eDNA monitoring of restoration effects on the River Severn (Phase II)

Unlocking the Severn LIFE project

February 2024

Natural England Commissioned Report NECR503

NATURAL ENGLAND

About Natural England

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

Further Information

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

Copyright

Catalogue code: NECR503

Report details

Authors

Nathan P. Griffiths
Bernd Hänfling
Demetra Andreou
Barbara Morrissey
Graham S. Sellers
Dasha Svoboda

Natural England Project Manager

Louise Whatley

Contractors

Dr Bernd Hänfling
Institute for Biodiversity and Freshwater Conservation
UHI Inverness
Email: bernd.haenfling.ic@uhi.ac.uk

Dr Demetra Andreou
Faculty of Science and Technology
Bournemouth University

Keywords

eDNA, metabarcoding, restoration, monitoring, River Severn, Shad

Citation

Griffiths *et al*, 2024., eDNA monitoring of restoration effects on the River Severn (Phase II). NECR503. Natural England.

Foreword

DNA-based methods offer a significant opportunity to change how we monitor and assess biodiversity. However, for most techniques, there is still development required before they can be used in routine monitoring. Natural England has been exploring the further use of these methods for environmental monitoring for several years, delivering a series of reports which focus on the development of DNA-based methods with potential in a particular area.

Natural England aims to make monitoring programmes more efficient, by developing and testing eDNA techniques for aquatic communities. River restoration schemes have the potential to lead to changes in fish communities, and eDNA techniques are being developed to monitor community composition and distribution of fishes following reconnection. This project is a continuation from the Natural England project 'eDNA methods for monitoring restoration effects on the River Severn' (Sellers, *et al*, unpublished).

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

Executive summary

The 'Unlocking the Severn' project has reconnected 158 miles of shad and sea lamprey habitat through the creation of fish passes at four weirs on the River Severn between Diglis and Lincomb. Reconnection schemes in river systems can lead to changes in the fish community (Magilligan et al. 2016), therefore monitoring community composition and distribution of fishes prior and post reconnection is required to assess the impact of such changes. The aim of this project was to generate biodiversity data for 2022 using eDNA metabarcoding approaches which can be used to investigate wider ecological effects of restoring river connectivity in the River Severn (Unlocking the Severn LIFE project). Specifically, we wished to map the distribution of shad (Alosa spp.) eDNA in relation to weirs in the Severn and investigate the possible presence of shad in the Thames; provide semi-quantitative data on fish community structure in the Severn and Thames; and compare the 2022 fish community data with data from previous metabarcoding surveys in 2021 and 2018 in the River Severn; and provide data on invertebrate diversity and community composition in the Severn and Thames respectively.

eDNA samples were collected at five sites in the River Severn and two sites in the river Thames, by Natural England (NE) and Canal and Rivers Trust (CRT) staff, with the help of CRT volunteers. Samples were collected over two dates during the putative shad spawning season at the same locations along the River Severn (9th May, 31st May) and Thames (12th May, 1st June) respectively. Replicate samples (250-900 ml each) were collected from mid-river at each location and individually filtered on site using Sterivex filters. Filters were sent to the laboratory for DNA extraction and sequencing using PCR with two different primer sets; fish specific primers targeting the mitochondrial (mt) 12S region (MiFish, Miya et al. 2015, 2020) and invertebrate specific primers targeting a section of the generic mt COI barcoding region (Leese et al. 2021). During the process PCR products were tagged with an index unique to each sample and then sequenced on an Illumina MiSeq using v3. Raw sequencing data were analysed using a reproducible metabarcoding bioinformatic protocols established at UHI Inverness.

During the first sampling event on the 9th May two of the sites at the downstream end of the sampling range tested positive for Alosa eDNA indicating that the species had not passed the Holt fish pass at Holt Fleet Bridge weir at this stage in the migration. During the second sampling event on the 31st May shad eDNA was found at all five sites indicating that the species had successfully negotiated all four weirs. No shad eDNA was detected at either of the two sites in the river Thames (Figure 3) at any of the sampling dates. A total of 25 and 20 fish taxa and 86 and 40 invertebrate species were detected in eDNA samples from the rivers Severn and Thames respectively. As expected, fish and invertebrate community structures differed considerably between the Severn and Thames. Fish species detected in this survey in the River Severn were largely the same as those detected in the 2021 eDNA metabarcoding survey. Shad DNA was detected above the Lincomb weir during the 2022 shad spawning run indicating that the shad are accessing the reconnected habitat in their first spawning event following reconnection. The results of previous surveys in the River Severn support the usefulness of eDNA metabarcoding in

monitoring changes in fish communities resulting for habitat restoration, and potentially more broadly to assess water quality through surveying invertebrate communities. It is recommended that a long-term monitoring programme is established with consistent spatio-temporal coverage to enable the effect of eDNA dispersal on spatial variation; research into incorporating hydraulic modelling approaches into lotic eDNA surveys to allow spatially more accurate predictions of species distributions; and review existing invertebrate eDNA specific metrics for the use in future monitoring of ecological status.

Contents

1 Introduction

Rivers across the world have been obstructed through navigation weirs and dams resulting in fragmentation and biodiversity losses (van Puijenbroek et al 2019). In recent years, there has been a concerted conservation effort to rewild rivers through extensive reconnection projects. Such reconnection schemes in river systems can lead to changes in the fish community (Magilligan et al. 2016) and monitoring community composition and distribution of fishes prior and post reconnection are required to assess the impact of such changes. Furthermore, the structural changes can alter flow, depth and habitat topologies which can lead to considerable differences in the fish and invertebrate assemblages.

The lower River Severn basin has been subjected to considerable river engineering in the 19th Century through a series of weirs to enable navigation in support of industry. These obstructions have inhibited migratory species including European shad (*Alosa spp.*) and sea lamprey (*Petromyzon marinus*), disrupting their spawning migration routes. It is also possible that it has affected the distribution of other fish species. In response to this, the 'Unlocking the Severn' project has reconnected 158 miles of shad and sea lamprey habitat through the creation of fish passes at four weirs on the river - Diglis weir (October 2020), Bevere (September 2020), Holt (January 2022) and Lincomb (June 2021). As a result of these fish passes, in May-June 2022, the shad had access to the reconnected habitat, with their spatial use of the river being monitored through radio tracking, audio moths, fish pass monitoring (viewing window) and eDNA based detection (described here).

Environmental DNA (eDNA) metabarcoding approaches are increasingly used for biodiversity monitoring of aquatic habitats. This approach combines eDNA with modern High-Throughput-Sequencing technology and allows the simultaneous characterization of entire biological communities. Previous research has shown that eDNA metabarcoding is more effective at detecting elusive fish species than established invasive surveying techniques such as electric fishing or fyke netting, and in both lentic and lotic habitats (Hänfling et al. 2016, Lawson Handley et al. 2019, Li et al. 2019, Griffiths et al. 2020, Pont et al. 2021). eDNA metabarcoding approaches have also been successfully used to characterise aquatic invertebrate communities (Deiner et al. 2016, Harper et al. 2021) and to detect cryptic non-native invertebrate species (e.g. Blackman et al. 2018). However, eDNA metabarcoding results for invertebrates are usually not directly comparable to conventional survey methods such as kick-sampling as the taxonomic overlap between the different approaches is often limited (Gleason et al. 2021, Harper et al. 2021). Correspondingly, the use of these data for ecological status monitoring requires a re-evaluation of the quality indices used (Hering et al. 2018). For fish, eDNA metabarcoding can also provide reliable semi-quantitative estimates and, in lentic habitats, the approach is now being deployed in routine monitoring programmes by the DEFRA agencies and to estimate the ecological status of lakes (Willby et al. 2019).

An eDNA metabarcoding approach was used in 2018 (samples collected during the shad run) and 2021 (samples collected after the shad run at two temporal scales) to monitor the distribution of fish species across the River Severn catchment, prior to and shortly after the

removal of major obstructions to fish passage (Antognazza et al 2021a, Sellers et al 2022). Furthermore, an invertebrate metabarcoding survey was carried out during 2021. These data provide the basis for monitoring changes in biological communities over the coming years.

2 Project Aims and objectives

The aim of this project was to generate biodiversity data for 2022 using eDNA metabarcoding approaches which can be used to investigate wider ecological effects of restoring river connectivity in the River Severn (Unlocking the Severn LIFE project). The project focused on five sites in the middle section of the River Severn which are located before, after and in-between a series of four major weirs. Additionally, samples from two putative shad spawning sites in the upstream of the river Thames were included.

Specific objectives were to:

1) Map the distribution of shad (*Alosa spp.*) eDNA in relation to weirs in the Severn and investigate the possible presence of shad in the Thames.

2) Provide semi-quantitative data on fish community structure in the Severn and Thames.

3) Compare the 2022 fish community data with data from previous metabarcoding surveys in 2021 and 2018 in the River Severn.

4) Provide data on invertebrate diversity and community composition in the Severn and Thames respectively.

3 Methods

3.1 Sample collection and DNA extraction

Water samples were collected by Natural England (NE) and Canal and Rivers Trust (CRT) staff, with the help of CRT volunteers. Samples were collected at two dates during the putative shad spawning season at the same locations along the River Severn (9th May, 31st May) and Thames (12th May, 1st June) respectively (Figure 1). For each location and date, three replicate samples (250-900 ml each) were collected from mid-river at each location and individually filtered on site using Sterivex filters. Samples were transported to the lab in cold storage within a few days of sampling and prior to DNA extraction all samples were stored in a fridge at -80°C for a maximum of 43 days. Individual sampling replicates underwent DNA extraction at the University of Bournemouth. The extracted DNA samples were transported to the UHI Inverness for sequencing (see Appendix 2A and B for detailed descriptions of filtration and DNA extraction).
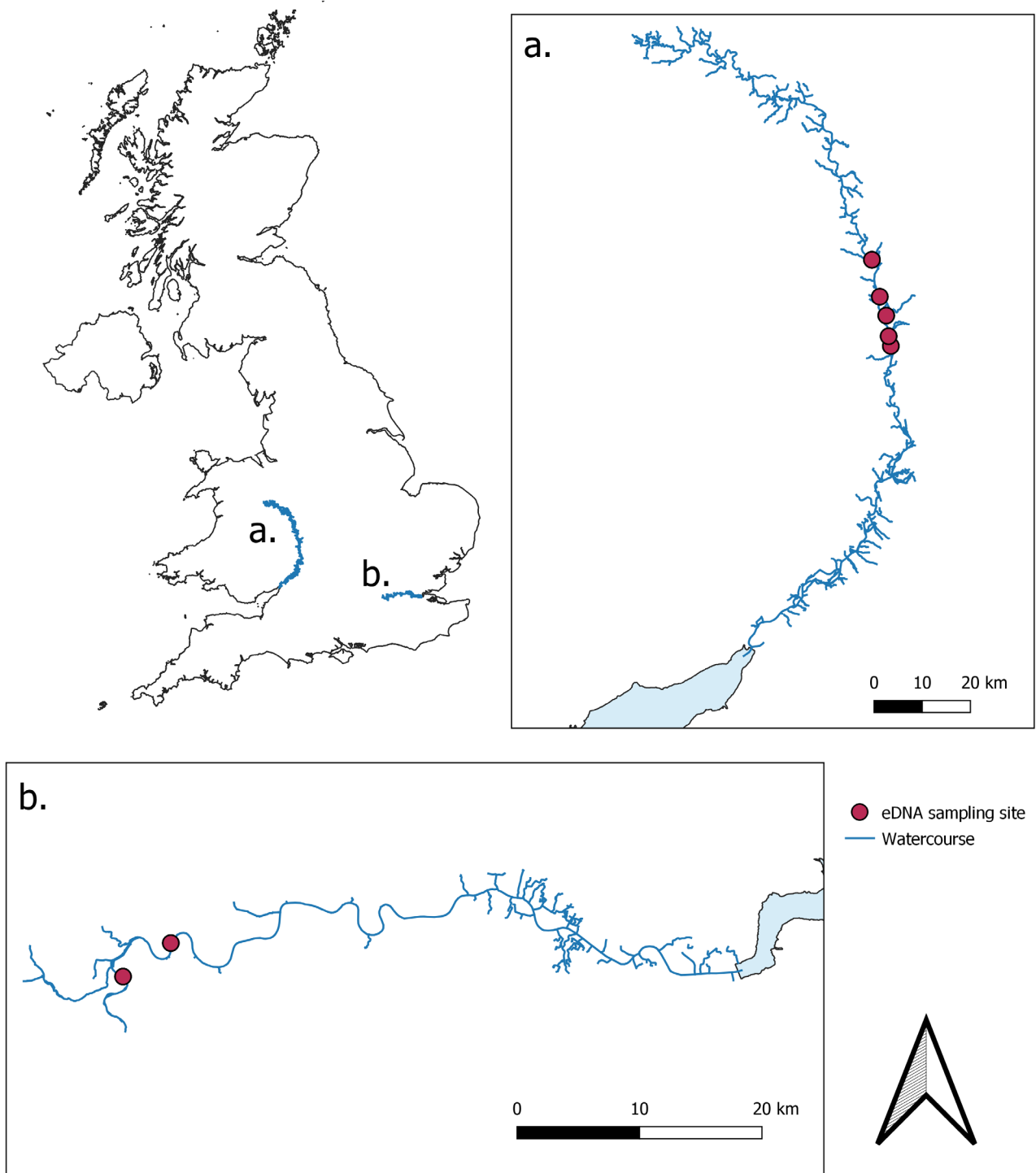
Figure 1. eDNA sample sites in the River Severn (a.) and River Thames (b.) in context of the UK map. Contains data from OS Open Rivers (© Ordnance Survey 2023).

3.2 PCR and DNA sequencing

DNA samples were sequenced following metabarcoding protocols established at UHI Inverness. First, DNA extractions were PCR-amplified in triplicate using two different primer sets; fish specific primers targeting the mitochondrial (mt) 12S region (MiFish, Miya et al. 2015, 2020) and invertebrate specific primers targeting a section of the generic mt COI barcoding region (Leese et al. 2021). During the process PCR products were tagged with an index unique to each sample and then sequenced on an Illumina MiSeq using v3 chemistry (see Appendix 2E and H for detailed description).

3.3 Bioinformatics and data analysis

Raw sequencing data were analysed using a reproducible metabarcoding bioinformatic workflow, *Tapirs* (https://github.com/EvoHull/Tapirs), developed at the University of Hull. Sequencing reads underwent a *BLAST* (Zhang et al. 2000) taxonomic assignment against curated reference databases for UK vertebrates and invertebrates respectively. A criterion of 98% (vertebrates) and 95% (invertebrates) identity with the reference sequences was applied (see Appendix 2F and J for detailed description). Following taxonomic assignment, a noise threshold of 0.1% of total reads per sample was applied when comparing vertebrates with data from previous years to remove low-frequency reads (Hänfling et al. 2016), while a minimum reads threshold was applied to invertebrate data comparisons removing any species with <20 reads assigned. Most reads were assigned to the species level, but as the fish specific molecular marker used here cannot distinguish species of the genera *Leuciscus* (common dace, *L. leuciscus* and ide, *L. idus*)*, Lampetra* (brook lamprey, *L. planeri* and river lamprey *L. fluviatilis*) and Alosa (twaite shad, *A. fallax* and Allis shad, *A. alosa*) reads of these species were assigned to the genus level. All remaining assignments to taxonomic levels higher than species were excluded from the analysis. Finally, reads assigned to positive controls, reads which could not be assigned to any taxon and samples with no taxonomically assignable reads were also excluded from the analysis. Any samples with less than 500 total reads are also considered to be of poor quality and removed.

The number of sequence reads assigned to species were converted to relative reads per sample (proportional reads) to create a standardised measure of eDNA community composition across samples (species reads / total sample reads). Previous research has found a strong correlation of fish read counts with actual recorded abundance and biomass of fish communities within UK freshwater systems (Li et al. 2019, Di Muri et al. 2020) and these data can therefore be interpreted as a proxy of relative abundance. No such relationship has been demonstrated for invertebrate metabarcoding data.

Species richness (total number of species) per site was used as a measure for α - diversity and total number of species for each river as measure for ɣ – diversity. Differences in community composition (β - diversity) was calculated as Bray-Curtis distance between

sites and visualised using non-metric multidimensional scaling (NMDS) ordination, differences between groupings were tested for significance using PERMANOVA.

To create a reliable comparison between the studies, the raw 2018 sequencing data from Antognazza et al. (2021) were reanalysed using the bioinformatics workflow above. Likewise, the raw 2021 sequencing data from Sellers et al. (2022, unpublished report to NE) were analysed following the same workflow as the present study. Fish biodiversity metrics were also generated in the same manner as above. Of the five sites sampled on the River Severn in the current study (2022), three were included in previous years: Diglis(S1) (2018, 2021), Sabrina(S2) (2021) and Stourport(S5) (2021). Therefore, only data from the aforementioned sites taken in previous years were compared directly in this report. When comparing between different years, all samples from each site are included together (i.e. temporal samples were merged). NMDS ordination was used to visualise differences in fish community estimates between year and site. All downstream data processing, analysis, and visualisation was carried out using R (R Core Team, 2022).

4 Results

4.1 Distribution of *Alosa* eDNA

During the first sampling event on the 9th May two of the sites at the downstream end of the sampling range tested positive for *Alosa* eDNA indicating that the species had not passed the Holt fish pass at Holt Fleet Bridge weir at this stage in the migration. During the second sampling event on the 31st May shad eDNA was found at all five sites indicating that the species had successfully negotiated all four weirs (Figure 2).
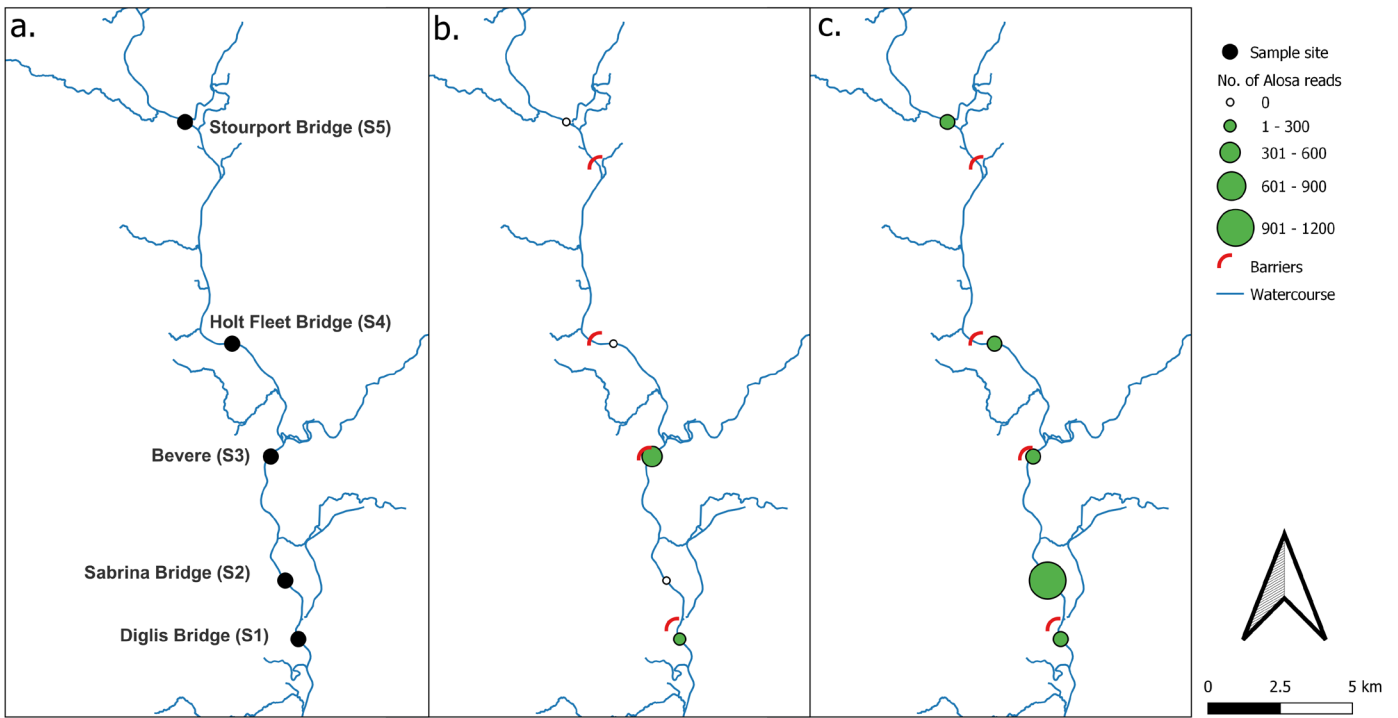
Figure 2. Average number of reads assigned to shad (*Alosa spp.*) at each site in the River Severn. Including site names and codes (a.), and number of reads detected in the first (b.) and second (c.) sampling event. Circle size represents quantitative read categories (see side legend). Contains data from OS Open Rivers (© Ordnance Survey 2023).

No shad eDNA was detected at either of the two sites in the river Thames (Figure 3) at any of the sampling dates.
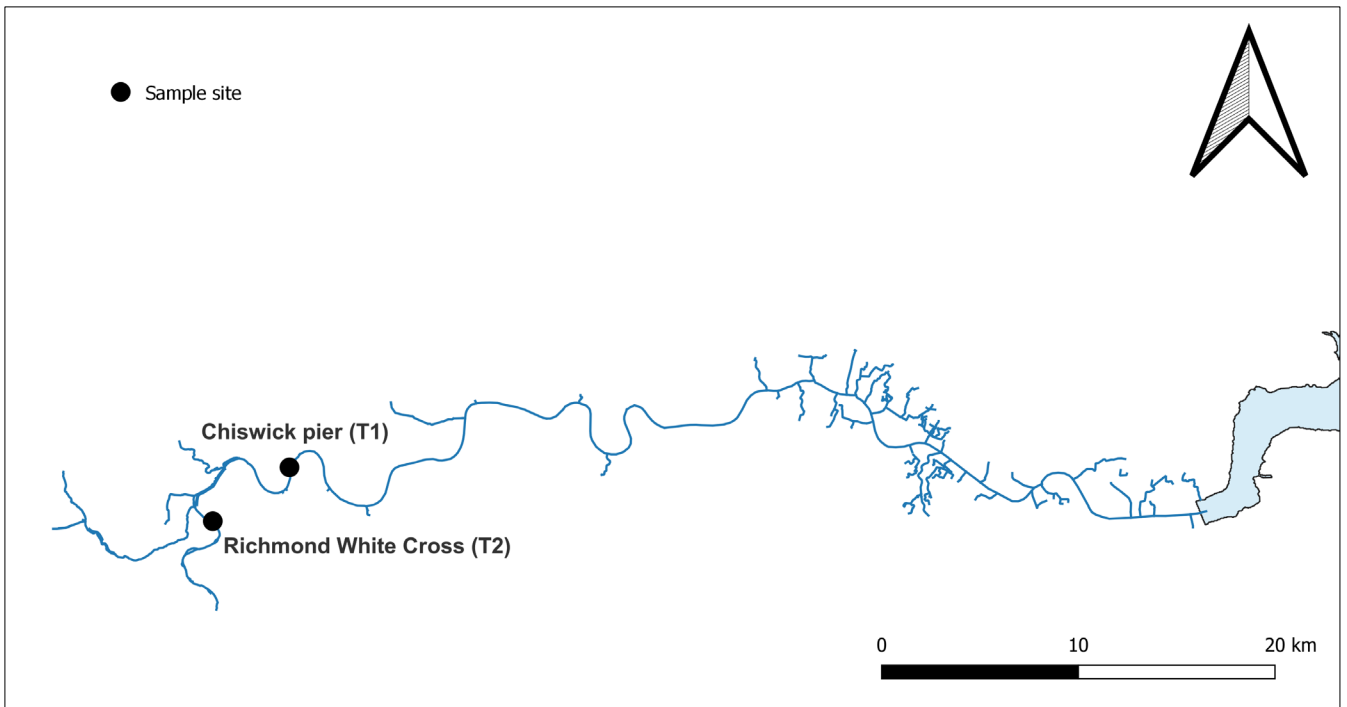

Figure 3. Sampling sites in the River Thames (no shad DNA detected). Contains data from OS Open Rivers (© Ordnance Survey 2023).

4.2 Fish community 2022

A total of 25 and 20 fish taxa were detected in eDNA samples from the rivers Severn and Thames respectively. The α-diversity ranged from 17 to 20 species in sites from the River Severn and from 6 to 15 species in the River Thames (Figure 4). Next to *Alosa* two additional migratory species were detected in the Severn, sea lamprey (*Peteromyzon marinus*) was found in sites S2 and S3 and Atlantic salmon (*Salmo salar*) was detected at all sites. *S. salar* was also detected at both sites in the River Thames in addition to one detection of European smelt (*Osmerus operlanus*).

As expected, the fish community structure differed considerably and significantly between the Severn and Thames (Figures 4 and 5; PERMANOVA, R^2 = 0.59, DF = 1, P = 0.001). The Severn fish community was characterised by a wide range of cyprinid species and predators while the community in the River Thames was dominated by a few species such as European eel (*Anquilla anguilla*), flounder (*Platichthys flesus*) and common bream (*Abramis brama*) (Figure 4).
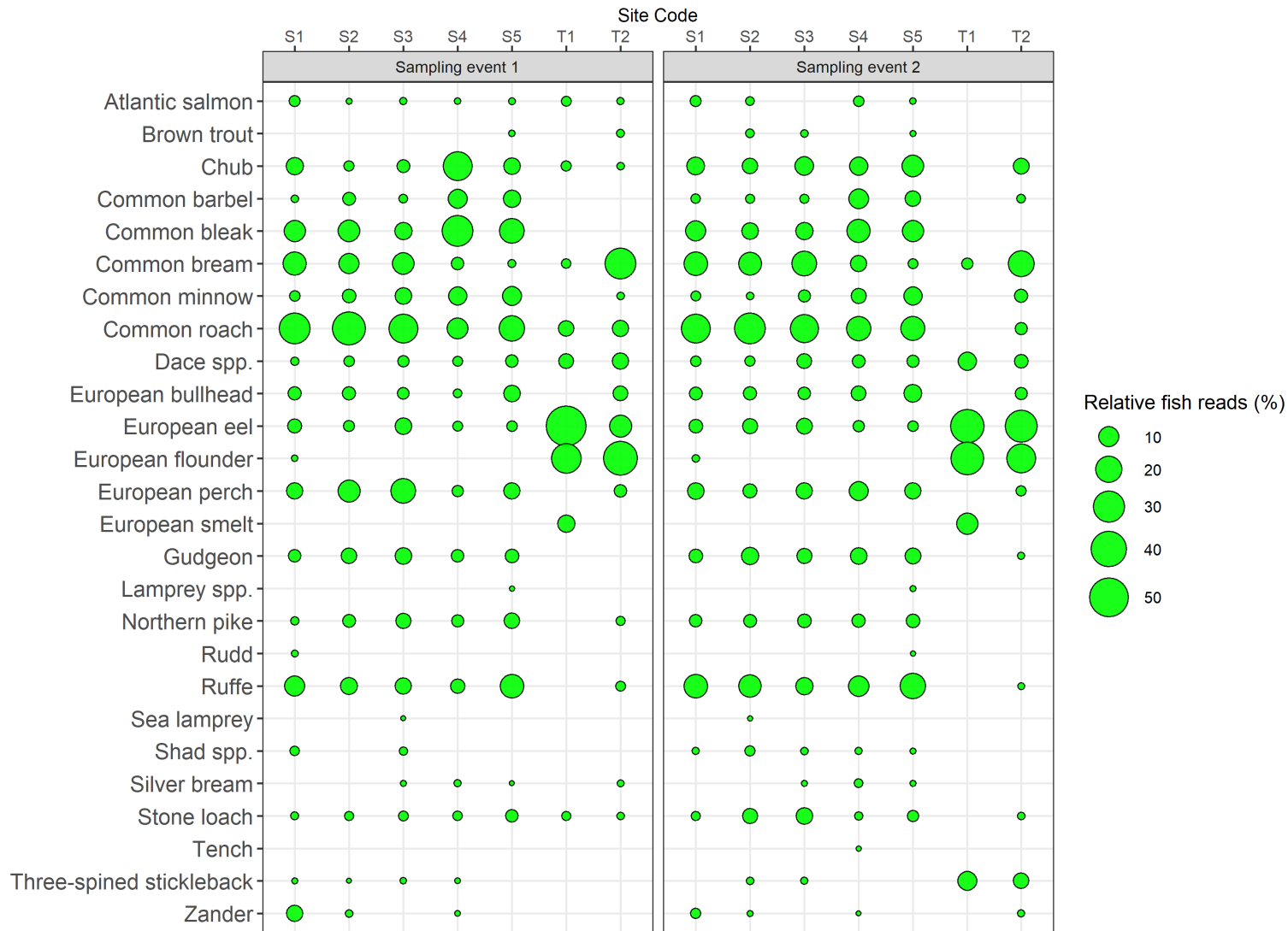
Figure 4. Relative fish reads assigned to each species at each site, split between the first and second sampling events of 2022. Site codes indicate sites from the rivers Severn and Thames.
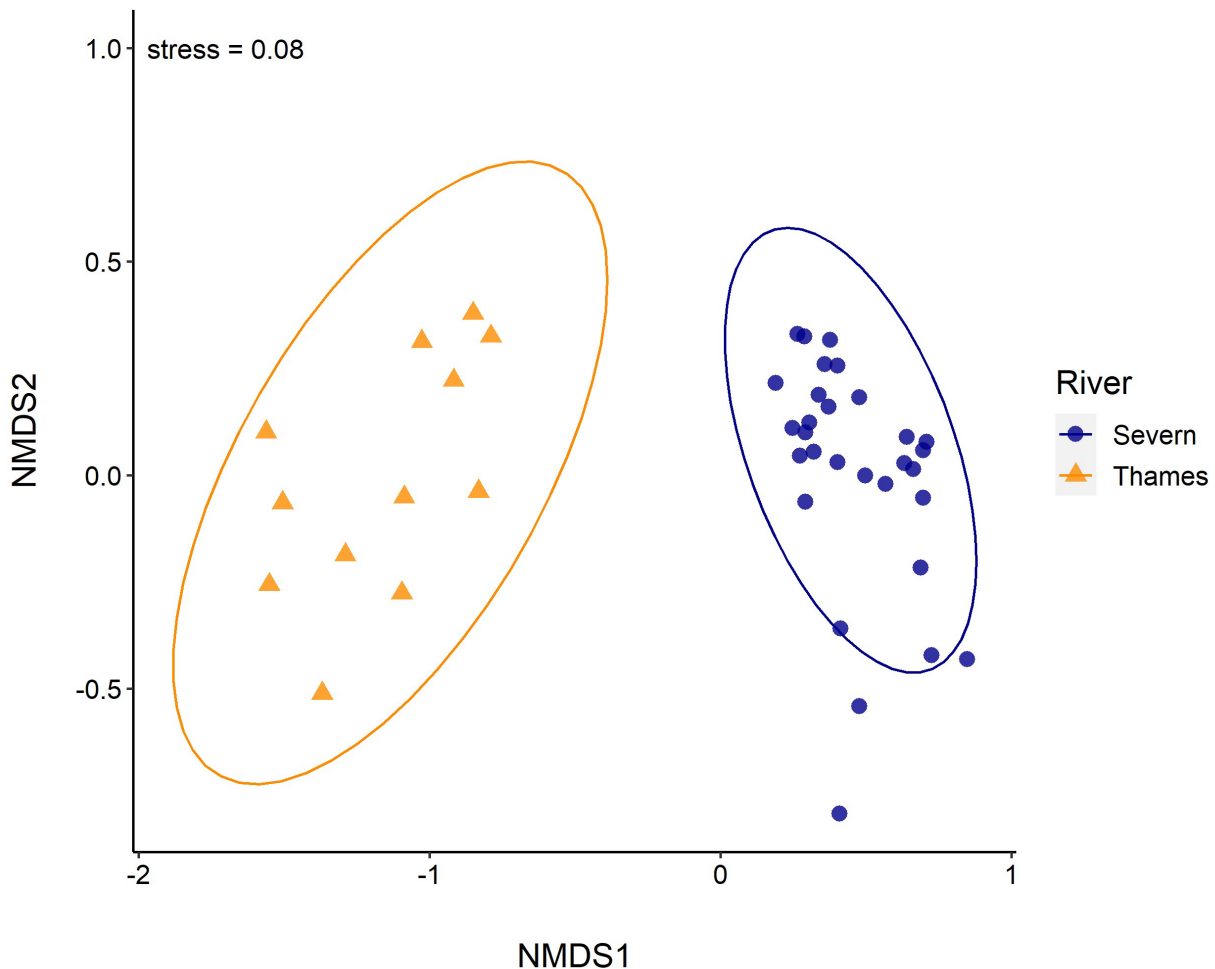
Figure 5. Non-metric multidimensional scaling (NMDS) of fish Bray-Curtis distances among individual samples from both river systems, showing similarity between samples and sample groups. The ellipses indicate the estimated multidimensional space of individual groups. Larger ellipses indicate more community dissimilarity among samples.

The fish community structure within the River Severn is relatively homogenous compared to the Thames (Figure 5). A separate analysis within the River Severn however indicates that the sampling site is a significant driver of changes in fish community (PERMANOVA, $R^2 = 0.47$, DF = 4, P = 0.001); the associated NMDS shows that the two most upstream sites S4 and S5 are distinct from the three downstream sites S1 S2 and S3 (Figure 6). There is no significant difference in the community composition between the two sampling events in the River Severn (PERMANOVA, $R^2 = 0.07$, DF = 1, P = 0.081).
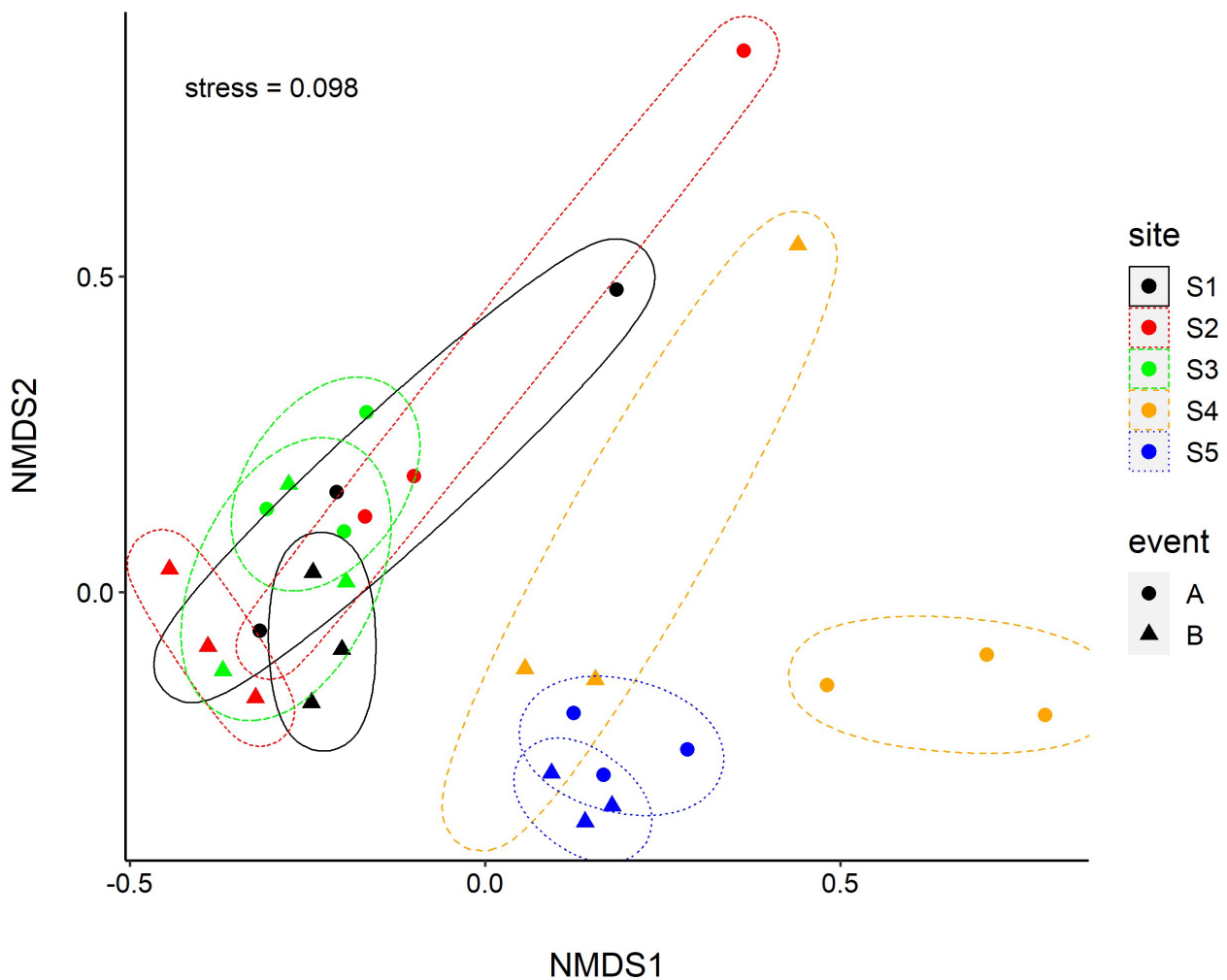
Figure 6. Non-metric multidimensional scaling (NMDS) of fish pairwise Bray-Curtis distances among individual samples within the River Severn showing similarity between samples and sample groups. The ellipses indicate the estimated multidimensional space of individual groups from the first sampling event (A) and second sampling event (B). Larger ellipses indicate more community dissimilarity among samples.

4.3 Invertebrates 2022

A total of 86 and 40 invertebrate species were detected in eDNA samples from the rivers Severn and Thames respectively (Figure 7). The α-diversity ranged from 25-46 species in sites from the River Severn and from 15-25 species in the River Thames. The majority of species detected in the River Severn were from the orders diptera (true flies), tubificida (aquatic worms) and trichoptera (caddisflies) with ephemeroptera (mayflies) and coleoptera (beetles) also represented with a good number of species. In contrast the species diversity in the River Thames was dominated by just two orders, diptera and tubificida.
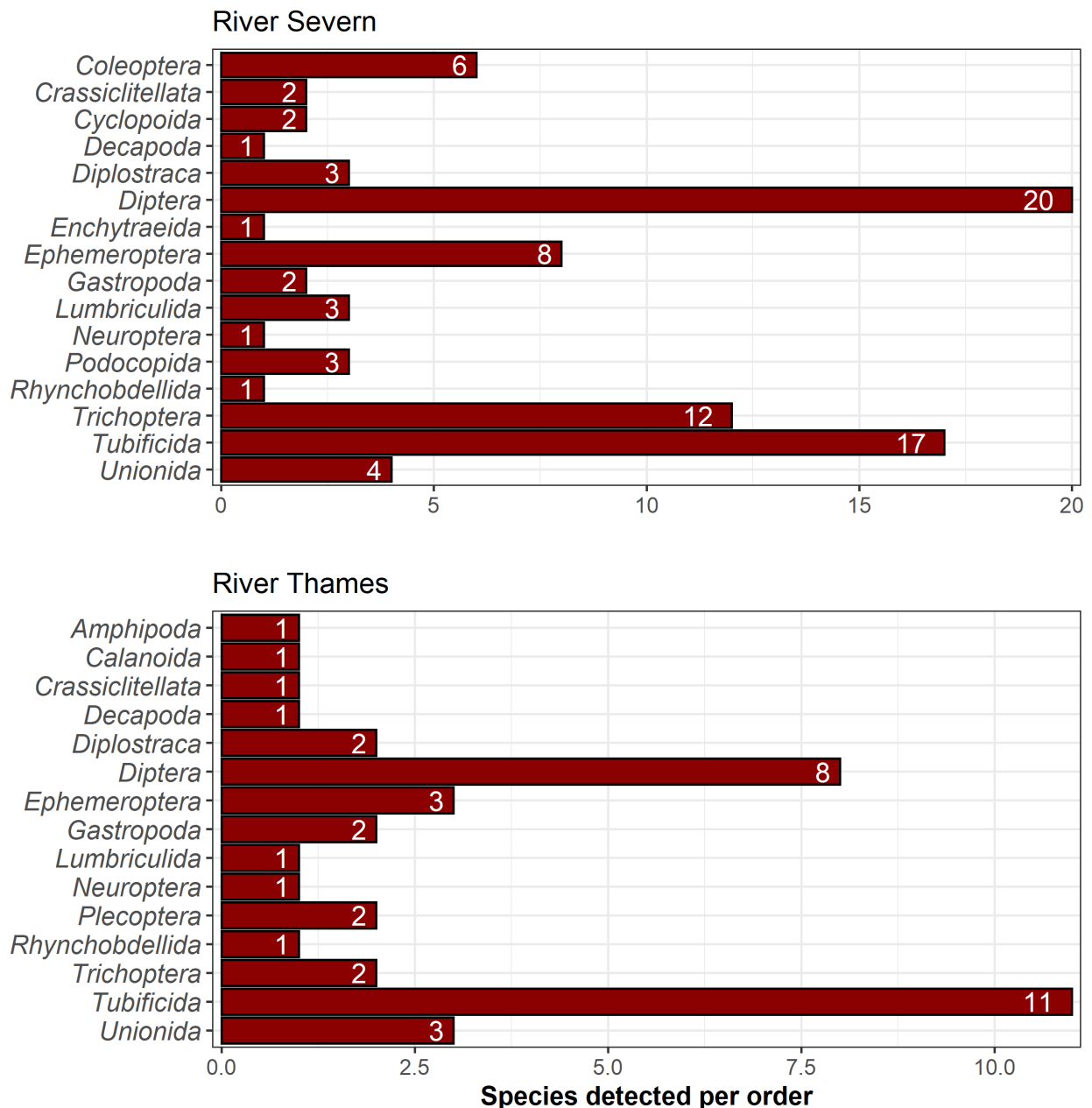
Figure 7. Number of invertebrate species detected by order in each river. Bars are proportional to the number of species assigned, white labels indicate specific number of species detected within each order.

In terms of read counts diptera, ephemeroptera and tubificida and unionidae (freshwater mussels) dominated the samples from the River Severn, whereas calanoida (order of copepods), diptera, gastropoda (snails) and tubificida were the most common sequences in the Thames samples.

As expected, the invertebrate community structure differed considerably and significantly between the Severn and Thames (Figures 8 and 9; PERMANOVA, R2 = 0.45, DF = 1, P = 0.001).
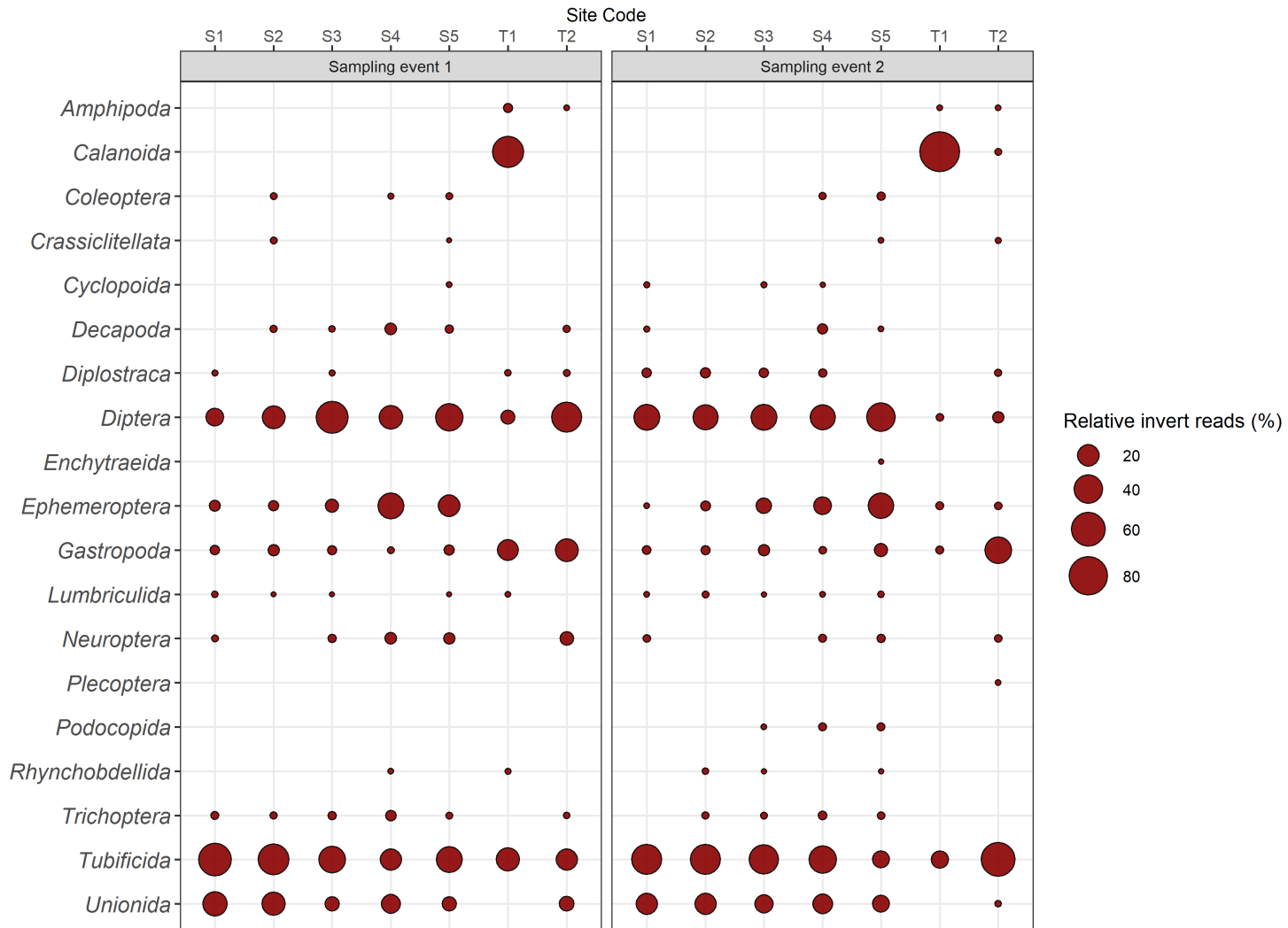
Figure 8. Relative invertebrate reads assigned to each species at each site, split between the first and second sampling events of 2022. Site codes indicate sites from the rivers Severn and Thames.

Figure 9. Non-metric multidimensional scaling (NMDS) of invertebrate Bray-Curtis distances among individual samples from both river systems, showing similarity between samples and sample groups. The ellipses indicate the estimated multidimensional space of individual groups. Larger ellipses indicate more community dissimilarity among samples.

The invertebrate community structure within the River Severn is relatively homogenous compared to the Thames (Figure 9). A separate analysis within the River Severn however indicates that the sampling site is a significant driver of changes in invertebrate community (PERMANOVA, $R^2 = 0.52$, DF = 4, P = 0.001); the associated NMDS shows that the most upstream site (S5) is distinct from the others (Figure 10). There is also a weak but significant difference in the invertebrate community composition between the two sampling events in the River Severn (PERMANOVA, $R^2 = 0.12$, DF = 1, P = 0.013)

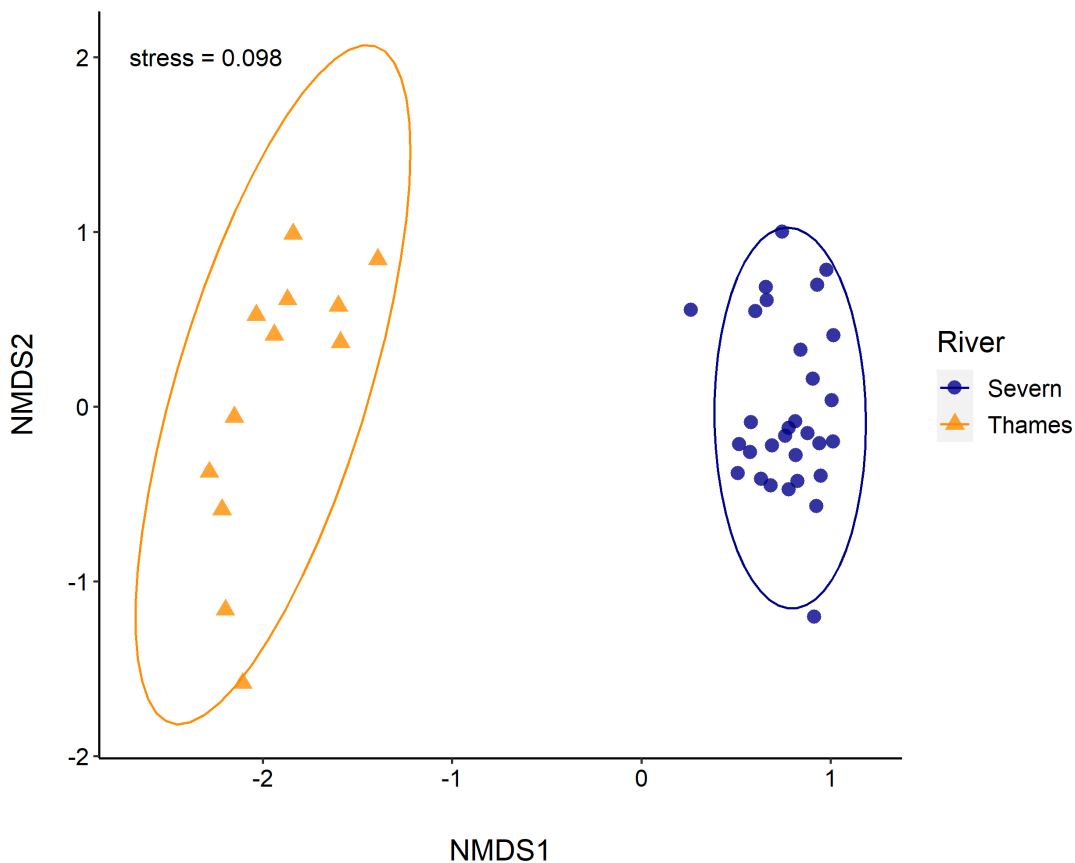**Figure 10. Non-metric multidimensional scaling (NMDS) of invertebrate pairwise Bray-Curtis distances among individual samples within the River Severn showing similarity between samples and sample groups. The ellipses indicate the estimated multidimensional space of individual groups. Larger ellipses indicate more community dissimilarity among samples.**

4.4 Fish community comparison (2018 - 2022)
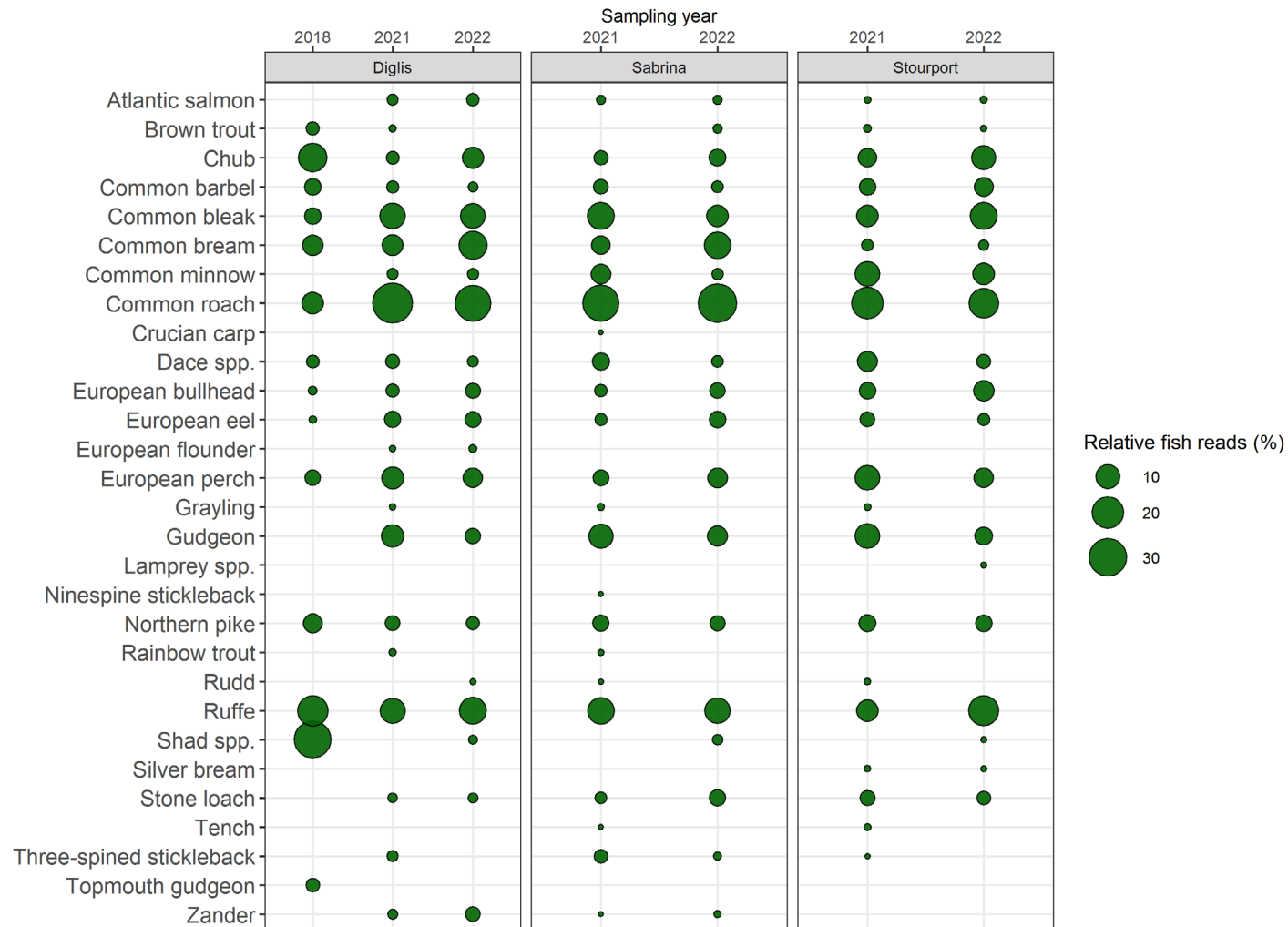
Figure 11. Relative fish reads assigned to each species at 3 sites, providing an overview of the fish community assemblages in different sampling years.

As expected, the fish community structure differed between sampling sites (PERMANOVA, $R^2 = 0.14$, DF = 2, P = 0.001) and repeat sampling years (PERMANOVA, $R^2 = 0.30$, DF = 2, P = 0.001) in the River Severn. However, the 2018 data were highly variable (Figure S3), thus 2021 and 2022 comparisons are visualised separately (Figure 12). Considering only the 2021 / 2022 sampling data a weaker but significant difference in fish community was observed between the years (PERMANOVA, $R^2 = 0.06$, DF = 1, P = 0.033) and fish community also varied between sites (PERMANOVA, $R^2 = 0.23$, DF = 2, P = 0.001).
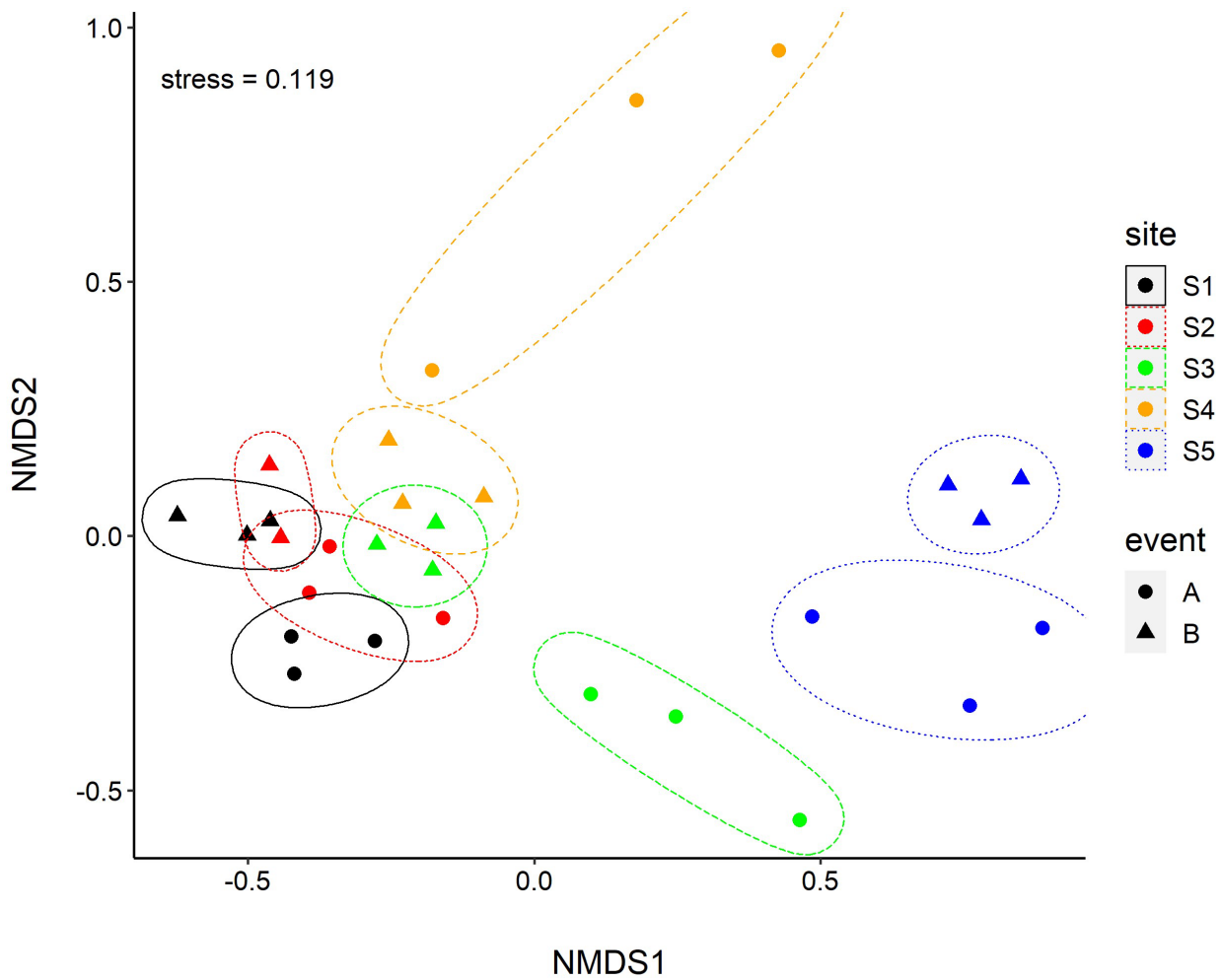

Figure 12. Non-metric multidimensional scaling (NMDS) of fish Bray-Curtis distances among individual samples within the River Severn covering sites sampled over different years. The ellipses indicate the estimated multidimensional space of individual groups. Larger ellipses indicate more community dissimilarity among samples.
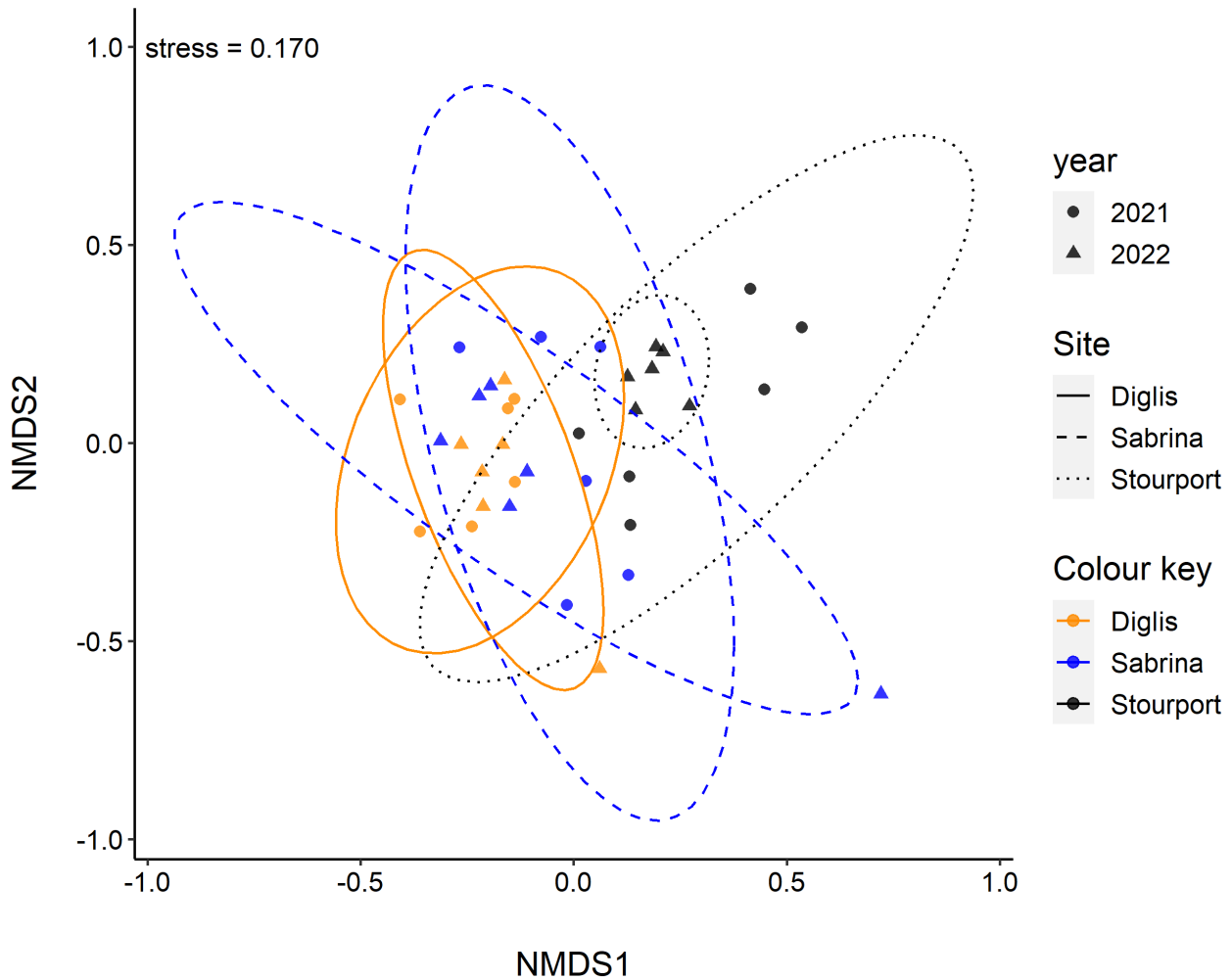
5 Discussion

5.1 Data quality and interpretation

Screening of different types of control samples including field blanks, extraction blanks, PCR positives and negatives showed no evidence of contamination in the vertebrate data. The only vertebrate reads detected across all process blanks were low-level human detections in one extraction blank (Figure S1). Therefore, when discussing and visualising the 12S fish data from 2022 all species >3 reads were considered positive. It is possible for low-level spurious reads to occur from environmental contamination, however in this instance we are confident that any fish DNA reported was present at site. Likewise, invertebrate process blanks were also mostly contamination free, with the exception of our positive control species (*Eudasyphora cyanicolor*) and *Baetis rhodani* presenting low-level reads (Figure S2). Nonetheless, positive control species are not considered in downstream analysis and the low-level detection of *B. rhodani* is not sufficient to influence our relative abundance metrics here. All invertebrate assignments from the 2022 data with >3 reads were therefore considered positive when discussing and visualising this data. All samples from 2022 had >500 reads assigned, and therefore were all considered of sufficient quality.

When comparing 2022 data between sampling sites, sampling events, or sampling years, a low frequency reads threshold of 0.1% was applied to all fish data and any invertebrate assignments <20 reads omitted before further analysis. Such threshold approaches have proven effective at eliminating low-level noise from eDNA metabarcoding data (Hänfling et al. 2016), and was applied in the previous report (Sellers et al. 2022, unpublished report to NE).

5.2 Fish community data

The 25 fish species detected in this survey in the River Severn were largely the same as those detected in the 2021 eDNA metabarcoding survey (Sellers et al. 2022, unpublished report to NE). Only three species which represented rare detections in 2021 were not detected in 2022, rainbow trout (*O. mykiss*), crucian carp (*C. carassius*) and grayling (*T. thymallus*). This could be due to sampling stochasticity as the detection probability for rare species is lower than for common species (Sellers et al. 2020) or reflect differences in the spatio-temporal sampling design of the two surveys. For example, the 2021 study included 15 sites across almost the entire length of the River Severn whereas this survey included only five sites in the middle section of the river, only three of which matched the previous survey sites. Furthermore, the 2021 survey included samples detected in both summer and autumn whereas the samples for the 2022 survey were collected during two sampling events in late spring with very similar environmental conditions. Previous studies have shown seasonal differences in detection probability of fish species in lakes due to differences in temperature and eDNA dispersal among seasons (Lawson-Handley et al. 2019) and highly variable water flow rates in rivers are likely to contribute further to such

seasonal differences (Sellers et al. 2022, unpublished report to NE). For example, during the 2021 survey grayling was only detected in the middle section during autumn. Additionally, previous studies have observed seasonal peaks in detectability as a result of species-specific ecology and lifecycle traits (i.e. spawning activity) (Di Muri et al. 2022).

Despite the lower spatio-temporal coverage in 2022 three additional species were detected, shad (*Alosa* spp*.*), sea lamprey (*P. marinus*) and lampreys of the genus *Lampetra* (brook and/or river lamprey). The absence of shad in the 2021 data can be explained by a mistiming of the survey in relation to the shad spawning season. The associated migration of shad and presence in freshwater is likely to end in July and the 2021 summer sampling was carried out in late July. The absence of the other two taxa in 2021 could also be due to stochasticity as both represented rare detections in 2022. However, interestingly at least two of the three taxa, shad and sea lamprey have a migratory lifestyle, and *Lampetra* includes both migratory and non-migratory forms. Therefore, these data could indicate a first sign of an improvement of the connectivity of the River Severn for migratory fish, but further monitoring is required to confirm this trend.

Shad DNA was detected above the Lincomb weir during the 2022 shad spawning run indicating that the shad are accessing the reconnected habitat in their first spawning event following reconnection. Detection of shad DNA was limited to sites 1 and 2 at the start of the shad spawning run, but extended to all sites by the end of the spawning run. This progressive use of spawning ground by the shad has also been documented for the River Teme (which is a tributary of the River Severn) (Antognazza et al 2021b). eDNA based detection of shad is a powerful tool to monitor the use of the reconnected habitat by this cryptic and endangered species.

Furthermore, count data from the Diglis fish pass viewing window indicated that by the first eDNA sampling event (9th May 2022), 67 shad had been seen passing upstream of the fish pass at Diglis (Figure S1, Appendix 1), and by the second sampling event (31st May 2022) this had increased to 708 shad (Figure 13). Shad tracking data also indicated that shad spend an average of 13.7 days upstream of Diglis up to a maximum of 32.2 days (Environment Agency, unpublished), resulting in high quantities of DNA in the river. This is a rare instance in eDNA sampling where count data is available for the target species and provides confidence in the results of this eDNA based sampling method. The fish pass at Diglis is monitored using 24 hour camera footage and motion detection and there is minimal leakage of fish travelling upstream that are not detected in the viewing window. It should be noted however that read counts do not correlate directly to abundance of shad, but give an indication of the quantity of shad DNA in the habitat at the time of sampling.
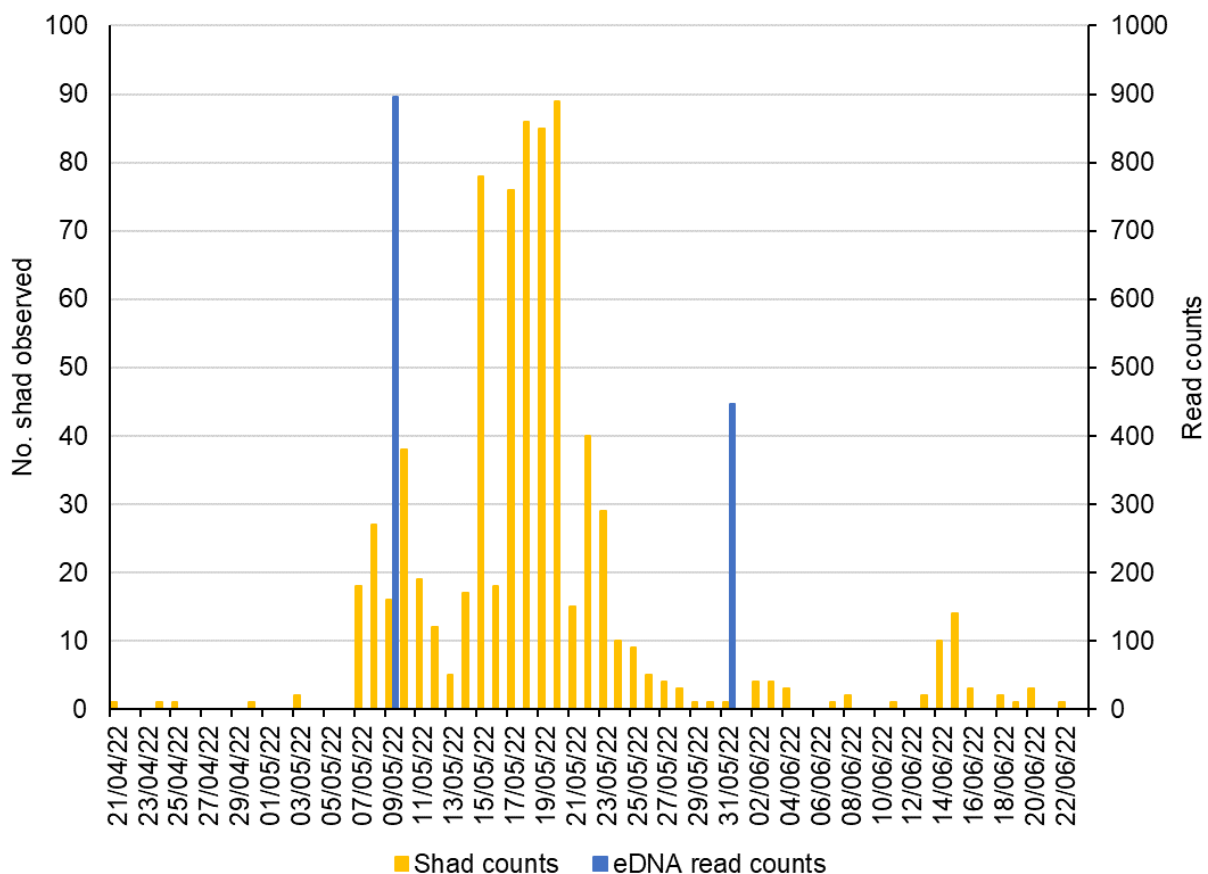
Figure 13. Shad count data and eDNA read counts at the Diglis Fish pass. Shad count data was captured using 24 hour video footage and automated motion detection from the Diglis viewing window from the first spring tide on 16th April 2022. eDNA samples were collected on the 9th and 31st May 2022. Data source: Environment Agency, unpublished.

The overall fish eDNA community showed small but significant changes when comparing the 2021 and 2022 surveys. As discussed above there was little difference in the species detected and therefore this differentiation is attributable to small changes in frequencies of individual species. However, given the considerable difference in sampling design across the two surveys it is difficult to draw biologically meaningful conclusions from these data.

No shad eDNA was discovered in the River Thames despite anecdotal evidence of the species being present there (Louise Whatley pers. comm.). Given the high detection probability of eDNA metabarcoding for fish in lotic systems (e.g. Pont et al 2018, Griffiths et al 2020) it is unlikely that shad eDNA was present at the two sites surveyed here. However, it is possible that the limited spatio-temporal coverage of the sampling was simply not suitable to capture the eDNA signal of this migratory species.

5.3 Invertebrate community data

The 2022 survey detected 86 invertebrate species, spanning 16 taxonomic orders in the River Severn and 40 species spanning 15 Taxonomic orders in the River Thames. As discussed in Sellers et al. (2022, unpublished report to NE) these primers were specifically

designed to be effective for macroinvertebrate taxa which are important in conventional approaches to biological monitoring of river health such as RIVPACS, (e.g., ephemeroptera, plecoptera, trichoptera) and other ecologically important groups such as diptera and tubificida (Leese et al 2021). The original publication provides a test of these primers in a limited geographical context in a single region in Germany. However, this survey and the surveys of 2021 (Sellers et al. 2022, unpublished report to NE) confirm that the primers are also effective for that purpose in UK rivers.

5.4 Future recommendations

The results of previous surveys in the River Severn support the usefulness of eDNA metabarcoding in monitoring changes in fish communities resulting for habitat restoration, and potentially more broadly to assess water quality through surveying invertebrate communities. However, the application is currently limited by the lack of a consistent survey programme across years and the difficulty to incorporate the effect of eDNA dispersal on spatial variation. Furthermore, the meaningful application of invertebrate metabarcoding requires further development of eDNA specific metrics. In order to maximise the benefit of these approaches we therefore recommend:

1) The establishment of long-term monitoring programmes with a consistent spatio-temporal sampling scheme across years of survey.

2) Investment in R&D to incorporate hydraulic modelling approaches into lotic eDNA surveys to allow spatially more accurate predictions of species distributions.

3) A detailed review of the existing invertebrate eDNA data and the development of a strategy on how these can be used for future monitoring of ecological status.

References

Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K (2018) Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in ecology and evolution / British Ecological Society 9:134–147.

Antognazza CM, Britton JR, Potter C, Franklin E, Hardouin EA, Gutmann Roberts C, Aprahamian M, Andreou D (2019) Environmental DNA as a non-invasive sampling tool to detect the spawning distribution of European anadromous shads (Alosa spp.). Aquatic conservation: marine and freshwater ecosystems 29:148–152.

Antognazza CM, Britton RJ, Read DS, Goodall T, Mantzouratou A, De Santis V, Davies P, Aprahamian M, Franklin E, Hardouin EA, Andreou D (2021) Application of eDNA metabarcoding in a fragmented lowland river: Spatial and methodological comparison of fish species composition. Environmental DNA 3:458–471.

Antognazza CM, Britton JR, De Santis V, Kolia K, Turunen OA, Davies P, Allen L, Hardouin EA, Crundwell C and Andreou D, 2021. Environmental DNA reveals the temporal and spatial extent of spawning migrations of European shad in a highly fragmented river basin. Aquatic Conservation: Marine and Freshwater Ecosystems, 31:2029-2040.

Blackman RC, Benucci M, Donnelly R, Hänfling B, Harper LR, Kimbell H, Sellers GS, Sheard AM, Watson HV, Lawson-Handley L (2018) *Targeting the invaders – targeted detection of four priority freshwater invasive non-native species using environmental DNA*. PeerJ Preprints. doi: 10.7287/peerj.preprints.27284v1

Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890.

Deakin CT, Deakin JJ, Ginn SL, Young P, Humphreys D, Suter CM, Alexander IE, Hallwirth CV (2014) Impact of next-generation sequencing error on analysis of barcoded plasmid libraries of known complexity and sequence. Nucleic acids research 42:e129.

Deiner K, Fronhofer EA, Mächler E, Walser J-C, Altermatt F (2016) Environmental DNA reveals that rivers are conveyer belts of biodiversity information. Nature communications 7:12544.

Di Muri, C., Lawson Handley, L., Bean, C. W., Benucci, M., Harper, L. R., James, B., Li, J., Winfield, I. J., & Hänfling, B. (2022). Spatio-temporal monitoring of lake fish spawning activity using environmental DNA metabarcoding. Environmental DNA, 00, 1–12.

Di Muri C, Lawson Handley L, Bean CW, Li J, Peirson G, Sellers GS, Walsh K, Watson HV, Winfield IJ, Hänfling B (2020) Read counts from environmental DNA (eDNA) metabarcoding reflect fish abundance and biomass in drained ponds. Metabarcoding and Metagenomics 4:e56959.

eDNA Society of Japan (2019) Environmental DNA Sampling and Experiment Manual Version 2.1 Available from http://ednasociety.org/

Gleason JE, Elbrecht V, Braukmann TWA, Hanner RH, Cottenie K (2021) Assessment of stream macroinvertebrate communities with eDNA is not congruent with tissue-based metabarcoding. Molecular ecology 30:3239–3251.

Griffiths NP, Bolland JD, Wright RM, Murphy LA, Donnelly RK, Watson HV, Hänfling B (2020) Environmental DNA metabarcoding provides enhanced detection of the European eel Anguilla anguilla and fish community structure in pumped river catchments. Journal of fish biology 97:1375–1384.

Hänfling B, Lawson Handley L, Read DS, Hahn C, Li J, Nichols P, Blackman RC, Oliver A, Winfield IJ (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular ecology 25:3101–3119.

Harper LR, Handley LL, Hahn C, Boonham N (2018) Ground-truthing environmental DNA (eDNA) metabarcoding for ecological hypothesis testing at the pondscape. bioRxiv.

Harper LR, Lawson Handley L, Sayer CD, Read DS, Benucci M, Blackman RC, Hill MJ, Hänfling B (2021) Assessing the impact of the threatened crucian carp (Carassius carassius) on pond invertebrate diversity: A comparison of conventional and molecular tools. Molecular ecology 30:3252–3269.

Hering D, Borja A, Jones JI, 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, Kelly M (2018) Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water research 138:192–205.

Hleap JS, Littlefair JE, Steinke D, Hebert PDN, Cristescu ME (2021) Assessment of current taxonomic assignment strategies for metabarcoding eukaryotes. Molecular ecology resources 21:2190–2203.

Kitson JJN, Hahn C, Sands RJ, Straw NA, Evans DM, Lunt DH (2019) Detecting host-parasitoid interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding. Molecular ecology 28:471–483.

Köster J, Rahmann S (2012) Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 28:2520–2522.

Lawson Handley L, Read DS, Winfield IJ, Kimbell H, Johnson H, Li J, Hahn C, Blackman R, Wilcox R, Donnelly R, Others (2019) Temporal and spatial variation in distribution of fish environmental DNA in England's largest lake. Environmental DNA 1:26–39.

Leese F, Sander M, Buchner D, Elbrecht V, Haase P, Zizka VMA (2021) Improved freshwater macroinvertebrate detection from environmental DNA through minimized nontarget amplification. Environmental DNA 3:261–276.

Li J, Hatton-Ellis TW, Lawson Handley L, Kimbell HS, Benucci M, Peirson G, Hänfling B (2019) Ground-truthing of a fish-based environmental DNA metabarcoding method for assessing the quality of lakes. The Journal of applied ecology 56:1232–1244.

Magilligan FJ, Nislow KH, Kynard BE, Hackman AM (2016) Immediate changes in stream channel geomorphology, aquatic habitat, and fish assemblages following dam removal in a small upland catchment. Geomorphology 252:158–170.

Merkes CM, McCalla SG, Jensen NR, Gaikowski MP, Amberg JJ (2014) Persistence of DNA in carcasses, slime and avian feces may affect interpretation of environmental DNA data. PloS one 9:e113346.

Miya M, Gotoh RO, Sado T (2020) MiFish metabarcoding: a high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. Fisheries science: FS 86:939–970.

Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, Kondoh M, 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 2:150088.

Pont D, Valentini A, Rocle M, Maire A, Delaigue O, Jean P, Dejean T (2021) The future of fish-based ecological assessment of European rivers: from traditional EU Water Framework Directive compliant methods to eDNA metabarcoding-based approaches. Journal of fish biology 98:354–366.

R Core Team (2021) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

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

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome research 22:939–946.

Sellers G, Andreou D, Britton R, Hänfling B (2022) eDNA methods for monitoring restoration effects in the River Severn. Unpublished report to Natural England

Schnell IB, Bohmann K, Gilbert MTP (2015) Tag jumps illuminated--reducing sequence-to-sample misidentifications in metabarcoding studies. Molecular ecology resources 15:1289–1303.

Willby N, Law A, Bull C, Hänfling B, Lawson Handley L, Winfield IJ (2019) A tool for classifying the ecological status of lake fish in Britain based on eDNA metabarcoding. *Report to the Scottish Environment Protection Agency, UK*.

van Puijenbroek PJ, Buijse AD, Kraak MH and Verdonschot PF (2019) Species and river specific effects of river fragmentation on European anadromous fish species. River Research and Applications. 35:68-77.

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. Journal of computational biology: a journal of computational molecular cell biology 7:203–214.

Appendices

Appendix 1

Figure S1. Contamination heatmap visualising the number of fish reads assigned to process blanks by species. Here, all species detected in the EB (extraction blanks), NC (PCR negative controls) and Pos (PCR positive controls) are visualised.

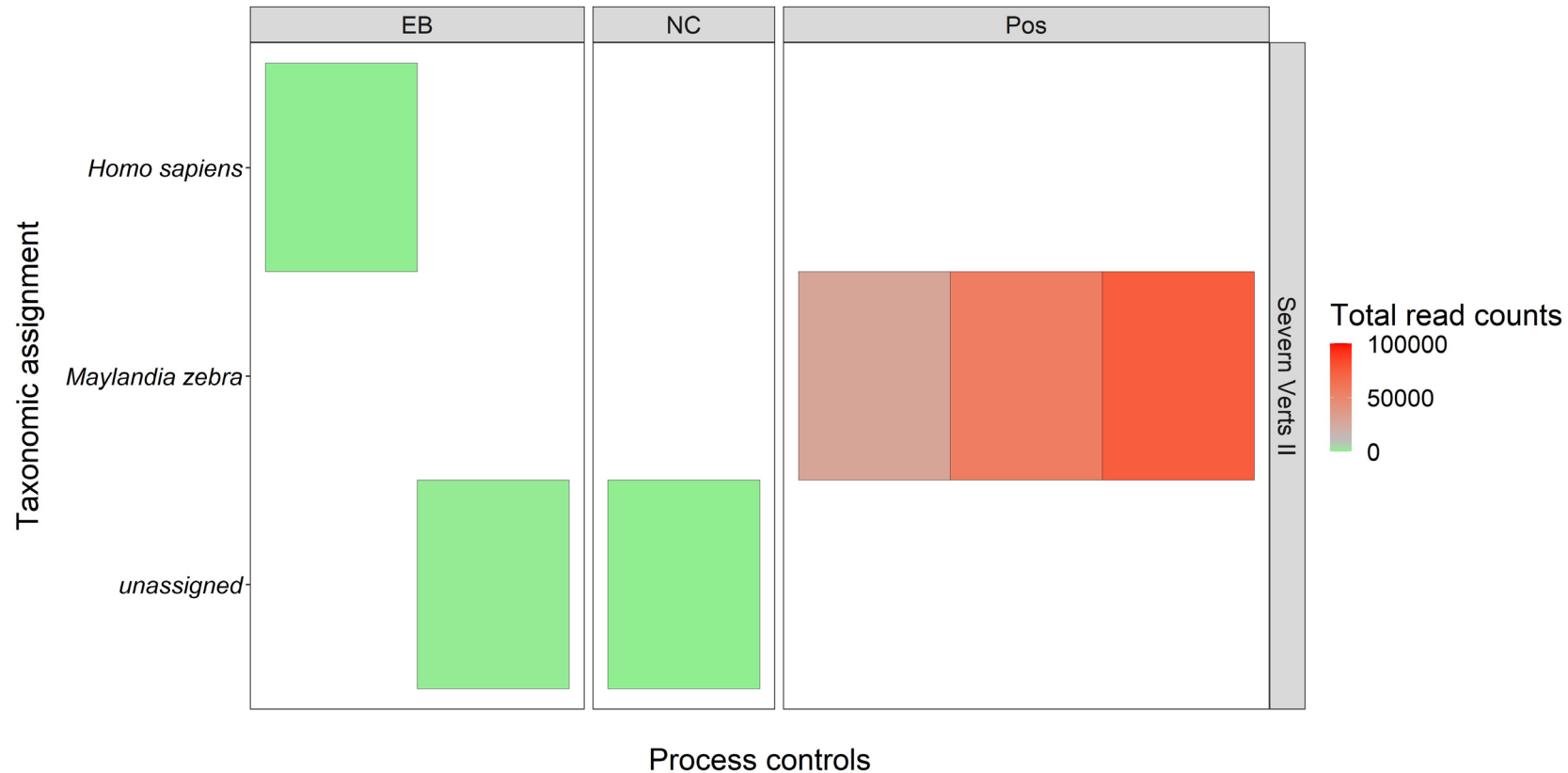eDNA monitoring of restoration effects on the River Severn (Phase II)
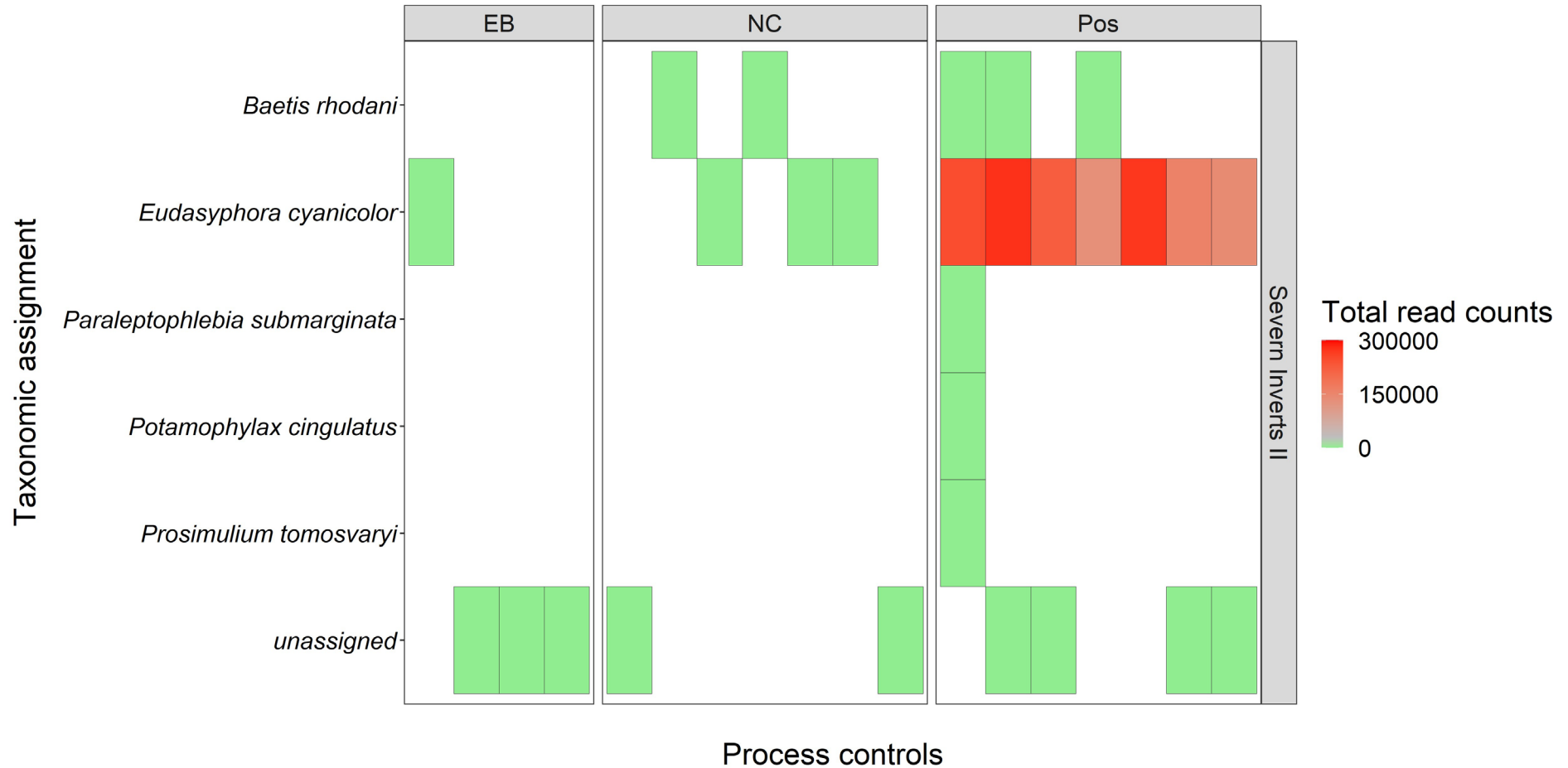
Figure S2. Contamination heatmap visualising the number of invertebrate reads assigned to process blanks by species. Here, all species detected in the EB (extraction blanks), NC (PCR negative controls) and Pos (PCR positive controls) are visualised.

Figure S3. Non-metric multidimensional scaling (NMDS) of fish Bray-Curtis distances among individual samples within the River Severn over different sampling years. The ellipses indicate the estimated multidimensional space of individual groups. Larger ellipses indicate more community dissimilarity among samples.

Appendix 2 - Detailed materials and methods

A. Sample collection

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

Sampling protocol

GLOVES must be worn at all times of sampling

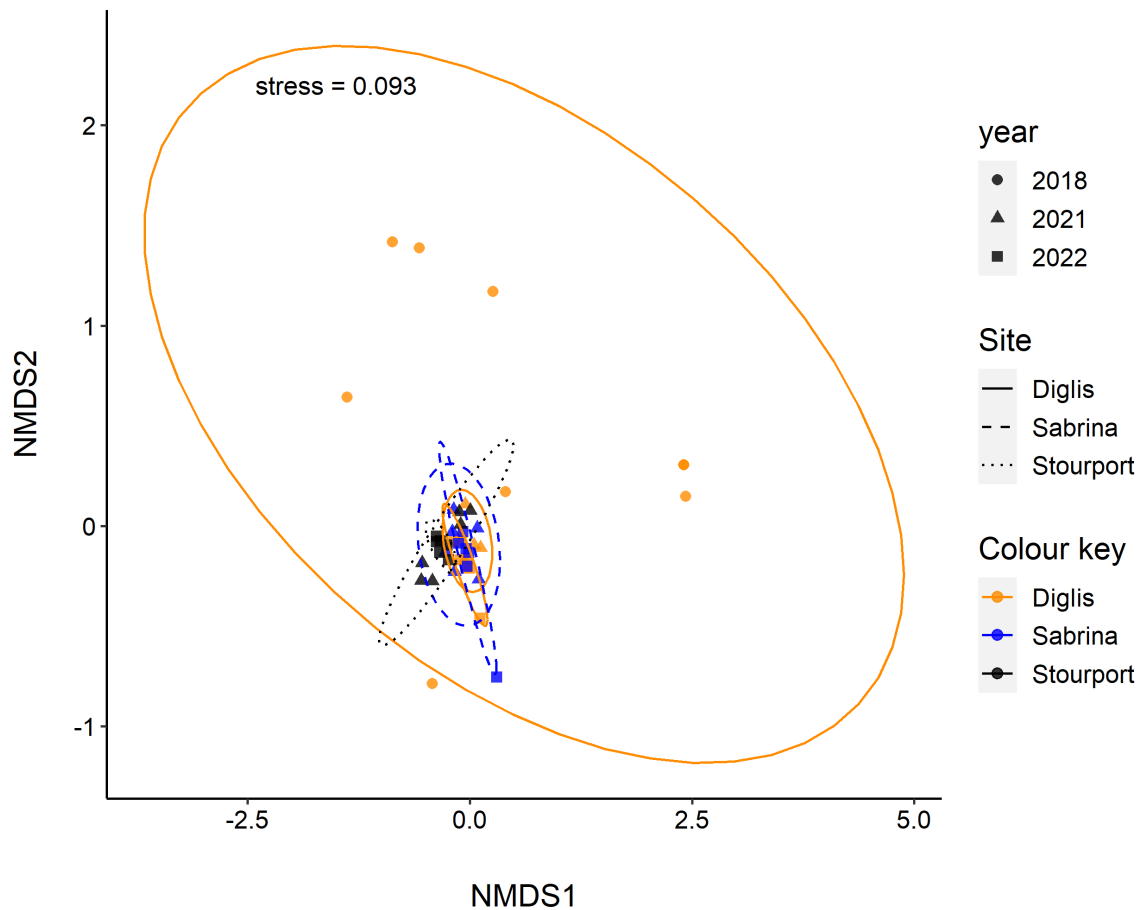**You will be working with bleach - wear protective eye gear and take care of your clothes.**

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

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

Step 4: Place filled whirl-pak bag in white bucket (the bag can stand). Carefully remove the cable tie (the cable ties are releasable). Ensure that the sampling unit is held high as the water will spill if it is not. Place sampler in clean white bucket to prevent contamination.	
Step 5: Use the syringe to draw up water	
Step 6: Attach syringe to Sterivex filter unit Take care NOT to overtighten. Perform pressure filtration until <u>all</u> the sampled water (approximately 1 l) has been filtered. If the Sterivex unit is blocked and you cannot filter the 1 l, then please note the approximate	
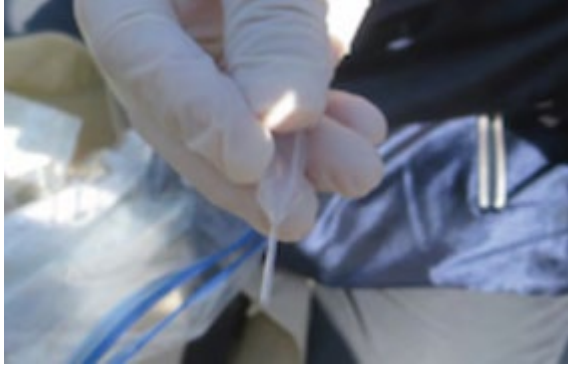
volume filtered on the table provided.	
Step 7: When the filtration is finished, remove the Sterivex filter unit and fill the syringe with air. Reattach the Sterivex filter unit to the syringe, and push out the residual moisture from the filter unit. Repeat this procedure several times until no water comes out of the filter unit.	
Step 8: Seal the outlet post of the Sterivex filter unit with parafilm, while the Sterivex filter unit remains attached to the syringe.	
Step 9: Pipette RNAlater from the microcentrifuge tube using a disposable pipette.	

Step 10: Inject the RNAlater to the Sterivex filter unit from the inlet port using the disposable pipette.	
Step 11: Seal the inlet port with parafilm.	
Step 12: Wipe the surface of the Sterivex filter unit using paper towel and write the necessary information, using a felt pen: Site Sample number Date Place in the prelabelled plastic bag (the one which contained the RNAlater, disposable pipette and parafilm) Place all individual Sterivex filter units in the Site sample bag and store immediately in cool box/bag.	

*Steps 8-12 are from the Japanese eDNA society; Environmental DNA Sampling and Experiment Manual Version 2.1 (published April 25, 2019)
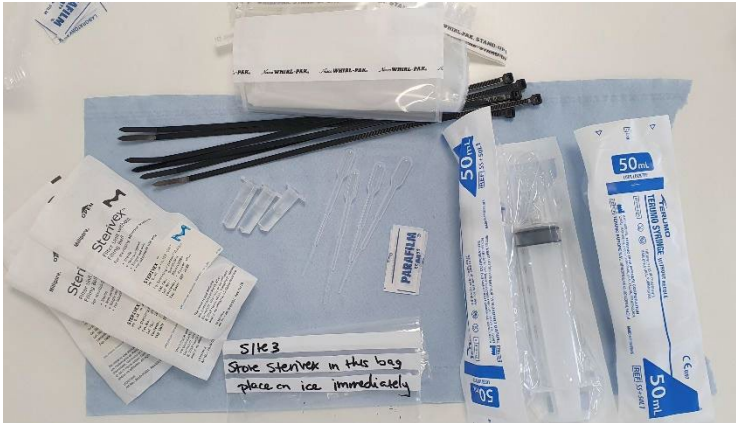
Disinfecting the water sampler and chain between sites.
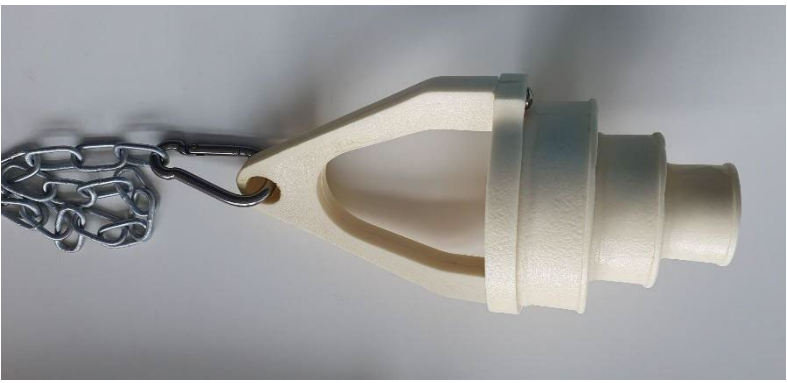
Place the sampler and chain in a white bucket. Prepare enough 10 % bleach solution to fully immerse them. Soak for a minimum of 10 minutes. You can leave to soak whilst travelling to the next site (a lid has been provided to prevent bleach solution from splashing).

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

GLOSSARY

Water sampler	
Syringe	
Sterivex filter unit, pore size 0.45µm	

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

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

B. eDNA capture and extraction

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

DNA extraction from a Sterivex filter cartridge

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

C. DNA sample quantification and purity

To assess DNA quantity and purity, 2µl aliquots of each DNA sample were analysed on a Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific). Sample purity curves (relative to A260 wavelength absorption) were visually compared to that of pure DNA to ascertain overall sample quality using a custom script in *R* (R Core Team 2021).

D. MiFish primers

The MiFish primers used, detailed in Miya et al. (2020), include 2 additional variations (MiFish-E and MiFish-U2) and the MiFish primers (MiFish-U) of Miya et al. (2015). Primers were combined in a 2:1:1 ratio of MiFish-U-F/R:MiFish-E-F/R:MiFish-U2-F/R. Primer sequences are given below.

MiFish-U-F   GTCGGTAAAACTCGTGCCAGC

MiFish-U-R   CATAGTGGGGTATCTAATCCCAGTTTG

MiFish-E-F   GTTGGTAAATCTCGTGCCAGC

MiFish-E-R   CATAGTGGGGTATCTAATCCTAGTTTG

MiFish-U2-F GCCGGTAAAACTCGTGCCAGC

MiFish-U2-R CATAGGAGGGTGTCTAATCCCCGTTTG

PCR amplification with an annealing temperature of 65°C using Q5 High-Fidelity DNA polymerase (New England Biolabs) greatly reduced non-specific amplification of a larger 255 bp fragment of bacterial ribosomal RNA abundant at lower annealing temperatures. This removed the need for gel extraction purification of the sequencing product for the River Severn samples and further reduced contamination risks. This is not the case for all water bodies and is dependent on the bacterial load in the water and, potentially, the DNA extraction method. Miya et al. do stipulate the need for gel extractions even with an annealing temperature of 65°C.

E. Fish eDNA metabarcoding library preparation

Dedicated rooms were available for pre-PCR and post-PCR processes. Pre-PCR processes were performed in the UHI eDNA facility, which has separate rooms for filtration, DNA extraction, and PCR preparation of sensitive environmental samples. PCR reactions were set up in an ultraviolet (UV) and bleach sterilised laminar flow hood. Post-PCR processes were performed in the UHI Inverness laboratory, which has rooms dedicated to pre-PCR of nonsensitive samples, PCR, agarose gel electrophoresis, PCR product purification, library quality control (Qubit, Agilent 2200 TapeStation, real-time quantitative PCR (qPCR)), and sequencing (Illumina MiSeq).

Libraries were prepared for sequencing using a nested metabarcoding workflow with a two step PCR protocol, where Multiplex Identification (MID) tags (unique 8-nucleotide sequences) were included in the first and second PCR for sample identification (Kitson et al. 2019). DNA extracts were PCR-amplified using fish-specific primers that target a 170 bp fragment of the mitochondrial 12S ribosomal RNA (rRNA) region in fish (Miya et al. 2020). The primers were modified for the present study to include MID tags, heterogeneity spacers, sequencing primers, and pre-adapters. There were 24 unique MID tags for the forward and 24 unique MID tags for the reverse primers. This allowed 24 samples (i.e. 22 samples plus controls) to each be labelled with a unique forward and a unique reverse primer to reduce barcode misassignment and tag jumps (Deakin et al. 2014, Schnell et al. 2015). During the first PCR, samples were processed in batches of up to 22 eDNA samples, one negative control and one positive control, keeping sampling replicates in the same batch where possible. The PCR positive control was zebra mbuna (*Maylandia zebra*) DNA (0.05 ng/µl). *M. zebra* is an exotic cichlid which is not found in UK freshwater habitats.

The first PCR was performed in triplicate for each sample/control to combat stochasticity arising from low target DNA concentrations. PCR replicates for each sample/control had the same tag combination. Eight-strip PCR tubes with individually attached lids were used for PCR reactions. PCR reactions were performed in 25 µl volumes, consisting of: 12.5 µl of Q5 High-Fidelity 2x Master Mix (New England Biolabs), 0.5 µl of Thermo Scientific Bovine Serum Albumin (Fisher Scientific UK Ltd.), 7 µl of MGW (Fisher Scientific UK Ltd.), 1.5 µl of each 10 µM tagged primer, and 2 µl of template DNA. PCR reactions were sealed with mineral oil (Sigma-Aldrich) droplets. PCR was performed on an Applied Biosystems Veriti Thermal Cycler (Life Technologies) with the following thermocycling profile: 95°C for 3 mins, 35 cycles of 98°C for 20 s, 65°C for 15 s, and a final elongation of 72°C for 5 mins then held at 4°C.

PCR products were stored at 4°C until PCR technical replicates for each sample/control were pooled. 2 µl of each pooled PCR product was visualised on 2% agarose gels. PCR products were deemed positive where there was amplification at the expected size (~310 bp) on the gel. PCR products were stored at -20°C until they were pooled according to lake and band strength (no/very faint band = 20 µl, faint band = 15 µl, bright band = 10 µl, very bright band = 5 µl) on gel (Alberdi et al. 2018) to create sub-libraries for a size selection bead purification protocol (Rohland & Reich 2012). A ratio of 0.9x bead solution to 100 µl of each sub-library was used for purification. Eluted DNA (25 µl) was stored at 4°C until second PCR amplification.

Inhibited samples, those that failed to amplify and had no primer dimer or exhibited extremely weak bands for both, were diluted 1:10 with TE buffer. These dilutions were processed as a separate and additional sub-library (following the above). The original inhibited counterparts were still processed as part of their initial sub-library.

The second PCR bound pre-adapters, MID tags, and Illumina adapters to the purified sub-libraries. 10 unique forward and reverse MID tag combinations were selected and applied to 6 sub-libraries. Two replicates were performed for each sub-library in 50 μl volumes, consisting of: 25 μl of Q5 High-Fidelity 2x Master Mix (New England Biolabs), 13 μl of MGW (Fisher Scientific UK Ltd.), 3 μl of each 10 μM tagged primer (final concentration 0.6 μM; Integrated DNA Technologies), and 4 μl of template DNA. PCR was performed on an Applied Biosystems Veriti Thermal Cycler (Life Technologies) with the following thermocycling profile: 95°C for 3 mins, 10 cycles of 98°C for 20 s, 72°C for 1 min, and a final elongation of 72°C for 5 mins then held at 4°C. PCR duplicates for each sub-library had the same tag combination.

PCR products were stored at 4°C until duplicates for each sub-library were pooled. 2 μl of each pooled PCR product was visualised on 2% agarose gels. PCR products were deemed positive where there was amplification at the expected size (~400 bp) on the gel. Sub-libraries were stored at 4°C until size selection bead purification. A ratio of 0.85x bead solution to 50 μl of each sub-library was used for purification. Eluted DNA (25 μl) was stored at 4°C until normalisation and final purification.

Sub-libraries were quantified on a Qubit 3.0 fluorometer using a dsDNA HS Assay Kit (Invitrogen) and normalised by pooling according to sample size and library concentration. The pooled library was purified using the same ratios, volumes, and protocol as the second PCR purification. Based on the Qubit™ concentration, the library was diluted to 4 nM. The library was checked with an Agilent 2200 TapeStation using High Sensitivity D1000 ScreenTape (Agilent Technologies) to verify a fragment of the expected size (385 bp) remained. The library was then quantified by qPCR using the NEBNext Library Quant Kit for Illumina (New England Biolabs). Based on the qPCR concentration, the library was adjusted to 4 nM and denatured following the Illumina MiSeq library denaturation and dilution guide. The final library was sequenced at 13 pM with 10% PhiX Control on an Illumina MiSeq using 2 x 300 bp V3 chemistry (Illumina).

F. Fish eDNA metabarcoding bioinformatics

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

Raw reads were quality trimmed from the tail with a 5 bp sliding window (qualifying phred score of Q30 and an average window phred score of Q30) using *fastp* (Chen et al. 2018), allowing no more than 40% of the final trimmed read bases to be below Q30. Primers were removed by trimming the

first 21 and 27 bp of forward and reverse reads respectively. Reads were then tail cropped to a maximum length of 170 bp and reads shorter than 90 bp were discarded.

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

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

The final clustered, non-redundant query sequences were then compared against a curated UK vertebrate reference database (Harper et al. 2018) using *BLAST* (Zhang et al. 2000). Taxonomic identity was assigned using a custom majority lowest common ancestor (MLCA) approach based on the top 2% query *BLAST* hit bit-scores, with at least 90% query coverage and a minimum identity of 98%. Of these filtered hits, 80% of unique taxonomic lineages therein had to agree at descending taxonomic rank (domain, phylum, class, order, family, genus, species) for it to be assigned a taxonomic identity. If a query had a single *BLAST* hit it was assigned directly to this taxon only if it met all previous MLCA criteria. Read counts assigned to each taxonomic identity were calculated from query cluster sizes. Lowest taxonomic rank was to species and assignments higher than order were classed as unassigned.

G. Invertebrate primers

The invertebrate primers used are detailed in Leese et al. (2021). Primer sequences are given below.

fwhF2   GGDACWGGWTGAACWGTWTAYCCHCC

EPTDr2n   CAAACAAATARDGGTATTCGDTY

Due to the number of degenerate bases in the primers, standard taq polymerases (e.g. Q5 High-Fidelity DNA polymerase (New England Biolabs)) do not work. Multiplex PCR Master Mix (QIAGEN Multiplex PCR Plus Kit) was used as recommended by Leese et al. (2021).

H. Invertebrate eDNA metabarcoding library preparation

Dedicated rooms were available for pre-PCR and post-PCR processes. Pre-PCR processes were performed in the UHI eDNA facility, which has separate rooms for filtration, DNA extraction, and PCR preparation of sensitive environmental samples. PCR reactions were set up in an ultraviolet (UV) and bleach sterilised laminar flow hood. Post-PCR processes were performed in the UHI Inverness laboratory, which has rooms dedicated to pre-PCR of nonsensitive samples, PCR, agarose gel electrophoresis, PCR product purification, library quality control (Qubit, Agilent 2200 TapeStation, real-time quantitative PCR (qPCR)), and sequencing (Illumina MiSeq).

Libraries were prepared for sequencing using a two step PCR protocol. The first PCR amplified the target region with Multiplex Identification (MID) tags (unique 8-nucleotide sequences) were included in the second PCR for sample identification. DNA extracts were PCR-amplified using primers that target a ~140 bp fragment of the cytochrome c oxidase subunit 1 (COI) mitochondrial region in invertebrates (Leese et al. 2021). The primers were modified for the present study to include heterogeneity spacers, sequencing primers, and pre-adapters. During all stages, samples were processed in the same batches of up to 22 eDNA samples, one negative control and one positive control, keeping sampling replicates in the same batch where possible. The PCR positive control was bluebottle fly (*Calliphora vicina*) genomic DNA (0.05 ng/μl).

The first PCR was performed in triplicate for each sample/control to combat stochasticity arising from low target DNA concentrations. Eight-strip PCR tubes with individually attached lids were used for PCR reactions. PCR reactions were performed in 25 μl volumes, consisting of: 12.5 μl of Multiplex PCR Master Mix (QIAGEN Multiplex PCR Plus Kit), 0.5 μl of Thermo Scientific Bovine Serum Albumin (Fisher Scientific UK Ltd.), 7 μl of MGW (Fisher Scientific UK Ltd.), 1.5 μl of each 10 μM tagged primer, and 2 μl of template DNA. PCR reactions were sealed with mineral oil (Sigma-Aldrich) droplets. PCR was performed on an Applied Biosystems Veriti Thermal Cycler (Life Technologies) with the following thermocycling profile: 10 cycles of 95°C for 30 s, 64°C for 90 s(- 1°C each cycle), 72°C for 30 s. 25 cycles of 95°C for 30 s, 54°C for 90 s, 72°C for 30 s, and a final elongation of 68°C for 10 mins then held at 4°C.

PCR products were stored at 4°C until PCR technical replicates for each sample/control were pooled. 5 μl of each pooled PCR product was visualised on 2% agarose gels. Of all the samples/controls processed, only the positive control was visibly amplified at the expected size (~260 bp) after PCR. All PCR products were individually purified with a size selection bead purification protocol (Rohland & Reich 2012). A ratio of 0.9x bead solution to 50 μl of each sample/control was used for purification. Eluted DNA (15 μl for samples, 50μl for controls) was stored at 4°C until second PCR amplification.

The second PCR bound pre-adapters, MID tags, and Illumina adapters to the purified PCR products. Each sample/control was treated with one of the 384 unique forward and reverse MID tag combinations (16 forward and 24 reverse). Two replicates were performed for each sample/control in 50 μl volumes, consisting of: 25 μl of Q5 High-Fidelity 2x Master Mix (New England Biolabs), 13 μl of MGW (Fisher Scientific UK Ltd.), 3 μl of each 10 μM tagged primer (final concentration 0.6 μM; Integrated DNA Technologies), and 4 μl of template DNA. PCR was performed on an Applied Biosystems Veriti Thermal Cycler (Life Technologies) with the following thermocycling profile: 95°C for 3 mins, 10 cycles of 98°C for 20 s, 72°C for 1 min, and a final elongation of 72°C for 5 mins then held at 4°C. PCR duplicates for each sample/control had the same tag combination.

PCR products were stored at 4°C until PCR technical replicates for each sample/control were pooled. 5 μl of each pooled PCR product was visualised on 2% agarose gels. PCR products were deemed positive where there was amplification at the expected size (~330 bp) on the gel. Each processed batch of PCR products were pooled according band strength (no/very faint band = 20 μl, faint band = 15 μl, bright band = 10 μl, very bright band = 5 μl) on gel (Alberdi et al. 2018) to create sub-libraries for a size selection bead purification protocol (Rohland & Reich 2012). A ratio of 0.9x bead solution to 100 μl of each sub-library was used for purification. Eluted DNA (25 μl) was stored at 4°C until normalisation and final purification.

Sub-libraries were quantified on a Qubit 3.0 fluorometer using a dsDNA HS Assay Kit (Invitrogen) and normalised by pooling according to sample size and library concentration. The pooled library was purified using the same ratios, volumes, and protocol as the second PCR purification. Based on the Qubit™ concentration, the library was diluted to 4 nM. The library was checked with an Agilent 2200 TapeStation using High Sensitivity D1000 ScreenTape (Agilent Technologies) to verify a fragment of the expected size (~330 bp) remained. The library was then quantified by qPCR using the NEBNext Library Quant Kit for Illumina (New England Biolabs). Based on the qPCR concentration, the library was adjusted to 4 nM and denatured following the Illumina MiSeq library denaturation and dilution guide. The final library was sequenced at 13 pM with 10% PhiX Control on an Illumina MiSeq using 2 x 300 bp V3 chemistry (Illumina).

I. UK invertebrate reference database creation

The reference database was created from the curated reference sequences used in Harper et al. (2021) (https://doi.org/10.5281/zenodo.3993125). Harper et al. did not include dipterans in their database due to GenBank records for Diptera missing important record features. After initial tests we discovered errors in assignments, the PCR positive (a fly) was assigned to a Coleoptera species. To remedy these issues, additional records for UK Diptera species were downloaded from The Barcode of Life Data System (BOLD) (https://www.boldsystems.org/index.php/). BOLD

records are more tightly curated than Genbank, and are focused on a few specific regions, in particular, COI. At present, the *Tapirs* workflow uses NCBI taxonomy, so only BOLD reference sequences with an associated NCBI accession number were used for database creation. An NCBI taxonomic id (taxid) was acquired for each reference sequence to create an accession-to-taxid map. The final *BLAST* database was created from the reference sequences using *makeblastdb* (see https://www.ncbi.nlm.nih.gov/books/NBK569841/) and included the accession-to-taxid map (-*taxid_map*) to allow for taxonomic assignment of OTUs for downstream *Tapirs* analysis.

J. Invertebrate eDNA metabarcoding bioinformatics

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

Raw reads were quality trimmed from the tail with a 5 bp sliding window (qualifying phred score of Q30 and an average window phred score of Q30) using *fastp* (Chen et al. 2018), allowing no more than 40% of the final trimmed read bases to be below Q30. Primers were removed by trimming the first 26 and 23 bp of forward and reverse reads respectively. Reads were then tail cropped to a maximum length of 142 bp and reads shorter than 100 bp were discarded.

Sequence read pairs were merged into single reads using *fastp*, provided there was a minimum overlap of 20 bp, no more than 5% mismatches and no more than 5 mismatched bases between pairs. Only forward reads were kept from read pairs that failed to be merged. A final length filter removed any reads longer than 160 bp to ensure sequence lengths approximated the expected fragment size (~142 bp).

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

The final clustered, non-redundant query sequences were then compared against the UK invertebrate reference database using *BLAST (Zhang et al. 2000)*. Taxonomic identity was assigned using a custom majority lowest common ancestor (MLCA) approach based on the top 2% query *BLAST* hit bit-scores, with at least 90% query coverage and a minimum identity of 95%. Of these filtered hits, 80% of unique taxonomic lineages therein had to agree at descending

taxonomic rank (domain, phylum, class, order, family, genus, species) for it to be assigned a taxonomic identity. If a query had a single *BLAST* hit it was assigned directly to this taxon only if it met all previous MLCA criteria. Read counts assigned to each taxonomic identity were calculated from query cluster sizes. Lowest taxonomic rank was to species and assignments higher than order were classed as unassigned.