# Appendix 2: Sample collection and protocols for eDNA extraction and metabarcoding

## 1.1 Sample collection

Twenty 2 L water samples, each comprised of 5 x 400 ml subsamples taken at 5 m intervals, were collected from the study lakes as described in Hänfling et al. (2016). Samples were collected from the shoreline at equidistant locations (access permitting) using sterile Gosselin™ HDPE plastic bottles (Fisher Scientific UK Ltd, UK) and disposable nitrile gloves (STARLAB, UK). A blank (2 L molecular grade water [MGW]) was transported alongside samples in insulated coolboxes with ice packs. Coolboxes and ice packs were sterilised with 10% v/v chlorine-based commercial bleach (Elliott Hygiene Ltd, UK) solution (containing ~3% sodium hypochlorite) before use.

## 1.2 eDNA capture and extraction

Water samples were vacuum-filtered within 24 hrs of collection in a dedicated eDNA facility at the University of Hull (UoH), the UoH mobile laboratory, or EA laboratories. Surfaces and equipment were sterilised before, during, and after set-up in all work areas. Surfaces and vacuum pumps were wiped with 10% bleach solution. Non-electrical equipment was immersed in 10% bleach solution for 10 minutes, followed by 5% v/v MicroSol detergent (Anachem, UK), and rinsed with purified water. Each 2 L water sample was filtered through sterile 0.45 µm mixed cellulose ester membrane filters with pads (47 mm diameter; Whatman, GE Healthcare, UK) using Nalgene™ filtration units. Two filters were used for each sample and 30 mins allowed for water to pass through each filter, totalling one hour of filtration time per sample. Blanks ($n$ = 39) transported alongside samples were filtered during the last round of filtration for each lake. Equipment was sterilised after each round of filtration. Filters were removed from pads using sterile tweezers and placed in sterile 5 ml polypropylene screw-cap tubes (Axygen™, Fisher Scientific UK Ltd, UK) or 50 mm petri dishes (Fisher Scientific UK Ltd, UK). Tubes or petri dishes were labelled, sealed with parafilm (Bemis™, Fisher Scientific UK Ltd, UK), placed in gripseal bags, and stored at -20 °C until DNA extraction. DNA extraction followed the Mu-DNA water protocol (Sellers et al. 2018). Briefly, captured DNA was liberated from the filter membranes via bead-milling lysis. The lysate underwent an inhibitor removal step prior to purification with silica membrane EZ-10 DNA Mini Spin Columns (NBS Biologicals, UK) and final elution (100 µl). An extraction blank ($n$ = 39), consisting only of extraction buffers, was extracted alongside samples for each lake. Aliquots of each DNA extract were taken for measurement with a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific) to assess yield and purity. DNA extracts were then frozen at -20 °C until PCR amplification.

## 1.3 eDNA metabarcoding

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).

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

For the first PCR, we processed lakes individually to minimise cross-contamination risk. Controls were PCR-amplified alongside eDNA samples to screen for sources of potential contamination. Consequently, the first PCR for each lake consisted of 20 eDNA samples (see Table 2 for exceptions), a sampling/filtration blank, an extraction blank, a PCR negative control which substituted MGW (Fisher Scientific UK Ltd, UK) for template DNA, and a PCR positive control where the template DNA was zebra mbuna (*Maylandia zebra*) DNA (0.05 ng/µL). *M. zebra* is an exotic cichlid which is not found in UK freshwater habitats.

DNA extracts were PCR-amplified using vertebrate-specific primers that target a 106 bp fragment of the mitochondrial 12S ribosomal RNA (rRNA) region in fish (Riaz et al. 2011; Kelly et al. 2014). Previous eDNA metabarcoding studies have shown that these primers have a low false negative rate for bony fishes in a marine mesocosm and natural coastal ecosystem (Kelly et al. 2014; Port et al. 2016). We validated these primers *in vitro* for 22 common UK freshwater fishes and *in situ* on three deep lakes in the English Lake District, and demonstrated their suitability for eDNA metabarcoding of UK lake fish communities (Hänfling et al. 2016). The primers were modified for the present study to include MID tags, heterogeneity spacers, sequencing primers, and pre-adapters (Appendix). There were 24 unique MID tags for the forward and 24 unique MID tags for the reverse primers. This allowed 24 samples (i.e. all samples from one lake 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 according to lake (i.e. 20 eDNA samples, one filtration blank, one extraction blank, one no-template control and one known-template control) to minimise cross-contamination risk between lakes.

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 possessed the same tag combination. Eight-strip PCR tubes with individually attached lids were used instead of 96-well plates to further minimise cross-contamination risk between samples (Port et al. 2016). PCR reactions were performed in 25 µl volumes, consisting of: 12.5 µl of Q5® High-Fidelity 2x Master Mix (New England Biolabs® Inc., MA, USA), 0.5 µl of Thermo Scientific™ Bovine Serum Albumin (Fisher Scientific UK Ltd, UK), 7 µl of MGW (Fisher Scientific UK Ltd, UK), 1.5 µL of each 10 µM tagged primer (final concentration 0.75 µM; Integrated DNA Technologies, Belgium), and 2 µl of template DNA. PCR reactions were sealed with mineral oil (Sigma-Aldrich Company Ltd, UK) droplets (Harper et al. 2018). PCR strips were transported in gripseal bags to the EvoHull laboratory, where the positive control DNA was added to reaction tubes. PCR was performed in our PCR room on an Applied Biosystems® Veriti Thermal Cycler (Life Technologies, CA, USA) with the following thermocycling profile: 98 °C for 5 mins, 35 cycles of 98 °C for 10 s, 58 °C for 20 s and 72 °C for 30 s, followed by a final elongation step at 72 °C for 7 mins.

PCR products were stored at 4 °C until PCR technical replicates for each sample/control were pooled and 2 µl of pooled PCR product added to 0.5 µl of 5x DNA Loading

Buffer Blue (Bioline®, UK). PCR product was visualised on 2% agarose gels (1.6 g Bioline® Agarose in 80 ml 1x sodium borate) stained with 10,000X GelRed Nucleic Acid Gel Stain (Cambridge Bioscience, UK), and gels imaged using Image Lab Software (Bio-Rad Laboratories Ltd, UK). A PCR product was deemed positive where there was an amplification band of the expected size (200-300 bp) on the gel. PCR products were stored at -20 °C until they were pooled according to lake and band strength (no/very faint band = 20 µl, faint band = 15 µl, bright band = 10 µl, very bright band = 5 µl) on gel (Alberdi et al. 2018) to create sub-libraries for purification with Mag-BIND® RxnPure Plus magnetic beads (Omega Bio-tek Inc, GA, USA), following the double DNA size selection protocol established by Bronner et al. (2009). Ratios of 0.9x and 0.15x magnetic beads to 100 µl of each sub-library were used. Eluted DNA (25 µl) was stored at 4 °C until second PCR amplification.

The second PCR bound pre-adapters, MID tags, and Illumina adapters to the purified sub-libraries. Sixteen unique MID tags were available for the forward and 24 unique MID tags were available for the reverse primers. This allows up to 40 sub-libraries to 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). In the present study, we selected 10 unique tag combinations to be applied to 39 sub-libraries split across four sequencing runs. 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® Inc., MA, USA), 13 µl of MGW (Fisher Scientific UK Ltd, UK), 3 µl of each 10 µM tagged primer (final concentration 0.6 µM; Integrated DNA Technologies, Belgium), and 4 µl of template DNA. PCR was performed on an Applied Biosystems® Veriti Thermal Cycler (Life Technologies, CA, USA) with the following thermocycling profile: 95 °C for 3 mins, 10 cycles of 98 °C for 20 s and 72 °C for 1 min, followed by a final elongation step at 72 °C for 5 mins. Reactions without primers or template DNA were prepared in the UoH eDNA facility. Strip tubes were transported in gripseal bags to the EvoHull pre-PCR room where primers were added in a UV and bleach sterilised laminar hood. PCR duplicates for each sub-library possessed the same tag combination. DNA from purified sub-libraries was then added in the EvoHull PCR room.

PCR products were stored at 4 °C until duplicates for each sub-library were pooled, and 2 µl of pooled product was added to 0.5 µl of 5x DNA Loading Buffer Blue (Bioline®, UK). PCR products were visualised on 2% agarose gels (1.6 g Bioline® Agarose in 80 ml 1x Sodium borate) stained with 10,000X GelRed Nucleic Acid Gel Stain (Cambridge Bioscience, UK), and gels imaged using Image Lab Software (Bio-Rad Laboratories Ltd, UK). Again, PCR products were deemed positive where there was an amplification band of the expected size (300-400 bp) on the gel. Sub-libraries were stored at 4 °C until purification with Mag-BIND® RxnPure Plus magnetic beads (Omega Bio-tek Inc, GA, USA), following the double DNA size selection protocol established by Bronner et al. (2009). Ratios of 0.7x and 0.15x magnetic beads to 50 µL of each sub-library were used. 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 Qubit™ dsDNA HS Assay Kit (Invitrogen, UK) and normalised by pooling according to sample size and concentration. The pooled library was purified using the same ratios, volumes, and protocol as second PCR purification. Based on the Qubit™ concentration, the library was diluted to 4 nM. The library was checked using an Agilent 2200 TapeStation and High Sensitivity D1000 ScreenTape (Agilent Technologies, CA, USA) to verify secondary product had been removed successfully and a fragment of the expected size (315 bp) remained. The library was then quantified by qPCR using the NEBNext® Library Quant Kit for Illumina® (New England Biolabs® Inc., MA, USA). Based on the qPCR concentration, the library was adjusted to 4 nM and

denatured following the Illumina MiSeq library denaturation and dilution guide. The library was sequenced at 13 pM with 10% PhiX Control (to improve clustering during initial sequencing) on an Illumina MiSeq® using 2 x 300 bp V3 chemistry (Illumina Inc., CA, USA).

### 1.4 Bioinformatics analysis

Raw sequence reads were demultiplexed using a custom Python script then processed using metaBEAT (metaBarcoding and Environmental Analysis Tool) v0.97.11 (https://github.com/HullUni-bioinformatics/metaBEAT). Raw reads were quality trimmed from the read ends (minimum per base phred score Q30) and across sliding windows (window size 5bp; minimum average phred score Q30) using Trimmomatic v0.32 (Bolger, Lohse, & Usadel, 2014). Reads were cropped to a maximum length of 110 bp and reads shorter than 90 bp after quality trimming were discarded. The first 18 bp of remaining reads were also removed to ensure no locus primer remained. Sequence pairs were merged into single high quality reads using FLASH v1.2.11 (Magoč & Salzberg, 2011), provided there was a minimum overlap of 10 bp and no more than 10% mismatch between pairs. Only forward reads were kept for pairs that could not be merged. A final length filter (106 bp ± 20%) was applied to ensure sequences were reflected the expected fragment size (106 bp). Retained sequences were screened for chimeric sequences against our custom reference database for UK fish (Hänfling et al. 2016) using the uchime algorithm (Edgar, Haas, Clemente, Quince, & Knight, 2011), as implemented in vsearch v1.1 (Rognes, Flouri, Nichols, Quince, & Mahé, 2016). Redundant sequences were removed by clustering at 100% identity ('--cluster_fast' option) in vsearch v1.1 (Rognes et al., 2016). Clusters were considered sequencing error and omitted from further processing if they were represented by less than three sequences. Non-redundant sets of query sequences were then compared against our UK fish reference database (Hänfling et al. 2016) using BLAST (Zhang et al. 2000). Putative taxonomic identity was assigned using a lowest common ancestor (LCA) approach based on the top 10% BLAST matches for any query that matched a reference sequence across more than 95% of its length at minimum identity of 100%. We also compared query sequences to our UK vertebrate reference database (Harper et al. 2018) using a minimum BLAST identity of 98%. Unassigned sequences from this comparison were subjected to a separate BLAST search against the complete NCBI nucleotide (nt) database at 98% identity.

### 1.5 Data Analysis

For the purposes of this report, we include only the fish sequence read data in downstream analyses. The data produced by metaBEAT were summarised as the number of sequence reads per fish species in each sample. Downstream analyses were performed in R v.3.4.3 (R Core Team, 2017). The total read counts per sample were calculated and used to calculate the proportional read counts for each fish species. We then applied a false positive sequence threshold of 0.001 (0.1%) to remove taxonomic assignments that may have resulted from contamination during library preparation or sequencing (De Barba et al. 2014; Hänfling et al. 2016; Port et al. 2016). For downstream analyses, the data was condensed to: (a) the number of sequence reads (i.e. read counts) per species in each lake, and (b) the proportion of sampling sites in which a given species was detected (i.e. site occupancy).