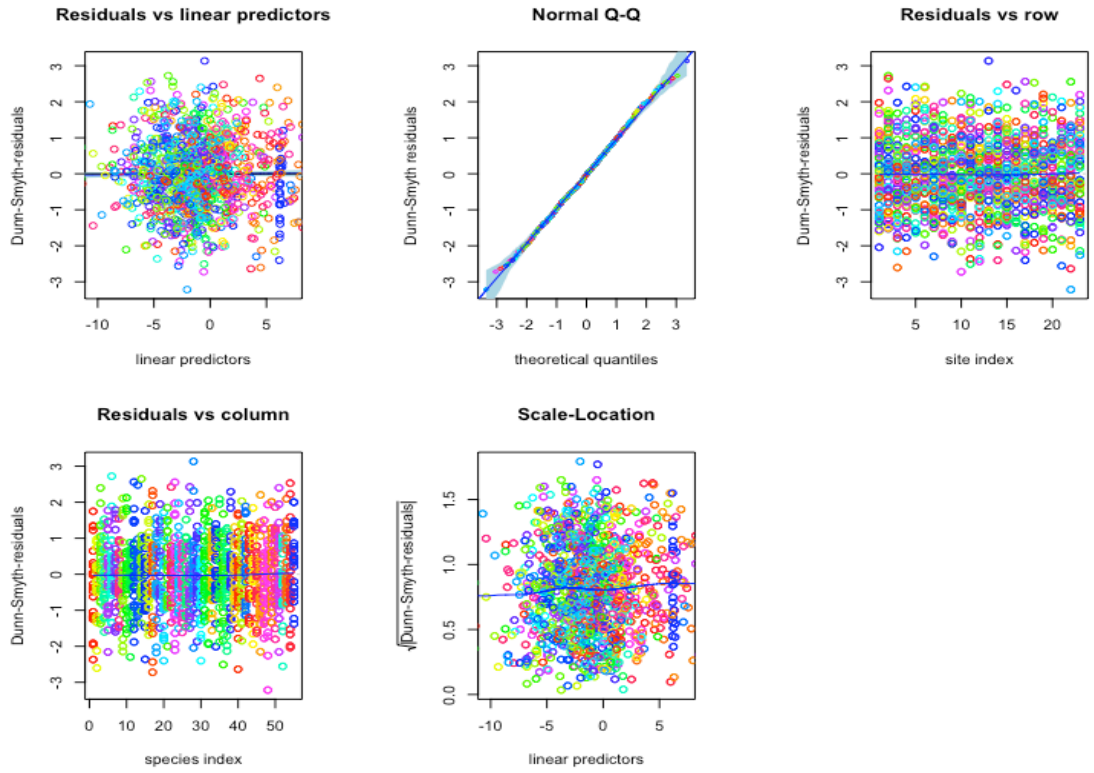
Supplementary Figure 7: Conditional density plot of the smoothed proportion of Shanny (*Lipophys pholis*) presence to absence in each tide for increases in tidal range. The faint grey x-axis tick marks indicate the actual distribution of tidal range (m) measurements.



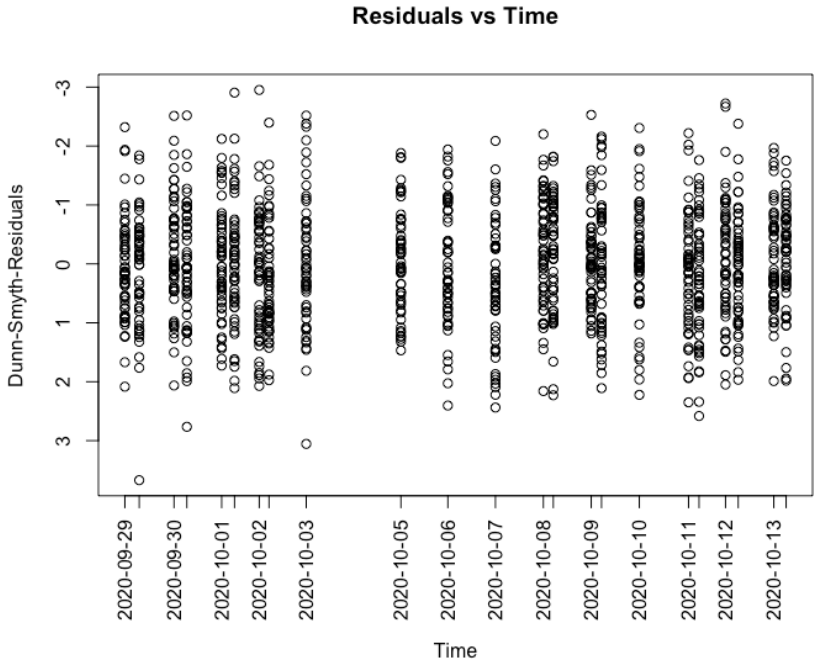
Supplementary Figure 8: Residuals plots for the negative binomial GLLVM of read counts (log link; two latent variables; 50 iterations; offset: log total fish reads per sample).



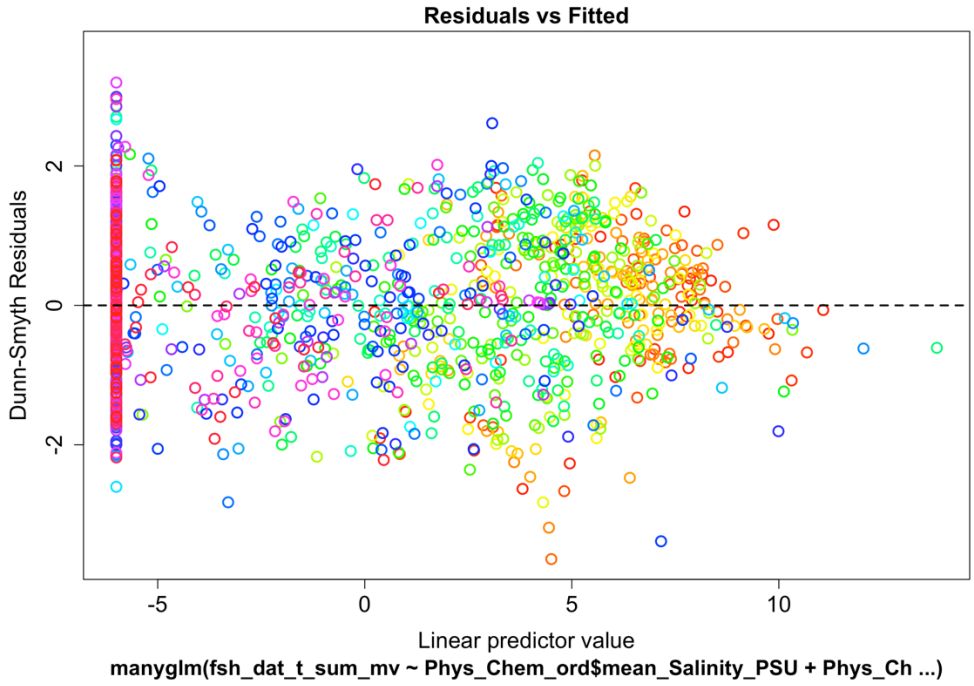
Supplementary Figure 9: Residuals plotted against time for the negative binomial GLLVM of read counts (log link; two latent variables; 50 iterations; offset: log total fish reads per sample).



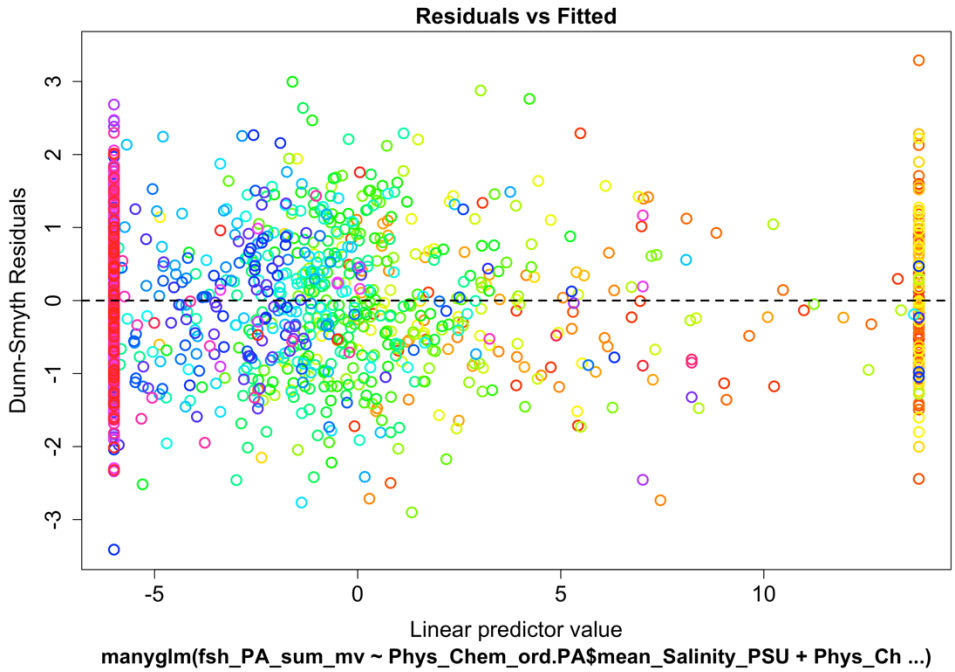
Supplementary Figure 10: Residual plots for species presence/absences binomial distribution (probit link) GLLVM with two latent variables (read depth as a fixed effect; 50 iterations).



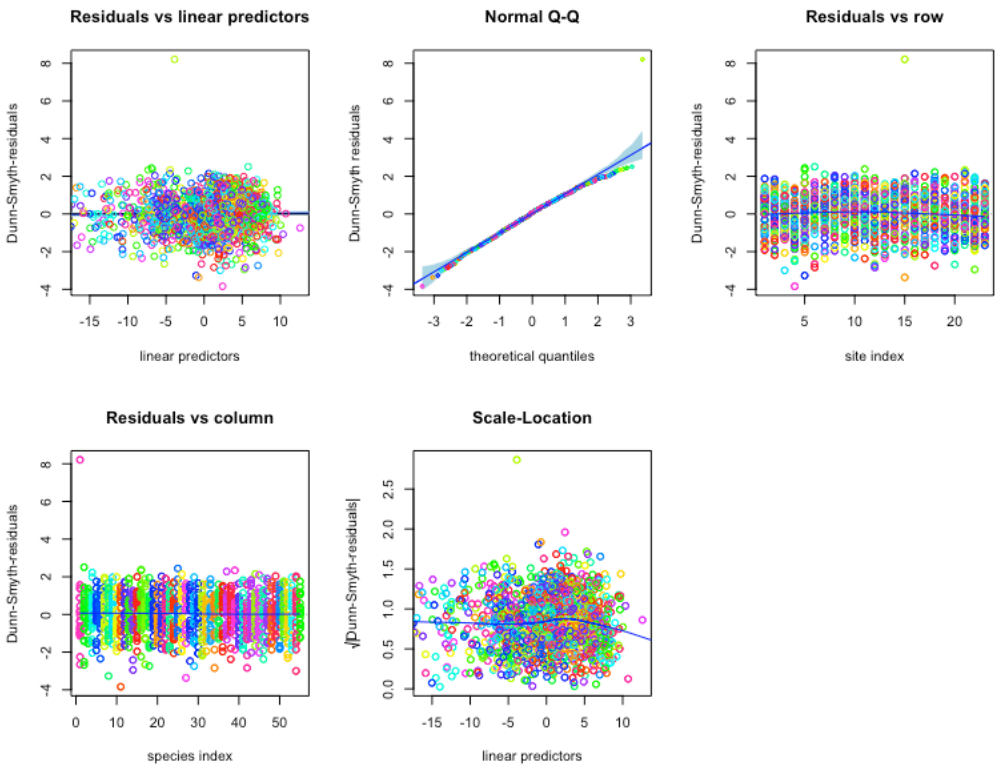
Supplementary Figure 11: Residual plots against time for species presence/absences binomial distribution (probit link) GLLVM with two latent variables (read depth as a fixed effect; 50 iterations).



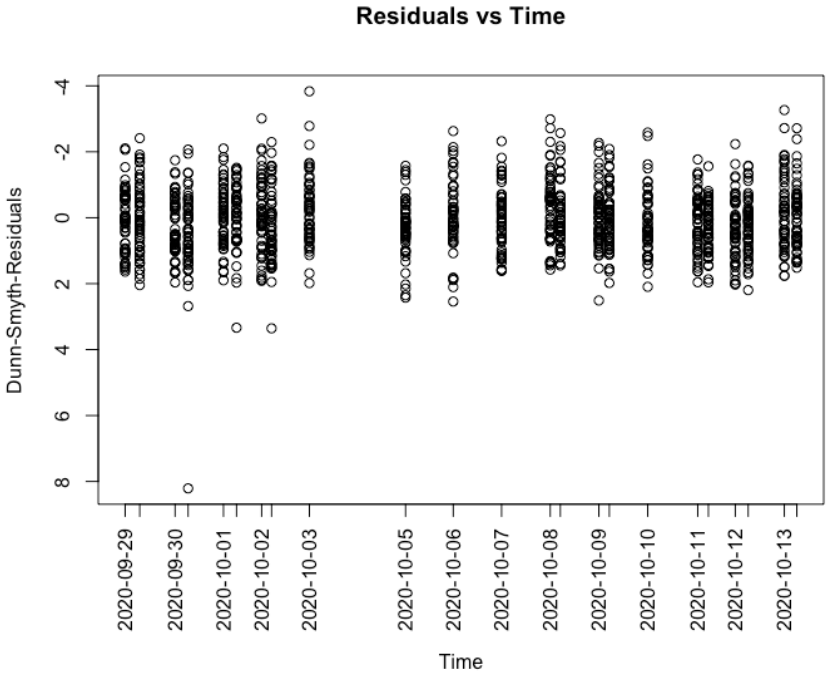
Supplementary Figure 12: Residuals plots for the negative binomial multivariate GLM of read counts (log link, offset: log total fish reads per sample).



Supplementary Figure 13: Residual plots for species presence/absences binomial distribution (logit link) multivariate GLM.



Supplementary Figure 14: Residuals plots for the negative binomial GLLVM of species read counts with trait interactions (log link; two latent variables; 50 iterations; offset: log total fish reads per sample).



Supplementary Figure 15: Residuals plotted against time for the negative binomial GLLVM of species read counts with trait interactions (log link; two latent variables; 50 iterations; offset: log total fish reads per sample).

Chapter 4.

Environmental DNA Metabarcoding Reveals Ecologically Relevant Variation in Fish Assemblages Between Macrotidal Estuaries and Across Seasons

Authors: Thomas I. Gibson, Charles Baillie, Rupert Collins, Owen S. Wangensteen, Laura Corrigan, Amy Ellison, Morton Heddell-Cowie, Hannah Westoby, Barry Byatt, Lori Lawson-Handley, Ana Soto, Simon Creer, Martin Genner and Stefano Mariani

Author Contributions

Project Phase	Author
Grant Acquisition	SM, MG
Study Design	SM, MG, OSW, MH-C/EA, RC
Field Sampling	OSW, AS, MH-C/EA
Laboratory Analysis	CB, OSW
Supporting Data Acquisition	TIG, EA
Bioinformatic Analysis	RC
Statistical Analysis	TIG
Manuscript Writing and Review	TIG, SC, SM, L-LH

Abstract

Estuarine ecosystems are threatened by numerous anthropogenic pressures. Fish assemblages are a dominant component of estuarine macrofauna, and represent higher trophic level organisms that can be assessed to effectively understand the health of these transitional water ecosystems. A growing number of studies have used environmental DNA (eDNA) metabarcoding to assess the biodiversity of fishes in estuaries. However, no studies have compared eDNA metabarcoding to existing fish sampling methodologies across multiple estuaries and seasons. The present study aimed to compare the fish assemblages detected via eDNA metabarcoding of surface water samples to contemporary sampling with fishing gears in three macrotidal estuaries in northeast England (UK), over two seasons: early summer and autumn. Following metabarcoding analysis using an established teleost-specific assay, we found the majority of species caught by fishing were also detected by eDNA. In addition, in certain instances eDNA detected substantially greater species richness estimates compared to fishing methods, including a non-native and various species of conservation interest. The assemblage composition reconstructed via eDNA analysis were significantly different from the assemblage recovered from fishing with seine nets. Importantly, when eDNA data was analysed at the sample level, it could effectively discriminate between estuaries and between seasons consistently across estuaries. The findings are discussed in the context of fish ecology and environmental monitoring requirements. Overall, the findings indicate that eDNA is suited to gathering large amounts of information on fish biodiversity, at a relatively low effort compared to established fishing methods.

Keywords: Estuaries, Fishes, Environmental DNA, Water Framework Directive, Metabarcoding, Biomonitoring.

1. Introduction

Estuaries are ecosystems with diverse physical characteristics, characterised by strong ecological gradients and higher physicochemical variability than other aquatic systems (reviewed in Whitfield & Elliott, 2011). Estuaries are categorised in legislation alongside other coastal brackish water ecosystems, such as lagoons, as transitional waters (EC, 2000). Estuaries are of substantial ecological importance, possessing high primary (Nixon, 1988) and secondary productivity (Allen, 1982; Costa *et al*., 2002), and are globally among the most valuable ecosystems in terms of their ecosystem service provision (Costanza *et al.*, 1997). Despite the ecological and economic importance of estuaries, estuarine ecosystems are heavily impacted by anthropogenic activities, chiefly habitat degradation and pollution (Kennish, 2002). Therefore, there are efforts to conserve and protect estuarine ecosystems, and their associated species and habitats via legislation. For example, within the European Union (EU) the key legislation relating to environmental protection is the Water Framework Directive (WFD; 2000/60/EC; EC, 2000) and the Habitats Directive (92/43/EEC; EC, 1992). In the United Kingdom (UK) these have been replaced by equivalent legislation directly derived from EU directives (JNCC, 2021; UK Parliament SI 2017/1012 and SI 2017/407).

Estuaries play a critical role in the ecology of numerous fish species, and fishes are one of the dominant macrofaunal groups in these environments (Martino and Able, 2003). In Europe, fishes mostly use estuaries as temporary habitats, as feeding or nursery grounds, and marine species are the dominant contributors to their biodiversity (Franco *et al.*, 2008). Estuaries also provide an environment for truly estuarine species and a migratory route for economically valuable diadromous species, such as Atlantic salmon (*Salmo salar*) and European eel (*Anguilla anguilla,* McLusky & Elliott, 2004). Aside from their intrinsic and economic value fishes are important indicators for the health of estuarine ecosystems (Christophe *et al.*, 2015; Whitfield & Elliot, 2002). Within the WFD, fish assemblages are one of the quality elements used to determine the ecological status of transitional waters (EC, 2000; Hatton-Ellis, 2008). Assessments are conducted by generation of multi-metric indices of fish assemblage health, which are compared to reference conditions to detect anthropogenic disturbances (Christophe *et al.*, 2015; Lepage *et al.*, 2016). This requires data on various aspects of fish assemblage structure and function (Coates *et al*., 2007; Harrison & Whitfield, 2004; Harrison & Kelly, 2013). Fish assessment methods show direct correlations with overall anthropogenic pressure (Lepage *et al.*, 2016), and with individual stressor categories, such as water pollution and oxygen depletion (Teichert *et al.*, 2016). Therefore, data on fish assemblage is highly useful when planning restoration measures (Teichert *et al.*, 2016), which are required in many transitional waters across the EU and UK (EEA, 2018a).

Regardless of the efficacy of fish assessments, the vast majority (74%) of transitional waters in the EU and the UK are not assessed on the basis of the ecological quality of their fish fauna (EEA 2018a, 2018b). This maybe partially due to the resource intensive nature of sampling with fishing gears. Metrics rely on data from capture based methods using a variety of fishing gear types (beam and otter trawls, seine and fyke nets etc.), deployed at multiple sampling stations (Coates *et al.*, 2007; Colclough *et al.*, 2002; Delpech *et al.*, 2010; Harrison & Kelly, 2013). Due to the sampling biases of different fishing gears, the use of multiple gear types to gain a comprehensive sample of the fish assemblage is often used (Coates *et al.,* 2007; Harrison & Kelly, 2013). This is an effective method (see Harrison & Kelly, 2013). However, due to the highly dynamic nature of estuarine environments, it is difficult to implement multi-gear surveys consistently and sample the entire fish assemblage in a cost efficient manner (Waugh *et al.*, 2019). Analysis of species richness between estuaries in England and Wales using Transitional and Coastal waters (TraC) fish survey data showed seine netting provided the widest spatial

coverage between, and within estuaries. Comparably, although an integral party of WFD fish assessment, fyke nets and beam trawls had not been deployed consistently enough across estuaries for the data to be useful (Waugh *et al.*, 2019). In an assessment context, this is likely to cause bias in estimates of fish assemblage health. Therefore, there is a clear requirement to develop methods that can be consistently applied to complement and enhance existing fish sampling designs.

In recent years there have been substantial advances in the analysis of Environmental DNA (eDNA) for surveying fish assemblages. Environmental DNA is defined as DNA isolated from an environmental sample without capturing the organism or including any obvious signs of biological material (Taberlet *et al.*, 2012; Thomsen and Willerslev, 2015). The application of eDNA metabarcoding (see Lawson Handley, 2015; Deiner *et al.*, 2017) is potentially well suited to WFD fish assessments in estuaries. It is a more sensitive method of assessing species richness than conventional methods, although not every species detected by conventional means can be detected by eDNA analysis (Chapter 2; Cole *et al*., 2022; Hallam *et al.*, 2021; Zou *et al.*, 2020). Analysis of eDNA can detect a different assemblage composition to conventional methods in estuaries (Chapter 2; Cole *et al.,* 2022; Hallam *et al.,* 2021). Changes in fish assemblage composition can also be detected by eDNA at various spatial scales within and between estuaries and adjacent ecosystems, from less than 1 to 100s of kilometres (Chapter 2; Cole *et al*., 2022; García-Machado *et al.*, 2021; Hallam *et al.*, 2021). In addition, spatial changes in assemblage structure can be correlated with physicochemical variables, such as salinity (Chapter 2; Ahn *et al.*, 2020). Seasonal changes in the composition of fish assemblages in estuaries can also be detected by metabarcoding (Stoeckle *et al.*, 2017; Zou *et al.*, 2020), but not in all studies (Chapter 2; Hallam *et al.*, 2021). Finally, eDNA metabarcoding can be used to assess variation in the structure of fish assemblages among multiple estuaries and coastal areas at the regional scale, and in relation to geographic and anthropogenic factors (Kume *et al.*, 2021).

It is increasingly clear that eDNA is a useful tool for the assessment of fish biodiversity in estuaries. However, there are caveats regarding data interpretation due to eDNA transport from outside the ecosystem (Chapter 2; Chapter 3; Deiner & Altermatt, 2014; Shaw *et al.*, 2016; Yamamoto *et al*., 2017). Uncertainty also exists around the link between eDNA metabarcoding read abundance, and the abundance, or biomass, of the fishes which produce the trace DNA (Hansen *et al.*, 2018; Lamb *et al.*, 2019; Rourke *et al.*, 2022). Regardless of recent advances there is still substantial requirement for additional comparative studies between biodiversity assessment achieved via eDNA analysis and fishing in estuaries. To the best of our knowledge, there have been no comparative studies across multiple estuaries, over multiple seasons using both eDNA metabarcoding and fishing gears. At the time of writing, temporal studies on these habitats tend to focus on one system (DiBattista *et al.*, 2022) and the only studies that have compared multiple estuaries have used eDNA metabarcoding data exclusively (Ahn *et al.*, 2020; Kume *et al.*, 2021). Comparative studies are essential to provide comprehensive assessments of the fish assemblage using eDNA analysis and fishing. Thereby facilitating further integration of eDNA into the wider set of methods used to study fishes in estuaries and coastal ecosystems.

1.2 Aims and Objectives

The overall aim of this study is to compare the fish assemblage detected via eDNA metabarcoding of surface water samples to contemporary sampling with fishing gears (seines, fykes and beam trawls) in three macrotidal estuaries (the Tees, Esk and Tweed, Northeast England, UK), over two seasons: autumn and early summer. Sampling was conducted initially

in the Tees and the Tweed in autumn 2016, and then across all estuaries and both seasons in 2017. A qualitative comparison was made between eDNA metabarcoding and all the species detected with fishing gears in the estuaries from 2007 to 2017. Firstly, we hypothesised that eDNA metabarcoding would detect more species in each estuary overall, and would detect a different assemblage composition (species presence/absence) to contemporary data from fishing gears. Secondly, we hypothesised that differences in assemblage composition would be detected between estuaries, seasons and different salinity zones, and that these trends would be consistent across eDNA and fishing methods. The second hypothesis was also tested for eDNA independently, using data at the sample level.

2. Methods

2.1 Study Location

Sampling was conducted in the marine freshwater transitional regions of the following rivers: the Tweed (Berwick-upon-Tweed), Tees (Middlesbrough) and Esk (Whitby). These estuaries are situated on the east cost of Britain and flow into the North Sea (figure 1). They are sites of significant urban and industrial development, particularly on the Tees. All these estuaries are macrotidal (tidal range: 4.1 to 4.6) and exhibit variation in their overall size, with the Tees being the largest (ABPmer and HR Wallingford, 2007). There is also variation in extent of habitats and the types of habitats within them (EA, 2016). The Tweed Estuary, and to a lesser extent the Tees Estuary, are covered by some degree of protected area legislation (EA, 2022b, 2022a).

2.2 Fish Sampling

Within each estuary the Environment Agency (EA) use a multi-method approach for monitoring fish assemblages broadly following Colclough *et al.* (2002). Sampling has used the current complement of fishing gears since 2011. Previous sampling used a sub-set of gear types since 2007. Prior to 2007, more sporadic sampling occurred which did not include the Esk. Surveys are conducted twice annually, in later spring/early summer (May/June; here after 'early summer') and the autumn (September/October). During each survey, double fyke nets (0.5 m high x 2.5 long; 10 mm mesh cod end joined by a 6 m long x 15 mm mesh ladder), seine nets (43 m long x 4 m deep; knotless mesh 6.5 mm on centre panel 14 mm and on wings) and beam trawls (1.5 m high x 4 m long, 1.2 m cod end, with a main knotless mesh size of 20 mm and a cod end mesh size of 8 mm knotless mesh) are deployed in shallow intertidal waters at a number of stations in each estuary. Fishing gear is typically deployed between 2 and 4 hrs after low tide and surveys generally coincided with neap tides. Surface water physiochemical parameters: dissolved oxygen (%), salinity (measured in practical salinity units), temperature (°C) and pH, are also recorded at each station alongside gear deployment (using a Pro Plus; YSI Inc.). Sampling for each gear and station is conducted as follows: seine nets are hauled twice at each station, beam trawls are towed once for 200 m at each station and double fyke nets are set below the low water mark and recovered after a full 12 hr tidal cycle. Currently in the Esk and Tweed seine nets are deployed at three stations in the lower, middle and upper region of each estuary. Whereas, in the Tees only one seine net station is present in the lower estuary. In the Tweed and Tees, beam trawling occurs at one seine netting station in the lower region of each estuary. In the Esk and Tweed, fyke nets are set at one independent station in the lower region of each estuary (figure 1). Prior to 2016, other stations in each estuary have been sampled. Otter trawling was also used to assess the outer Tees in autumn 2016. However, this did not coincide with eDNA sampling and is not considered when making direct comparisons between eDNA and fishing in 2016 and 2017.

2.3 eDNA Sampling

When possible, water samples were collected on the same days as the EA netting surveys, allowing direct comparisons between fishing data and eDNA data. On seven sampling events at individual stations it was not possible to sample on the same day as fishing (supplementary table 1). On these occasions, eDNA sampling occurred within a day of fishing at the same tidal state. Sampling was carried out at all stations where fishing was conducted in the three estuaries. In the Tees, an additional upper estuary seine netting station, which was sampled until 2015, was also sampled with eDNA. Surveys were conducted for the Esk and Tees in October 2016, May-June 2017 and September-October 2017. The Tweed was only sampled in June and September 2017 (supplementary table 1). In total, four stations were sampled for eDNA on the Esk and Tweed respectively, and two on the Tees, although the upper most station on the Tweed was not sampled in summer 2017 (supplementary table 2). During each survey, triplicate 2 L surface water samples (0-1 m depth) were collected in HDPE bottles (Nalgene Nunc International, Rochester, NY, USA) cleaned previously with 10% bleach, each covered with a 250-µm nylon mesh to prevent the influx of debris, immediately before fishing commenced. In total, 36 samples were collected in the Esk and 21 samples were collected in the Tees and Tweed respectively. Following collection, samples were packed in individual sterile plastic bags, placed on ice and transported to the laboratory. Field blanks of the same volume of water were taken into the field for the Esk sampling and treated in the same way as samples. Following arrival at the laboratory (within 5 hrs after collection), water samples were filtered using PES 0.22 µm Sterivex filter units (Merck Millipore, Burlington, MA, USA) using a 100 mL polypropylene syringe and cleared of water. When the full 2 L could not be passed through the filter due to clogging, the volume of water was recorded. The mean filtered volume per sample was 400 ml (sd: 200 ml). Following filtration, the filters were stored at −20°C.

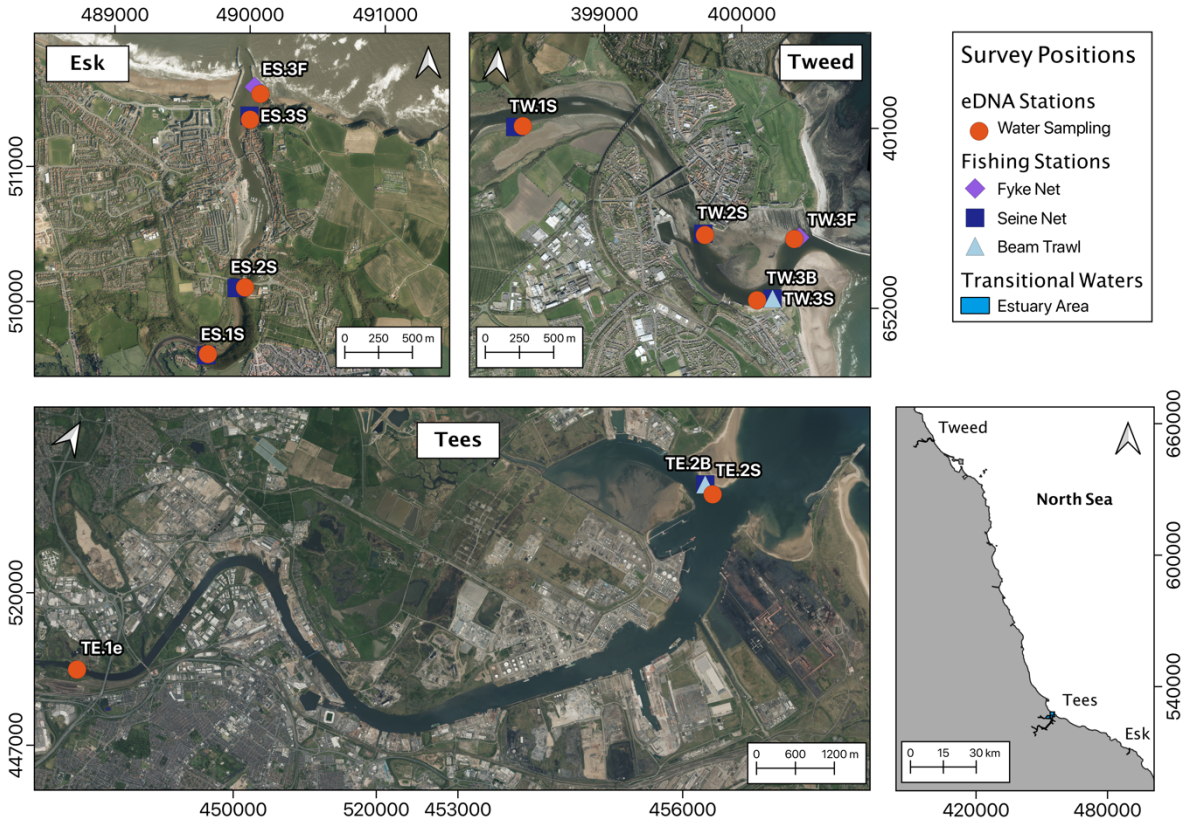


Figure 1: Map of the Esk, Tweed and Tees estuaries, giving their geographic location within Britain and the distribution of eDNA sampling stations in autumn 2017 relative to fishing gear sampling stations, as an example of the sampling design. Sampling stations are labelled with each estuary, a general site number and then sampling method at the exact station. Coordinate System: British National Grid (EPSG:27700) axis in eastings/northings (m). British Coastline (Wessel & Smith, 1996 and 2017), Satellite Photography (Getmapping PLC, 2014), Estuary position and extent (EA, 2021).

2.4 Lab Methodology

DNA extraction and pre-PCR preparations were carried out in a dedicated pre-PCR laboratory, separate from post-PCR procedures, in separate rooms with equipment and surfaces regularly cleaned using a 10% commercial bleach solution, and subjected to UV irradiation at the end of each session. Total eDNA was extracted from filters using DNeasy PowerSoil kits (QIAGEN, Hilden, Germany; see Collins *et al.*, 2019). Field and extraction blank controls were processed in parallel. Following extraction, a ~167 bp fragment of the 12S rRNA region was amplified using the fish specific Tele02 primers (Miya *et al.*, 2015; Taberlet *et al.*, 2018). Each primer possessed a unique 8-bp index tag to allow sample identification following sequencing thereby facilitating use of a one-step PCR protocol. PCR reactions for each DNA extract were conducted in triplicate along with sterile H_2O blanks. The total reaction volume was 26 µl containing: 16 µl Amplitaq Gold Master Mix (Applied Biosystems, Waltham, MA, USA), 0.16 µl of Bovine Serum Albumin, 2 µl of DNA extract, 1 µl of each forward and reverse primer (5 µM) and 5.84 µl H_2O. The thermocycler profile was as follows: 95°C for 10 mins, followed by 40 cycles of 95°C for 30 s, 60°C for 45 s, and 72°C for 30 s, with a final extension at 72°C for 5 mins. Following amplification, amplicons, including field blanks (n = 4), extraction blanks (n = 4), PCR blanks (n = 2) and well blanks (unused tag combinations, n = 10), were combined into three separate pools. Primer dimer was removed using a HighPrep^{TM} PCR clean-up (MagBio Genomics, Gaithersburg, MD, USA), with 1X paramagnetic beads to pool ratio. Three, one for each pool, PCR-free dual-indexed libraries were prepared using the KAPA

Hyper Prep Kit (Roche, Basel, Switzerland). This followed the manufacturer instructions but with a prolonged adaptor ligation step of 90 mins and at a lowered temperature of 37°C. Libraries were then quantified using qPCR, pooled in equimolar concentrations and loaded onto an Illumina MiSeq platform at a concentration of 8pM and sequenced using V2 chemistry (2x150-bp paired-end; Illumina, San Diego, CA, USA).

2.5 Bioinformatics

Bioinformatic processing followed the approach in Collins *et al.* (2019). Briefly, this entailed sample demultiplexing using cutadapt v2.10 (Martin, 2011), sequence denoising and dereplication using dada2 v1.18.0 (Callahan *et al.*, 2016), homology filtering of ASVs (amplicon sequence variants) using hidden Markov models with hmmer v3.1 (Eddy, 2011). Reference library sequences were used as priors during dada2 denoising to avoid erroneously discarding rare sequences. Following initial processing, approximate taxonomic assignment was conducted using SINTAX (Edgar, 2016) and NCBI RefSeq v205 (O'Leary *et al.*, 2016) to exclude non-fish ASVs. Taxa were then matched against a curated reference library for British Isles fishes (Collins, 2021a; Collins *et al.*, 2021) using phylogenetic placement with epa-ng v0.3.7 (Barbera *et al.*, 2019) and sequence similarity blastn v2.10.1 (Camacho *et al.*, 2009); see Collins *et al.* (2019) for details on the similarity thresholds used. To control for potential cross-contamination, an exclusion list of sequences generated from another concurrent lab project was used. For a few species, haplotypes were shared between close relatives. In this case, a species level identification was not possible and a higher level taxonomic assignment was given e.g. genus or family. In the case of dace (*Leuciscus leuciscus*) and orfe (*L. idus*), which shared a haplotype, detections were treated as *L. leuciscus* rather than *L. idus*, as the latter is restricted to ornamental ponds and commercial coarse fisheries and was absent from the TraC fish data from 2007 to 2017. In the case of the shared haplotype between whiting (*Merlangius merlangus*) and haddock (*Melanogrammus aeglefinus*), these were treated as *M. merlangus* given that *M. aeglefinus* had never been detected in the study ecosystems from 2007 to 2017. Finally, the assignment for big-scale sand smelt (*Atherina boyeri*) is most likely to be sand smelt (*A. presbyter*), as a reference sequence was absent for the latter species. *A. boyeri* is also relatively rare in British estuaries, while *A. presbyter* in known to occur in the Esk and Tees from TraC Fish data. For ease, all taxa are referred to as "species" from here on as this was correct in the majority of cases.

To account for contamination, a conservative per species read threshold cut-off was calculated using an adaption of the approach in Yamamoto *et al.* (2017). Species which contributed equal to or less than 0.08 % of the total target fish reads in a sample were considered absent. This cut-off was generated by taking the total reads of each species in the negative controls, dividing it by the number of negative controls (12), then taking the mean, to give a per species, per control, contamination value (17). This value was multiplied by the number of water samples (78) to give the total potential per taxa contamination (1364). Total potential per species contamination was then divided by the total target fish reads in the water samples (1774611) to give the threshold cut-off (0.00077 ≈ 0.08 %). Following cleaning, sample based rarefaction curves were appraised to determine which samples had not plateaued (supplementary figure 1 and 2). Samples which had not plateaued were then visualised using nMDS (Bray-Curtis/Sørensen; presence/absence) to identify potential outliers in species composition (supplementary figure 3).

In addition to taxonomic assignment, species were assigned to estuarine use functional guilds using the guild classification system for European estuaries (Franco *et al.*, 2008). These describe the overall ecological use of an estuary by a species and its links between the estuary

and marine, and freshwater areas (Elliott *et al.*, 2007). Guild assignments were based on the European system initially, but the classifications in Elliott & Hemingway (2002) and Elliott & Dewailly (1995) were used when multiple guild assignments were present for individual species. Briefly, these estuarine use guilds include the Marine Stragglers (MS), Marine Migrants (MM), Estuarine Species (ES), Anadromous Species (A), Catadromous Species (C), Freshwater Species (F) (Franco *et al*., 2008). Where no guild assignment was found, species were counted as Unassigned (UA). This was also performed where taxa could not be identified to species level.

2.6 Statistical Analysis

For comparisons of eDNA with fishing, each gear type station was paired with the closest corresponding eDNA station using nearest neighbour analysis via the Distance Matrix tool in QGIS (coordinate reference system: OSGB 1936; QGIS Development Team, 2019). Stations without a matching eDNA sample within 200 m were discarded, to exclude the unpaired eDNA station in the upper Tees. The mean distance between fishing and eDNA stations was 70 m (sd: 43 m). Environmental data from fishing stations were assigned to eDNA stations using the same method with the mean distance being 114 m (sd: 164 m). Given eDNA sampling was not always concurrent with fishing, the *exact* physiochemical readings would not be representative given their high temporal variation (see Chapter 3). Instead, a broad categorisation of the salinity at each station was included in the analysis following the Venice System (1959): Euhaline (salinity > 30), Polyhaline (18 - 30), Mesohaline (5 - 18), Oligohaline (0.5 to 5.0) and Limnetic (< 0.5; McLusky, 1992). Salinity was not colinear with other variables (temperature, DO, pH; supplementary figure 4).

Species detected with eDNA were first checked to see if they had been detected in each estuary by TraC Fish surveys from 2007 to 2017. This period was chosen as it encompassed the time from which fish sampling started in the Esk and covered the eDNA survey period. Following this species lists were generated for each estuary for eDNA and fishing, for each season, using all the data for 2016 and 2017. The taxonomic composition of these lists was compared using UpSet plots (Conway *et al.*, 2017). Where species were detected to a higher taxonomic resolution by fishing than by eDNA. Only a single match was counted between eDNA and fishing, to fully represent the diversity of species caught by fishing gears.

Overall species richness estimates were summarised for each estuary, for eDNA and fishing method per season using rarefaction and extrapolation (R/E) sampling curves generated with the iNEXT package (Chao *et al.*, 2014; Hsieh *et al.*, 2016). R/E curves were generated for twice the sample size of each method type (Hsieh *et al.*, 2016) and 95% confidence intervals (CI) and standard errors calculated using 1000 bootstrap replicates. Curves were generated for each estuary to compare among the full eDNA data, the eDNA filtered by species detected in TraC Fish surveys from 2007 to 2017, and the fish capture data. All the available data for each estuary was used for these comparisons and the autumn data across years was summed. Both the asymptotic species richness, and the species richness estimate per 8 samples was calculated. In addition, comparisons of asymptotic species richness were made between different estuaries. Species richness R/E curves were calculated for each estuary using only the data for 2017, pooled between seasons, one for each method as before. Comparisons between these overall estimates from eDNA and fishing within each estuary were also made.

For spatio-temporal comparisons of assemblage composition (presence/absence) subsets of the full eDNA and fishing gear data were compared at the level of the sampling station. Firstly, to account for the difference in sampling design between eDNA and fishing gears, samples

from each sampling event at a station were aggregated together and treated as an independent replicate. Aggregation accounted for the spatio-temporal non-independence of eDNA samples, or multiple seine tows, within each sampling event at a station (Hurlbert, 1984). To take account of variation in the taxonomic resolution between metabarcoding and fishing for some taxa, the fishing data was reduced to the resolution of the eDNA data. This is because exact differences in species richness were less important here than differences in composition. The presence/absence of species was compared between eDNA stations and fishing stations using ordinations generated with generalised linear latent variable models (GLLVM; R Package: gllvm; Niku *et al.*, 2019). The binomial distribution (probit link) was used, with 50 initial runs including random variation (jitter = 0.01) to minimise sensitivity to the chosen starting values (Niku *et al.*, 2019). GLLVMs were fitted with 2 to 5 latent variables and model fit assessed using the AICc to check 2-D ordinations were appropriate. Prior to ordination with fishing gears the eDNA data was ordinated on its own, with sample read count included as a fixed factor. Model selection using AICc confirmed if sample read count had a substantial influence on assemblage composition. Secondly, due to constraints in the spatial deployment of fishing gears in each TraC Fish survey, the deployment of gears was not consistent within and between estuaries (figure 1). To provide comparisons of fishing method composition against eDNA, the eDNA and fishing data were subdivided into three groups and analysed separately. Each group consisted of stations sampled with each gear type (i.e. beam trawls, seine and fykes nets) and their nearest neighbour eDNA stations. Nearest neighbour stations were confirmed, as before, using QGIS (QGIS Development Team, 2019). For each analysis the assemblage composition was determined graphically for each data subset using GLLVM as described above. Differences in assemblage composition between methods were tested explicitly using multivariate GLMs (binomial distribution, logit link) in the mvabund package (Wang *et al.*, 2012). For the seine net comparison, the fixed effects of the factors sampling method, estuary, season, salinity zone and year on assemblage composition were tested. In addition, the interaction between method and each environmental variable (estuary, season, salinity zone, year) was tested to determine environmental effects were consistent across methodologies. The analysis was also repeated after omitting the 2016 data ensuring differential sampling of estuaries across years had not influenced the overall conclusions. For comparisons with fyke nets and beam trawls no interactions terms were included and only the fixed effects of method, estuary, season and salinity zone were included in the initial model due to small sample sizes. Model selection was applied to each initial model using backwards selection and assessing the AIC (Zuur *et al.*, 2007). Explanatory variables that did not improve model fit were dropped. For the final model, an analysis of variance test was conducted using the Wald test statistic and assuming correlation between species response variables (see Wang *et al.*, 2012). *P*-values were calculated using the PIT-trap resampling method (5000 bootstrap iterations), while *p*-values for univariate tests for each species response variable were adjusted for multiple testing (Wang *et al.*, 2012).

Fish assemblage composition across estuaries and seasons was further analysed using sample level eDNA data from 2017. Assemblage composition was studied using the same ordination and a multivariate GLM approach, as above. For the multivariate GLM the main effect of estuary, seasons and salinity zone were examined, with an interaction term between estuary and season in the initial model. To account for non-independence between samples collected together at a station (Hurlbert, 1984), sampling event was used as a blocking factor allowing resampling between groups of samples. This allowed valid inferences to be made across independent blocks of correlated sets of observations to account for any within-sampling event species correlations (Wang *et al.*, 2012). This approach was not available for the initial comparison between eDNA and fishing because blocks must contain balanced sample sizes.

3. Results

3.1 Sequencing Results

A total of 1,786,739 target fish reads were obtained from samples and blanks with 95 species detected in the samples overall. A total of 12,128 target fish reads (0.68 % of the total fish reads) were detected in the blanks from 57 species (table 1; supplementary table 3). The median number of reads per sample was 20,210, but this was highly variable (inter quartile range, IQR: 27,362). After applying the 0.08 % per sample cut-off threshold, the total number of species detected was 76, with a median per sample species richness of 28 (IQR: 19; see supplementary figure 1). Sample level rarefaction curves showed the vast majority of samples had plateaued, or were beginning to, with no extreme outliers in assemblage composition (supplementary figures 1, 2 and 3).

3.2 Qualitative Comparison of eDNA with TraC Fish Surveys

All the 76 taxa that passed the contamination filter had the potential to occur in UK estuaries, and 70 taxa had previously been categorised into estuarine use guilds (table 1 and supplementary table 3). The two species which were not categorised were the anadromous invasive pink salmon (*Oncorhynchus gorbuscha)* and the marine redband fish (*Cepola macrophthalma*), which was probably a previously uncategorised Marine Straggler. The other four higher level taxa were not possible to categorise as they contained species with differing guild assignments. A substantial proportion of the species (47.4%) detected using eDNA across estuaries, had never been detected in any of the estuaries in TraC fish surveys from 2007 to 2017. The majority of novel species were Marine Stragglers (47.2 %), with Freshwater Species (22.2 %) and Marine Migrant (16.7 %) species also making up a substantial proportion. Similarly, when comparing these novel eDNA detections for each estuary individually: 62.3 % of the 61 species in the Esk, 51.5 % of the 66 species in the Tees, and 54.9 % of the 51 species in the Tweed, were novel detections (table 1 and supplementary table 3).

Overall, qualitative comparisons between the species detected by eDNA and all fishing gears over both seasons in 2016 and 2017, showed that most species detected by fishing gears in the contemporary surveys could be detected with eDNA. Overall, in the Esk 13 of the 15 species (86.7 %) detected by fishing were detected by eDNA (figure 2). Eelpout (*Zoarces viviparus*) could not be detected by eDNA, and while both *Pollachius pollachius* and *P. virens* were detected by fishing, these could not be differentiated by eDNA, so only one detection was counted. In the Tees, 7 of the 8 (87.5 %) species detected by fishing were detected by eDNA (figure 3). Two species of the Ammodytidae were detected by fishing and these could not be differentiated by eDNA and so again only a single detection was counted. For the Tweed, 18 of the 20 species (90 %) detected by fishing were detected by eDNA (figure 4). The ninespine stickleback (*Pungitius pungitus*) was detected by fishing in autumn, whereas *Z. vivparus* was detected by fishing in summer. Neither *Z. viviparus* or *P. pungitus* were present in the eDNA data prior to cleaning hence the applied cut-off threshold did not negatively influence these comparisons. Similarly, when the whole species list for TraC Fishing surveys from 2007 to 2017 was compared to the species lists from the eDNA sampling surveys in 2016 and 2017, the majority of the species were detected by eDNA. Of the 30 total species detected by TraC Fishing on the Esk, 23 species (76.7 %) were detected by eDNA. On the Tees, of the 40 species detected by fishing, 32 (80.0 %) were detected by eDNA. On the Tweed, of the 32 species detected by fishing, 23 (71.9 %) were detected by eDNA.

Table 1: Species detected by eDNA in each estuary (ordered by combined read abundance across estuaries)

Species	Common Name	Guilds	Esk			Tees			Tweed		
			Detections	Reads	TRAC 2007 - 2017	Detections	Reads	TRAC 2007 - 2017	Detections	Reads	TRAC 2007 - 2017
Salmo trutta †	Brown trout	A	36	242351	●	16	18237	●	17	22044	●
Clupea harengus †	Atlantic herring	MM	36	95931	●	20	71537	●	16	37702	●
Pleuronectes platessa †	European plaice	MM	29	65728	●	19	64142	●	12	43663	●
Phoxinus phoxinus †	Eurasian minnow	F	36	52109	○	12	1073	○	21	112728	●
Sprattus sprattus †	European sprat	MM	35	29967	●	21	56058	●	14	13472	●
Gadus morhua †	Atlantic cod	MM	35	82717	●	15	4212	●	9	1632	●
Platichthys flesus †	European flounder	MM	34	27738	●	21	39655	●	19	11549	●
Salmo salar †	Atlantic salmon	A	35	53759	●	13	2400	○	19	10228	●
Ammodytidae †	Sand lances	UA	28	10767	■	18	47060	■	10	4112	■
Sardina pilchardus †	European pilchard	MM	25	23969	○	15	16150	○	12	19729	○
Limanda limanda †	Dab	MM	10	2742	●	13	54678	●	2	12	○
Gasterosteus aculeatus †	Three-spined stickleback	A	27	3614	●	8	532	●	21	49512	●
Scomber scombrus †	Atlantic mackerel	MS	33	29322	○	13	10268	○	10	10685	○
Barbatula barbatula †	Stone loach	F	33	23391	○	9	258	○	21	25459	○
Merlangius merlangus †	Whiting	MM	27	5348	●	21	34215	●	9	1844	●
Zeugopterus punctatus †	Topknot	MS	21	12262	○	13	8555	○	8	8872	○
Trachurus trachurus †	Atlantic horse mackerel	MS	24	12585	○	14	7036	○	9	7703	○
Pomatoschistus microps †	Common goby	ES	22	11720	●	11	5837	●	12	8875	●
Anguilla anguilla †	European eel	C	25	5780	●	6	2838	●	19	9877	●
Labrus bergylta †	Ballan wrasse	MS	20	5676	○	13	4509	○	8	4370	○
Dicentrarchus labrax †	European seabass	MM	24	3144	○	15	8702	●	9	2394	○
Molva molva †	Ling	MS	30	10733	○	10	758	○	9	927	○
Atherina presbyter †	Sand smelt	MM	20	5787	●	11	2934	●	8	3600	○
Gobio gobio	Gudgeon	F	13	7014	○	0	0	○	5	184	○

Table 1 Continued

Species	Common Name	Guilds	Esk			Tees			Tweed		
			Detections	Reads	TRAC 2007 - 2017	Detections	Reads	TRAC 2007 - 2017	Detections	Reads	TRAC 2007 - 2017
Lampetra sp.	Lampreys	A	11	5842	●	0	0	○	7	891	●
Pomatoschistus minutus †	Sand goby	ES	24	3943	●	8	798	●	10	1479	●
Rutilus rutilus	Roach	F	1	2	○	3	5136	●	1	342	○
Pollachius sp.	Pollack/Saithe	UA	10	3877	●	2	212	●	3	1104	●
Triglidae †	Sea robins	UA	26	2808	○	11	1055	●	10	1009	○
Abramis brama	Freshwater bream	F	0	0	○	5	4808	●	0	0	○
Thymallus thymallus	Grayling	F	3	24	○	0	0	○	11	4504	○
Oncorhynchus mykiss †	Rainbow trout	F	20	1241	○	13	1043	○	13	1957	○
Taurulus bubalis †	Longspined bullhead	MS	23	2287	●	10	542	●	9	685	●
Chelon ramada/labrosus †	Grey mullet	UA	18	685	●	13	2080	○	8	510	○
Trisopterus minutes †	Poor cod	MS	21	1012	○	12	1014	●	8	579	○
Agonus cataphractus	Hooknose	ES	0	0	○	1	2413	●	0	0	○
Ciliata mustela	Fivebeard rockling	MM	2	290	●	3	1049	●	1	1007	●
Leuciscus leuciscus	Common dace	F	1	1	○	5	2339	●	0	0	○
Belone belone †	Garfish	MM	17	821	○	8	501	○	7	644	○
Myoxocephalus scorpius	Shorthorn sculpin	ES	2	1242	●	2	458	●	1	24	●
Perca fluviatilis	European perch	F	1	1	○	3	1562	○	0	0	○
Symphodus bailloni †	Baillon's wrasse	MS	15	523	○	6	231	○	6	270	○
Esox lucius	Northern pike	F	0	0	○	1	959	○	0	0	○
Echiichthys vipera †	Lesser weever	MS	7	924	○	2	6	●	0	0	○
Symphodus melops †	Corkwing wrasse	ES	16	454	○	4	151	○	6	261	○
Cottus gobio	Bullhead	F	1	1	○	3	827	○	0	0	○

Table continued in supplementary table 3. **Detections:** Total presence/absence per estuary. **TraC 2007 to 2017:** species present in fishing data**. Presence Cat.:** Exact Species Present: ●; Species present within clade: ■; No Species Present: ○. †: indicates detected in blanks.

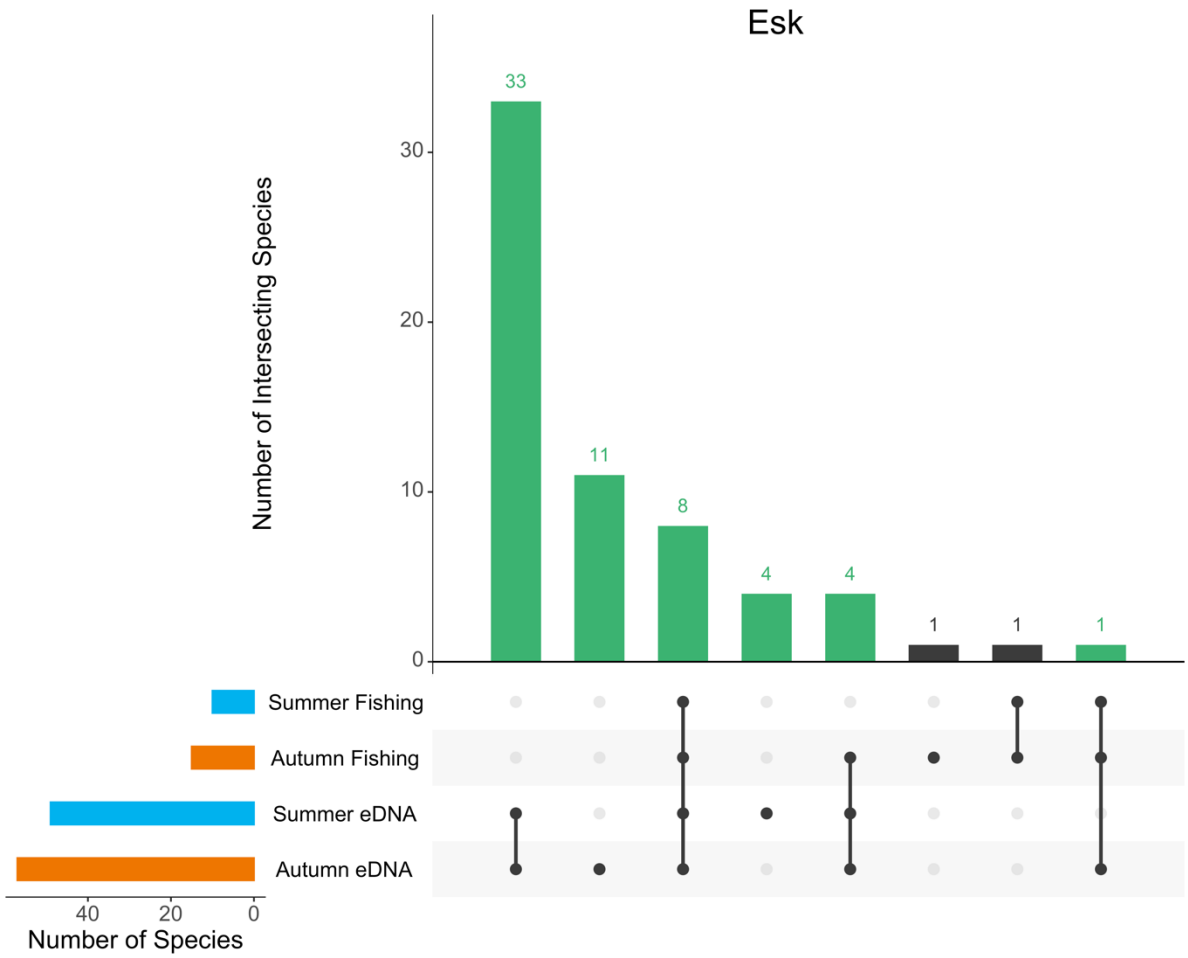


Figure 2: UpSet plot (Conway *et al.*, 2017) showing the number of intersecting species between each species list generated for each dataset for the Esk estuary (2016 and 2017). Species lists were generated for eDNA and fishing from early summer and autumn. The bottom panel shows each dataset and which intersections between species lists they contributed to. Single black dots indicate no intersections between species lists. Black dots connected by lines indicate which lists shared species. The top bar graph gives the number of species shared between each species list in each intersection. Green bars indicate intersections which contained species detected via eDNA, black bars indicate intersections where only fishing contributed. The lower left-hand bar chart shows the total number of species within each species list, blue bars indicate summer data, orange bars indicate autumn data.

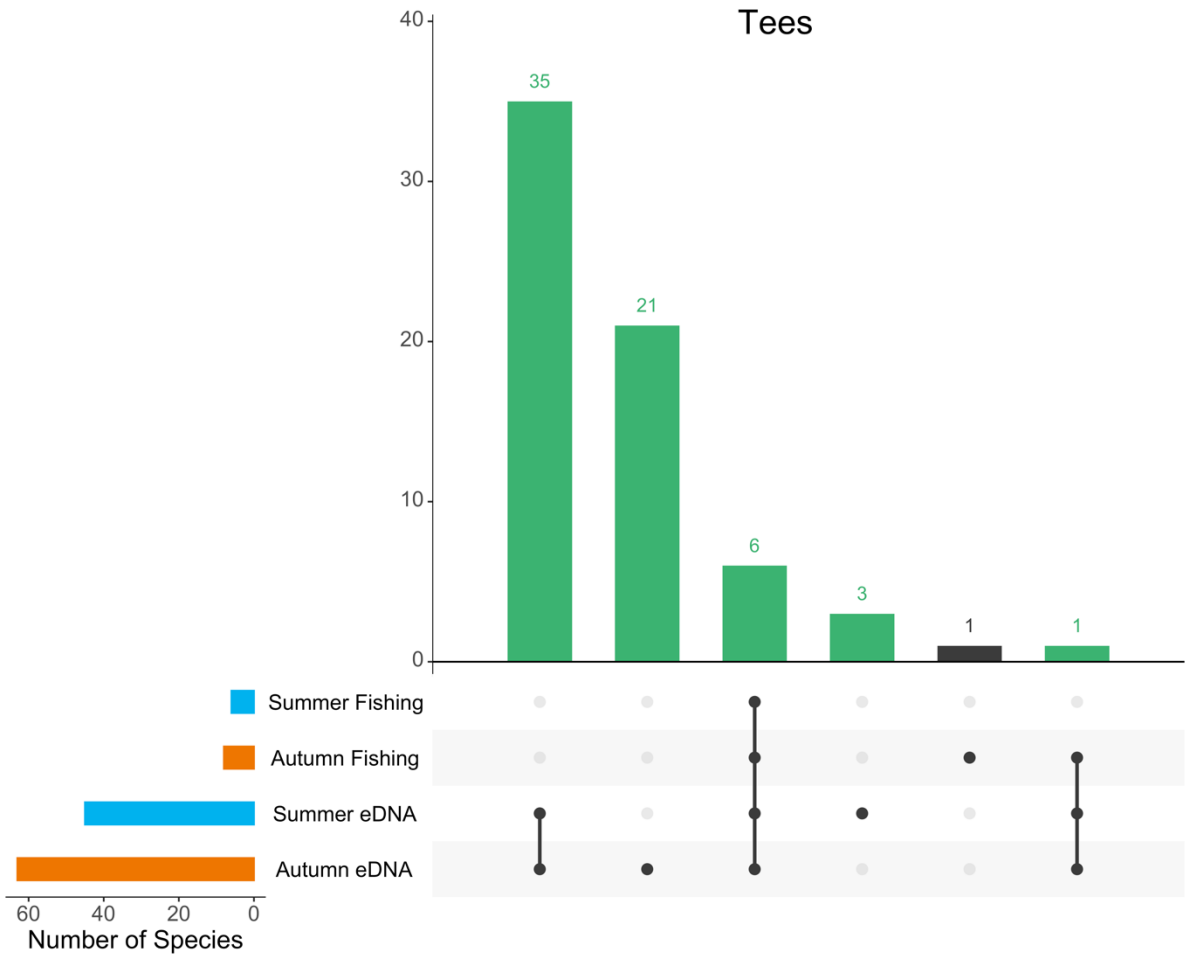


Figure 3: UpSet plot (Conway *et al.*, 2017) showing the number of intersecting species between each species list generated for each dataset for the Tees estuary (2016 to 2017). Species lists were generated for eDNA and fishing from early summer and autumn. The bottom panel shows each dataset and which intersections between species lists they contributed to. Single black dots indicate no intersections between species lists. Black dots connected by lines indicate which lists shared species. The top bar graph gives the number of species shared between each species list in each intersection. Green bars indicate intersections which contained species detected via eDNA, black bars indicate intersections where only fishing contributed. The lower left-hand bar chart shows the total number of species within each species list, blue bars indicate summer data, orange bars indicate autumn data.

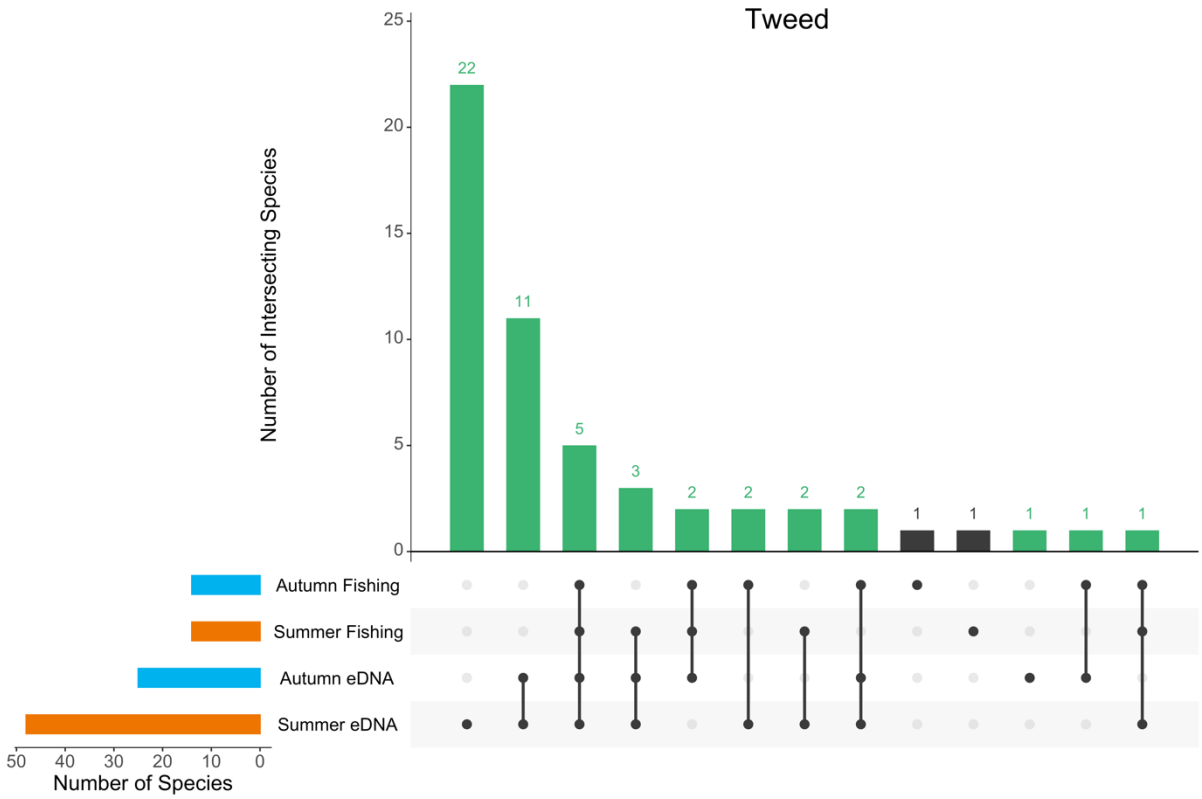


Figure 4: UpSet plot (Conway *et al.*, 2017) showing the number of intersecting species between each species list generated for each dataset for the Tweed estuary (2017). Species lists were generated for eDNA and fishing from early summer and autumn. The bottom panel shows each dataset and which intersections between species lists they contributed to. Single black dots indicate no intersections between species lists. Black dots connected by lines indicate which lists shared species. The top bar graph gives the number of species shared between each species list in each intersection. Green bars indicate intersections which contained species detected via eDNA, black bars indicate intersections where only fishing contributed. The lower left-hand bar chart shows the total number of species within each species list, blue bars indicate summer data, orange bars indicate autumn data.

3.3 Overall Species Richness from eDNA and Fishing

Species richness R/E curves calculated using all data for each estuary, per season, showed unfiltered eDNA data detected a higher asymptotic species richness compared to fishing on the Esk in summer and autumn (CI none-overlapping). On the Tees in autumn, both unfiltered and eDNA data filtered by TraC Fish detections (2007 - 2017) detected a higher species richness than fishing. There was no difference in asymptotic species richness between eDNA and fishing on the Tweed (figure 5; CI overlapping; supplementary table 4). Estimating the species richness per eight samples showed the filtered and unfiltered eDNA data consistently detected more species than fishing gears, for a given sample size. Except for the Tweed where this difference was only present in summer for unfiltered data (supplementary table 5). Therefore, eDNA is a more efficient method of detecting species richness for a given sample size. On the Esk and Tweed there was no difference in asymptotic species richness between seasons, both for filtered and unfiltered eDNA data. In the Tees, asymptotic species richness was higher in autumn than early summer in both the unfiltered and filtered eDNA data. There were no statistically significant differences in asymptotic species richness between seasons, per estuary for fishing gears (supplementary table 4).

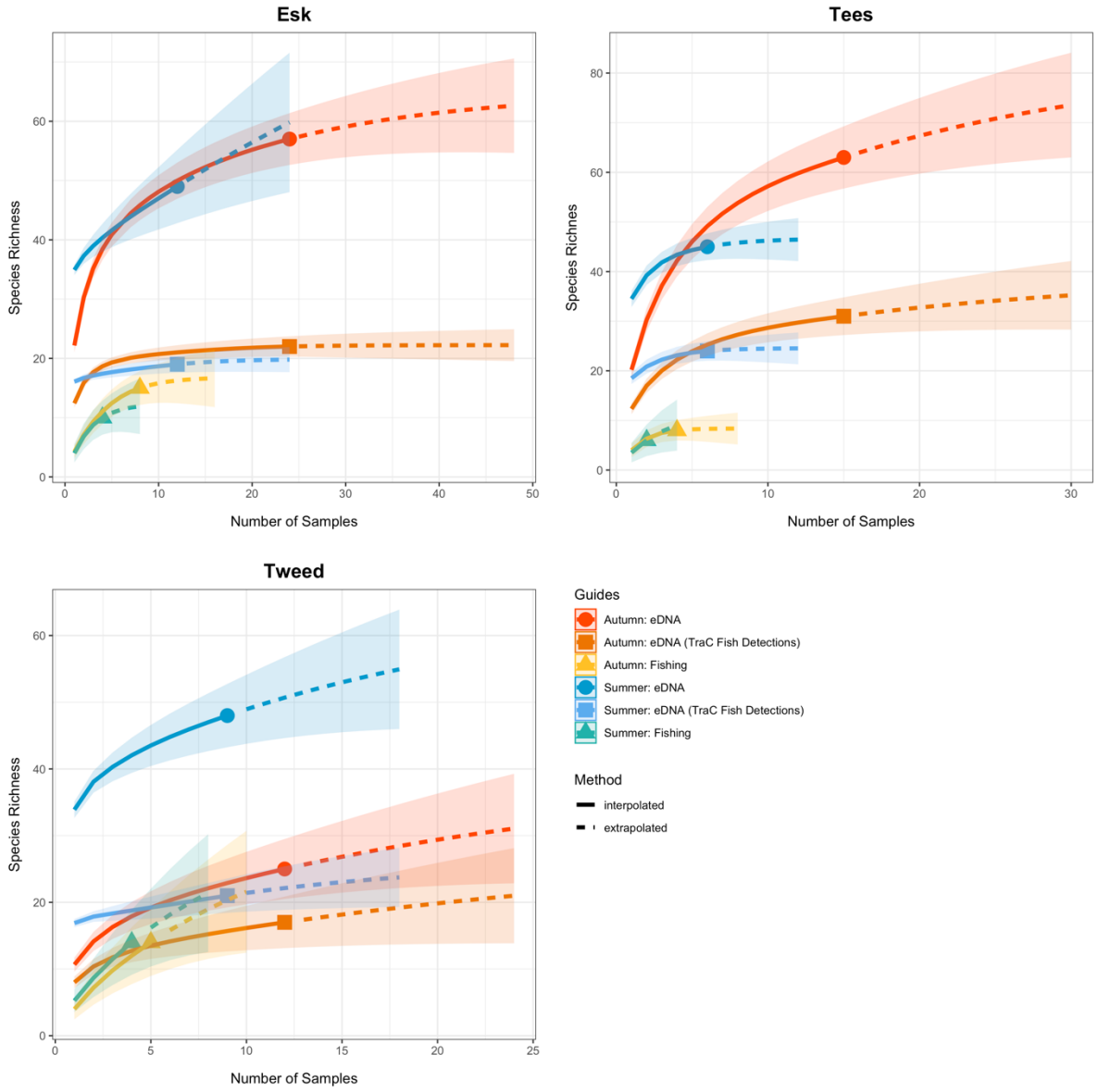


Figure 5: Species richness Rarefaction and Estimation (R/E) curves and 95% confidence intervals calculated on species presence/absence data for eDNA, eDNA data filtered by species detected in fishing (2007 to 2017), and fishing gears (combined across methods) for each estuary, per season. All available data was used for each estuary. Bright colours indicate autumn, cooler colours represent early summer sampling, shapes indicate dataset. R/E curves were calculated, to double the observed sample size, using the iNEXT software (Hsieh *et al.*, 2016), confidence intervals were calculated using 1000 bootstrap iterations.

Species richness R/E curves calculated by pooling the data across both seasons from 2017 showed the unfiltered eDNA detections appeared to be approaching, but had not reached, an asymptote in the Esk and Tweed, but were closer to reaching an asymptote in the Tees. Overall, the eDNA data filtered by TraC fish detections across *all* estuaries, appeared to be closer to reaching an asymptote than the unfiltered data. The fishing data were close to approaching an asymptote in the Tees and Esk, but not in the Tweed (figure 6). Neither the unfiltered or filtered eDNA data showed a statistically significant difference in asymptotic species richness between estuaries (figure 6, supplementary table 6). In the fishing data the Tweed showed a significantly higher asymptotic species richness than the Esk and Tees (supplementary table 6). In the Esk and Tees eDNA detected a significantly higher species richness than fishing methods, both for

the filtered and unfiltered eDNA data. There was no statistically significant differences in asymptotic species richness in the Tweed between methods (figure 5, supplementary table 6).

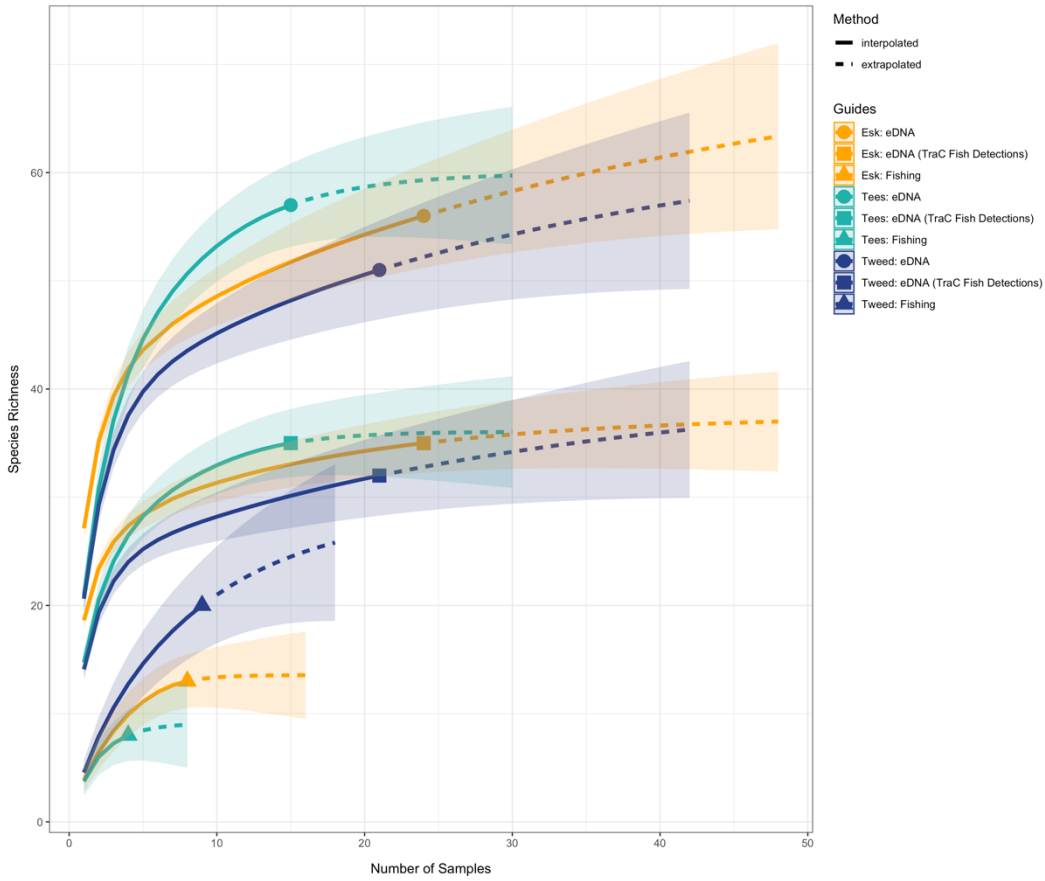


Figure 6: Species richness Rarefaction and Estimation (R/E) curves and 95% confidence intervals calculated on species presence/absence data for eDNA, eDNA data filtered by species detected in fishing (2007 to 2017), and fishing gears (combined across methods) for each estuary. Different estuaries are indicated by different colours, shapes indicate dataset. Only the 2017 data was used to make comparisons between estuaries valid. R/E curves were calculated, to double the observed sample size, using the iNEXT software (Hsieh *et al.*, 2016), confidence intervals were calculated using 1000 bootstrap iterations.

3.4 Station Level Assemblage Composition between Methods

The binomial GLLVM ordination of all eDNA and fishing gears at sampling stations, across both years, showed a clear difference in assemblage composition (presence/absence) between eDNA and all fishing methods (figure 7; model residuals: supplementary figure 6). Overall, the composition of fishing methods appeared relatively similar across stations, seasons, and estuaries, but a subtle gradient was present in assemblage composition between the Tweed, Esk and Tees. The eDNA stations showed substantially greater variability in assemblage composition. Firstly, there was a general separation in composition between eDNA stations in autumn and summer. In summer, the assemblage composition appeared more similar between stations in different estuaries, whereas in autumn there appeared to be greater separation between the estuaries (figure 7). However, this may have been driven by the inclusion 2016 data (supplementary figure 7). Interestingly this separation between estuaries appeared to mirror the subtle change seen in fishing gears. Ordination of only eDNA stations showed no improvement in model fit when per sample read depth was included (AIC increased from 1532 to 1627), suggesting read depth had limited influence on assemblage composition.

Analysis of assemblage composition using binomial multivariate GLMs between eDNA and each gear type was carried out individually, given the variation in the spatial distribution of fishing gears. Firstly, for comparisons of assemblage composition between eDNA and seine nets, the best fitting model contained the explanatory variables: method, season, estuary and year. There was a highly statistically significant effect of method (Wald: 8.008, $p = < 2 \times 10^{-16}$) on assemblage composition, whereas year, season, and estuary had no statistically significant effect (table 2; residuals: supplementary figure 8). There was no change in the statistical significance of the effects of method, season and estuary in an alternative model using only the 2017 data (supplementary table 7). Differences in assemblage composition between methods was confirmed by GLLVM ordination (supplementary figure 9). For method, six species were detected significantly more frequently in eDNA than in seine nets, across three estuarine use guilds. This included Atlantic Salmon (*Salmo salar*), another Anadromous species, an Estuarine Species, a Marine Migrant species and a marine taxa *Pollachius sp*. (table 3). Secondly, for comparisons of eDNA and fyke net composition, the best fitting model contained only method as an explanatory variable and its effect was not statistically significant (supplementary table 8; supplementary figure 10). Thirdly, the best fitting model comparing eDNA against beam trawls contained method, season and estuary as explanatory variables, but their effects were not statistically significant (supplementary table 9; supplementary figure 11). However, in both cases there was separation in assemblage composition between eDNA and the method in the GLLVM ordination (supplementary figure 6). Therefore, it seems likely the non-statistically significant effects were due to low sample sizes between eDNA and these two gear types.

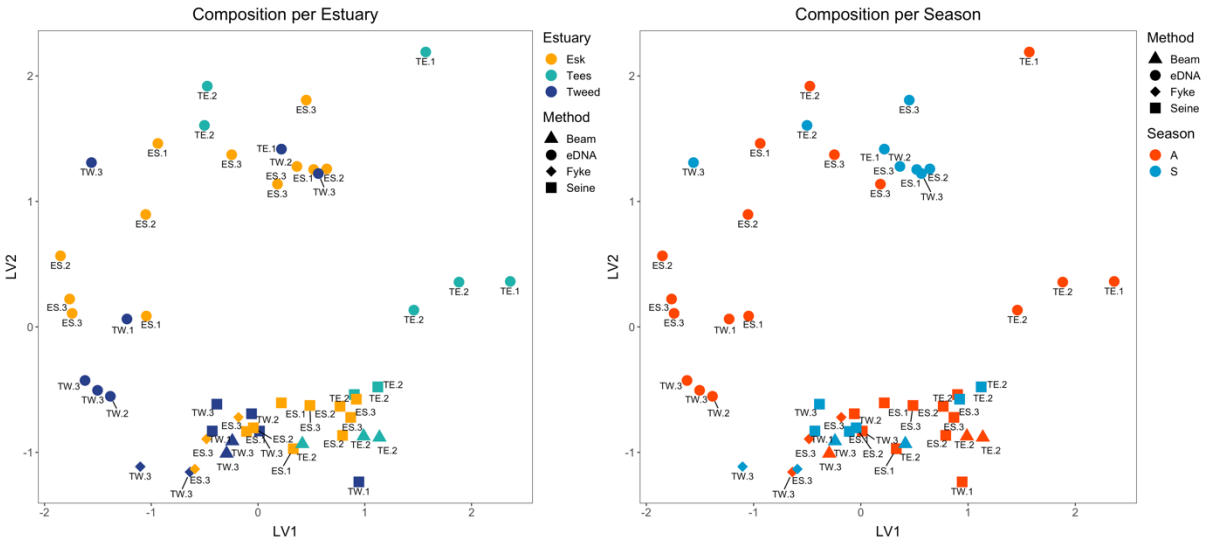


Figure 7: Presence/absence of fish species per station for eDNA and all fishing methods (seine nets, fyke nets and beam trawls) modelled using a binomial GLLVM (probit link; two latent variables; 50 iteration). In the left-hand panel colours indicate the estuary, whereas in the righthand panel they indicate season (A = autumn, S = early summer). Across both panels, shapes indicate the different sampling method. Point labels show estuary and general sampling site within which stations were situated.

Table 2: ANOVA results for the comparison between eDNA and seine nets.

AIC: 1301			
Final Model: Species Presence/Absence ~ Method + Season + Estuary + Year			
Explanatory Variable	**Residual DF**	**Wald-Test**	**P-Value**
Intercept	31		
Method	30	8.008	$< 2 \times 10^{-16}$
Season	29	6.766	0.082
Estuary	27	6.226	0.447
Year	26	3.213	0.513

Table 3: ANOVA results for species with a statistically significant association ($p > 0.05$; adjusted for multiple testing) for Method from the comparison between eDNA and seine nets.

	Method (Seine vs. eDNA)	
Species (Guild)	Wald Statistic	*p*-value
Gasterosteus aculeatus (A)	3.435	0.002
Pleuronectes platessa (MM)	3.217	0.004
Pomatoschistus microps (ES)	2.693	0.019
Pomatoschistus minutus (ES)	2.974	0.008
Salmo salar (A)	3.708	$< 2 \times 10^{-16}$
Pollachius sp. (UA)	2.575	0.039

3.5 Sample Level Assemblage Composition between Seasons and Estuaries

Binomial GLLVM ordination of only the eDNA samples taken in 2017 without comparison to fishing gears, showed a clear difference in assemblage composition between the different estuaries. There was also a substantial difference in assemblage composition between summer and autumn in each estuary. There also appeared to be slightly greater separation between the three estuaries in autumn compared to summer (figure 8; residuals: supplementary figure 12). The ordination showed no improvement in model fit, when per sample read depth was included as an explanatory variable (AICc increased from 2487 to 2514), suggesting read depth had limited influence on assemblage composition overall. In the binomial multivariate GLM sampling station was used as a blocking variable to account for sample none-independence at the station level. The initial exploratory model using only data for samples which had salinity zone classifications showed that model fit improved when zone (along with the interaction term) was dropped from the model (initial AIC: 2239, final AIC: 1771). Therefore, a second model was run including the stations in the upper Tees which lacked salinity classifications, and without zone as an explanatory variable. Only season and estuary were retained by model selection using AIC, the interaction term between them did not improve model fit. Both season and estuary showed a strongly statistically significant effect on assemblage composition (table 4, residuals: supplementary figure 13), supporting the evidence from the ordination (figure 8). In addition, seven species across the guild classifications: Estuarine Species and Marine Straggler species showed a significantly higher incidence in summer compared to autumn. Three other species, *Limanda limanda* (MM), *Molva molva* (MS), *Anguilla anguilla* (C) and *Scomber scombrus* (MS) showed statistically significant differences in incidence across estuaries (table 5), with the exact relationship depending on the species (supplementary figure 14).

Table 4: ANOVA results for seasons and estuaries for eDNA in 2017

AIC: 2012			
Final Model (Block - Station): Species Presence/Absence ~ Season + Estuary			
Explanatory Variable	**Residual DF**	**Wald-Test**	**P-Value**
Intercept	59		
Season	58	12.66	$< 2 \times 10^{-16}$
Estuary	56	11.31	0.001

Table 5: ANOVA results per species with a statistically significant association ($p > 0.05$; adjusted for multiple testing) with Season and Estuary.

	Season		**Estuary**	
Species (Guild)	Wald Statistic	*p*-value	Wald Statistic	*p*-value
Atherina presbyter (ES)	4.674	0.003		
Chelon ramada/labrosus (UA)	4.787	0.002		
Pomatoschistus microps (ES)	4.273	0.013		
Pomatoschistus minutus (ES)	4.138	0.020		
Trisopterus minutus (MS)	4.674	0.003		
Zeugopterus punctatus (MS)	4.691	0.003		
Labrus bergylta (MS)	4.787	0.002		
Scomber scombrus (MS)			3.624	0.008
Limanda limanda (MM)			3.477	0.014
Molva molva (MS)			3.609	0.008
Anguilla anguilla (C)			3.666	0.008

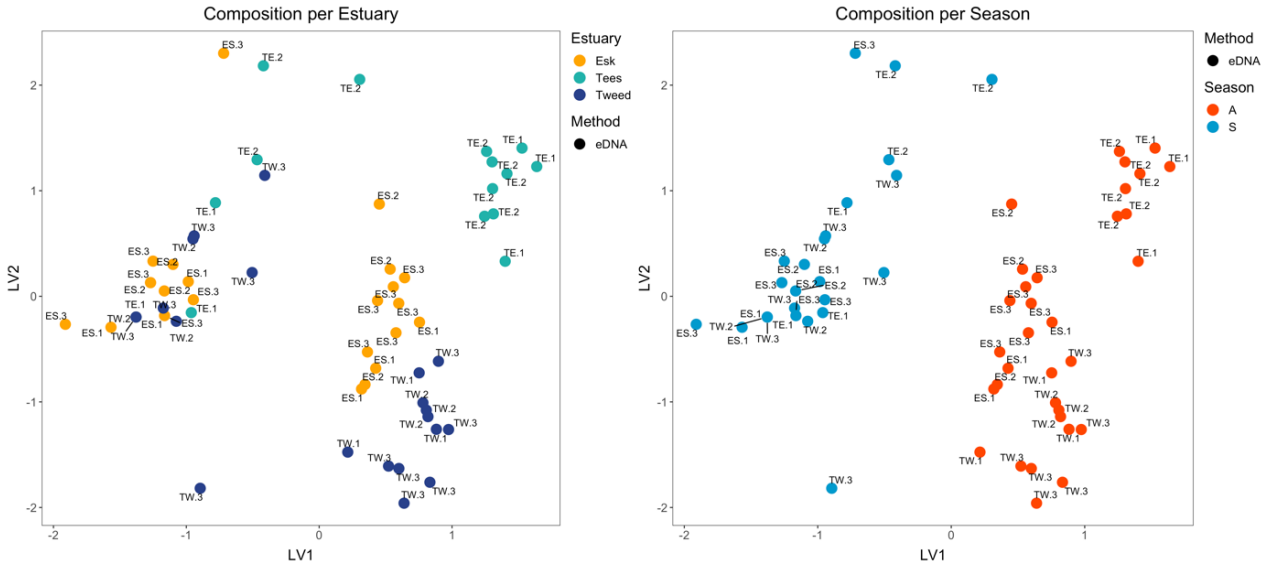


Figure 8: Presence/absence of fish species per sample for eDNA in 2017, modelled using a binomial GLLVM (probit link; two latent variables; 50 iteration). In the left-hand panel colours indicate the estuary, whereas in the righthand panel they indicate season (A = autumn, S = early summer). Across both panels, shapes indicate the different sampling method. Point labels show estuary and general sampling site within which sampling stations, and samples were situated.

4. Discussion

4.1 Overview

This is one of the first studies to compare the fish assemblages detected via eDNA metabarcoding of surface water samples with conventional fishing gears, across multiple macrotidal estuaries and seasons. Previous studies in estuaries have either focused on single ecosystems, not provided direct comparison with fishing data, or not investigated seasonal trends across estuaries. The majority of species caught by fishing were detected by eDNA and species richness estimates for each estuary were greater using eDNA, in several cases. Numerous previously unrecorded species were detected by the eDNA analysis, which included both endangered species and a none-native. In addition, when focusing the analysis on the sample level, eDNA metabarcoding data resolved clear spatio-temporal changes in assemblage composition. Whereas comparative analyses of both eDNA and fishing data at the station level did not resolve these spatio-temporal differences.

4.2 Detection and Composition between eDNA and Fishing Gears

Overall, eDNA detected 76 species, approximately half of which were novel detections, mostly consisting of Marine Stragglers. The species of conservation interest detected by eDNA were European eel (*Anguilla anguilla*), European smelt (*Osmerus eperlanus*), a species of shad (*Alosa sp.*), a species of lamprey (*Lampetra sp.*) and Atlantic salmon (*Salmo salar*). We acknowledge that the *S. salar* detection could be environmental contamination from wastewater effluent, or genuine detections. In addition, it was possible to detect the potentially invasive pink salmon (*O. gorbuscha*) in the Tees (NNSS, 2019). Therefore, eDNA metabarcoding is an effective way of detecting species of conservation interest and non-natives in estuaries, as concluded elsewhere (Kume *et al.*, 2021). Overall, eDNA detected 86.7 % to 90.0% of the species caught by fishing, depending on the estuary, between 2016 and 2017. This is higher than comparable studies in UK estuaries where eDNA detected 71% (Chapter 2) and 72% (Hallam *et al.*, 2021) of the species detected by fishing gears. Therefore, although the greatest coverage of fish species would be provided by a combination of eDNA and fishing (as elsewhere; Chapter 2; Hallam *et al*., 2021; Zou *et al.*, 2020), only a few species were missed by the eDNA analysis. Species absent from the eDNA data were eelpout (*Z. viviparus*) and ninespine stickleback (*P. pungitus*), both present in the reference database. However, the reference sequence for *Z. viviparus* was short (135 bp), which probably caused a failure of detection. More of an issue was the inability of the 12S fragment generated by the Tele02 primers to resolve the various species of the Ammodytidae, and the congeneric pollock and saithe (*Pollachius* spp.), causing an underestimation of species diversity. Although data cleaning did not affect these comparisons, it probably had an influence on comparisons of composition at the sampling station level. A wide range of species were present in field, extraction and PCR blanks, but at a low number of reads overall. A threshold cut off, a standard approach in eDNA metabarcoding (Sepulveda *et al.*, 2020) was therefore used to address contamination, using an adaption of the methodology in Yamamoto *et al.* (2017).

Environmental DNA metabarcoding could detect a higher asymptotic species richness than fishing gears in different seasons in 2016 and 2017 (barring the Tees in summer), and across data for both seasons in 2017, for the Esk and Tees. Calculating the expected species richness for a given sample size showed that eDNA was generally more efficient at detecting species than fishing. Therefore, there was support for the initial hypothesis that eDNA would detect more species in an estuary overall. This was comparable to results from the Dee (Wales, UK) where eDNA detected a greater species richness compared to seine and fyke nets (Chapter 2),

and studies that have shown a greater species richness than conventional methods (Zou *et al.*, 2020; Hallam *et al.*, 2021; Cole *et al.*, 2022). However, eDNA did not detect a higher asymptotic species richness than fishing in the Tweed. It is notable that the Tweed was the only estuary to be sampled with the full complement of fishing gears (seine, fyke and beam trawls) used in the multi-method survey. It is likely therefore that the non-significant differences in estimated species richness was because a greater component of the fish community was being sampled by the multiple gear types. This is comparable to the tidal Thames where eDNA showed no statistically significant differences in overall species richness estimates compared to a multimethod fishing survey (CI overlapping; Hallam *et al.*, 2021).

The assemblage composition of stations sampled with eDNA metabarcoding and seine nets showed a clear, statistically significant difference. For comparisons between eDNA and fyke and seine nets, this difference was only obvious from ordination and was not statistically significant. This was probably due to the small sample sizes used in these comparisons. In combination with the evidence from the qualitative analysis, this supports the assertion that eDNA metabarcoding and fishing gears capture different components of a given assemblage (Aglieri *et al.*, 2021). A different assemblage composition between eDNA and seine nets has been reported previously (Chapter 2). Differences in assemblage composition are comparable to those reported between eDNA and BRUVs (Baited Remove Underwater Videos; Cole *et al.*, 2022) and a multimethod netting technique on the Thames (Hallam *et al.*, 2021). As with Chapter 2, the species detected more frequently using metabarcoding were also detected in seine nets to some degree in 2016 and 2017. Specifically, these were three-spined stickleback (*Gasterosteus aculeatus*), European plaice (*Pleuronectes platessa*), common goby (*Pomatoschistus microps*), sand goby (*Pomatoschistus minutus*), saithe/pollock (*Pollachius* sp.) and Atlantic salmon (*S. salar*). All these species, except *S. salar*, are known to have been detected at least once in every estuary from 2007 to 2017. Overall, this suggests eDNA metabarcoding is a reliable method for detecting many species commonly found in estuaries. This is a more problematic conclusion for species which are widely eaten by humans e.g. *S. salar*. Further research is required to determine the influence of wastewater on the false positive detection of fish species.

4.3 Comparison between Seasons and Estuaries

Initial analyses using eDNA data aggregated at the level of the sampling station and fishing data were unable to detect assemblage differences between estuaries and seasons. Although it was obvious from visual inspection of ordinations that such differences existed. Interestingly, the differences between estuaries captured by the eDNA data, appeared to mirror subtle variations in assemblage composition obvious from fishing gears. Comparably, when analysed independently in 2017, and taking into account sample non-independence, the eDNA data showed a clear difference in assemblage composition between estuaries and seasons. This supported the hypothesis that eDNA would show differences in composition between seasons and estuaries, but not for salinity zones. Regarding seasonal differences there was a clear shift in assemblage composition between early summer (May/June) and autumn (October/September) which was consistent across estuaries. This is comparable to eDNA metabarcoding studies that have detected seasonal changes in the fish assemblage composition within estuaries (Stoeckle *et al.*, 2017; Zou *et al.*, 2020) and in coastal ecosystems (Sigsgaard *et al.*, 2017). Seasonal changes in the fish assemblage structure in temperate estuaries are well documented (Maes *et al.*, 2005; Henderson and Bird, 2010; Selleslagh *et al.*, 2012). Seasonal ichthyofaunal changes are caused by sequential immigration and emigration of marine, estuarine, diadromous and freshwater fishes. These changes in species composition are controlled by spawning times, and the time needed for juveniles and larvae to recruit into the

estaury (Maes *et al.*, 2005; Teichert *et al.*, 2018b). Sampling over early summer and autumn in TraC fish surveys accounts for this migration (Waugh *et al.*, 2019). Despite sampling multiple seasons, differences in asymptotic species richness between seasons were harder to identify. Only the Tees showed a greater species richness in autumn. The lack of overt differences between summer and autumn sampling is surprising, as we expected a higher species richness in autumn compared to early summer in Britain (e.g. in the Severn Estuary, Henderson & Bird, 2010).

Regarding composition, seven species occurred more frequently in early summer than in autumn. Two of the species, the Marine Stragglers topknot (*Zeugopterus punctatus*) and ballan wrass (*Labrus bergylta*), were novel detections. In addition, a species of mullet *Chelon labrosus* or *C. ramada* was also detected. *C. ramada* has previously been caught by fishing on the Esk. It is notable that for five of the seven species detected more frequently in autumn, their spawning times in the waters around Britain overlap with the sampling period in early summer (supplementary table 7). It is possible that spawning within, or outside of the estuaries, depending on the ecological guild, caused greater detection in early summer. For species such as *Pomatoschistus minutus*, and other *Pomatoschistus spp.,* adult abundance is generally higher in autumn, rather than early summer (Maes *et al.*, 2005; Henderson and Bird, 2010). This supports the assumption that it was spawning rather than adult abundance driving this pattern. The exception to this was Poor cod (*Trisopeterus minutus*) which spawns from February to March (and can be more abundant in autumn, Henderson & Bird, 2010) and *C. labrosus* which spawns from July to August (supplementary table 10). However, in the case of *C. labrosus* it is notable that juveniles move into estuaries in April-June (Kottelat and Freyhof, 2007) whereas *C. ramda* does spawn in June (Maitland and Campbell, 1992). Therefore, changes in abundance across various life history stages, depending on species, probably drove the differences in assemblage composition in eDNA rather than spawning alone.

In addition to the changes between seasons there was a clear shift in assemblage composition between estuaries, over both seasons. Regarding the species level results, *Molva molva* (MS) and *Scomber scombrus* (MS) had never been detected by fishing gears from 2007 to 2017 across estuaries. It is intriguing that *L. limanda* had the lowest incidence from eDNA metabarcoding in the Tweed, in which this species was never detected by fishing from 2007 to 2017. Therefore, it is possible the eDNA data reflected differences in abundance between systems. *A. anguilla* had the lowest incidence from eDNA data in the Tees. This estuary has a history of extremely high levels of industrial chemical pollution (e.g. Polycyclic Aromatic Hydrocarbons; Law *et al.*, 1997; Woodhead *et al.*, 1999) and individuals of *A. anguilla* in the Tees are heavily contaminated with PAH (Ruddock *et al*., 2003). Therefore, it is entirely possible that environmental pollutants may influence the abundance of eels in the Tees and subsequent eDNA detections. General differences in assemblage composition could be due to differing environmental factors between the estuaries. For example, there may be differences in the type and amount of available habitat types among the estuaries (EA, 2016). Subtidal soft sediments (Pihl *et al.*, 2002) and intertidal areas are important fish habitat in estuaries (Nicolas *et al.*, 2010b; Teichert *et al.*, 2018a), and the composition of fish assemblages is influenced by the structure of intertidal seascapes (Teichert *et al.*, 2018a). As mentioned above, variation in the concentrations of chemical pollutants between estuaries may be another factor. From 2013 to 2016 the Esk consistently failed WFD assessments relating to chemical pollution and the presence of priority hazardous substances (EA, 2022a), while the Tees failed in every year except 2015 (EA, 2022b). By comparison, the Tweed showed generally 'good' water quality in terms of chemical pollutants (barring 2013 and 2014) and priority hazardous substances across several years (EA, 2022c). Contamination by heavy metals has been shown to decrease the density of individuals of juveniles of Marine Migrants species and organic contaminants

have a negative effect on the probability of detecting these fishes and their overall species richness (Courrat *et al.*, 2009). More general environmental factors such as variation in the size of the different estuaries studied may have influenced the structure of fish assemblages, e.g. the Tees was the largest estuary (ABPmer and HR Wallingford, 2007). In the North-East Atlantic, larger estuaries, with wider entrance widths typically have a higher fish species richness (Nicolas *et al.*, 2010b, 2010a; Waugh *et al.*, 2019). There was also no statistically significant difference in asymptotic species richness between the estuaries detected using eDNA in 2017. The higher asymptotic species richness recovered by fishing methods in the Tweed compared to the Esk and Tees was probably a function of the different gear combinations used. The fact that a higher species richness in the Tweed could not be detected using eDNA may be because of the influence of eDNA transport, because the 'true' fish species richness does not actually vary, or because of other stochastic factors influencing eDNA detection.

Salinity zone did not improve the fit of models studying assemblage composition. This is in contrast to other spatially replicated eDNA studies which have found a clear effect of salinity on fish assemblage composition (Chapter 2; Ahn *et al.*, 2020). Salinity, is one of the key environmental variables influencing the spatial structure of estuarine fish assemblages (Nicolas *et al.*, 2010b; Selleslagh *et al.*, 2009; Whitfield *et al.*, 2012). It is likely that the use of broad salinity zone classifications rather than exact salinity measurements produced this effect, as the latter is considered a better measure of studying salinity effects on the nekton (Greenwood, 2007). Therefore, concurrent sampling of salinity and other physiochemical parameters alongside eDNA is desirable to take their influence into account. Another issue with the present study is one of sampling design. It was not possible to sample all estuaries in early summer in 2016 or both seasons in this year. Therefore, the temporally unbalanced design reduced the utility of the 2016 data. In addition, the spatial distribution of sampling stations was not consistent within each estuary, particularly in the Tees. Thus spatial heterogeneity in eDNA sampling may have influenced the observation of differences in assemblage composition between estuaries. This was a consequence of the eDNA design generally mirroring the fishing surveys. Future eDNA surveys for biomonitoring should aim for spatially consistent designs, given that eDNA sampling has less deployment constraints than fishing gears (see Chapter 2).

4.4 Implications for Management

This study has provided a direct comparison of overall species richness and assemblage composition between eDNA metabarcoding and fishing gears, in multiple estuaries and over two seasons. Further research should use the data generated from eDNA metabarcoding to calculate metrics for the TFCI which require estimates of species richness (Coates *et al.*, 2007). These metrics can then be used to investigate differences in the anthropogenic pressures such as levels of chemical contamination between ecosystems. To increase the power of these analyses, sampling should be expanded to a larger number of estuaries that are currently sampled using TraC fish surveys in the United Kingdom (see Waugh *et al.*, 2019). A clear advantage here is that eDNA sampling designs are relatively easy to standardise between estuaries, compared to multigear fishing methods, and therefore eDNA is likely able to provide greater comparability and discriminatory power between ecosystems. It was striking from this analysis that sample level eDNA data could detect clear differences between estuaries compared to analyses using station level data. Further research on the relationship between reads from eDNA metabarcoding and abundance is also required in estuaries. This has not been addressed here but it may be that combining the relative abundance data from metabarcoding with qPCR to provide concentration estimates for multiple species would be useful (Bleijswijk *et al.*, 2020). Another area that needs to be resolved in the utility of eDNA metabarcoding, is

how to incorporate the wealth of additional, novel species detections that typically arise from eDNA analysis. The novel species detections will be a mix of species which are less likely to be detected by fishing gears but are present within the estuary, and species that are detected as a result of eDNA transport, both from upstream and the open sea. Clearly transport is an issue as spatially specific data is required for assessments of ecosystem health. Research into eDNA transport incorporating particle tracking models along with occupancy and process based models allowing false positives to be accounted for, is required (see Burian *et al.*, 2021). Caveats notwithstanding, even at the current development stage, the identification of a recent none-native (e.g. pink salmon) and endangered taxa (e.g. lamprey, shad) illustrates the utility of eDNA to provide immediate evidence to influence management actions.

5. Conclusion

Overall, it can be concluded that eDNA metabarcoding is an effective way to assess the biodiversity of fishes between multiple macrotidal estuaries and seasons. It is increasingly clear that this technique can play a role alongside existing methods for the assessment of ecosystem health in estuaries.

Acknowledgments

This work is dedicated to the memory of Dr Laura Corrigan, who sadly passed away in 2020. Laura was critical to the initial collaboration and coordination of this project between the 'SeaDNA' team and the Environment Agency. We would also like to acknowledge Dr Naiara Sales, Riccardo Lollobrigidi, Dr Sandra Garces-Pastor, and the Environment Agency staff for all their estuarine sampling efforts over the years. This study has been funded by Grant NE/N005759/1 from the UKRI Natural Environment Research Council.

Chapter 4: Supplementary Material

Supplementary Table 1: Sampling dates for fishing and eDNA

Estuary	Year	Season	Fishing Dates	eDNA Dates
Esk	2016	Autumn	25/10/2016	26/10/2016
			26/10/2016	26/10/2016
	2017	Autumn	17/10/2017	17/10/2017
		Summer	22/06/2017	22/06/2017
			23/06/2017	22/06/2017
Tees	2016	Autumn	31/10/2016	31/10/2016
			13/11/2016	31/10/2016
	2017	Autumn	27/09/2017	27/09/2017
		Summer	31/05/2017	30/05/2017
Tweed	2017	Autumn	19/09/2017	19/09/2017
			20/09/2017	19/09/2017
		Summer	08/06/2017	07/06/2017

Note: the later date range for the Tees in autumn relates to otter trawling.
This was not directly considered in this study.

Supplementary Table 2: sampling effort for eDNA

Estuaries	Year	Season	Station N.	Sample N.
Esk	2016	Autumn	4	12
	2017	Spring	4	12
		Autumn	4	12
Tees	2016	Autumn	2	6
	2017	Spring	2	6
		Autumn	3	9
Tweed	2017	Spring	3	9
		Autumn	4	12

Supplementary Table 3: continuation of table 1 - detected species in each estuary.

Species	Common Name	Guilds	Esk			Tees			Tweed		
			Detections	Reads	TRAC 2007-2017	Detections	Reads	TRAC 2007-2017	Detections	Reads	TRAC 2007-2017
Pholis gunnellus	Rock gunnel	ES	1	1	●	2	11	●	2	760	●
Gobius paganellus †	Rock goby	ES	14	239	○	8	216	○	9	277	●
Trisopterus luscus †	Pouting	MM	6	580	○	4	74	●	0	0	○
Scophthalmus maximus †	Turbot	MM	3	195	○	3	320	●	0	0	○
Squalius cephalus	Chub	F	2	4	○	6	467	○	1	10	○
Alosa sp. †	River herring	A	12	262	○	3	75	○	4	88	○
Zeus faber †	Dory	MS	4	88	○	8	306	○	2	18	○
Syngnathus rostellatus †	Nilsson's pipefish	ES	11	168	○	3	57	●	7	160	●
Ciliata septentrionalis †	Northern rockling	MS	3	296	○	3	45	○	0	0	○
Aphia minuta †	Transparent goby	ES	3	94	○	6	173	○	3	59	○
Hippoglossoides platessoides	Long rough dab	MS	0	0	○	3	172	○	0	0	○
Solea solea †	Common sole	MM	2	27	○	4	94	●	1	4	○
Engraulis encrasicolus	European anchovy	MM	5	117	○	0	0	○	0	0	○
Spinachia spinachia	Sea stickleback	ES	0	0	○	0	0	○	1	69	●
Oncorhynchus gorbuscha	Pink salmon	UA	0	0	○	1	67	○	0	0	○
Osmerus eperlanus	European smelt	A	1	57	●	0	0	○	0	0	○
Arnoglossus laterna †	Scaldfish	MS	0	0	○	2	4	○	1	50	○
Labrus mixtus †	Cuckoo wrasse	MS	0	0	○	1	52	○	0	0	○
Microstomus kitt †	Lemon sole	MS	2	41	○	1	2	●	0	0	○
Crystallogobius linearis †	Crystal goby	MS	1	4	○	1	37	○	0	0	○
Cyclopterus lumpus	Lumpsucker	MM	0	0	○	0	0	○	1	27	○
Lophius piscatorius †	Angler	MS	0	0	○	4	19	○	0	0	○
Chelon auratus †	Golden grey mullet	MM	2	13	○	0	0	○	0	0	○
Cepola macrophthalma †	Red bandfish	UA	1	4	○	1	2	○	0	0	○

Continuation of Supplementary Table 3: detected species in each estuary.

Species	Common Name	Guilds	Esk			Tees			Tweed		
			Detections	Reads	TRAC 2007-2017	Detections	Reads	TRAC 2007-2017	Detections	Reads	TRAC 2007-2017
Buglossidium luteum †	Solenette	MS	1	5	○	0	0	○	0	0	○
Callionymus lyra	Common dragonet	MS	0	0	○	0	0	○	1	5	○
Ctenolabrus rupestris	Goldsinny-wrasse	MS	0	0	○	1	5	○	0	0	○
Spondyliosoma cantharus †	Black sea bream	MM	0	0	○	1	4	○	0	0	○
Micrenophrys lilljeborgii	Norway bullhead	MS	0	0	○	1	2	○	0	0	○
Mullus surmuletus †	Striped red mullet	MS	0	0	○	1	1	○	0	0	○

Continuation of Table 1. **Detections:** Total presence/absence per estuary. **TraC: 2007 to 2017** indicates presence in fishing data**. Presence Cat.:** Exact Species Present: ●; Species present within clade: ■; No Species Present: ○. †: indicates detected in blanks.

Supplementary Table 4: Estimated asymptotic richness for each estuary, per season (2016 and 2017 data)

Estuary	Site	Season	Species Richness				
			Observed	Estimator	SE	95% CI	
						Lower CI	Upper CI
Tweed	eDNA	Summer	48	66.00	18.00	51.52	140.03
		Autumn	25	39.67	15.16	27.76	102.97
	eDNA (TraC Fish Detections)	Summer	21	26.33	6.49	21.82	55.55
		Autumn	17	28.46	15.74	18.52	103.22
	Fishing	Summer	14	32.75	18.31	17.77	107.22
		Autumn	14	34.00	19.49	18.03	113.23
Tees	eDNA	Summer	45	46.67	2.27	45.22	57.39
		Autumn	63	84.00	15.06	68.94	137.22
	eDNA (TraC Fish Detections)	Summer	24	24.56	1.13	24.05	30.81
		Autumn	31	39.40	9.53	32.41	80.91
	Fishing	Summer	6	12.25	8.75	6.81	54.15
		Autumn	8	8.38	0.83	8.03	13.12
Esk	eDNA	Summer	49	109.50	36.80	69.15	230.69
		Autumn	57	64.99	6.64	58.93	90.08
	eDNA (TraC Fish Detections)	Summer	19	19.92	2.13	19.06	32.23
		Autumn	22	22.24	0.71	22.01	26.59
	Fishing	Summer	10	12.34	2.85	10.36	25.15
		Autumn	15	16.82	2.24	15.28	26.98

CI = Confidence Interval. SE = Standard Error.

Supplementary Table 5: Estimated species richness and confidence intervals, per 8 samples, per estuary and season.

Estuary	Dataset	Season	Method	SC	Species Richness		
					Estimate	95 % CI	
						Lower CI	Upper CI
Esk	eDNA	Summer	interpolated	0.97	44.96	41.02	48.91
		Autumn	interpolated	0.95	45.83	43.25	48.41
	eDNA (TraC Fish Detections)	Summer	interpolated	0.99	18.32	17.19	19.44
		Autumn	interpolated	0.98	20.34	19.41	21.27
	Fishing	Summer	extrapolated	0.96	11.92	6.80	17.04
		Autumn	observed	0.89	15.00	12.17	17.83
Tees	eDNA	Summer	extrapolated	0.99	45.82	42.53	49.10
		Autumn	interpolated	0.91	53.82	49.29	58.34
	eDNA (TraC Fish Detections)	Summer	extrapolated	1.00	24.34	21.91	26.77
		Autumn	interpolated	0.94	27.28	24.51	30.04
	Fishing	Summer	extrapolated	0.93	11.42	3.67	19.17
		Autumn	extrapolated	1.00	8.36	5.36	11.36
Tweed	eDNA	Summer	interpolated	0.97	47.00	42.91	51.09
		Autumn	interpolated	0.93	22.11	18.82	25.41
	eDNA (TraC Fish Detections)	Summer	interpolated	0.97	20.56	18.05	23.06
		Autumn	interpolated	0.94	15.22	12.65	17.79
	Fishing	Summer	extrapolated	0.75	21.39	12.66	30.11
		Autumn	extrapolated	0.66	18.97	11.91	26.03

CI = Confidence Interval. SC = Sample Coverage.

Supplementary Table 6: Estimated asymptotic richness for each estuary in 2017

Dataset	Estuary	Species Richness				
		Observed	Estimator	SE	95% CI	
					Lower CI	Upper CI
eDNA	Esk	56	71.97	14.25	59.57	127.57
	Tees	57	59.99	2.89	57.61	71.68
	Tweed	51	63.86	11.89	53.75	111.03
eDNA (TraC Fish Detections)	Esk	35	37.56	3.36	35.36	53.18
	Tees	35	36.07	1.54	35.13	43.53
	Tweed	32	40.57	9.71	33.44	82.87
Fishing	Esk	13	13.56	1.02	13.05	18.95
	Tees	8	9.13	1.80	8.12	18.23
	Tweed	20	28.89	7.50	22.11	57.43

CI = Confidence Interval. SE = Standard Error.

Supplementary Table 7: ANOVA results for the comparison between eDNA and seine nets in 2017

AIC: 976			
Final Model: Species Presence/Absence ~ Method + Season + Estuary			
Explanatory Variable	**Residual DF**	**Wald-Test**	**P-Value**
Intercept	23		
Method	22	6.909	<2e-16 ***
Season	21	5.051	0.527
Estuary	19	3.683	0.611

Supplementary Table 8: ANOVA results for the comparison between eDNA and fyke nets

AIC: 469			
Final Model: Species Presence/Absence ~ Method			
Explanatory Variable	**Residual DF**	**Wald-Test**	**P-Value**
Intercept	9		
Method	8	2.688	0.519

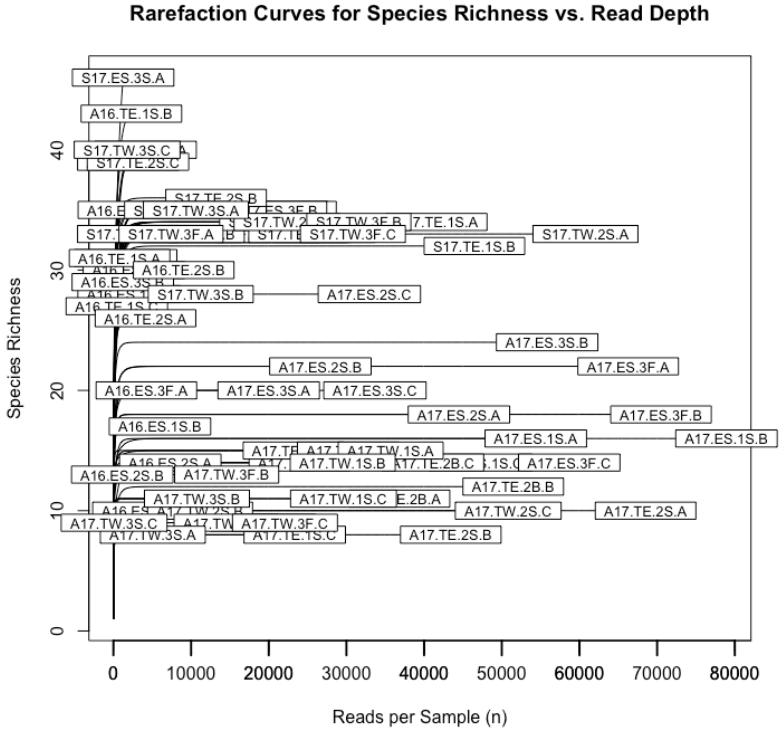
Supplementary Table 9: ANOVA results for the comparison between eDNA and beam trawls

AIC: 653			
Final Model: Species Presence/Absence ~ Method + Season + Waterbody			
Explanatory Variable	**Residual DF**	**Wald-Test**	**P-Value**
Intercept	9		
Method	8	3.686	0.154
Season	7	1.949	0.536
Estuary	6	1.620	0.152

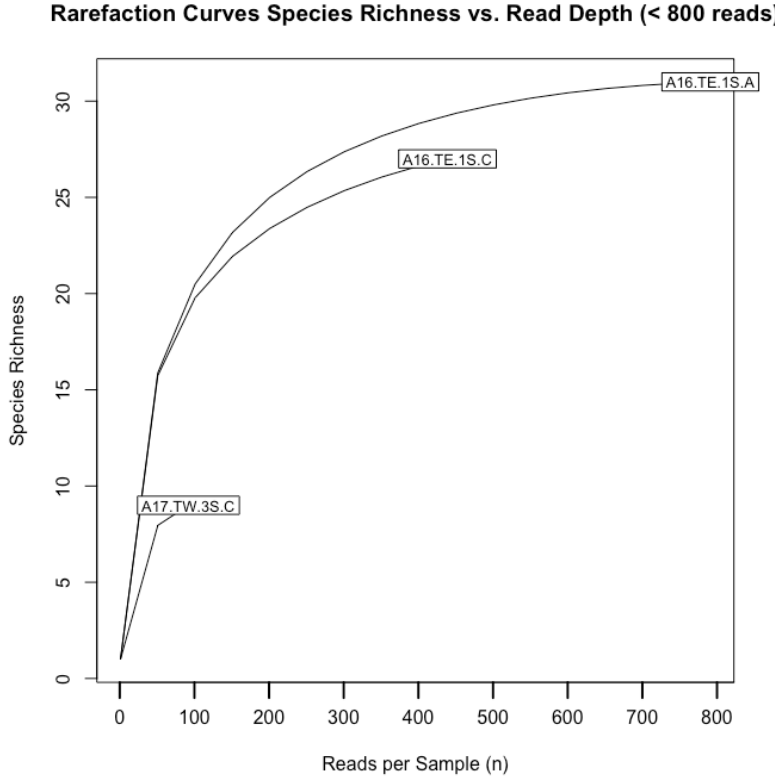
Supplementary Table 10: Spawning times of selected fish species

Species (Guild)	**Spawning Time**	**Geographical Location**	**Reference**
***Atherina boyeri* (ES)**	May to July	Aberthaw Lagoon, Bristol Channel, UK	Creech (1992)
***Atherina presbyter* (MM)**	April to July	English Channel, UK	Muus & Neilsen (1999)
Chelon labrosus* (MM*)	July to August	English Channel and Irish Waters, UK	Muus & Neilsen (1999)
***Chelon ramada* (C)**	June to August	British Isles, UK	Maitland & Campbell (1992)
***Pomatoschistus microps* (ES)**	April to August	UK	Maitland & Campbell (1992)
***Pomatoschistus minutus* (ES)**	December to June	Ythan Estuary, Scotland, UK	Claridge *et al.* (1985)
***Trisopterus minutus* (MS)**	February to March	English Channel, UK	Cohen *et al.* (1990)
***Zeugopterus punctatus* (MS)**	March to June	n/a	Nielsen (1986)
***Labrus bergylta* (MS)**	May to August	North Sea	Quignard & Pras (1986)

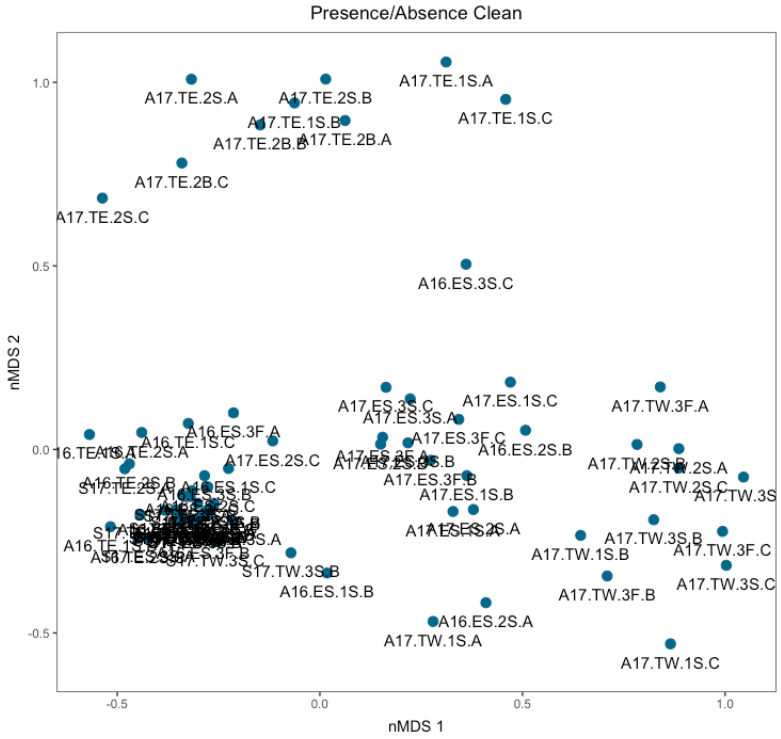
Data compiled via FishBase (Froese and Fauly, 2021)



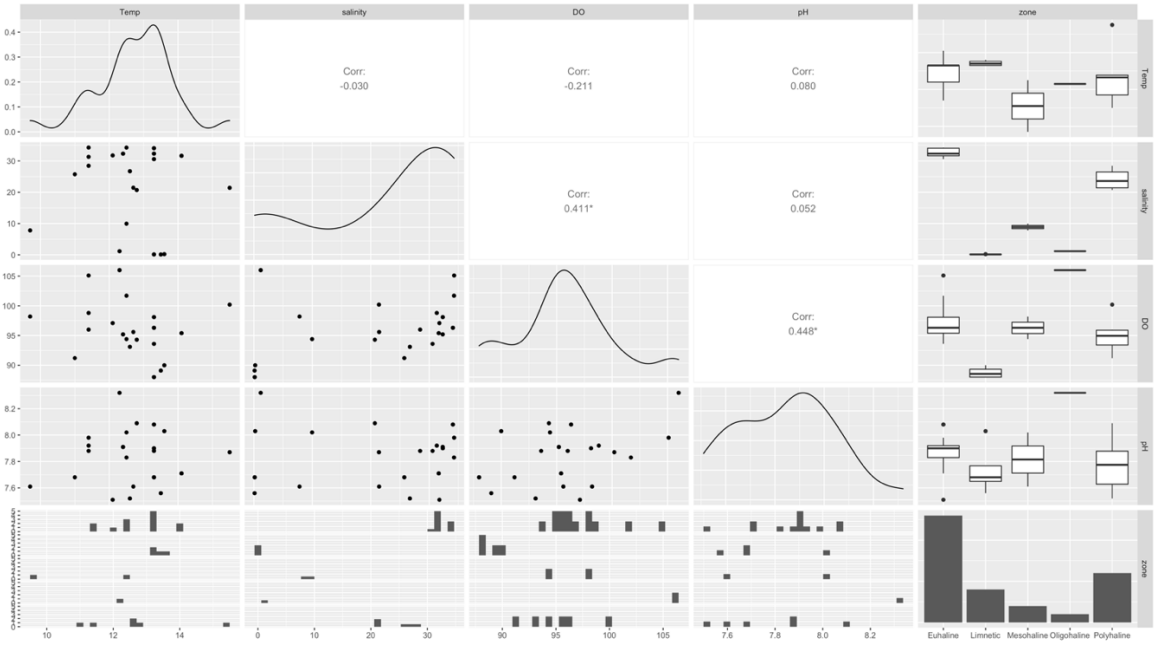
Supplementary Figure 1: Rarefaction curve of fish taxonomic richness (≈ species richness) against fish read depth for each individual sample following the 0.08% read contribution cut-off. This creates a flat profile relative to the raw data (not shown) due to the removal of rare species. Calculated using rarecurve(step = 50) in Vegan (Oksanen *et al.*, 2015).



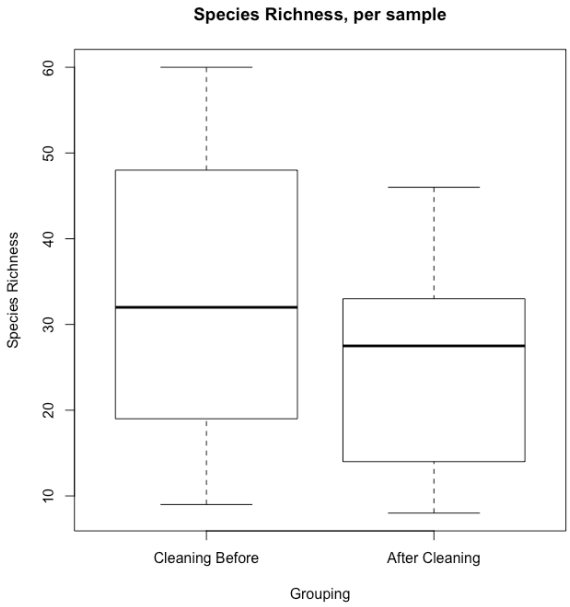
Supplementary Figure 2: Rarefaction curve of fish taxonomic richness (≈ species richness) against fish read depth for each individual sample following the 0.08% cut-off, showing only samples with below 800 reads. Calculated using rarecurve (step = 50) in Vegan (Oksanen *et al.*, 2015).



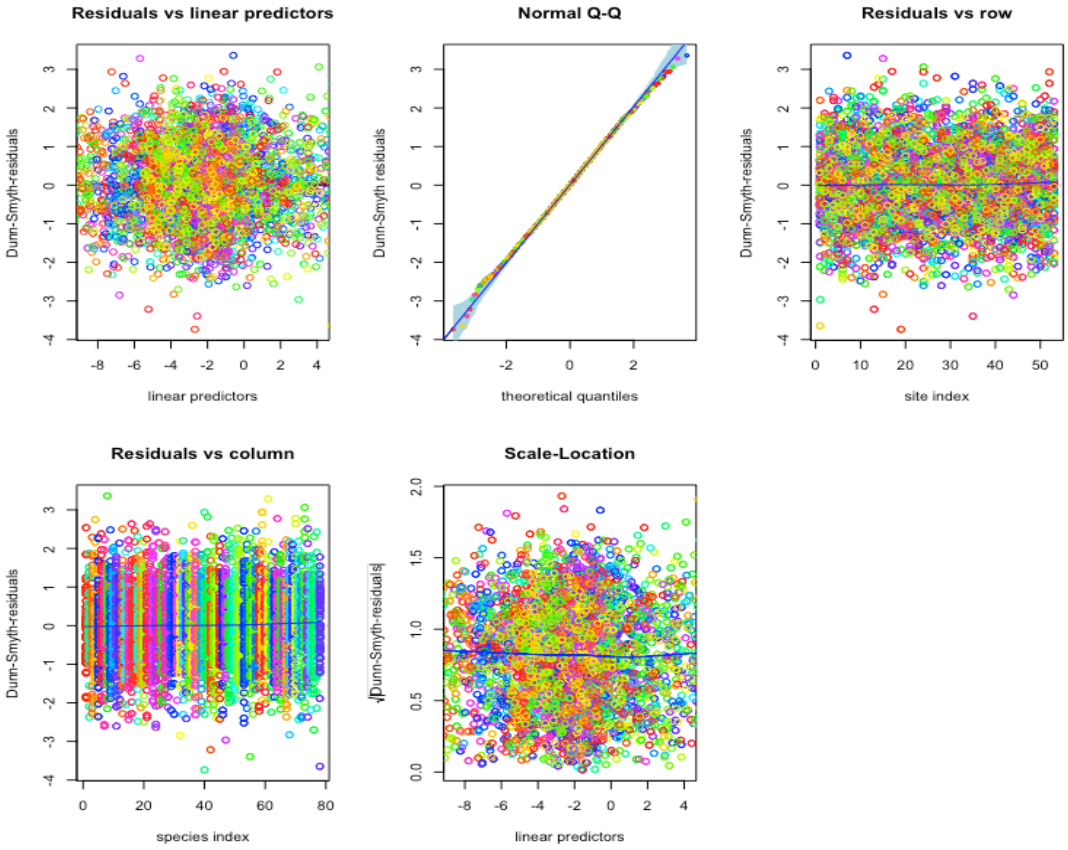
Supplementary Figure 3: An nMDS ordination for eDNA metabarcoding data generated using Bray-Curtis distances calculated on species presence/absence following the 0.008 % read contribution cut-off.



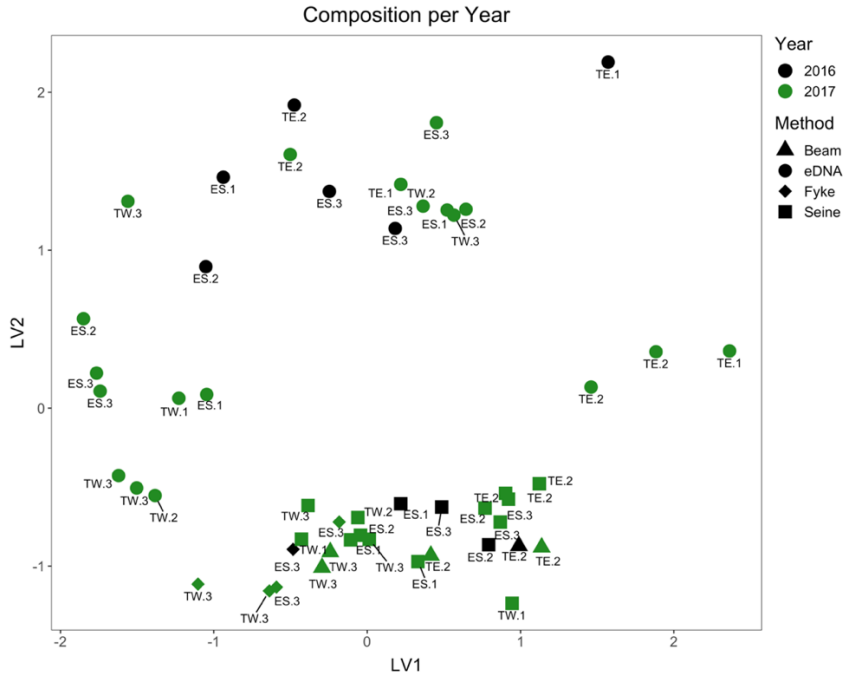
Supplementary Figure 4: Correlation between each combination of environmental variables for the data collected alongside fishing, with Pearson correlation coefficients calculated between continuous variables.



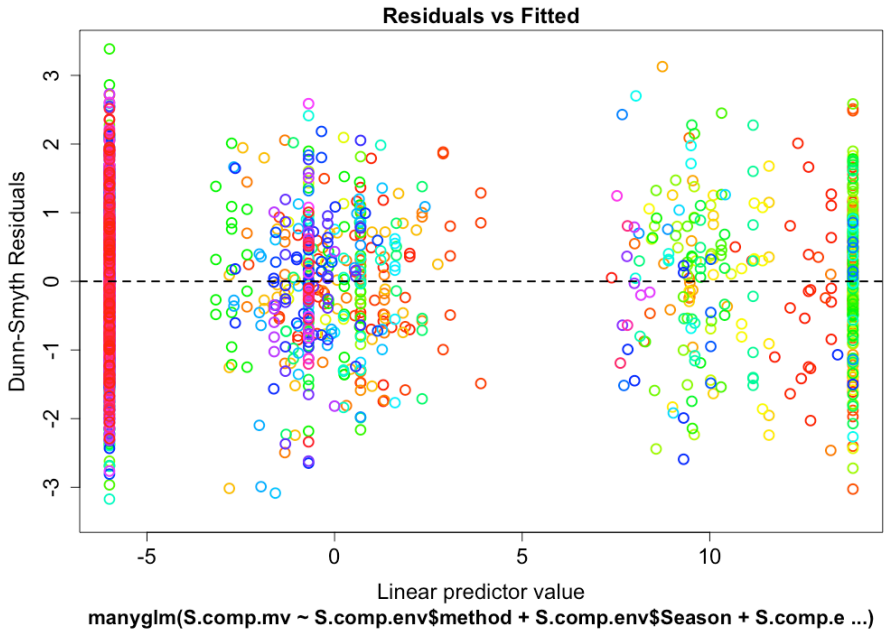
Supplementary Figure 5: The distribution of species richness values per sample, before and after applying the cleaning threshold of 0.08 % per sample read cut off.



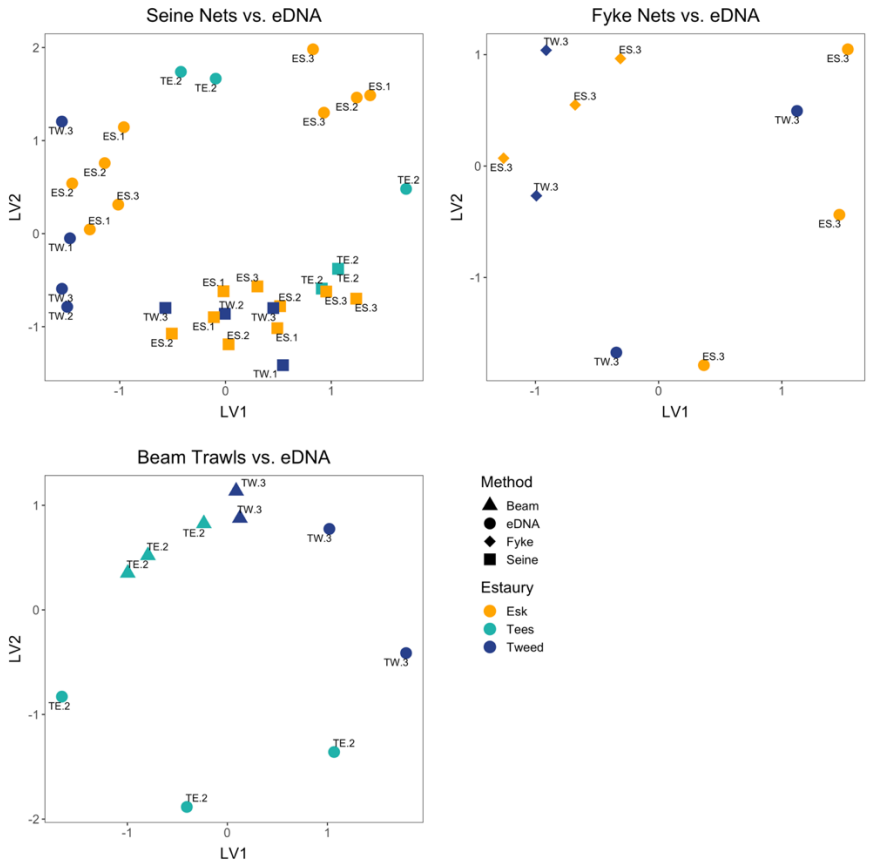
Supplementary Figure 6: Residual plots from the GLLVM ordination of eDNA and all fishing gear (fyke, beam and seine) data per station.



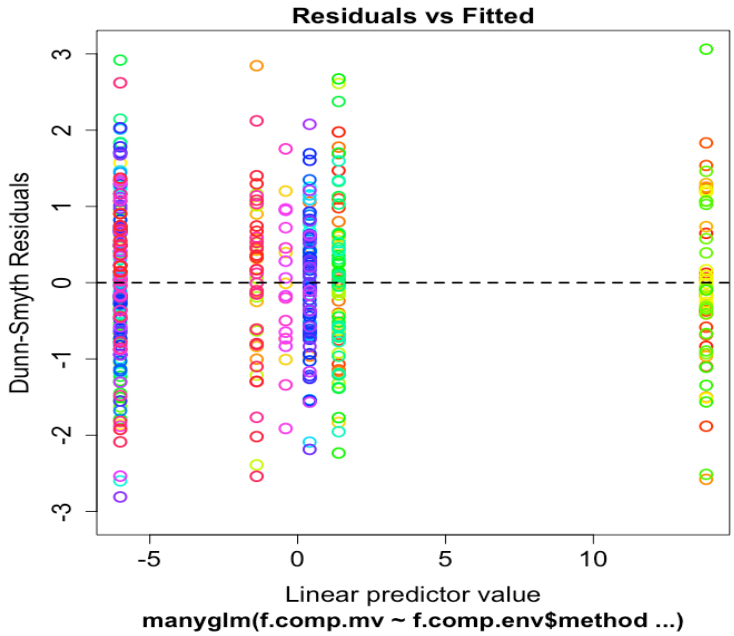
Supplementary Figure 7: Presence/absences of fish species per station for eDNA and all fishing methods (seine nets, fyke nets and beam trawls) modelled using a binomial GLLVM (probit link; two latent variables; 50 iteration). Colours indicate data from different years, shapes indicate the different sampling methods.



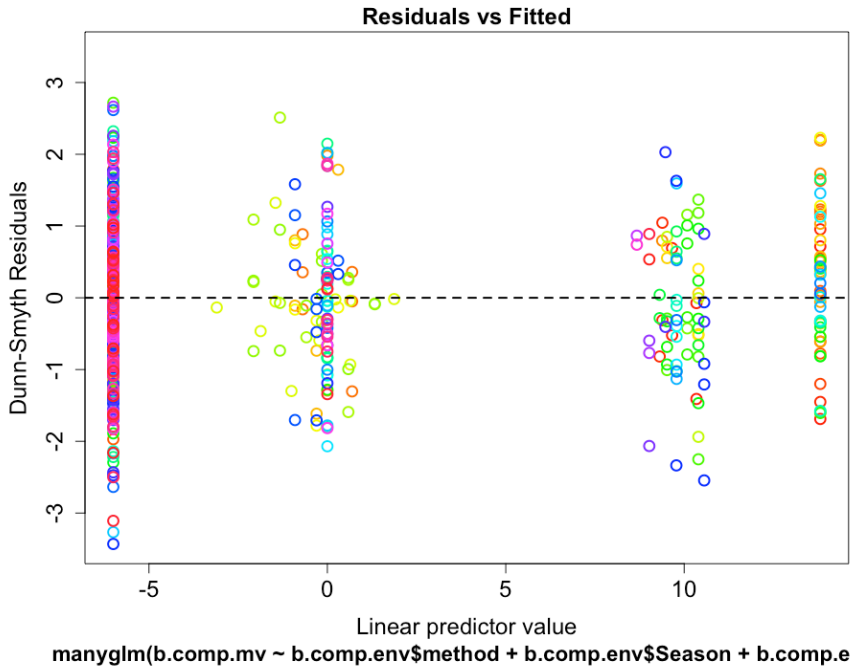
Supplementary Figure 8: Residuals vs. fit from the multivariate GLM comparing eDNA with seine nets.



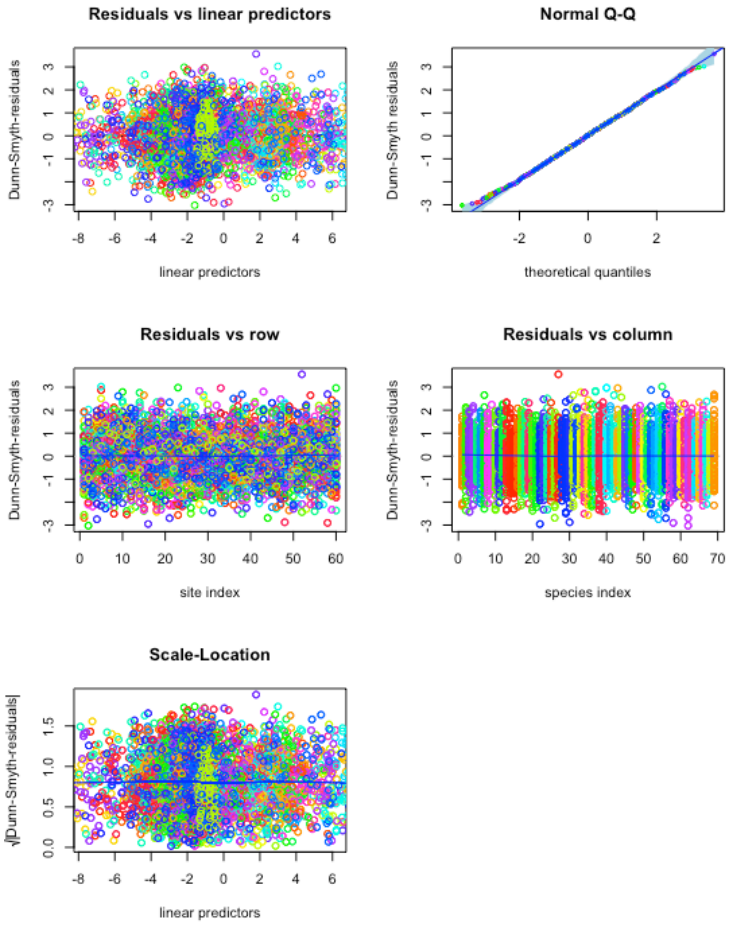
Supplementary Figure 9: Presence/absence of fish species per station for seine netting stations vs. eDNA, fyke netting stations vs. eDNA, and beam trawling stations vs. eDNA modelled using a binomial GLLVM (probit link; two latent variables; 50 iteration). In all panels colours indicate the estuary, whereas shapes indicate the different method used.



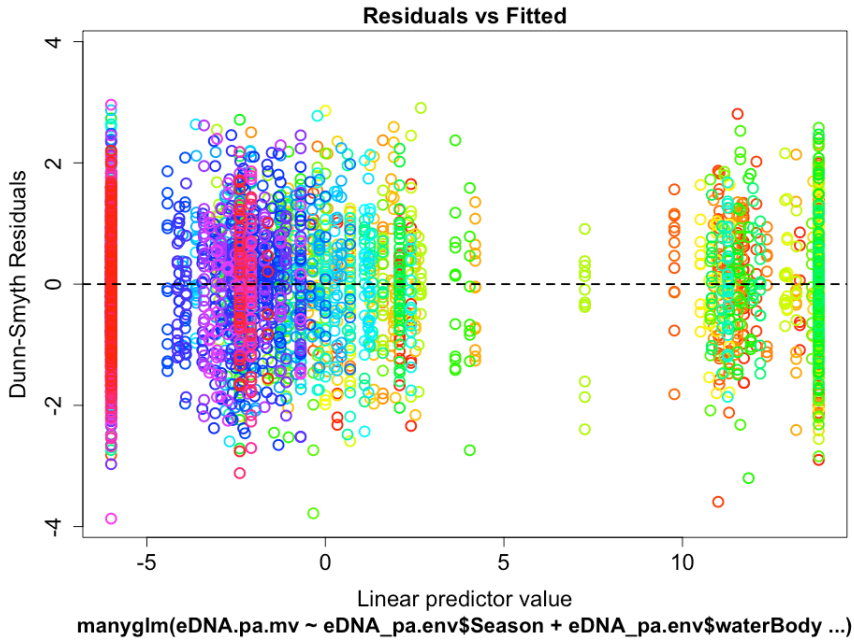
Supplementary Figure 10: Residuals vs. fit from the multivariate GLM comparing eDNA with Fyke nets.



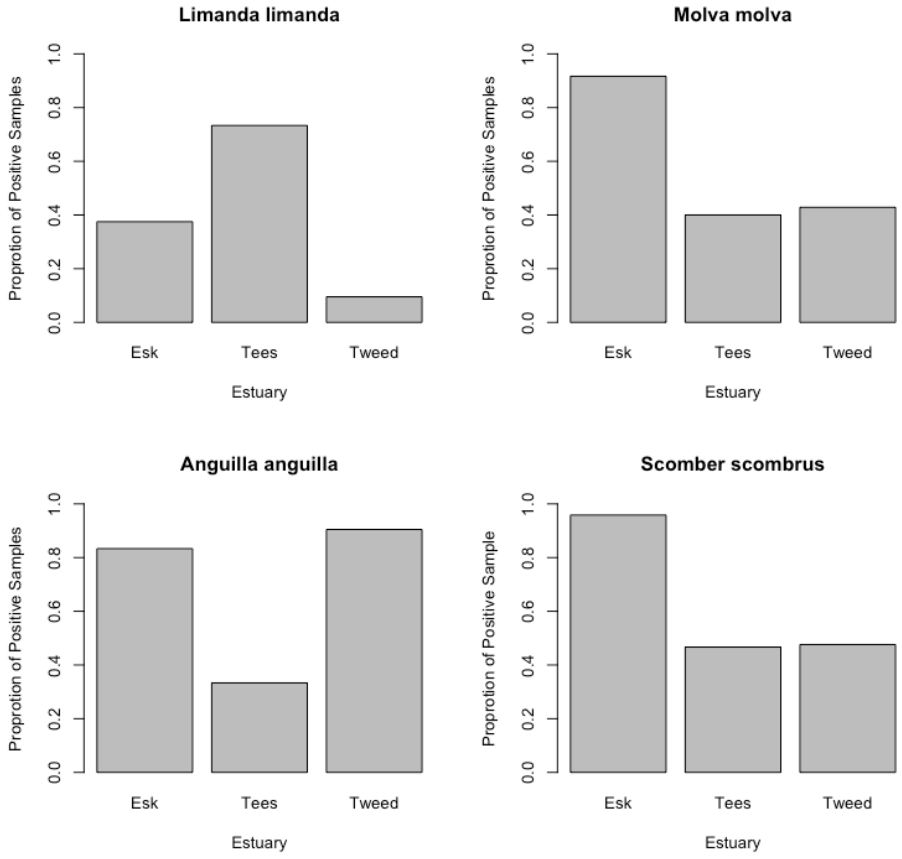
Supplementary Figure 11: Residuals vs. fit from the multivariate GLM comparing eDNA with beam trawls.



Supplementary Figure 12: Residuals from the GLLVM of all eDNA data analysed at the sample level for 2017.



Supplementary Figure 13: Residuals vs. fit from the multivariate GLM of all eDNA data 2017 at the sample level.



Supplementary Figure 14: Proportion of samples positive for the detection of each species in each estuary, identified to show differences in presence/absence between estuaries by multivariate GLM.

Chapter 5.

The Implications of Fish eDNA Metabarcoding for Biomonitoring and the Requirement for Future Research

Author Contributions

Phase	Author
Research and Writing	TIG
Review	SC

1. Introduction

The aim of this thesis was to conduct the basic ecological research to allow the future development of an eDNA metabarcoding tool to assess the biodiversity of fishes in estuaries. This has been addressed via three studies: assessing the spatial and temporal (both seasonal and short-term tidal) variation within estuaries (Chapter 2 and 3) and then comparing assemblages between estuaries, and seasons, in three separate ecosystems (Chapter 4). These studies used comparable sampling designs, laboratory, bioinformatic and statistical methodologies to provide a comprehensive assessment of the use of eDNA metabarcoding for the study of fish assemblages. In addition, species detections were compared to previous, and, or contemporary fish sampling to contextualise the results in relation to existing approaches. The classification of species into functional estuarine use guilds (Elliott *et al.*, 2007) added to the ecological insights from the research. Clearly, there are methodological issues related to the present study, particularly relating to laboratory methods (Chapter 2 and 3). However, the overall results are clear. The aim of this discussion is to summarise the findings of the PhD, discuss how this data can be used in a monitoring context and outline future research.

2. Overview of Findings

Comparisons between eDNA analysis and the existing fishing approaches in the Dee, Esk, Tees and Tweed showed that most species caught by fishing gears can be detected by eDNA metabarcoding; including species of conservation interest. In addition, *generally*, eDNA detected a greater species richness at the level of the estuary than fishing, although this was not the case for every gear type, or combination of gears, in every estuary (Chapter 2 and 4). Furthermore, eDNA clearly had a different assemblage composition to seine nets (Chapter 2 and 4). A proportion of the species detected by eDNA were novel detections, never detected by WFD fishing surveys in each ecosystem before (Chapter 2, 3 and 4). The detection of novel species, and, or species which are widely eaten, needs to be considered carefully given the influence of eDNA transport (Deiner & Altermatt, 2014; Shaw *et al*., 2016; Yamamoto *et al.*, 2017). The results from the comparative studies (Chapter 2 and 4) were similar to other comparative studies of fish assemblages in estuaries using metabarcoding (Zou *et al.*, 2020; Hallam *et al.*, 2021; Cole *et al.*, 2022). Overall, these results show eDNA analysis is an efficient way of detecting fishes in a biomonitoring context. Encouragingly, the detection of European smelt (*Osmerus eperlanus*) in the Dee is of direct use to Natural England (Chapter 2). Smelt is a 'Priority Species' (UK post-2010 Biodiversity Framework) for conservation. However, in the Dee Site of Special Scientific Interest (SSSI) they are not a reportable or specifically monitored feature (Graham *et al.*, 2021), nor has a Marine Conservation Zone (MCZ) been designated for them. Metabarcoding data will be used make the case for further conservation projects on *O. eperlanus* and may lead to an MCZ being designated to protect them (R. Horner, *pers. comm.*).

Regarding the relationship between the eDNA assemblage and ecological parameters. There was a clear correlation between spatial (Chapter 2) and short-term temporal (Chapter 3) variation in salinity (and its correlates) and eDNA assemblage composition. The incidence of four Marine Migrant and an Estuarine Species in eDNA and seine net samples increased with salinity in the Dee, and for two Marine Migrants in eDNA and all gear type samples. Importantly this relationship was, *fairly*, consistent between both eDNA and fish sampling (Chapter 2). Similarly, short-term temporal variation in salinity had a positive effect on the relative read counts of species from the Marine Straggler, Estuarine Species and Marine Migrant estuarine use guilds (Chapter 3). Although this was not confirmed by contemporary fishing. The findings are consistent with the evidence that salinity is a key environmental variable influencing the structure of fish assemblages (Marshall and Elliott, 1998; Selleslagh *et al.*, 2009; Whitfield *et al.*, 2012). The salinity relationship was not apparent in the study of multiple estuaries, but this was probably an artefact of the methodology regarding salinity classifications (Chapter 4; Greenwood, 2007). In addition, spatial structure in the fish assemblage was detected in the Dee using eDNA. Marine species were indicators of the assemblages found in the lower estuary and a freshwater species was an indicator of the assemblage in the upper estuary, with some overlap at intermediate stations (Chapter 2), as would be expected in a European estuary (Nicolas *et al.*, 2010b). In addition, eDNA can detect differences in assemblage composition, *between* estuaries, although the factors driving inter-estuary heterogeneity are yet to be investigated (Chapter 4).

Temporal variation in fish eDNA assemblage composition could be detected at both the short-term tidal scale (Chapter 3) and seasonally between early summer and autumn (Chapter 4), but not consistently (Chapter 2). A differentiation in assemblage composition could be detected between tidal states, prior to and around a spring tide. Assemblage composition became more similar before and following a neap tide, also coinciding with an increase in the amount of freshwater entering the estuary (Chapter 3). At the seasonal scale, across the Tees, Tweed and Esk differentiation in composition between early summer and autumn may have been related to spawning times, and or the migration of species into the estuary (Chapter 4). Changes in fish assemblage structure are known to occur at both the short-term tidal scales (Greenwood & Hill, 2003; Wilson & Sheaves, 2001) and on seasonal time scales (Maes *et al.*, 2005; Henderson and Bird, 2010; Selleslagh *et al.*, 2012). Overall, these results suggest eDNA metabarcoding can provide temporally and spatially ecological coherent results. However, it is difficult to determine if correlations with ecological parameters, are due to changes in the structure of the fish assemblage itself or changes in the distribution, transport and dilution of their eDNA. The most likely answer is both (Chapter 2 and 3), findings that have clear implications for survey design. An eDNA survey to compare fish composition between estuaries, in a specific season, would incorporate the following. Sampling should be spatially replicated, e.g. a systematic longitudinal design along each estuary. The optimal number of samples could be estimated from Chapter 2 and 4. Sampling should occur as close to a set tidal state as feasible, on a standardised point in the spring-neap cycle. Other less easily controlled factors, such as physicochemical parameters and river flow should be measured and included as covariates in any analysis. All these points are best practice in surveys using fishing gears (Hemingway and Elliot, 2002). These recommendations only apply to surface water samples and investigations of the vertical gradients of fish eDNA in estuaries of varying levels of stratification (see McLusky and Elliott, 2004) are required. Vertical gradients in eDNA can be substantial in Fjords, for example (Jeunen *et al.*, 2020).

3. Main Areas for Improvement

The largest area for improvement is related to the methodology of Chapter 2 and 3 (Chapter 4 is not considered here). It was possible to develop a modification of the DNA extraction methodology using DNeasy Blood and Tissue Kits (QIAGEN, Hilden, Germany; Spens and Evans et al., 2017) by adding a flocculation step (Sellers *et al.*, 2018) followed by PCR using Multiplex Master Mix *Taq* polymerase (QIAGEN). This allowed *fairly* consistent sample amplification (but see Chapter 3). However, the method development was extremely time and cost inefficient. An alternative and more efficient approach would have been to modify the protocol of an extraction kit already designed to remove inhibitors from environmental samples e.g. DNeasy Power Water Kit (QIAGEN), DNeasy Power Soil (QIAGEN, see Collins *et al.*, 2019) or equivalents. Although, these are comparatively expensive to Blood and Tissue kits, the margins narrow after the modifications required by the adaption of the Spens and Evans *et al.* (2017) method, used here, are costed (T. I. Gibson, *pers. obs.*). Test extractions should have then been amplified using several *Taq* polymerase master mixes and treated with an inhibitor removal kit, if required. It should also be noted a comparable project using the Blood and Tissue kit and Multiplex Master Mix (QIAGEN) also suffered sample inhibition in the Thames estuary (Hallam *et al.*, 2021). Therefore, PCR inhibition needs to be addressed in a robust manner before eDNA is used for biomonitoring in estuaries. Comparably, in Chapter 4, DNeasy Power Soil kits were used and had no issues with PCR inhibition, to my knowledge. These particular technical issues also reduced the time available to optimise the Elas02/MiFish-E primers (Miya *et al.*, 2015; Taberlet *et al.*, 2018), which were not included for this reason. Therefore, analyses focused mainly on the Teleost assemblage. However, elasmobranchs are only caught by monitoring infrequently in the studied ecosystems, and this did not reduce comparability with the fishing methodology (T. I. Gibson, *pers. obs.*). Another issue was the low yield of fish sequences after sequencing (Chapter 2 and 3). It is not clear why such low yields of fish DNA were retrieved (see Chapter 3 for a discussion). It may have been more PCR optimisation using the Tele02 primers could have improved the results. The Tele02 primers were integrated into the methodology at a late stage, to maintain full comparability with Chapter 4 and following *in silico* analysis (Collins *et al.*, 2019). In addition, although non-specific amplification was excluded from sequencing via gel slicing, alternative quality control measures are available. A more precise methodology such as a BluePippin System (Sage Science, Inc., Beverly, MA, USA), or additional bead cleaning prior to sequencing would have more efficiently removed lower or higher molecular weight material and *might* have improved the results. However, it is difficult to know if a technical fix would have improved fish sequence yields given the clear propensity of Tele02 primers to amplify none-teleost vertebrate DNA (Chapter 2 and 3).

The sub-optimal lab methodology did have implications on the data analysis. Failure to yield data from all samples at stations or sampling events (compounded with a technical laboratory error in Chapter 2), meant incorporation of sample level non-independence in analyses of assemblage composition was not possible (see Wang *et al.*, 2012). Comparably in Chapter 4, sample level analyses substantially improved the results. More generally, no station or sample level analyses of species richness were conducted. This was because the primary focus was on species richness estimates at the estuary level, of most relevance to management, and due to time constraints. Finer scale analyses using a Poisson distributed GLMM with sampling station/sampling event as a random effect (random intercept) would be insightful (Zuur *et al.*, 2007; Harrison *et al.*, 2018). Regarding experimental design, greater levels of sampling in Chapter 2 and 3 would have improved inference. Particularly for Chapter 2 where eDNA sampling alongside every gear deployment would have provided the most powerful dataset. However, this was due to budgetary constraints and the practicalities of working in a

small sampling team, e.g. one to two personnel. Finally, the statistical methods used did not account for the imperfect detection of species i.e. false negatives. For this an occupancy modelling framework is required and has been developed for eDNA metabarcoding in coastal ecosystems (McClenaghan *et al.*, 2020). Occupancy models are hierarchical models which determine the probability of detecting a species at several levels. For eDNA these evaluate the occupancy probability (probability of species presence at a site), detection probability (probability of detection in a technical replicate, e.g. PCR, if present) and capture probability (probability of capture in a field replicate, e.g. water sample, if present, Burian *et al.*, 2021).

4. Implications for Management

Environmental DNA analysis can provide data on the structure (e.g. species richness, composition) and function of fish assemblages (estuarine use guilds) and allows detection of species of conservation interest. Therefore, what is the best way, at present, to integrate eDNA metabarcoding in monitoring and calculate metrics of the health of fish assemblages in estuaries? Existing assessment metrics for estuarine fish use information on species diversity, composition, the ecological and feeding guilds of species, the presence of selected indicator species and species abundance (Coates *et al.*, 2007; Harrison & Whitfield, 2004; Harrison & Kelly, 2013). From here on, the focus will be the Estuarine Multi-metric Fish Index (EMFI) used in Northern Ireland (Harrison & Kelly, 2013) and Wales, and the Transitional Fish Classification Index (TFCI) which is used in England and Scotland (Coates *et al.*, 2007). The taxonomic richness and composition data from environmental DNA could calculate 8 out of 14 metrics from the EMFI (Harrison & Kelly, 2013) and 8 out of the 10 metrics from the TFCI (Coates *et al.*, 2007). The other metrics require the use of relative abundance data (Coates *et al.*, 2007; Harrison *et al.*, 2017), and are more problematic. A related issue is three metrics in the TFCI and four in the EMFI relate to feeding guilds (Elliott *et al.*, 2007). Calculation of these guilds using eDNA is challenging for species with ontogenetic shifts in diet e.g. Atlantic cod (*Gadus morhua,* Elliott *et al.*, 2007), given eDNA is blind to changes in life history. In a practical context a knowledge of the diet of the predominant size category, or life cycle stage, present within the estuary is required (Elliott *et al.*, 2007), requiring direct capture methods e.g. fishing or visual methods. Ecological changes associated with developmental stages are less of an issue with estuarine use guilds, as the guilds describe the overall ecological use of an estuary by a given species in its life cycle (Elliott *et al.*, 2007). Aside from these considerations, the most problematic barriers to eDNA metric generation are how to incorporate the potential influence of eDNA transport and how to use eDNA in a quantitative manner.

Transport of eDNA could affect all metrics generated by eDNA data. For example, transport from the sea, river or wastewater could artificially inflate species richness estimates calculated for an estuary. Although, some of these issues matter less if a larger spatial scale, e.g. the river catchment, is being considered. Chapter 2 and 3 found clear evidence for eDNA detecting species known to occur upstream. Neither Chapter 3 or 4 accounted for this, but Chapter 2 focused the analysis on only those species with previous detections from fishing gears over a 16 year period. The removal of species not previously detected by historical fishing was an extremely crude approach, but only removed species level detections in the Freshwater estuarine use guild, potentially more likely to be transported from the river. Fundamentally, eDNA data possesses inherent detection uncertainty and it is not possible to prove *definitively* that a fish was present when the sample was taken (Jerde, 2021). Focusing analyses on prior detections, or the most abundant species in an ecosystem (not addressed here) increases the confidence managers can place in results. In addition, the majority of novel species in Chapter 4 were also Marine Stragglers. In the EMFI Marine Straggler and Freshwater Straggler species (a subdivision of the Freshwater guild) are not included in metric calculation as neither group

are dependent on estuaries (Harrison & Kelly, 2013). Novel species in these guilds are perhaps more likely to be detections from eDNA transport given their core distributions are outside of the estuary (Elliott *et al.*, 2007). Therefore, by focusing on previously detected species and certain ecological guilds for metric generation it may be possible to reduce the potential influence of eDNA transport. Clearly further research is required. In time process-based models, using a mechanistic understanding of the dynamics of eDNA concentrations in estuaries, will allow systematically occurring false positives to be accounted for. The output of process-based models may also be included as prior information, influencing the likelihood of species presence, for parameter estimation in Bayesian occupancy models (Burian *et al.*, 2021). Presence/absence results for specific species from Chapter 2 and 4 indicated detection probability of eDNA was higher than fishing. Direct analyses of the greater detection probability of eDNA vs. fishing gears is also required to inform this process, and managers (Jerde, 2021). It may also be possible to use previous detection by fishing (or abundance) and estuarine use guild classification as prior information in a Bayesian occupancy model (Burian *et al.*, 2021).

Regarding species abundance, the data from Chapter 2 and 4 could be used to calculate correlations of relative abundance for individual species between specific gear types and eDNA data, but this has not been assessed due to time constraints. Relative read counts could be modelled using a negative binomial distribution (log link) and using the (log) total number of fish reads per sample as a model offset (Zuur *et al.*, 2009) against fishing percentage abundance (in fishing) as a fixed effect. Multi-species correlations could be made using a mixed model with a random slope for fishing percentage abundance and random intercept for each species, to take into account differences between relative reads and the relative abundance of caught fish, between species (Zuur *et al.*, 2007; Harrison *et al.*, 2018). Alternatively, calculating a spearman's rank correlation between the proportions of each species in the eDNA sample vs. netting and adjusting the *p*-value for multiple testing using "fdr" (Benjamini & Hochberg, 1995) would provide a rough indication of the overall relationship. However, in this thesis mock communities were not included, nor were fish weighed to quantify biomass. Mock communities have been suggested as a useful approach to explicitly test the relationship between read abundance and input biomass *in vitro*, to improve inference in the field (Lamb *et al.*, 2019). Additionally, a measure of the size distributions of individuals of a given fish species should be used as a covariate in more sophisticated analysis, thereby considering size-based variation in DNA shedding rates (Rourke *et al.*, 2022). This would require extensive fishing. An additional difficulty could be the variability in fish relative reads over short time scales (Chapter 3). It was not known if the variability in fish reads was down to changes in actual fish abundance, or changes in the hydrographic conditions. Therefore, it is probably easier to use a single fishing gear for the generation of abundance metrics (for a restricted set of species). Then use eDNA to generate the other metrics for the overall assemblage. Potentially, this may allow the relationship between relative reads and abundance/biomass to be established per survey. This could then be used to generate quantitative data from all eDNA samples for the whole assemblage. In situations where fishing gear is not available, a smaller range of metrics could simply be generated.

There is clearly a case to use detections from eDNA metabarcoding alongside data from conventional gears for routine biomonitoring. But how can eDNA contribute to biomonitoring in a practical manner, now? Contemporary fish sampling is *extremely* patchy. In Wales EMFI surveys are currently carried out using a relatively high sampling effort within each estuary, but each survey only occurs in autumn, once every three years. In this context it would be advantageous to continue the fish sampling in the current manner and sample every autumn using eDNA to fill in the gaps. If the data was necessary, sampling could continue in early

summer using eDNA. Sampling using eDNA alongside the fish survey would also be useful for determining false positives and negatives in an occupancy modelling approach. The fishing data could be used as a 'high credibility' dataset to establish true positive and negative results, which would be compared to the concurrent eDNA data. This would allow the estimation of false positive rates with high confidence and allow the occurrence of false positives across the eDNA dataset to be established (Burian *et al.*, 2021). Although, this may be challenging if the true false positive rate changes over years. In England, the estuarine monitoring context is different; fish surveys still occur in each estuary twice a year in early summer and autumn. However, as demonstrated in Chapter 4 the sampling effort per survey is extremely low, and the survey designs and combination of gears used within each system is inconsistent. The data from beam trawls and fyke nets is of particularly limited value for comparisons between ecosystems in a scientific, and therefore bioassessment context (Waugh *et al.*, 2019). Comparisons between estuaries are therefore probably biased. The most efficient sampling regime in this context would be to reduce the sampling campaigns down to a single gear type, e.g. seine nets (comparable to French surveys, Delpech *et al.*, 2010). Then sample the overall fish assemblage for each survey using eDNA. More generally, given it is relatively easy to deploy, eDNA sampling could also be expanded to waterbodies which are not currently assessed. Across the EU and the UK the majority (74%) of transitional waters are not assessed on the basis of the ecological quality of their fish assemblages (EEA 2018a, 2018b). Nevertheless, this number will include some transitional lagoons where the assessment of fishes assemblages is not considered appropriate (WFD - UKTAG, 2009). In England only 29% of transitional waters (excluding lagoons) are assessed on the basis of their fish fauna (EA, 2022b). In Wales, only 24% of transitional waters (including lagoons) are assessed for fishes (NRW, 2019). This is quite a low level of sampling considering fish classification methods are effective indicators of anthropogenic pressure (Lepage *et al.*, 2016; Teichert *et al.*, 2016).

How therefore will the costs of complementary eDNA sampling be financed? The UK is the 5th largest economy in the world (The World Bank, 2020). Overall public sector spending on 'Environmental Protection' by central government has increased since 2016/2017 to 2020/2021 (not accounting for inflation). However, annual spending on 'Protection of Biodiversity and Landscape' has fallen from 418 to 356 million pounds, a decline of around 15% (HM Treasury, 2021). Clearly, this spending is a political choice and will wax and wane with the agendas of successive governments, the salience of environmental issues and the political will required to address them. However, against the background of current decline, monitoring budgets are under massive pressure. The most efficient use of resources would be to use other sampling campaigns as platforms of opportunity to collect water samples to limit sampling costs. For example, routine sampling of phytoplankton occurs in more transitional waters than fishing sampling (EA, 2022b; NRW, 2019). Phytoplankton sampling occurs on a monthly basis, with a minimum of 10 samples collected per year (WFD - UKTAG, 2014), see online data for the Esk (EA, 2022). Although, to target the widest range of transitional waters, sampling alongside regular chemical monitoring would be preferable (EA, 2022b; NRW, 2019). Regarding analysis costs, these could be minimised by coordinating with other stakeholders, e.g. Natural England and NGOs, to survey other species of interest. In this thesis the focus has been the fish assemblage. However, the sequence data from Chapter 4 have shown bird and mammals species can be detected from the surrounding habitats, including species of conservation importance (Mariani *et al.*, 2021). An important future objective using the data in Chapter 2 and 3 would be to identify if other vertebrates of conversation interest can be detected. Considering the eDNA 'bycatch species' would build the case for the utility of eDNA datasets to other stakeholders. However, where additional primers sets are required for

effectively targeting other taxa, the associated consumables and potentially sequencing costs would increase. Even in this context, it is likely that costs would fall overall as the cost of filters, DNA extraction and labour costs per stakeholder would be lower per sample.

5. Other Further Research

A variety of different potential research requirements have been outlined above. However, several key research areas warrant further discussion. Firstly, particle tracking studies are required to understand the dynamics of fish eDNA within estuaries to better understand the spatio and temporal scale which eDNA represents. These should incorporate information on eDNA degradation rates from laboratory studies (Collins *et al.*, 2018). These studies are critical to developing the process-based models which will allow false positive detection to be accounted for (Burian *et al.*, 2021). Several estuaries such as the Conwy (Robins *et al.*, 2014), already have well developed hydrographic models. Research into this area is already being conducted in the Mersey estuary to support eDNA sampling (P. Robins, *pers. comm.*). A comparable study has been conducted into the relative importance of the key environmental processes influencing viral dispersal by using hydrodynamic modelling in a river-estuary-coast system (Robins *et al.*, 2019). The main difference between viral particle tracking approaches and eDNA transport models is that eDNA will be continuously produced by fish present throughout the different waterbodies, while the fish themselves are highly mobile (see Chapter 3). In modelling of pathogenic viruses from wastewater sources, viral input can be comparably well defined (Robins *et al.*, 2019). Therefore, it is likely such studies would require substantial ground truthing with eDNA time-series sampling, while DNA from a non-native species, or a synthetic sequence, could act as a tracer. A key point here is that such models will likely require estimates of eDNA concentration (Robins *et al.*, 2019). Therefore, relative abundance data from metabarcoding will have to be combined with qPCR to provide concentration estimates for multiple species (see Bleijswijk *et al.*, 2020), both to inform and allow the outputs of models to be useful. It is also critical to investigate the potential role of wastewater on confounding species detection (Chapter 2 to 4). There is currently no research on the effects of wastewater effluents on fish detection. Studies of pathogenic viruses provides a useful model to follow (Farkas *et al.*, 2018b, 2018a). Sampling of untreated influent and treated effluent could be conducted as a time-series, either over a season, or over a year at a lower temporal resolution, at waste-water treatment sites. A location in an estuary could be sampled concurrently to determine if the detection of specific species in the estuary was correlated with their detection in wastewater (comparable to Farkas *et al.*, 2018a). As above, spiking of a suitable tracer into wastewater would be useful. Although a synthetic tracer may not degrade in the same manner as eDNA released from fishes in the environment.

Secondly, direct correlation of metrics of fish assemblage health with anthropogenic pressures, informed by previous studies on fish (Courrat *et al.*, 2009; Teichert *et al.*, 2016), are of fundamental importance. The data required would be an extensive multi-estuary survey (see above) ideally covering a wide variety of estuaries with varying levels of anthropogenic impact. Ideally this would include all 27 estuaries across England and Wales monitored for fishes (Waugh *et al.*, 2019). A detailed survey of anthropogenic pressures (see McLusky and Elliott, 2004) in each estuary would have to be conducted, using the extensive background monitoring data e.g. on water quality etc. (EA, 2022b). Following this, the influence of anthropogenic pressure on a selection of metrics generated from the eDNA data (Courrat *et al.*, 2009), or an ecological quality ratio generated from all metrics (Teichert *et al.*, 2016) would be studied. Metric selection and generation could use the guidelines discussed above. Analyses should also take into account environmental variables, such as salinity (Chapter 2 and 3; Courrat *et al.*, 2009). The relationship between the response variable(s) and measures of individual groups of

anthropogenic pressures could then be established. For example, synthetic variables generated by Principle Component Analysis of heavy metal and organic pollution could be used (see Courrat *et al.*, 2009). Alternatively, the relationship between the response variable(s), and *all* pressures, and the interaction between pressures, could be determined using a random forest model (Teichert *et al.*, 2016). Overall, this type of study would allow the complex detail around eDNA transport to be side-stepped to some extent, because the focus would be on determining how metrics of fish health respond to anthropogenetic pressures. Transport of eDNA would influence the results as an additional source of variability to the relationship between metrics and pressures. However, if statistically significant signals in the relationship between metrics and pressures could be determined relative to eDNA transport, and other sources of noise, then this would still achieve the overall aim. The exact provenance of species could then be addressed in detail, later, using the more advanced statistical methodologies to account for false positive detection discussed above. Alternatively useful results may only be achievable from incorporating these methods from the outset.

6. Conclusion

In conclusion, this thesis has met its overall aim of providing the research on the utility of eDNA metabarcoding for the surveying of the fish assemblage in estuaries. The work has shown, despite methodological challenges, that eDNA analysis is an effective way of assessing the fish fauna of an estuary and the 'eDNA fish assemblage' responds to key ecological variables in a similar manner to the fish assemblage. If used cautiously, in combination with fishing, it is likely eDNA will meet the challenges of assessing the health of fish assemblages in estuaries and therefore inform measures of ecosystem health more generally. A wide variety of further research is required mainly on eDNA transport, its quantitative nature and establishing the link between anthropogenic pressures and eDNA derived metrics of fish assemblage health. However, it is likely eDNA analysis will be a key tool in informing the substantial restoration measures that are required to restore and maintain the biodiversity, ecological function and services of estuarine ecosystems, globally, during the 21st century.

Notes

My observations relevant to eDNA sampling design, discussed during this chapter, have been included as a contribution to the book chapter Perry *et al*., 'Design considerations for eDNA metabarcoding surveys', in *Applied Environmental Genomics*, eds: Jarman, S. *et al*., *in prep*. I am a co-author on this work, which has now passed peer review.

Referances

Able, K.W. (2005) 'A re-examination of fish estuarine dependence: Evidence for connectivity between estuarine and ocean habitats', *Estuarine, Coastal and Shelf Science*, 64(1), pp. 5–17. Available at: https://doi.org/10.1016/j.ecss.2005.02.002.

ABPmer and HR Wallingford (2007) *The Estuary-Guide: A website based overview of how to identify and predict morphological change within estuaries*, *Website prepared for the joint Defra/EA Flood and Coastal Erosion Risk Management R&D Programme.* Available at: http://www.estuary-guide.net (Accessed: 2 June 2022).

Aglieri, G., Baillie, C., Mariani, S., Cattano, C., Calò, A., Turco, G., Spatafora, D., Di Franco, A., Di Lorenzo, M., Guidetti, P. and Milazzo, M. (2021) 'Environmental DNA effectively captures functional diversity of coastal fish communities', *Molecular Ecology*, 30(13), pp. 3127–3139. Available at: https://doi.org/10.1111/mec.15661.

Ahn, H., Kume, M., Terashima, Y., Ye, F., Kameyama, S., Miya, M., Yamashita, Y. and Kasai, A. (2020) 'Evaluation of fish biodiversity in estuaries using environmental DNA metabarcoding', *PLoS ONE*. Available at: https://doi.org/10.1371/journal.pone.0231127.

Allen, L.G. (1982) 'Seasonal abundance, composition, and productivity of the littoral fish assemblage in Upper Newport Bay, California', *Fishery Bulletin*, 80(4), pp. 769–790.

Andrews, S. (2010) 'FastQC: A Quality Control Tool for High Throughput Sequencing Data'. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Andruszkiewicz, E.A., Sassoubre, L.M. and Boehm, A.B. (2017a) 'Persistence of marine fish environmental DNA and the influence of sunlight', *PLoS ONE*, 12(9), pp. 1–18. Available at: https://doi.org/10.1371/journal.pone.0185043.

Andruszkiewicz, E.A., Starks, H.A., Chavez, F.P., Sassoubre, L.M., Block, B.A. and Boehm, A.B. (2017b) 'Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding', *PLoS ONE*, 12(4), pp. 1–20. Available at: https://doi.org/10.1371/journal.pone.0176343.

Attrill, M.J. and Rundle, S.D. (2002) 'Ecotone or ecocline: Ecological boundaries in estuaries', *Estuarine, Coastal and Shelf Science*, 55(6), pp. 929–936. Available at: https://doi.org/10.1006/ecss.2002.1036.

Bain, M.B., Harig, A.L., Loucks, D.P., Goforth, R.R. and Mills, K.E. (2000) 'Aquatic ecosystem protection and restoration: advances in methods for assessment and evaluation', *Environmental Science & Policy*, 3(1), pp. 89–98. Available at: https://doi.org/10.1016/S1462-9011(00)00029-0.

Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T. and Stamatakis, A. (2019) 'EPA-ng: Massively parallel evolutionary placement of genetic sequences', *Systematic Biology*, 68(2), pp. 365–369. Available at: https://doi.org/10.1093/sysbio/syy054.

Barbier, E.B., Hacker, S.D., Kennedy, C., Koch, E.W., Stier, A.C. and Silliman, B.R. (2011) 'The value of estuarine and coastal ecosystem services', *Ecological Monographs*, 81(2), pp. 169–193. Available at: https://doi.org/10.1890/10-1510.1.

Barnes, M.A. and Turner, C.R. (2016) 'The ecology of environmental DNA and implications for conservation genetics', *Conservation Genetics*, 17(1), pp. 1–17. Available at: https://doi.org/10.1007/s10592-015-0775-4.

Barnes, M.A., Turner, C.R., Jerde, C.L., Renshaw, M.A., Chadderton, W.L. and Lodge, D.M. (2014) 'Environmental conditions influence eDNA persistence in aquatic systems', *Environmental Science and Technology*, 48(3), pp. 1819–1827. Available at: https://doi.org/10.1021/es404734p.

Beck, M.W. and Airoldi, L. (2007) 'Loss, status and trends for coastal marine habitats of Europe', *Oceanography and Marine Biology: An Annual Review, Vol 48*, 45, pp. 345–405. Available at: https://doi.org/10.1201/9781420050943.ch7.

Beck, M.W., Heck, K.L., Able, K.W., Childers, D.L., Eggleston, D.B., Gillanders, B.M., Halpern, B., Hays, C.G., Hoshino, K., Minello, T.J., Orth, R.J., Sheridan, P.F. and Weinstein, M.P. (2001) 'The identification, conservation, and management of estuarine and marine nurseries for fish and invertebrates', *BioScience*, 51(8), p. 633. Available at: https://doi.org/10.1641/0006-3568(2001)051[0633:TICAMO]2.0.CO;2.

Becker, A., Cowley, P.D. and Whitfield, A.K. (2010) 'Use of remote underwater video to record littoral habitat use by fish within a temporarily closed South African estuary', *Journal of Experimental Marine Biology and Ecology*, 391(1–2), pp. 161–168. Available at: https://doi.org/10.1016/j.jembe.2010.06.028.

Becker, A., Holland, M., Smith, J.A. and Suthers, I.M. (2016a) 'Fish movement through an estuary mouth is related to tidal flow', *Estuaries and Coasts*, 39(4), pp. 1199–1207. Available at: https://doi.org/10.1007/s12237-015-0043-3.

Becker, A., Whitfield, A.K., Cowley, P.D., Cole, V.J. and Taylor, M.D. (2016b) 'Tidal amplitude and fish abundance in the mouth region of a small estuary', *Journal of Fish Biology*, 89(3), pp. 1851–1856. Available at: https://doi.org/10.1111/jfb.13056.

Beckley, L.E. (1985) 'Tidal exchange of ichthyoplankton in the Swartkop estuary mouth, South Africa', *African Zoology*, 20(1), pp. 15–20.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the false discovery rate: A practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), pp. 289–300. Available at: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Berger, C.S., Hernandez, C., Laporte, M., Côté, G., Paradis, Y., Kameni T., D.W., Normandeau, E. and Bernatchez, L. (2020) 'Fine-scale environmental heterogeneity shapes fluvial fish communities as revealed by eDNA metabarcoding', *Environmental DNA*, 2(4), pp. 647–666. Available at: https://doi.org/10.1002/edn3.129.

Blaber, S.J.M. and Blaber, T.G. (1980) 'Factors affecting the distribution of juvenile estuarine and inshore fish', *Journal of Fish Biology*, 17(2), pp. 143–162. Available at: https://doi.org/10.1111/j.1095-8649.1980.tb02749.x.

Blaber, S.J.M., Cyrus, D.P., Albaret, J.J., Ching, C.V., Day, J.W., Elliott, M., Fonseca, M.S., Hoss, D.E., Orensanz, J., Potter, I.C. and Silvert, W. (2000) 'Effects of fishing on the structure and functioning of estuarine and nearshore ecosystems', *ICES Journal of Marine Science*, 57(3), pp. 590–602. Available at: https://doi.org/10.1006/jmsc.2000.0723.

Bleijswijk, J.D.L., Engelmann, J.C., Klunder, L., Witte, H.J., Witte, J.I. and Veer, H.W. (2020) 'Analysis of a coastal North Sea fish community: Comparison of aquatic environmental DNA concentrations to fish catches', *Environmental DNA*, 2(4), pp. 429–445. Available at: https://doi.org/10.1002/edn3.67.

Bolaños, R., Brown, J.M., Amoudry, L.O. and Souza, A.J. (2013) 'Tidal, riverine, and wind influences on the circulation of a macrotidal estuary', *Journal of Physical Oceanography*, 43(1), pp. 29–50. Available at: https://doi.org/10.1175/JPO-D-11-0156.1.

Boyes, S.J. and Elliott, M. (2016) 'Brexit: The marine governance horrendogram just got more horrendous!', *Marine Pollution Bulletin*, 111(1–2), pp. 41–44. Available at: https://doi.org/10.1016/j.marpolbul.2016.08.020.

British Oceaographic Data Centre, Environment Agency and National Oceanography Centre (2020) *UK National Tide Gauge Network, Llandudno*. Available at: https://www.bodc.ac.uk/data/hosted_data_systems/sea_level/uk_tide_gauge_network/ (Accessed: 10 October 2021).

Burian, A., Mauvisseau, Q., Bulling, M., Domisch, S., Qian, S. and Sweet, M. (2021) 'Improving the reliability of eDNA data interpretation', *Molecular Ecology Resources*, 21(5), pp. 1422–1433. Available at: https://doi.org/10.1111/1755-0998.13367.

Buxton, A.S., Groombridge, J.J. and Griffiths, R.A. (2017) 'Is the detection of aquatic environmental DNA influenced by substrate type?', *PLoS ONE*, 12(8), pp. 1–14. Available at: https://doi.org/10.1371/journal.pone.0183371.

Cáceres, M. De and Legendre, P. (2009) 'Associations between species and groups of sites: indices and statistical inference', *Ecology*, 90(12), pp. 3566–3574. Available at: https://doi.org/10.1890/08-1823.1.

De Cáceres, M., Legendre, P. and Moretti, M. (2010) 'Improving indicator species analysis by combining groups of sites', *Oikos*, 119(10), pp. 1674–1684. Available at: https://doi.org/10.1111/j.1600-0706.2010.18334.x.

Callahan, B. (no date) *DADA2 Pipeline Tutorial (1.16)*. Available at: https://benjjneb.github.io/dada2/tutorial_1_8.html.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. and Holmes, S.P. (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*. Available at: https://doi.org/10.1038/nmeth.3869.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) 'BLAST+: architecture and applications', *BMC Bioinformatics*, 10(1), p. 421. Available at: https://doi.org/10.1186/1471-2105-10-421.

Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. and Ellison, A.M. (2014) 'Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies', *Ecological Monographs*, 84(1), pp. 45–67. Available at: https://doi.org/10.1890/13-0133.1.

Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) 'fastp: an ultra-fast all-in-one FASTQ preprocessor', *Bioinformatics*, 34(17), pp. i884–i890. Available at: https://doi.org/10.1093/bioinformatics/bty560.

Childs, A.R., Cowley, P.D., Næsje, T.F., Booth, A.J., Potts, W.M., Thorstad, E.B. and Økland, F. (2008) 'Do environmental factors influence the movement of estuarine fish? A case study using acoustic telemetry', *Estuarine, Coastal and Shelf Science*, 78(1), pp. 227–236. Available at: https://doi.org/10.1016/j.ecss.2007.12.003.

Christophe, M., Rachid, A. and Mario, L. (2015) 'Fish as reference species in different water masses', in *Aquatic Ecotoxicology*. Elsevier, pp. 309–331. Available at: https://doi.org/10.1016/B978-0-12-800949-9.00013-9.

Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J.C., Bonin, A., Taberlet, P. and Pont, D. (2016) 'Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system', *PLoS ONE*, 11(6), pp. 1–19. Available at: https://doi.org/10.1371/journal.pone.0157366.

Claridge, P.N., Hardisty, M.W., Potter, I.C. and Williams, C. V. (1985) 'Abundance, life history and ligulosis in the Gobies (Teleostei) of the inner Severn Estuary', *Journal of the Marine Biological Association of the United Kingdom*, 65(4), pp. 951–968. Available at: https://doi.org/10.1017/S0025315400019433.

Clarke, K.R., Somerfield, P.J. and Gorley, R.N. (2008) 'Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage', *Journal of Experimental Marine Biology and Ecology*, 366(1–2), pp. 56–69. Available at: https://doi.org/10.1016/j.jembe.2008.07.009.

Clarke, K.R. and Warwick, R.M. (2001) *Change in marine communities: An approach to statistical analysis and interpretation*. 2nd edn. Plymouth: PRIMER-E Ltd.

Coates, S., Waugh, A., Anwar, A. and 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), pp. 225–240. Available at: https://doi.org/10.1016/j.marpolbul.2006.08.029.

Cognetti, G. and Maltagliati, F. (2000) 'Biodiversity and adaptive mechanisms in brackish water fauana', *Marine Pollution Bulletin*, 40(1), pp. 7–14.

Cohen, D.M., Inada, T., Iwamoto, T. and Scialabba, N. (1990) 'Gadiform fishes of the world (Order Gadiformes). An annotated and illustrated catalogue of cods, hakes, grenadiers and other gadiform fishes known to date.', in *FAO species catalogue, FAO Fish. Synop.* Rome: FAO, p. 442.

Coissac, E., Riaz, T. and Puillandre, N. (2012) 'Bioinformatic challenges for DNA metabarcoding of plants and animals', *Molecular Ecology*, 21(8), pp. 1834–1847. Available at: https://doi.org/10.1111/j.1365-294X.2012.05550.x.

Colclough, S.R., Gray, G., Bark, A. and Knights, B. (2002) 'Fish and fisheries of the tidal Thames: Management of the modern resource, research aims and future pressures', *Journal of Fish Biology*, 61(Supplement A), pp. 64–73. Available at: https://doi.org/10.1006/jfbi.2002.2068.

Cole, V.J., Harasti, D., Lines, R. and Stat, M. (2022) 'Estuarine fishes associated with intertidal oyster reefs characterized using environmental DNA and baited remote underwater video', *Environmental DNA*, 4(1), pp. 50–62. Available at: https://doi.org/10.1002/edn3.190.

Collins, R. (2021a) *Meta-Fish-Lib: A genralised,dynamic DNA reference library for metabarcoding of fishes.* Available at: https://github.com/genner-lab/meta-fish-lib (Accessed: 1 January 2020).

Collins, R. (2021b) *Meta-Fish-Pipe*. Available at: https://github.com/genner-lab/meta-fish-pipe (Accessed: 1 January 2020).

Collins, R.A., Bakker, J., Wangensteen, O.S., Soto, A.Z., Corrigan, L., Sims, D.W., Genner, M.J. and Mariani, S. (2019) 'Non-specific amplification compromises environmental DNA metabarcoding with COI', *Methods in Ecology and Evolution*, 10(11), pp. 1985–2001. Available at: https://doi.org/10.1111/2041-210X.13276.

Collins, R.A., Trauzzi, G., Maltby, K.M., Gibson, T.I., Ratcliffe, F.C., Hallam, J., Rainbird, S., Maclaine, J., Henderson, P.A., Sims, D.W., Mariani, S. and Genner, M.J. (2021) 'Meta-Fish-Lib: A generalised, dynamic DNA reference library pipeline for metabarcoding of fishes', *Journal of Fish Biology*, p. jfb.14852. Available at: https://doi.org/10.1111/jfb.14852.

Collins, R.A., Wangensteen, O.S., O'Gorman, E.J., Mariani, S., Sims, D.W. and Genner, M.J. (2018) 'Persistence of environmental DNA in marine systems', *Communications Biology*, 1(1), pp. 1–11. Available at: https://doi.org/10.1038/s42003-018-0192-6.

Conway, J.R., Lex, A. and Gehlenborg, N. (2017) 'UpSetR: an R package for the visualization of intersecting sets and their properties', *Bioinformatics*, 33(18), pp. 2938–2940. Available at: https://doi.org/10.1093/bioinformatics/btx364.

Copernicus (2019) *Sentinel 2 natural colour (Bands 432) 10 m*, *European Earth Observation and Monitoring Programme* (Accessed: 1 April 2021).

Costa, M.J., Cabral, H.N., Drake, P., Economou, A.N., Fernandez-Delgado, C., Gordo, L., Marchand, J. and Thiel, R. (2002) 'Recruitment and Production of Commercial Species in Estuaries', in M. Elliot and K.L. Hemingway (eds) *Fishes in Estuaries*. Oxford: Blackwell Science Ltd, pp. 54–123.

Costanza, R., D'Age, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., Raski, R.G., Sutton, P., Van Den Belt, M., D'Arge, R., De Groot, R., Limburg, K., Costanza, R., Van Den Belt, M., Farber, S., Sutton, P., Raskin, R.G., O'Neill, R. V., Grasso, M., Naeem, S., D'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., Raskin, R.G., Sutton, P. and Van Den Belt, M. (1997) 'The value of the world's ecosystem services and natural capital', *Nature*, 387(6630), pp. 253–260. Available at: https://doi.org/10.1038/387253a0.

Costanza, R., Kemp, M., Kemp, W.M. and Boynton, W.R. (1993) 'Predictability, scale, and biodiversity in coastal and estuarine ecosystems: implications for management', *Ambio*, 22(2–3), pp. 88–96. Available at: https://doi.org/10.1016/0006-3207(94)90629-7.

Courrat, A., Lobry, J., Nicolas, D., Laffargue, P., Amara, R., Lepage, M., Girardin, M. and Le Pape, O. (2009) 'Anthropogenic disturbance on nursery function of estuarine areas for marine species', *Estuarine, Coastal and Shelf Science*, 81(2), pp. 179–190. Available at: https://doi.org/10.1016/j.ecss.2008.10.017.

Creech, S. (1992) 'A study of the population biology of Atherina boyeri Risso, 1810 in Aberthaw Lagoon, on the Bristol Channel, in South Wales', *Journal of Fish Biology*, 41(2), pp. 277–286. Available at: https://doi.org/10.1111/j.1095-8649.1992.tb02657.x.

Cyrus, D.P. and Blaber, S.J.M. (1992) 'Turbidity and salinity in a tropical northern Australian estuary and their influence on fish distribution', *Estuarine, Coastal and Shelf Science*, 35(6), pp. 545–563. Available at: https://doi.org/10.1016/S0272-7714(05)80038-1.

Czech, L., Barbera, P. and Stamatakis, A. (2020) 'Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data', *Bioinformatics*, 36(10), pp. 3263–3265. Available at: https://doi.org/10.1093/bioinformatics/btaa070.

Czech, L. and Stamatakis, A. (2019) 'Scalable methods for analyzing and visualizing phylogenetic placement of metagenomic samples.', *PLoS ONE*, 14(5), p. e0217050. Available at: https://doi.org/https://doi.org/10.1371/journal. pone.0217050.

Davies-Colley, R.J. and Smith, D.G. (2001) 'Turbidity, suspended sediment and water clarity: A review', *Journal of the American Water Resources Association*, 37(5), pp. 1085–1101. Available at: https://doi.org/10.1111/j.1752-1688.2001.tb03624.x.

Day, J.W., Gunn, J.D., Folan, W.J., Yáñez-Arancibia, A. and Horton, B.P. (2007) 'Emergence of complex societies after sea level stabilized', *Eos*, 88(15), pp. 169–176. Available at: https://doi.org/10.1029/2007EO150001.

Day, J.W., Gunn, J.D., Folan, W.J., Yáñez-Arancibia, A. and Horton, B.P. (2012) 'The influence of enhanced post-glacial coastal margin productivity on the emergence of complex societies', *Journal of Island and Coastal Archaeology*, 7(1), pp. 23–52. Available at: https://doi.org/10.1080/15564894.2011.650346.

Deegan, L.A. (1993) 'Nutrient and energy transport between estuaries and coastal marine ecosystems by fish migration', *Canadian Journal of Fisheries and Aquatic Sciences*, 50, pp. 74–79. Available at: https://doi.org/10.1139/f93-009.

Deiner, K. and Altermatt, F. (2014) 'Transport distance of invertebrate environmental DNA in a natural river', *PLoS ONE*, 9(2). Available at: https://doi.org/10.1371/journal.pone.0088786.

Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., Pfrender, M.E. and Bernatchez, L. (2017a) 'Environmental DNA metabarcoding: Transforming how we survey animal and plant communities', *Molecular Ecology*, 26(21), pp. 5872–5895. Available at: https://doi.org/10.1111/mec.14350.

Deiner, K., Renshaw, M.A., Li, Y., Olds, B.P., Lodge, D.M. and Pfrender, M.E. (2017b) 'Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA', *Methods in Ecology and Evolution*, 8(12), pp. 1888–1898. Available at: https://doi.org/10.1111/2041-210X.12836.

Delpech, C., Courrat, A., Pasquaud, S., Lobry, J., Le Pape, O., Nicolas, D., Boët, P., Girardin, M. and Lepage, M. (2010) 'Development of a fish-based index to assess the ecological quality of transitional waters: The case of French estuaries', *Marine Pollution Bulletin*, 60(6), pp. 908–918. Available at: https://doi.org/10.1016/j.marpolbul.2010.01.001.

DiBattista, J.D., Fowler, A.M., Riley, I.J., Reader, S., Hay, A., Parkinson, K. and Hobbs, J.-P.A. (2022) 'The use of environmental DNA to monitor impacted coastal estuaries', *Marine Pollution Bulletin*, 181, p. 113860. Available at: https://doi.org/10.1016/j.marpolbul.2022.113860.

Doi, H., Inui, R., Akamatsu, Y., Kanno, K., Yamanaka, H., Takahara, T. and Minamoto, T. (2017) 'Environmental DNA analysis for estimating the abundance and biomass of stream fish', *Freshwater Biology*, 62(1), pp. 30–39. Available at: https://doi.org/10.1111/fwb.12846.

Dong, D., Yan, A., Liu, H., Zhang, X. and Xu, Y. (2006) 'Removal of humic substances from soil DNA using aluminium sulfate', *Journal of Microbiological Methods*, 66(2), pp. 217–222. Available at: https://doi.org/10.1016/j.mimet.2005.11.010.

Dufrene, M. and Legendre, P. (1997) 'Species assemblages and indicator species: The need for a flexible asymmetrical approach', *Ecological Monographs*, 67(3), pp. 345–366. Available at: https://doi.org/https://doi.org/10.1890/0012-9615(1997)067[0345:SAAIST]2.0.CO;2.

EC (European Commission) (1992) 'Council Directive 92/43/ECC on the conservation of natural habitats and of wild fauna and flora', *Official Journal of the European Union*, L 94(1259), pp. 40–52. Available at: https://doi.org/2004R0726 - v.7 of 05.06.2013.

EC (European Commission) (2000) 'Directive 2000/60/EC of the European Parliment and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy', *Official Journal of the European Communities*, L 269, pp. 1–93.

EC (European Commission) (2010) 'Directive 2009/147/EC of the European Parliament and of the Council on the conservation of wild birds', *Official Journal of the European Union*, L 20, pp. 7–25.

Eddy, S.R. (2011) 'Accelerated profile HMM searches', *PLoS Computational Biology*, 7(10), p. e1002195. Available at: https://doi.org/10.1371/journal.pcbi.1002195.

Edgar, R.C. (2016) 'SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences', *bioRxiv*, 074161. Available at: https://doi.org/10.1101/074161.

EEA (European Environment Agency) (2018a) *European Waters: Assessment of status and pressures 2018*, *EEA Report*. Luxembourg: Publications Office of the European Union, 201. Available at: https://www.eea.europa.eu/publications/state-of-water.

EEA (European Environment Agency) (2018b) *Quality element status*, *WISE-Freshwater WFD visualisation tool*. Available at: https://www.eea.europa.eu/themes/water/european-waters/water-quality-and-water-assessment/water-assessments/quality-elements-of-water-bodies (Accessed: 6 June 2022).

Eichmiller, J.J., Best, S.E. and Sorensen, P.W. (2016) 'Effects of temperature and trophic state on degradation of environmental DNA in lake water', *Environmental Science and Technology*, 50(4), pp. 1859–1867. Available at: https://doi.org/10.1021/acs.est.5b05672.

Elliott, M. and Dewailly, F. (1995) 'The structure and components of europeans estuarine fish assemblages', *Netherlands Journal of Aquatic Ecology*, 29, pp. 397–417. Available at: https://doi.org/10.1007/bf02084239.

Elliott, M. and Hemingway, K. (2002) *Fishes in Estuaries*. Edited by M. Elliot and K. Hemingway. Oxford: Blackwell Science Ltd. Available at: https://doi.org/10.2134/jeq2003.3750.

Elliott, M. and McLusky, D.S. (2002) 'The need for definitions in understanding estuaries', *Estuarine, Coastal and Shelf Science*, 55(6), pp. 815–827. Available at: https://doi.org/10.1006/ecss.2002.1031.

Elliott, M. and Quintino, V. (2007) 'The Estuarine Quality Paradox, environmental homeostasis and the difficulty of detecting anthropogenic stress in naturally stressed areas', *Marine Pollution Bulletin*, 54(6), pp. 640–645. Available at: https://doi.org/10.1016/j.marpolbul.2007.02.003.

Elliott, M. and Whitfield, A.K. (2011) 'Challenging paradigms in estuarine ecology and management', *Estuarine, Coastal and Shelf Science*, 94(4), pp. 306–314. Available at: https://doi.org/10.1016/j.ecss.2011.06.016.

Elliott, M., Whitfield, A.K., Potter, I.C., Blaber, S.J.M., Cyrus, D.P., Nordlie, F.G. and Harrison, T.D. (2007) 'The guild approach to categorizing estuarine fish assemblages: A global review', *Fish and Fisheries*, 8(3), pp. 241–268. Available at: https://doi.org/10.1111/j.1467-2679.2007.00253.x.

EMODnet Bathymetry Consortium (2018) *EMODnet Digital Bathymetry (DTM)*. Available at: https://doi.org/https://doi.org/10.12770/18ff0d48-b203-4a65-94a9-5fd8b0ec35f6 (Accessed: 1 October 2021).

Environment Agency (2016) *WFD Water Body Summary Table*, *Water Framework Directive assessment: estuarine and coastal waters*. Available at: https://www.gov.uk/guidance/water-framework-directive-assessment-estuarine-and-coastal-waters (Accessed: 15 June 2022).

Environment Agency (2021) *WFD Transitional Water Bodies Cycle 1*. Available at: https://data.gov.uk/dataset/a4d3a885-0130-410a-8621-9183a57cbc47/wfd-transitional-water-bodies-cycle-1 (Accessed: 1 June 2022).

Environment Agency (2022a) *Esk (E) Water Body*, *Catchment Data Explorer*. Available at: https://environment.data.gov.uk/catchment-planning/WaterBody/GB510402703400?cycle=2 (Accessed: 1 July 2022).

Environment Agency (2022b) *Esk At Lower Habour - Wfd Trac Waters*. Available at: https://environment.data.gov.uk/water-quality/view/sampling-point/NE-49000384 (Accessed: 6 July 2022).

Environment Agency (2022c) *Tees Water Body*, *Catchment Data Explorer*. Available at: https://environment.data.gov.uk/catchment-planning/WaterBody/GB510302509900 (Accessed: 6 June 2022).

Environment Agency (2022d) *Tweed Water Body*, *Catchment Data Explorer*. Available at: https://environment.data.gov.uk/catchment-planning/WaterBody/GB510202110000?cycle=2 (Accessed: 1 July 2022).

Environment Agency (2022e) *Water Quality Data Archive*. Available at: https://environment.data.gov.uk/water-quality/view/landing (Accessed: 1 July 2022).

Environment Agency (2022f) *WFD Transitional and Coastal Water Bodies Cycle 2 Classification 2019*. Available at: https://data.gov.uk/dataset/0f6c2aee-3f8e-476c-93df-e629881bd985/wfd-transitional-and-coastal-water-bodies-cycle-2-classification-2019 (Accessed: 1 June 2022).

Evans, N.T. and Lamberti, G.A. (2017) 'Freshwater fisheries assessment using environmental DNA: A primer on the method, its potential, and shortcomings as a conservation tool', *Fisheries Research*, 197, pp. 60–66. Available at: https://doi.org/10.1016/j.fishres.2017.09.013.

Evans, N.T., Shirey, P.D., Wieringa, J.G., Mahon, A.R. and Lamberti, G.A. (2017) 'Comparative cost and effort of fish distribution detection via environmental DNA analysis and electrofishing', *Fisheries*, 42(2), pp. 90–99. Available at: https://doi.org/10.1080/03632415.2017.1276329.

Ewels, P., Magnusson, M., Lundin, S. and Käller, M. (2016) 'MultiQC: summarize analysis results for multiple tools and samples in a single report', *Bioinformatics*, 32(19), pp. 3047–3048. Available at: https://doi.org/10.1093/bioinformatics/btw354.

Farkas, K., Cooper, D.M., McDonald, J.E., Malham, S.K., de Rougemont, A. and Jones, D.L. (2018a) 'Seasonal and spatial dynamics of enteric viruses in wastewater and in riverine and estuarine receiving waters', *Science of The Total Environment*, 634, pp. 1174–1183. Available at: https://doi.org/10.1016/j.scitotenv.2018.04.038.

Farkas, K., Marshall, M., Cooper, D., McDonald, J.E., Malham, S.K., Peters, D.E., Maloney, J.D. and Jones, D.L. (2018b) 'Seasonal and diurnal surveillance of treated and untreated wastewater for human enteric viruses', *Environmental Science and Pollution Research*, 25(33), pp. 33391–33401. Available at: https://doi.org/10.1007/s11356-018-3261-y.

Fausch, K.D., Lyons, J., Karr, J.R. and Angermeier, P.L. (1990) 'Fish communities as indicators of environmental degradation', *American Fisheries Society Symposium*, 8, pp. 123–144.

Franco, A., Elliott, M., Franzoi, P. and Torricelli, P. (2008) 'Life strategies of fishes in European estuaries: The functional guild approach', *Marine Ecology Progress Series*, 354, pp. 219–228. Available at: https://doi.org/10.3354/meps07203.

Froese, R. and Fauly, D. et al. (2021) *FishBase (v.: 08/2021)*. Available at: www.fishbase.org (Accessed: 1 June 2022).

Gamito, R., Pasquaud, S., Courrat, A., Drouineau, H., Fonseca, V.F., Gonçalves, C.I., Wouters, N., Costa, J.L., Lepage, M., Costa, M.J. and Cabral, H.N. (2012) 'Influence of sampling effort on metrics of fish-based indices for the assessment of estuarine ecological quality', *Ecological Indicators*, 23, pp. 9–18. Available at: https://doi.org/10.1016/j.ecolind.2012.03.009.

García-Machado, E., Laporte, M., Normandeau, E., Hernández, C., Côté, G., Paradis, Y., Mingelbier, M. and Bernatchez, L. (2021) 'Fish community shifts along a strong fluvial environmental gradient revealed by eDNA metabarcoding', *Environmental DNA*, p. edn3.221. Available at: https://doi.org/10.1002/edn3.221.

GB Non-Native Species Secretariat (2019) *Pink Salmon (Oncorhynchus gorbuscha)*, *Risk Assessment Summary Sheet*. Available at: https://www.nonnativespecies.org/non-native-species/information-portal/view/2419 (Accessed: 1 June 2022).

Getmapping PLC (2014) *High Resolution (25 cm) Vertical Aerial Imagery*, *Using: EDINA Aerial Digimap Service*. Available at: https://digimap.edina.ac.uk/aerial (Accessed: 1 June 2022).

Getmapping PLC (2018) *High Resolution (25cm) Vertical Aerial Imagery*, *Using: EDINA Aerial Digimap Service*. Available at: https://digimap.edina.ac.uk/aerial (Accessed: 1 October 2021).

Gillson, J. (2011) 'Freshwater flow and fisheries production in estuarine and coastal systems: Where a drop of rain is not lost', *Reviews in Fisheries Science*, 19(3), pp. 168–186. Available at: https://doi.org/10.1080/10641262.2011.560690.

Goldberg, C.S., Strickler, K.M. and Pilliod, D.S. (2015) 'Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms', *Biological Conservation*, 183, pp. 1–3. Available at: https://doi.org/10.1016/j.biocon.2014.11.040.

Grabemann, I. and Krause, G. (2001) 'On different time scales of suspended matter dynamics in the Weser estaury.', *Estuaries*, 24, pp. 688–698.

Graham, J., Rowland, C., Ribbens, J. and Colclouth, S. (2021) *European smelt (Osmerus eperlanus (L.)) Recovery Management Plan for the Solway Firth Marine Conservation Zone (MCZ)*. Natural England.

Green, M. and Sambrook, J. (2012) *Molecular Cloning: A Laboratory Manual.* 4th Ed. New York: Cold Spring Harbor Laboratory.

Greenwood, M.F.D. (2007) 'Nekton community change along estuarine salinity gradients: Can salinity zones be defined?', 30(3), pp. 537–542.

Greenwood, M.F.D. and Hill, A.S. (2003) 'Temporal, spatial and tidal influences on benthic and demersal fish abundance in the Forth estuary', *Estuarine, Coastal and Shelf Science*, 58(2), pp. 211–225. Available at: https://doi.org/10.1016/S0272-7714(03)00071-4.

Guest, M.A., Connolly, R.M. and Loneragan, N.R. (2003) 'Seine nets and beam trawls compared by day and night for sampling fish and crustaceans in shallow seagrass habitat', *Fisheries Research*, 64(2–3), pp. 185–196. Available at: https://doi.org/10.1016/S0165-7836(03)00109-7.

Hallam, J., Clare, E.L., Jones, J.I. and Day, J.J. (2021) 'Biodiversity assessment across a dynamic riverine system: A comparison of eDNA metabarcoding versus traditional fish surveying methods', *Environmental DNA*, 3(6), pp. 1247–1266. Available at: https://doi.org/10.1002/edn3.241.

Hänfling, B., Lawson Handley, L., Read, D.S., Hahn, C., Li, J., Nichols, P., Blackman, R.C., Oliver, A. and Winfield, I.J. (2016) 'Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods', *Molecular Ecology*, pp. 3101–3119. Available at: https://doi.org/10.1111/mec.13660.

Hansen, B.K., Bekkevold, D., Clausen, L.W. and Nielsen, E.E. (2018) 'The sceptical optimist: challenges and perspectives for the application of environmental DNA in marine fisheries', *Fish and Fisheries*, 19(5), pp. 751–768. Available at: https://doi.org/10.1111/faf.12286.

Harmelin-Vivien, M.L. and Francour, P. (1992) 'Trawling or visual censuses? methodological bias in the assessment of fish populations in seagrass beds', *Marine Ecology*, 13(1), pp. 41–51. Available at: https://doi.org/10.1111/j.1439-0485.1992.tb00338.x.

Harrison, J.B., Sunday, J.M. and Rogers, S.M. (2019) 'Predicting the fate of eDNA in the environment and implications for studying biodiversity', *Proceedings of the Royal Society B: Biological Sciences*, 286(1915), p. 20191409. Available at: https://doi.org/10.1098/rspb.2019.1409.

Harrison, T.D., Armour, N.D., McNeill, M.T. and Moorehead, P.W. (2017) 'A comparative study of Northern Ireland's estuaries based on the results of beam trawl fish surveys', *Estuarine, Coastal and Shelf Science*, 198, pp. 172–182. Available at: https://doi.org/10.1016/j.ecss.2017.09.010.

Harrison, T.D. and Kelly, F.L. (2013) 'Development of an estuarine multi-metric fish index and its application to Irish transitional waters', *Ecological Indicators*, 34, pp. 494–506. Available at: https://doi.org/10.1016/j.ecolind.2013.06.018.

Harrison, T.D. and Whitfield, A.K. (2004) 'A multi-metric fish index to assess the environmental condition of estuaries', *Journal of Fish Biology*, 65(3), pp. 683–710. Available at: https://doi.org/10.1111/J.1095-8649.2004.00477.x.

Harrison, X.A., Donaldson, L., Correa-Cano, M.E., Evans, J., Fisher, D.N., Goodwin, C.E.D., Robinson, B.S., Hodgson, D.J. and Inger, R. (2018) 'A brief introduction to mixed effects modelling and multi-model inference in ecology', *PeerJ*, 6, p. e4794. Available at: https://doi.org/10.7717/peerj.4794.

Hatton-Ellis, T. (2008) 'The Hitchhiker's Guide to the Water Framework Directive', *Aquatic Conservation: Marine and Freshwater Ecosystems*, 18, pp. 111–116.

Hemingway, K.L. and Elliot, M. (2002) 'Field Methods', in M. Elliot and K.L. Hemingway (eds) *Fishes in Estuaries*. Oxford: Blackwell Science Ltd., pp. 410–509.

Henderson, P.A. and Bird, D.J. (2010) 'Fish and macro-crustacean communities and their dynamics in the Severn Estuary', *Marine Pollution Bulletin*, 61(1–3), pp. 100–114. Available at: https://doi.org/10.1016/j.marpolbul.2009.12.017.

Hering, D., Borja, A., Jones, J.I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakare, S., Hänfling, B., Kahlert, M., Leese, F., Meissner, K., Mergen, P., Reyjol, Y., Segurado, P., Vogler, A. and Kelly, M. (2018) 'Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive', *Water Research*, 138, pp. 192–205. Available at: https://doi.org/10.1016/j.watres.2018.03.003.

Hinlo, R., Furlan, E., Suitor, L. and Gleeson, D. (2017) 'Environmental DNA monitoring and management of invasive fish: comparison of eDNA and fyke netting', *Management of Biological Invasions*, 8(1), pp. 89–100. Available at: https://doi.org/10.3391/mbi.2017.8.1.09.

HM Treasury (2021) *Public Expenditure Statistical Analyses 2021*. London: UK Government. Available at: https://www.gov.uk/government/statistics/public-expenditure-statistical-analyses-2021.

Hsieh, T.C., Ma, K.H. and Chao, A. (2016) 'iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers)', *Methods in Ecology and Evolution*, 7(12), pp. 1451–1456. Available at: https://doi.org/10.1111/2041-210X.12613.

Hurlbert, S.H. (1984) 'Pseudoreplication and the design of ecological field experiments', *Ecological Monographs*. Available at: https://doi.org/10.2307/1942661.

Jafar-Sidik, M., Gohin, F., Bowers, D., Howarth, J. and Hull, T. (2017) 'The relationship between Suspended Particulate Matter and Turbidity at a mooring station in a coastal environment: consequences for satellite-derived products', *Oceanologia*, 59(3), pp. 365–378. Available at: https://doi.org/10.1016/j.oceano.2017.04.003.

Jane, S.F., Wilcox, T.M., Mckelvey, K.S., Young, M.K., Schwartz, M.K., Lowe, W.H., Letcher, B.H. and Whiteley, A.R. (2015) 'Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams', *Molecular Ecology Resources*, 15(1), pp. 216–227. Available at: https://doi.org/10.1111/1755-0998.12285.

Jerde, C.L. (2021) 'Can we manage fisheries with the inherent uncertainty from eDNA?', *Journal of Fish Biology*, 98(2), pp. 341–353. Available at: https://doi.org/10.1111/jfb.14218.

Jeunen, G., Lamare, M.D., Knapp, M., Spencer, H.G., Taylor, H.R., Stat, M., Bunce, M. and Gemmell, N.J. (2020) 'Water stratification in the marine biome restricts vertical environmental DNA (eDNA) signal dispersal', *Environmental DNA*, 2(1), pp. 99–111. Available at: https://doi.org/10.1002/edn3.49.

Jeunen, G.J., Knapp, M., Spencer, H.G., Lamare, M.D., Taylor, H.R., Stat, M., Bunce, M. and Gemmell, N.J. (2019) 'Environmental DNA (eDNA) metabarcoding reveals strong discrimination among diverse marine habitats connected by water movement', *Molecular Ecology Resources*, 19(2), pp. 426–438. Available at: https://doi.org/10.1111/1755-0998.12982.

JNCC (2021) *B7. Surface water status*. Available at: https://jncc.gov.uk/our-work/ukbi-b7-surface-water-status/ (Accessed: 1 May 2022).

Jo, T., Murakami, H., Masuda, R., Sakata, M.K., Yamamoto, S. and Minamoto, T. (2017) 'Rapid degradation of longer DNA fragments enables the improved estimation of distribution and biomass using environmental DNA', *Molecular Ecology Resources*, 17(6), pp. 25–33. Available at: https://doi.org/10.1111/1755-0998.12685.

Jo, T., Murakami, H., Yamamoto, S., Masuda, R. and Minamoto, T. (2019) 'Effect of water temperature and fish biomass on environmental DNA shedding, degradation, and size distribution', *Ecology and Evolution*, 9(3), pp. 1135–1146. Available at: https://doi.org/10.1002/ece3.4802.

Karr, J.R. (1981) 'Assessment of Biotic Integrity using fish communities', *Fisheries*, 6(6), pp. 21–27. Available at: https://doi.org/10.1577/1548-8446(1981)006<0021:AOBIUF>2.0.CO;2.

Katoh, K. (2013) *MAFFT ver.7*. Available at: https://mafft.cbrc.jp/alignment/software/algorithms/algorithms.html.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002) 'MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform', *Nucleic Acids Research*, 30(14), pp. 3059–3066. Available at: https://doi.org/10.1093/nar/gkf436.

Katoh, K. and Standley, D.M. (2013) 'MAFFT Multiple Sequence Alignment software version 7: Improvements in performance and usability', *Molecular Biology and Evolution*, 30(4), pp. 772–780. Available at: https://doi.org/10.1093/molbev/mst010.

Kelly, R.P., Gallego, R. and Jacobs-Palmer, E. (2018) 'The effect of tides on nearshore environmental DNA', *PeerJ*, 6, pp. 1–20. Available at: https://doi.org/10.7717/peerj.4521.

Kennish, M.J. (2002) 'Environmental threats and environmental future of estuaries', *Environmental Conservation*, 29(1), pp. 78–107. Available at: https://doi.org/10.1017/S0376892902000061.

Kottelat, M. and Freyhof, J. (2007) *Handbook of European freshwater fishes*. Berlin: Kottelat, Cornol and Freyhof.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A. (2019) 'RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference', *Bioinformatics*, 35(21), pp. 4453–4455. Available at: https://doi.org/10.1093/bioinformatics/btz305.

Kubečka, J., Godø, O.R., Hickley, P., Prchalová, M., Říha, M., Rudstam, L. and Welcomme, R. (2012) 'Fish sampling with active methods', *Fisheries Research*, 123–124, pp. 1–3. Available at: https://doi.org/10.1016/j.fishres.2011.11.013.

Kubečka, J., Hohausová, E., Matěna, J., Peterka, J., Amarasinghe, U.S., Bonar, S.A., Hateley, J., Hickley, P., Suuronen, P., Tereschenko, V., Welcomme, R. and Winfield, I.J. (2009) 'The true picture of a lake or reservoir fish stock: A review of needs and progress', *Fisheries Research*, 96(1), pp. 1–5. Available at: https://doi.org/10.1016/j.fishres.2008.09.021.

Kume, M., Lavergne, E., Ahn, H., Terashima, Y., Kadowaki, K., Ye, F., Kameyama, S., Kai, Y., Henmi, Y., Yamashita, Y. and Kasai, A. (2021) 'Factors structuring estuarine and coastal fish communities across Japan using environmental DNA metabarcoding', *Ecological Indicators*, 121, p. 107216. Available at: https://doi.org/10.1016/j.ecolind.2020.107216.

Lamb, P.D., Hunter, E., Pinnegar, J.K., Creer, S., Davies, R.G. and Taylor, M.I. (2019) 'How quantitative is metabarcoding: A meta-analytical approach', *Molecular Ecology*, 28(2), pp. 420–430. Available at: https://doi.org/10.1111/mec.14920.

Lance, R., Klymus, K., Richter, C., Guan, X., Farrington, H., Carr, M., Thompson, N., Chapman, D. and Baerwaldt, K. (2017) 'Experimental observations on the decay of environmental DNA from bighead and silver carps', *Management of Biological Invasions*, 8(3), pp. 343–359. Available at: https://doi.org/10.3391/mbi.2017.8.3.08.

Laramie, M.B., Pilliod, D.S. and Goldberg, C.S. (2015) 'Characterizing the distribution of an endangered salmonid using environmental DNA analysis', *Biological Conservation*, 183, pp. 29–37. Available at: https://doi.org/10.1016/j.biocon.2014.11.025.

Laroche, O., Wood, S.A., Tremblay, L.A., Ellis, J.I., Lejzerowicz, F., Pawlowski, J., Lear, G., Atalah, J. and Pochon, X. (2016) 'First evaluation of foraminiferal metabarcoding for monitoring environmental impact from an offshore oil drilling site', *Marine Environmental Research*, 120, pp. 225–235. Available at: https://doi.org/10.1016/j.marenvres.2016.08.009.

Law, R.J., Dawes, V.J., Woodhead, R.J. and Matthiessen, P. (1997) 'Polycyclic aromatic hydrocarbons (PAH) in seawater around England and Wales', *Marine Pollution Bulletin*, 34(5), pp. 306–322. Available at: https://doi.org/10.1016/S0025-326X(96)00096-3.

Lawson Handley, L. (2015) 'How will the 'molecular revolution' contribute to biological recording?', *Biological Journal of the Linnean Society*, 115, pp. 750–766. Available at: http://onlinelibrary.wiley.com/doi/10.1111/bij.12516/full.

Lepage, M., Harrison, T., Breine, J., Cabral, H., Coates, S., Galván, C., García, P., Jager, Z., Kelly, F., Mosch, E.C., Pasquaud, S., Scholle, J., Uriarte, A. and Borja, A. (2016) 'An approach to intercalibrate ecological classification tools using fish in transitional water of the North East Atlantic', *Ecological Indicators*, 67, pp. 318–327. Available at: https://doi.org/10.1016/j.ecolind.2016.02.055.

Lindsay, P., Balls, P.W. and West, J.R. (1996) 'Influence of tidal range and river discharge on Suspended Particulate Matter fluxes in the Forth Estuary (Scotland)', *Estuarine, Coastal and Shelf Science*, 42(1), pp. 63–82. Available at: https://doi.org/10.1006/ecss.1996.0006.

Lonsdale, J., Nicholson, R., Weston, K., Elliott, M., Birchenough, A. and Sühring, R. (2018) 'A user's guide to coping with estuarine management bureaucracy: An Estuarine Planning Support System (EPSS) tool', *Marine Pollution Bulletin*, 127(2018), pp. 463–477. Available at: https://doi.org/10.1016/j.marpolbul.2017.12.032.

Lonsdale, J.A., Weston, K., Barnard, S., Boyes, S.J. and Elliott, M. (2015) 'Integrating management tools and concepts to develop an estuarine planning support system: A case study of the Humber Estuary, Eastern England', *Marine Pollution Bulletin*, 100(1), pp. 393–405. Available at: https://doi.org/10.1016/j.marpolbul.2015.08.017.

Lotze, H.K., Lenihan, H.S., Bourque, B.J., Bradbury, R.H., Cooke, R.G., Kay, M.C., Kidwell, S.M., Kirby, M.X., Peterson, C.H., Jackson, J.B.C. and Bay, M. (2006) 'Depletion, degradation, and recovery potential of Estuaries and Coastal Seas', *Science*, 312, pp. 1806–1809. Available at: https://doi.org/10.1126/science.1128035.

van der Maarel, E. (1990) 'Ecotones and ecoclines are different', *Journal of Vegetation Science*, 1(1), pp. 135–138.

Mächler, E., Deiner, K., Spahn, F. and Altermatt, F. (2016) 'Fishing in the water: Effect of sampled water volume on environmental DNA-based detection of macroinvertebrates', *Environmental Science and Technology*, 50(1), pp. 305–312. Available at: https://doi.org/10.1021/acs.est.5b04188.

Maes, J., Stevens, M. and Ollevier, F. (2005) 'The composition and community structure of the ichthyofauna of the upper Scheldt estuary: synthesis of a 10-year data collection (1991-2001)', *Journal of Applied Ichthyology*, 21(2), pp. 86–93. Available at: https://doi.org/10.1111/j.1439-0426.2004.00628.x.

Maitland, P.S. and Campbell, R.N. (1992) *Freshwater fishes of the British Isles*. London: HarperCollins Publishers.

Majaneva, M., Diserud, O.H., Eagle, S.H.C., Boström, E., Hajibabaei, M. and Ekrem, T. (2018) 'Environmental DNA filtration techniques affect recovered biodiversity', *Scientific Reports*, 8(4682), pp. 1–11. Available at: https://doi.org/10.1038/s41598-018-23052-8.

Mariani, S., Harper, L.R., Collins, R.A., Baillie, C., Wangensteen, O.S., McDevitt, A.D., Heddell-Cowie, M. and Genner, M.J. (2021) 'Estuarine molecular bycatch as a landscape-wide biomonitoring tool', *Biological Conservation*, 261, p. 109287. Available at: https://doi.org/10.1016/j.biocon.2021.109287.

Marshall, S. and Elliott, M. (1998) 'Environmental influences on the fish assemblage of the Humber estuary, U.K.', *Estuarine, Coastal and Shelf Science*, 46(2), pp. 175–184. Available at: https://doi.org/10.1006/ecss.1997.0268.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*, 17(1), p. 10. Available at: https://doi.org/10.14806/ej.17.1.200.

Martino, E.J. and Able, K.W. (2003) 'Fish assemblages across the marine to low salinity transition zone of a temperate estuary', *Estuarine, Coastal and Shelf Science*, 56(5–6), pp. 969–987. Available at: https://doi.org/10.1016/S0272-7714(02)00305-0.

McClenaghan, B., Compson, Z.G. and Hajibabaei, M. (2020) 'Validating metabarcoding-based biodiversity assessments with multi-species occupancy models: A case study using coastal marine eDNA', *PLoS ONE*, 15(3), pp. 1–17. Available at: https://doi.org/10.1371/journal.pone.0224119.

McKenna, J.E., Abbett, R., Waldt, E., David, A. and Snyder, J. (2013) 'Hybrid seine for full fish community collections', *Journal of Freshwater Ecology*, 28(1), pp. 125–131. Available at: https://doi.org/10.1080/02705060.2012.695752.

McLeod, C., M, Y., Brown, A., Burn, A., Hopkins, J. and Way, S. (2005) *The Habitats Directive: selection of Special Areas of Conservation in the UK.* 2nd Ed. Peterborough: Joint Nature Conservation Committee. Available at: www.jncc.gov.uk/SACselection (Accessed: 12 March 2018).

McLusky, D.S. (1992) 'Marine and estuarine gradients', *Marine pollution bulletin*, 24(1), pp. 55–56. Available at: https://doi.org/10.1016/0025-326x(92)90318-z.

McLusky, D.S. (1993) 'Marine and estuarine gradients - An overview', *Netherlands Journal of Aquatic Ecology*, 27(2–4), pp. 489–493. Available at: https://doi.org/10.1007/BF02334809.

McLusky, D.S. and Elliott, M. (2004) *The Estuarine Ecosystem: Ecology, Threats, and Management*. 3rd edn. Oxford: Oxford University Press. Available at: https://doi.org/10.1017/CBO9781107415324.004.

McLusky, D.S. and Elliott, M. (2007) 'Transitional waters: A new approach, semantics or just muddying the waters?', *Estuarine, Coastal and Shelf Science*, 71(3–4), pp. 359–363. Available at: https://doi.org/10.1016/j.ecss.2006.08.025.

Merkes, C.M., McCalla, S.G., Jensen, N.R., Gaikowski, M.P. and Amberg, J.J. (2014) 'Persistence of DNA in carcasses, slime and avian feces may affect interpretation of environmental DNA data', *PLoS ONE*, 9(11). Available at: https://doi.org/10.1371/journal.pone.0113346.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M. and Iwasaki, W. (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), pp. 1–33. Available at: https://doi.org/10.1098/rsos.150088.

Muus, B.J. and Neilsen, J.G. (1999) *Sea Fish*. Hedehusene (Denmark): Scandinavian Fishing Year Book.

Næsje, T.T.F., Cowley, P.P.D., Diserud, O.O.H., Childs, A.R.A., Kerwath, S.E.S. and Thorstad, E.B.E. (2012) 'Riding the tide: estuarine movements of a sciaenid fish, *Argyrosomus japonicus*', *Marine Ecology Progress Series*, 460, pp. 221–232. Available at: https://doi.org/10.3354/meps09780.

Nakagawa, H., Yamamoto, S., Sato, Y., Sado, T., Minamoto, T. and Miya, M. (2018) 'Comparing local- and regional-scale estimations of the diversity of stream fish using eDNA metabarcoding and conventional observation methods', *Freshwater Biology*, 63(6), pp. 569–580. Available at: https://doi.org/10.1111/fwb.13094.

Natural Resources Wales (2019) *Water Framework Directive (WFD) Transitional Waterbodies Cycle 2*, *© Natural Resources Wales and Database Right. All rights Reserved. Ordnance Survey Licence number 100019741. Crown Copyright and Database Right.* Available at: https://data.gov.uk/dataset/7ddae3f7-de97-4547-8bde-706e6d47e535/water-framework-directive-wfd-transitional-waterbodies-cycle-2 (Accessed: 1 June 2022).

NCBI (2021) *Nucleotide ('nt') BLAST Database*. Available at: https://ftp.ncbi.nlm.nih.gov/blast/db/ (Accessed: 1 October 2021).

Nicolas, D., Lobry, J., Lepage, M., Sautour, B., Le Pape, O., Cabral, H., Uriarte, A. and Boët, P. (2010a) 'Fish under influence: A macroecological analysis of relations between fish species richness and environmental gradients among European tidal estuaries', *Estuarine, Coastal and Shelf Science*, 86(1), pp. 137–147. Available at: https://doi.org/10.1016/j.ecss.2009.11.006.

Nicolas, D., Lobry, J., Le Pape, O. and Boët, P. (2010b) 'Functional diversity in European estuaries: Relating the composition of fish assemblages to the abiotic environment', *Estuarine, Coastal and Shelf Science*, 88(3), pp. 329–338. Available at: https://doi.org/10.1016/j.ecss.2010.04.010.

Nielsen, J.G. (1986) 'Scophthalmidae', in P.J.P. Whitehead, M.-L. Bauchot, J.-C. Hureau, J. Nielsen, and E. Tortonese (eds) *Fishes of the North-eastern Atlantic and the Mediterranean*. Paris: UNESCO, pp. 1287–1293.

Niku, J., Hui, F.K.C., Taskinen, S. and Warton, D.I. (2019) 'gllvm: Fast analysis of multivariate abundance data with generalized linear latent variable models in R', *Methods in Ecology and Evolution*, 10(12), pp. 2173–2182. Available at: https://doi.org/10.1111/2041-210X.13303.

Nixon, S.W. (1988) 'Physical energy inputs and the comparative ecology of lake and marine ecosystems', *Limnology and Oceanography*, 33(4part2), pp. 1005–1025. Available at: https://doi.org/10.4319/lo.1988.33.4part2.1005.

Northcott, S.J. (1991) 'A comparison of circatidal rhythmicity and entrainment by hydrostatic pressure cycles in the rock goby, Gobius paganellus L. and the shanny, Lipophrys pholis (L.)', *Journal of Fish Biology*, 39(1), pp. 25–33. Available at: https://doi.org/10.1111/j.1095-8649.1991.tb04338.x.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., Pruitt, K.D., *et al.* (2016) 'Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation', *Nucleic Acids Research*, 44(1), pp. 733–745. Available at: https://doi.org/10.1093/nar/gkv1189.

Oksanen, A.J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., Hara, R.B.O., Simpson, G.L., Solymos, P., Stevens, M.H.H. and Wagner, H. (2015) 'Package "vegan". Available at: https://doi.org/ISBN 0-387-95457-0.

Olin, M. and Malinen, T. (2003) 'Comparison of gillnet and trawl in diurnal fish community sampling', *Hydrobiologia*, 506–509(1999), pp. 443–449. Available at: https://doi.org/10.1023/B:HYDR.0000008545.33035.c4.

Ordnance Survey (2021) *OS Terrain 50*, *Crown copyright and database right 2022*. Available at: https://osdatahub.os.uk/downloads/open/Terrain50 (Accessed: 1 October 2021).

Palmer, M.A. and Febria, C.M. (2012) 'The heartbeat of ecosystems', *Science*, 336(6087), pp. 1393–1394. Available at: https://doi.org/10.1126/science.1223250.

Pattrick, P. and Strydom, N. (2014) 'Recruitment of fish larvae and juveniles into two estuarine nursery areas with evidence of ebb tide use', *Estuarine, Coastal and Shelf Science*, 149, pp. 120–132. Available at: https://doi.org/10.1016/j.ecss.2014.08.003.

Paulson, A., Curl, H., and Feely, R.. (1993) 'The biogeochemistry of nutrients and trace metals in Hood Canal, a Puget Sound fjord', *Marine Chemistry*, 43(1–4), pp. 157–173. Available at: https://doi.org/10.1016/0304-4203(93)90222-A.

Petit, F., Craquelin, S., Guespin-Michel, J. and Buffet-Janvresse, C. (1999) 'Nucleic acid extraction from polluted estuarine water for detection of viruses and bacteria by PCR and RT-PCR analysis', *Research in Microbiology*, 150(2), pp. 143–151. Available at: https://doi.org/10.1016/S0923-2508(99)80031-X.

Pihl, L., Codling, I., Mathieson, S., Mclusky, D.S. and Roberts, C. (2002) 'Habitat use by fishes in estuaries and other brackish areas', in M. Elliott and K. Hemingway (eds) *Fishes in Estuaries*. Oxford: Blackwell Science Ltd, pp. 10–53. Available at: https://doi.org/10.1002/9780470995228.ch2.

Pilliod, D.S., Goldberg, C.S., Arkle, R.S. and Waits, L.P. (2014) 'Factors influencing detection of eDNA from a stream-dwelling amphibian', *Molecular Ecology Resources*, 14(1), pp. 109–116. Available at: https://doi.org/10.1111/1755-0998.12159.

Pilliod, D.S., Goldberg, C.S., Arkle, R.S., Waits, L.P. and Richardson, J. (2013) 'Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples', *Canadian Journal of Fisheries and Aquatic Sciences*, 70(8), pp. 1123–1130. Available at: https://doi.org/10.1139/cjfas-2013-0047.

Piñol, J., Mir, G., Gomez-Polo, P. and Agustí, N. (2015) 'Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods', *Molecular Ecology Resources*, 15(4), pp. 819–830. Available at: https://doi.org/10.1111/1755-0998.12355.

Pochon, X., Wood, S.A., Keeley, N.B., Lejzerowicz, F., Esling, P., Drew, J. and Pawlowski, J. (2015) 'Accurate assessment of the impact of salmon farming on benthic sediment enrichment using foraminiferal metabarcoding', *Marine Pollution Bulletin*, 100(1), pp. 370–382. Available at: https://doi.org/10.1016/j.marpolbul.2015.08.022.

Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A., Roset, N., Schabuss, M., Zornig, H. and Dejean, T. (2018) 'Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation', *Scientific Reports*, 8(1), pp. 1–13. Available at: https://doi.org/10.1038/s41598-018-28424-8.

Port, J.A., O'Donnell, J.L., Romero-Maraccini, O.C., Leary, P.R., Litvin, S.Y., Nickols, K.J., Yamahara, K.M. and Kelly, R.P. (2016) 'Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA', *Molecular Ecology*, 25(2), pp. 527–541. Available at: https://doi.org/10.1111/mec.13481.

Project Jupyter (2020) 'Jupyter Notebook'. Project Jupyter. Available at: jupyter.org.

Promega (2018) *Technical Manual: ProNex Size-Selective Purification System*, *TM508*. Available at: https://www.promega.co.uk/products/sequencing/ngs-library-prep/pronex-size-selective-purification-system/?catNum=NG2001.

QGIS Development Team (2019) 'QGIS Geographic Information System'. Open Source Geospatial Foundation. Available at: https://www.qgis.org/.

QIAGEN (2020) *DNeasy Blood & Tissue Handbook*. Available at: https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/dneasy-blood-and-tissue-kit/.

Quignard, J.-P. and Pras, A. (1986) 'Labridae', in P.J.P. Whitehead, M.-L. Bauchot, J.-C. Hureau, J. Nielsen, and E. Tortonese (eds) *Fishes of the North-eastern Atlantic and the Mediterranean*. Paris: UNESCO, pp. 919–942.

R Core Team (2021) 'A Language and Environment for Statistical Computing', *R Foundation for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, p. https://www.R-project.org. Available at: http://www.r-project.org.

Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R.M. and Gough, K.C. (2014) 'The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology', *Journal of Applied Ecology*, 51(5), pp. 1450–1459. Available at: https://doi.org/10.1111/1365-2664.12306.

Remane, A. (1934) 'Die Brackwasserfauna.', *Verhandlungen Der Deutschen Zoologischen Gesellschaft*, 36, pp. 34–74.

Ribeiro, J., Bentes, L., Coelho, R., Gonçalves, J.M.S., Lino, P.G., Monteiro, P. and Erzini, K. (2006) 'Seasonal, tidal and diurnal changes in fish assemblages in the Ria Formosa lagoon (Portugal)', *Estuarine, Coastal and Shelf Science*, 67(3), pp. 461–474. Available at: https://doi.org/10.1016/j.ecss.2005.11.036.

Ringwood, A.H. and Keppler, C.J. (2002) 'Water quality variation and clam growth : Is pH really a non-issue in Estuaries ?', 25(5), pp. 901–907.

Robins, P.E., Farkas, K., Cooper, D., Malham, S.K. and Jones, D.L. (2019) 'Viral dispersal in the coastal zone: A method to quantify water quality risk', *Environment International*, 126, pp. 430–442. Available at: https://doi.org/10.1016/j.envint.2019.02.042.

Robins, P.E., Lewis, M.J., Simpson, J.H. and Malham, S.K. (2014) 'Future variability of solute transport in a macrotidal estuary', *Estuarine, Coastal and Shelf Science*, 151, pp. 88–99. Available at: https://doi.org/10.1016/j.ecss.2014.09.019.

Robson, H.L.A., Noble, T.H., Saunders, R.J., Robson, S.K.A., Burrows, D.W. and Jerry, D.R. (2016) 'Fine-tuning for the tropics: application of eDNA technology for invasive fish detection in tropical freshwater ecosystems', *Molecular Ecology Resources*, 16(4), pp. 922–932. Available at: https://doi.org/10.1111/1755-0998.12505.

Rotherham, D., Johnson, D.D., Kesby, C.L. and Gray, C.A. (2012) 'Sampling estuarine fish and invertebrates with a beam trawl provides a different picture of populations and assemblages than multi-mesh gillnets', *Fisheries Research*, 123–124, pp. 49–55. Available at: https://doi.org/10.1016/j.fishres.2011.11.019.

Rourke, M.L., Fowler, A.M., Hughes, J.M., Broadhurst, M.K., DiBattista, J.D., Fielder, S., Wilkes Walburn, J. and Furlan, E.M. (2022) 'Environmental DNA (eDNA) as a tool for assessing fish biomass: A review of approaches and future considerations for resource surveys', *Environmental DNA*, 4(1), pp. 9–33. Available at: https://doi.org/10.1002/edn3.185.

Roussel, J.M., Paillisson, J.M., Tréguier, A. and Petit, E. (2015) 'The downside of eDNA as a survey tool in water bodies', *Journal of Applied Ecology*, 52(4), pp. 823–826. Available at: https://doi.org/10.1111/1365-2664.12428.

Rozas, L.P. and Minello, T.J. (1997) 'Estimating densities of small fishes and decapod crustaceans in shallow estuarine habitats: A review of sampling design with focus on gear selection', *Estuaries*, 20(1), pp. 199–213. Available at: https://doi.org/10.2307/1352731.

RStudioTeam (2019) 'Integrated development for R.', *RStudio*. Boston, MA: RStudio. Available at: https://support.rstudio.com/hc/en-us/articles/206212048-Citing-RStudio%0Ahttps://rstudio.com.

Ruddock, P.J., Bird, D.J., McEvoy, J. and Peters, L.D. (2003) 'Bile metabolites of polycyclic aromatic hydrocarbons (PAHs) in European eels *Anguilla anguilla* from United Kingdom estuaries', *Science of The Total Environment*, 301(1–3), pp. 105–117. Available at: https://doi.org/10.1016/S0048-9697(02)00292-9.

Sakamoto, T. and Ortega, J.M. (2021) 'Taxallnomy: an extension of NCBI Taxonomy that produces a hierarchically complete taxonomic tree', *BMC Bioinformatics*, 22(1), p. 388. Available at: https://doi.org/10.1186/s12859-021-04304-3.

Sales, N.G., Wangensteen, O.S., Carvalho, D.C., Deiner, K., Præbel, K., Coscia, I., McDevitt, A.D. and Mariani, S. (2021) 'Space-time dynamics in monitoring neotropical fish communities using eDNA metabarcoding', *Science of the Total Environment*. Available at: https://doi.org/10.1016/j.scitotenv.2020.142096.

Samedy, V., Wach, M., Lobry, J., Selleslagh, J., Pierre, M., Josse, E. and Boët, P. (2015) 'Hydroacoustics as a relevant tool to monitor fish dynamics in large estuaries', *Fisheries Research*, 172, pp. 225–233. Available at: https://doi.org/10.1016/j.fishres.2015.07.025.

Schrader, C., Schielke, A., Ellerbroek, L. and Johne, R. (2012) 'PCR inhibitors - occurrence, properties and removal', *Journal of Applied Microbiology*, 113(5), pp. 1014–1026. Available at: https://doi.org/10.1111/j.1365-2672.2012.05384.x.

Schwentner, M., Zahiri, R., Yamamoto, S., Husemann, M., Kullmann, B. and Thiel, R. (2021) 'eDNA as a tool for non-invasive monitoring of the fauna of a turbid, well-mixed system, the Elbe estuary in Germany', *PLoS ONE*. Available at: https://doi.org/10.1371/journal.pone.0250452.

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*, 2, pp. 1–11. Available at: https://doi.org/10.3897/mbmg.2.24556.

Selleslagh, J. and Amara, R. (2008) 'Environmental factors structuring fish composition and assemblages in a small macrotidal estuary (eastern English Channel)', *Estuarine, Coastal and Shelf Science*, 79(3), pp. 507–517. Available at: https://doi.org/10.1016/j.ecss.2008.05.006.

Selleslagh, J., Amara, R., Laffargue, P., Lesourd, S., Lepage, M. and Girardin, M. (2009) 'Fish composition and assemblage structure in three Eastern English Channel macrotidal estuaries: A comparison with other French estuaries', *Estuarine, Coastal and Shelf Science*, 81(2), pp. 149–159. Available at: https://doi.org/10.1016/j.ecss.2008.10.008.

Selleslagh, J., Lobry, J., N'Zigou, A.R., Bachelet, G., Blanchet, H., Chaalali, A., Sautour, B. and Boët, P. (2012) 'Seasonal succession of estuarine fish, shrimps, macrozoobenthos and plankton: Physico-chemical and trophic influence. The Gironde estuary as a case study', *Estuarine, Coastal and Shelf Science*, 112, pp. 243–254. Available at: https://doi.org/10.1016/j.ecss.2012.07.030.

Sepulveda, A.J., Hutchins, P.R., Forstchen, M., Mckeefry, M.N. and Swigris, A.M. (2020) 'The elephant in the lab (and field): Contamination in aquatic environmental DNA studies', *Frontiers in Ecology and Evolution*, 8. Available at: https://doi.org/10.3389/fevo.2020.609973.

Seymour, M., Durance, I., Cosby, B.J., Ransom-Jones, E., Deiner, K., Ormerod, S.J., Colbourne, J.K., Wilgar, G., Carvalho, G.R., de Bruyn, M., Edwards, F., Emmett, B.A., Bik, H.M. and Creer, S. (2018) 'Acidity promotes degradation of multi-species environmental DNA in lotic mesocosms', *Communications Biology*, 1(4), pp. 1–8. Available at: https://doi.org/10.1038/s42003-017-0005-3.

Shaw, J.L.A., Clarke, L.J., Wedderburn, S.D., Barnes, T.C., Weyrich, L.S. and Cooper, A. (2016) 'Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system', *Biological Conservation*, 197, pp. 131–138. Available at: https://doi.org/10.1016/j.biocon.2016.03.010.

Shen, W., Le, S., Li, Y. and Hu, F. (2016) 'SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation', *PLOS ONE*. Edited by Q. Zou, 11(10), p. e0163962. Available at: https://doi.org/10.1371/journal.pone.0163962.

Sigsgaard, E.E., Nielsen, I.B., Carl, H., Krag, M.A., Knudsen, S.W., Xing, Y., Holm-Hansen, T.H., Møller, P.R. and Thomsen, P.F. (2017) 'Seawater environmental DNA reflects seasonality of a coastal fish community', *Marine Biology*, 164(6), pp. 1–15. Available at: https://doi.org/10.1007/s00227-017-3147-4.

Simpfendorfer, C.A., Kyne, P.M., Noble, T.H., Goldsbury, J., Basiita, R.K., Lindsay, R., Shields, A., Perry, C. and Jerry, D.R. (2016) 'Environmental DNA detects critically endangered largetooth sawfish in the wild', *Endangered Species Research*, 30(1), pp. 109–116. Available at: https://doi.org/10.3354/esr00731.

Song, J.W., Small, M.J. and Casman, E.A. (2017) 'Making sense of the noise: The effect of hydrology on silver carp eDNA detection in the Chicago area waterway system', *Science of the Total Environment*, 605–606, pp. 713–720. Available at: https://doi.org/10.1016/j.scitotenv.2017.06.255.

Spens, J., Evans, A.R., Halfmaerten, D., Knudsen, S.W., Sengupta, M.E., Mak, S.S.T., Sigsgaard, E.E. and Hellström, M. (2017) 'Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter', *Methods in Ecology and Evolution*, 8(5), pp. 635–645. Available at: https://doi.org/10.1111/2041-210X.12683.

Stoeckle, M.Y., Soboleva, L. and Charlop-Powers, Z. (2017) 'Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary', *PLoS ONE*, 12(4), pp. 1–15. Available at: https://doi.org/10.1371/journal.pone.0175186.

Strickler, K.M., Fremier, A.K. and Goldberg, C.S. (2015) 'Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms', *Biological Conservation*, 183, pp. 85–92. Available at: https://doi.org/10.1016/j.biocon.2014.11.038.

Stroud, D., Chambers, D., Cook, S., Buxton, N., Fraser, B., Clement, P., Lewis, P., McLean, I., Baker, H. and Whitehead, S. (eds) (2001) *The UK SPA network: its scope and content*. Peterborough: Joint Nature Conservation Committee. Available at: http://jncc.defra.gov.uk/page-1412 (Accessed: 12 March 2018).

Taberlet, P., Bonin, A., Zinger, L. and Coissac, E. (2018) *Environmental DNA: For Biodiversity Research and Monitoring*. Oxford: Oxford University Press. Available at: https://doi.org/10.1093/oso/9780198767220.001.0001.

Taberlet, P., Coissac, E., Hajibabaei, M. and Rieseberg, L.H. (2012) 'Environmental DNA', *Molecular Ecology*, 21(8), pp. 1789–1793. Available at: https://doi.org/10.1111/j.1365-294X.2012.05542.x.

Tamura, K. and Nei, M. (1993) 'Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees.', *Molecular Biology and Evolution*. Available at: https://doi.org/10.1093/oxfordjournals.molbev.a040023.

Teichert, N., Borja, A., Chust, G., Uriarte, A. and Lepage, M. (2016) 'Restoring fish ecological quality in estuaries: Implication of interactive and cumulative effects among anthropogenic stressors', *Science of The Total Environment*, 542, pp. 383–393. Available at: https://doi.org/10.1016/j.scitotenv.2015.10.068.

Teichert, N., Carassou, L., Sahraoui, Y., Lobry, J. and Lepage, M. (2018a) 'Influence of intertidal seascape on the functional structure of fish assemblages: Implications for habitat conservation in estuarine ecosystems', *Aquatic Conservation: Marine and Freshwater Ecosystems*, 28(4), pp. 798–809. Available at: https://doi.org/10.1002/aqc.2914.

Teichert, N., Lepage, M., Chevillot, X. and Lobry, J. (2018b) 'Environmental drivers of taxonomic, functional and phylogenetic diversity (alpha, beta and gamma components) in estuarine fish communities', *Journal of Biogeography*, 45(2), pp. 406–417. Available at: https://doi.org/10.1111/jbi.13133.

Telesh, I. V. and Khlebovich, V. V. (2010) 'Principal processes within the estuarine salinity gradient: A review', *Marine Pollution Bulletin*, 61(4–6), pp. 149–155. Available at: https://doi.org/10.1016/j.marpolbul.2010.02.008.

The World Bank (2020) *GDP (current US$)*, *World Bank national accounts data, and OECD National Accounts data files.* Available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD (Accessed: 1 July 2022).

Thiel, R., Sepúlveda, A., Kafemann, R. and Nellen, W. (1995) 'Environmental factors as forces structuring the fish community of the Elbe Estuary', *Journal of Fish Biology*, 46(1), pp. 47–69. Available at: https://doi.org/10.1111/j.1095-8649.1995.tb05946.x.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Møller, P.R., Rasmussen, M. and Willerslev, E. (2012) 'Detection of a diverse marine fish fauna using environmental DNA from seawater samples', *PLoS ONE*, 7(8), pp. 1–9. Available at: https://doi.org/10.1371/journal.pone.0041732.

Thomsen, P.F. and Willerslev, E. (2015) 'Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity', *Biological Conservation*, 183, pp. 4–18. Available at: https://doi.org/10.1016/j.biocon.2014.11.019.

Tillotson, M.D., Kelly, R.P., Duda, J.J., Hoy, M., Kralj, J. and Quinn, T.P. (2018) 'Concentrations of environmental DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales', *Biological Conservation*, 220, pp. 1–11. Available at: https://doi.org/10.1016/j.biocon.2018.01.030.

Turner, C.R., Barnes, M.A., Xu, C.C.Y., Jones, S.E., Jerde, C.L. and Lodge, D.M. (2014) 'Particle size distribution and optimal capture of aqueous macrobial eDNA', *Methods in Ecology and Evolution*, 5(7), pp. 676–684. Available at: https://doi.org/10.1111/2041-210X.12206.

Turner, C.R., Uy, K.L. and Everhart, R.C. (2015) 'Fish environmental DNA is more concentrated in aquatic sediments than surface water', *Biological Conservation*, 183, pp. 93–102. Available at: https://doi.org/10.1016/j.biocon.2014.11.017.

UK Parliament (SI 2017 No. 1012) *The Conservation of Habitats and Species Regulations 2017*. Available at: https://www.legislation.gov.uk/uksi/2017/1012/contents/made.

UK Parliament (SI 2017 No. 407) *The Water Environment (Water Framework Directive)(England and Wales) Regulations 2017*. Available at: https://www.legislation.gov.uk/uksi/2017/407/made.

Valdivia-Carrillo, T., Rocha-Olivares, A., Reyes-Bonilla, H., Domínguez-Contreras, J.F. and Munguia-Vega, A. (2021) 'Integrating eDNA metabarcoding and simultaneous underwater visual surveys to describe complex fish communities in a marine biodiversity hotspot', *Molecular Ecology Resources*, 21(5), pp. 1558–1574. Available at: https://doi.org/10.1111/1755-0998.13375.

Venice System (1959) 'Final resolution of the symposium on the classification of brackish waters', *Archivo di Oceanografia e Limnologia*, 11, pp. 243–248.

Wang, Y., Naumann, U., Wright, S.T. and Warton, D.I. (2012) 'Mvabund- an R package for model-based analysis of multivariate abundance data', *Methods in Ecology and Evolution*, 3(3), pp. 471–474. Available at: https://doi.org/10.1111/j.2041-210X.2012.00190.x.

Warry, F.Y., Reich, P., Hindell, J.S., Mckenzie, J. and Pickworth, A. (2013) 'Using new electrofishing technology to amp-up fish sampling in estuarine habitats', *Journal of Fish Biology*, 82(4), pp. 1119–1137. Available at: https://doi.org/10.1111/jfb.12044.

Waugh, A., Elliott, M. and Franco, A. (2019) 'Debunking paradigms in estuarine fish species richness', *Marine Ecology Progress Series*, 613, pp. 125–138. Available at: https://doi.org/10.3354/meps12883.

Wessel, P. and Smith, W.H.F. (1996) 'A global, self-consistent, hierarchical, high-resolution shoreline database', *Journal of Geophysical Research: Solid Earth*, 101(B4), pp. 8741–8743. Available at: https://doi.org/10.1029/96JB00104.

Wessel, P.W. and Smith, W.H.. (2017) *A Global Self-consistent, Hierarchical, High-resolution Geography Database. Version 2.3.7*. Available at: https://doi.org/https://doi.org/10.1029/96JB00104.

WFD - UKTAG (2009) *UKTAG Transitional Water Assessment Methods Fish Fuana Transitional Fish Classification Index*. Edinburgh, Scotland. Available at: www.wfduk.org.

WFD - UKTAG (2014) *UKTAG Transitional Water Assessment Method Phytoplankton. Transitional Water Phytoplankton Tool*. Stirling: Water Framework Directive – United Kingdom Advisory Group.

Whitaker, D. and Christman, M. (2015) 'Package "clustsig"', *R topics documented*, pp. 1–7. Available at: http://www.douglaswhitaker.com.

Whitfield, A. and Elliott, M. (2011) 'Ecosystem and biotic classifications of Estuaries and Coasts', in E. Wolanski and D.S. McLusky (eds) *Treatise on Estuarine and Coastal Science*. Waltham: Academic Press, pp. 99–124. Available at: https://doi.org/10.1016/B978-0-12-374711-2.00108-X.

Whitfield, A.K. and Elliot, M. (2002) 'Fishes as indicators of environmental and ecological changes within estuaries: a review of progress and some suggestions for the future', *Journal of Fish Biology*, 61, pp. 229–250. Available at: https://doi.org/10.1006/jfbi.2002.2079.

Whitfield, A.K., Elliott, M., Basset, A., Blaber, S.J.M. and West, R.J. (2012) 'Paradigms in estuarine ecology - A review of the Remane diagram with a suggested revised model for estuaries', *Estuarine, Coastal and Shelf Science*, 97, pp. 78–90. Available at: https://doi.org/10.1016/j.ecss.2011.11.026.

Wilcox, T.M., McKelvey, K.S., Young, M.K., Lowe, W.H. and Schwartz, M.K. (2015) 'Environmental DNA particle size distribution from Brook Trout (*Salvelinus fontinalis*)', *Conservation Genetics Resources*, 7(3), pp. 639–641. Available at: https://doi.org/10.1007/s12686-015-0465-z.

Wilcox, T.M., McKelvey, K.S., Young, M.K., Sepulveda, A.J., Shepard, B.B., Jane, S.F., Whiteley, A.R., Lowe, W.H. and Schwartz, M.K. (2016) 'Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char *Salvelinus fontinalis*', *Biological Conservation*, 194, pp. 209–216. Available at: https://doi.org/10.1016/j.biocon.2015.12.023.

Wilson, J. and Sheaves, M. (2001) 'Short-term temporal variations in taxonomic composition and trophic structure of a tropical estuarine fish assemblage', *Marine Biology*, 139(4), pp. 787–796. Available at: https://doi.org/10.1007/s002270100624.

Wolanski, E. (2007) *Estuarine Ecohydrology*. 1st edn. Amsterdam: Elsevier. Available at: https://doi.org/10.1016/B978-0-444-53066-0.X5001-6.

Woodhead, R.J., Law, R.J. and Matthiessen, P. (1999) 'Polycyclic Aromatic Hydrocarbons in surface sediments around England and Wales, and their possible biological significance', *Marine Pollution Bulletin*, 38(9), pp. 773–790. Available at: https://doi.org/10.1016/S0025-326X(99)00039-9.

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., Minamoto, T. and Miya, M. (2017) 'Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea', *Scientific Reports*, 7(40368), pp. 1–12. Available at: https://doi.org/10.1038/srep40368.

Yamamoto, S., Minami, K., Fukaya, K., Takahashi, K., Sawada, H., Murakami, H., Tsuji, S., Hashizume, H., Kubonaga, S., Horiuchi, T., Hongo, M., Nishida, J., Okugawa, Y., Fujiwara, A., Fukuda, M., Hidaka, S., Suzuki, K.W., Miya, M., Araki, H., Yamanaka, H., Maruyama, A., Miyashita, K., Masuda, R., Minamoto, T. and Kondoh, M. (2016) 'Environmental DNA as a "snapshot" of fish distribution: A case study of Japanese jack mackerel in Maizuru Bay, Sea of Japan', *PLoS ONE*, 11(3), pp. 1–18. Available at: https://doi.org/10.1371/journal.pone.0149786.

Yeoh, D.E., Valesini, F.J., Hallett, C.S., Abdo, D.A. and Williams, J. (2017) 'Diel shifts in the structure and function of nearshore estuarine fish communities', *Journal of Fish Biology*, 90(4), pp. 1214–1243. Available at: https://doi.org/10.1111/jfb.13222.

Zhang, S., Lu, Q., Wang, Y., Wang, X., Zhao, J. and Yao, M. (2020a) 'Assessment of fish communities using environmental DNA: Effect of spatial sampling design in lentic systems of different sizes', *Molecular Ecology Resources*, 20(1), pp. 242–255. Available at: https://doi.org/10.1111/1755-0998.13105.

Zhang, S., Zhao, J. and Yao, M. (2020b) 'A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish', *Methods in Ecology and Evolution*, 11(12), pp. 1609–1625. Available at: https://doi.org/10.1111/2041-210X.13485.

Zou, K., Chen, J., Ruan, H., Li, Z., Guo, W., Li, M. and Liu, L. (2020) 'eDNA metabarcoding as a promising conservation tool for monitoring fish diversity in a coastal wetland of the Pearl River Estuary compared to bottom trawling', *Science of The Total Environment*, 702, p. 134704. Available at: https://doi.org/10.1016/j.scitotenv.2019.134704.

Zuur, A.F., Ieno, E.N. and Smith, G.M. (2007) *Analysing Ecological Data*. New York: Springer.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A. and Smith, G.M. (2009) *Mixed effects models and extensions in ecology*. New York: Springer.