



LJMU Research Online

Collins, RA, Bakker, J, Wangensteen, OS, Soto, AZ, Corrigan, L, Sims, DW, Genner, MJ and Mariani, S

Non-specific amplification compromises environmental DNA metabarcoding with COI

<http://researchonline.ljmu.ac.uk/id/eprint/11575/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Collins, RA, Bakker, J, Wangensteen, OS, Soto, AZ, Corrigan, L, Sims, DW, Genner, MJ and Mariani, S (2019) Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution*. ISSN 2041-210X

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

<http://researchonline.ljmu.ac.uk/>



LJMU Research Online

Collins, RA, Bakker, J, Wangensteen, OS, Soto, AZ, Corrigan, L, Sims, DW, Genner, MJ and Mariani, S

Non-specific amplification compromises environmental DNA metabarcoding with COI

<http://researchonline.ljmu.ac.uk/id/eprint/11575/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Collins, RA, Bakker, J, Wangensteen, OS, Soto, AZ, Corrigan, L, Sims, DW, Genner, MJ and Mariani, S Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution*. ISSN 2041-210X

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

<http://researchonline.ljmu.ac.uk/>

1 **Non-specific amplification compromises** 2 **environmental DNA metabarcoding with COI**

3 **Rupert A. Collins*¹, Judith Bakker*^{2,3}, Owen S. Wangensteen^{3,4}, Ana Z. Soto³,**
4 **Laura Corrigan⁵, David W. Sims^{6,7}, Martin J. Genner¹, and Stefano Mariani^{3,8}**

5 ¹**School of Biological Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue,**
6 **Bristol BS8 1TQ, UK**

7 ²**Department of Biological Sciences, Florida International University, 11200 S.W., 8th Street, Miami,**
8 **Florida, 33199, USA**

9 ³**Ecosystems & Environment Research Centre, School of Environment & Life Sciences, University**
10 **of Salford, Salford M5 4WT, UK**

11 ⁴**Norwegian College of Fishery Science, UiT The Arctic University of Norway, N-9037, Tromsø,**
12 **Norway**

13 ⁵**Environment Agency, Tyneside House, Skinnerburn Road, Newcastle upon Tyne NE4 7AR, UK**

14 ⁶**Marine Biological Association of the United Kingdom, The Laboratory, Citadel Hill, Plymouth PL1**
15 **2PB, UK**

16 ⁷**Ocean and Earth Science, University of Southampton, National Oceanography Centre**
17 **Southampton, European Way, Southampton SO14 3ZH, UK**

18 ⁸**School of Natural Sciences & Psychology, Liverpool John Moores University, Liverpool, L3 3AF,**
19 **UK**

20 Corresponding author:

21 Stefano Mariani

22 Email address: s.mariani@salford.ac.uk

*Both authors contributed equally to this work.

23 **ABSTRACT**

24 1. Metabarcoding extra-organismal DNA from environmental samples is now a key technique in aquatic
25 biomonitoring and ecosystem health assessment. However, choice of genetic marker and primer set is a
26 critical consideration when designing experiments, especially so when developing community standards
27 and legislative frameworks. Mitochondrial cytochrome *c* oxidase subunit I (COI), the standard DNA barcode
28 marker for animals, with its extensive reference library, taxonomic discriminatory power, and predictable
29 sequence variation, is the natural choice for many metabarcoding applications such as the bulk sequencing
30 of invertebrates. However, the overall utility of COI for environmental sequencing of targeted taxonomic
31 groups has yet to be fully scrutinised.

32
33 2. Here, by using a case study of marine and freshwater fishes from the British Isles, we quanti-
34 fy the *in silico* performance of twelve mitochondrial primer pairs from COI, cytochrome *b*, 12S and 16S,
35 in terms of reference library coverage, taxonomic discriminatory power, and primer universality. We
36 subsequently test *in vitro* three COI primer pairs and one 12S pair for their specificity, reproducibility, and
37 congruence with independent datasets derived from traditional survey methods at five estuarine and coastal
38 sites in the English Channel and North Sea coast.

39
40 3. Our results show that for aqueous extra-organismal DNA at low template concentrations, both
41 metazoan and fish-targeted COI primers perform poorly in comparison to 12S, exhibiting low levels of
42 reproducibility due to non-specific amplification of prokaryotic and non-target eukaryotic DNAs.

43
44 4. An ideal metabarcode would have an extensive reference library for which custom primer sets
45 can be designed for either broad assessments of biodiversity or taxon specific surveys, but unfortunately,
46 low primer specificity hinders the use of COI, while the paucity of reference sequences is problematic for
47 12S. The latter, however, can be mitigated by expanding the concept of DNA barcodes to include whole
48 mitochondrial genomes generated by genome-skimming existing tissue collections.

49
50 [Keywords: 12S, COI, eDNA, Environmental DNA, metabarcoding, primer design.]

51 INTRODUCTION

52 DNA barcoding and metabarcoding techniques are now established and indispensable tools for the assessment
53 and monitoring of past and present ecosystems (Valentini et al., 2016; Leray and Knowlton, 2015; Thomsen and
54 Willerslev, 2015; Pedersen et al., 2015), and are being increasingly incorporated into policy and management
55 decisions (Kelly et al., 2014b; Mariani et al., 2015; Rees et al., 2014; Hering et al., 2018). A remarkably wide
56 range of biological substrates can now be sequenced to identify presence of a particular species or reconstruct
57 communities, and can include restaurant sushi meals (Vandamme et al., 2016), deep sea sediments (Guardiola
58 et al., 2015), permafrost ice cores (Willerslev et al., 2003), terrestrial insect collections (Ji et al., 2013), animal
59 faeces (Kartzinel et al., 2015) and seawater samples (Thomsen et al., 2012a).

60 The term “DNA metabarcoding” encompasses two distinct methodologies: (i) bulk sample metabarcoding,
61 which is the direct amplification of a concentrated mixture of organisms, from for example, plankton
62 (Clarke et al., 2017), mass arthropod collections (Yu et al., 2012) or gut material (Leray et al., 2013); or (ii)
63 “environmental DNA (eDNA) metabarcoding”, which is indirect amplification via extra-organismal DNA
64 in water, sediments, or soils (Taberlet et al., 2012). This latter methodology involves first isolating and
65 concentrating DNA using filters, rather than homogenising entire organisms or parts of organisms (Macher
66 et al., 2018; Yu et al., 2012; Spens et al., 2017). The detection of macrobial fauna such as vertebrates and
67 insects using aquatic eDNA has been recognised as a highly sensitive survey technique and a key use-case of
68 metabarcoding (Valentini et al., 2016; Rees et al., 2014). However, DNA from environmental samples such as
69 seawater is likely to be degraded (Collins et al., 2018), and also have a significant quantity of co-extracted
70 microbial DNA that may co-amplify with the targeted metazoan DNA molecules (Andújar et al., 2018; Stat
71 et al., 2017).

72 Early eDNA metabarcoding studies targeting fishes used the cytochrome *b* gene (Thomsen et al., 2012b,a;
73 Minamoto et al., 2012), but more recent studies have used the 12S ribosomal rRNA locus (Kelly et al., 2014a;
74 Port et al., 2016; Hänfling et al., 2016; Stoeckle et al., 2017; Ushio et al., 2018; Yamamoto et al., 2017), and
75 also 16S rRNA (Berry et al., 2017; Bylemans et al., 2018; Shaw et al., 2016; Stat et al., 2018; Jeunen et al.,
76 2018). Various regions of 12S have been proposed as metabarcoding markers, including a ca. 63 bp fragment
77 (Valentini et al., 2016), a ca. 106 bp fragment (Riaz et al., 2011; Kelly et al., 2014a), and a ca. 171 bp fragment
78 (Miya et al., 2015). Modified versions of some of these primers have also been published by Taberlet et al.
79 (2018). Ribosomal genes such as 12S and 16S offer the advantage of conserved priming sites (Deagle et al.,
80 2014; Valentini et al., 2016), and amplification across a broad range of fish taxa (Bylemans et al., 2018; Miya
81 et al., 2015). However, taxonomic resolution can be low (Hänfling et al., 2016; Andruszkiewicz et al., 2017;
82 Miya et al., 2015), with relatively short length ribosomal markers being unable to distinguish commercially
83 important species of the cod family Gadidae (Thomsen et al., 2016), for example. A problem for studies
84 using ribosomal markers are the reference libraries, which are usually poorly populated, and often have to
85 be developed for each project on an ad hoc basis (Thomsen et al., 2016; Stoeckle et al., 2017; Miya et al.,
86 2015). Assembling reference libraries for ribosomal genes is further complicated by frequently-used primer
87 sets amplifying different regions, so any two given 12S references from GenBank, for example, may not be
88 homologous.

89 For animals, the primary DNA barcode is the 5′ “Folmer” region of COI, the cytochrome *c* oxidase subunit
90 I gene (Folmer et al., 1994; Hebert et al., 2003). In comparison to ribosomal markers, the advantages of

91 COI are high interspecific variability (Ward, 2009), an extensive reference database (BOLD; Barcode of Life
92 Database; Ratnasingham and Hebert, 2007), and due to the protein-coding constraints of the gene, more
93 straightforward bioinformatic procedures such as alignment and denoising (Andújar et al., 2018). Inside of
94 the 5' Folmer fragment, multiple primer sets have been developed, targeting shorter regions in the 100–400 bp
95 range, which are more suitable than a full length barcode (ca. 658 bp) for analyses of degraded DNA, or for
96 sequencing on short read platforms such as Illumina (Elbrecht and Leese, 2017; Leray et al., 2013; Shokralla
97 et al., 2015). However, due its nucleotide variation, finding conserved priming regions within the Folmer
98 fragment is difficult, and concerns have been raised about the suitability of some COI primers in terms of
99 species-specific primer-template mismatches, which can result in inefficient, biased amplifications that may
100 hinder quantitative analyses (Deagle et al., 2014). Addressing this issue with bias requires incorporating a
101 high degree of degeneracy into COI primers (Leray et al., 2013; Marquina et al., 2019), particularly by the
102 use of multiple inosine sites (Elbrecht and Leese, 2017; Shokralla et al., 2015; Wangenstein et al., 2018).
103 Despite this problem, Andújar et al. (2018) argue that COI should be the standard marker for metabarcoding,
104 and COI markers are increasingly being used for eDNA metabarcoding (Stat et al., 2017; Kelly et al., 2017;
105 Bakker et al., 2017; Macher et al., 2018; Jeunen et al., 2018; Singer et al., 2019). However, studies comparing
106 efficacy markers have done so in a bulk-sample metabarcoding context (Clarke et al., 2017; Elbrecht and
107 Leese, 2017), or have compared only ribosomal markers for vertebrate eDNA applications (Bylemans et al.,
108 2018). Therefore, there lacks a clear assessment of how degenerate COI primers compare to 12S and 16S
109 rRNA when used on low-template-concentration environmental samples, where non-target DNA molecules
110 are found in abundance.

111 Given the importance of marker choice in metabarcoding studies (Alberdi et al., 2018), and the need
112 to thoroughly scrutinise the utility of COI in comparison with the widely used ribosomal markers (Deagle
113 et al., 2014; Andújar et al., 2018), we use a case study of fishes from the British Isles—a well studied and
114 important group in terms of ecosystem health and human food security—to ask the following questions: (i)
115 can COI primer sets be used as eDNA metabarcoding markers appropriate for aquatic vertebrate biodiversity
116 assessment; and (ii) how do they compare to alternative markers including 12S, 16S and cytochrome *b*? We
117 survey a range of published primer sets both *in silico* and *in vitro*, and include a degenerate metazoan COI
118 primer pair as well as novel fish-targeted COI sets with reduced degeneracy. Using *in silico* methods we assess
119 a number of factors: (i) the reference database coverage for the individual fragments, i.e. how many species
120 and individuals of each species are represented in public databases; (ii) the taxonomic discrimination of
121 each fragment, i.e. is each unique DNA sequence unambiguously associated with a single species name; and
122 (iii) the universality of the primer set, i.e. are all species of the target taxonomic group predicted to amplify
123 equally well. Then, we test using a series of water samples taken from locations with corresponding data from
124 traditional fish survey methods, three COI primer sets against a best performing alternative set, as based upon
125 the results of the *in silico* analyses. By PCR amplifying and sequencing these water samples we compare: (i)
126 the specificity of the primer set, i.e. the proportion of the reads that came from the target taxonomic group; (ii)
127 the power of the primer set, i.e. the total species richness estimated; (iii) the reproducibility of the primer set,
128 i.e. are the same species consistently represented in replicate water samples and PCRs; and (iv) the congruence
129 of the primer set, i.e. are the same species detected in the traditional surveys as the eDNA surveys.

130 METHODS

131 *In silico* analyses

132 *Reference library construction*

133 A list of fish species recorded from the marine and freshwater environments of the British Isles was
134 compiled from three sources: (i) the Global Biodiversity Information Facility (<https://www.gbif.org>; *rg-*
135 *bif v1.1.0*; Chamberlain and Boettiger, 2017); (ii) FishBase (<https://www.fishbase.org>); and (iii) the Eu-
136 ropean Water Framework Directive United Kingdom Technical Advisory Group list of transitional fish
137 species (<https://www.wfduk.org/resources/transitional-waters-fish>; Annex 1). These species were then cross-
138 referenced for all synonyms using *rfishbase v3.0.0* (Boettiger et al., 2012). The subsequent list of valid
139 species names and all their synonyms was then searched using *rentrez v1.2.1* (Winter, 2017) against NCBI
140 GenBank release 230 (nucleotide database; <https://www.ncbi.nlm.nih.gov/nucleotide/>) for any of the following
141 terms: “COI, 12S, 16S, rRNA, ribosomal, cytb, COI, cox1, cytochrome, subunit, COB, CYB, mitochondrial,
142 mitochondrion”. The Barcode of Life Database BOLD (<http://www.boldsystems.org/>) was also searched for
143 the same species using *bold v0.8.6* (Chamberlain, 2018).

144 Hidden Markov models of the alignments of each primer set were then constructed using *HMMER*
145 *v3.1b2* (<http://hmmer.org/>; Eddy, 1998) and the fish mitochondrial genome database ([http://mitofish.aori.u-](http://mitofish.aori.u-tokyo.ac.jp/)
146 [tokyo.ac.jp/](http://mitofish.aori.u-tokyo.ac.jp/); Iwasaki et al., 2013). These profiles were used to extract homologous regions of nucleotides
147 from the total mitochondrial data obtained from the GenBank and BOLD searches. The resulting sequences
148 were then annotated with metadata using *traits v0.3.0.9310* (Chamberlain et al., 2018). A phylogenetic quality
149 control step was then carried out by aligning the sequences in *MAFFT v7.271* (Katoh and Standley, 2013)
150 and constructing a maximum likelihood tree using *RAxML v8.2.12* (Stamatakis et al., 2008). Sequences with
151 putatively spurious annotations—i.e. those indicative of misidentifications—were filtered out if the following
152 criteria were met: (i) individual(s) of species *x* being identical to or nested within a cluster of sequences
153 of species *y*, but with other individuals of species *x* forming an independent cluster; and (ii) the putatively
154 spurious sequences coming from a single study, while the putatively correct sequences of species *x* and *y*
155 coming from multiple studies. Records flagged by NCBI as “unverified” were also omitted. The full reference
156 library and code to reproduce it can be found at <https://doi.org/10.6084/m9.figshare.7464521>.

157 *Primer design*

158 We designed two new COI metabarcoding primers targeting fishes (Table 1): “SeaDNA-short” and “SeaDNA-
159 mid”, which share a forward primer, and are internal to the Folmer fragment. The new primer pairs were
160 designed manually in *Geneious v8.8.1* (Kearse et al., 2012) using the same fish mitochondrial genome dataset
161 as described above, with the assistance of *Primer3* (Untergasser et al., 2012) and the sliding window functions
162 in *spider v1.3.0* (Boyer et al., 2012; Brown et al., 2012). The primers were tested on a range of fish tissue
163 extractions from elasmobranchs and actinopterygians, and produced strong clean PCR amplicons of the
164 expected size.

165 *In silico* PCR and taxonomic discrimination

166 Primers were evaluated using a subset of 955 unique sequences from 184 species obtained in the British Isles
167 fish reference library construction step, for which full mitochondrial genomes were available. Twelve primer
168 pairs were chosen for the *in silico* PCRs, representing COI, cytochrome *b*, ribosomal 12S and ribosomal 16S

Table 1. Primer sets assessed in this study. The approximate fragment length is based upon the length of that region in the *Anguilla anguilla* mitochondrial genome (AP007233.1). The asterisks represent the sequences of the Leray-XT primer set that were simplified by changing inosines to double-base ambiguities to allow an *in silico* assessment with *MFEprimer*. The standard DNA barcode marker for fishes (Ward et al., 2005) is presented for reference.

Primer set	Locus	Primer names	Oligonucleotide 5'–3'	Fragment length (bp)	Reference
Leray-XT	COI	mlCOIintF-XT mlCOIintF-XT* jgHCO2198 jgHCO2198*	GGWACWRGWTGRACWITITAYCCYCC GGWACWRGWTGRACWGTYTAYCCYCC TAIACYTCIGGRTGICCRARAAYCA TAKACYTCWGGRTGRCCARAAYCA	313	Wangensteen et al. (2018)
SeaDNA-short		coi.175f coi.226r	GGAGGCTTTGGMAAYTGRT GGGGGAAGAARYCARAARCT	55	This study
SeaDNA-mid		coi.175f coi.345r	GGAGGCTTTGGMAAYTGRT TAGAGGRGGGTARACWGTYCA	130	This study
Ward-barcode		FishF1 FishR1	TCAACCAACCACAAAGACATTGGCAC TAGACTTCTGGGTGGCCAAAGAATCA	655	Ward et al. (2005)
Minamoto-fish	Cytb	L14912-CYB H15149-CYB	TTCCTAGCCATACAYTAYAC GGTGCKCCTCAGAGGACATTTGKCCYCA	235	Minamoto et al. (2012)
MiFish-U	12S	MiFish-U-F MiFish-U-R	GTCGGTAAACTCGTGCCAGC CATAGTGGGGTATCTAATCCAGTTG	171	Miya et al. (2015)
MiFish-E		MiFish-E-F MiFish-E-R	GTTGGTAAATCTCGTGCCAGC CATAGTGGGGTATCTAATCCAGTTG	171	Miya et al. (2015)
Taberlet-tele02		Tele02-f Tele02-r	AAACTCGTGCCAGCCACC GGGTATCTAATCCCAGTTG	167	Taberlet et al. (2018)
Taberlet-elas02		Elas02-f Elas02-r	GTTGGTHAATCTCGTGCCAGC CATAGTAGGGTATCTAATCCAGTTG	171	Taberlet et al. (2018)
Valentini-tele01		L1848 H1913	ACACCGCCCGTCACTCT CTTCCGGTACACTTACCATG	63	Valentini et al. (2016)
Riaz-V5		12S-V5f 12S-V5r	ACTGGGATTAGATACCCC TAGAACAGGCTCCTCTAG	106	Riaz et al. (2011)
Berry-fish	16S	Fish16sF/D 16s2R	GACCCTATGGAGCTTTAGAC CGCTGTTATCCCTADRGTAACT	219	Berry et al. (2017)

169 (Table 1). *MFEprimer* v2.0 (Qu et al., 2012) was used to perform the *in silico* PCR on the untagged primers.
 170 Amplification universality was estimated using the Primer Pair Coverage (PPC) statistic from *MFEprimer*,
 171 where $PPC = \frac{Fm}{Fl} \times \frac{Rm}{Rl} \times (1 - CVfr)$, with Fl and Rl the length of the forward and reverse primers, and $CVfr$
 172 the coefficient of variability of matched lengths Fm and Rm to the template. Therefore, a PPC value of 100%
 173 indicates complete binding of both primers to a template. The highest PPC value was then selected for each
 174 species, and averaged over all species to provide the PPC for each primer set. Predicted non-amplifications
 175 with a default 5 bp 3' binding stability of $> 0\Delta G$ were set to a PPC of 0%. In order for sufficient RAM to
 176 be available to complete the analysis of the highly degenerate Leray-XT primer set, the inosine sites were
 177 simplified to double-base ambiguities. This was achieved by choosing the most frequent base combination
 178 in the mitogenome alignment. None of the altered inosine sites were within 8 bp of the 3' end of the primer
 179 (Table 1).

180 Taxonomic discrimination (= resolution) was assessed first using all available species from the British
 181 Isles fish reference library for each primer set individually, and then secondly on a subset of species for which
 182 sequences were present for all of the primer sets. Discrimination as a proportion of the total number of species
 183 was calculated following Ficetola et al. (2010): “A taxon unambiguously identified by a primer pair owns a
 184 barcode sequence associated to this pair that is not shared by any other taxa”.

185 **Primer evaluation *in vitro***

186 ***Field sites and traditional fish survey***

187 Five locations in the United Kingdom were surveyed for fishes using eDNA and traditional methods between
188 October and November of 2016. These included: the River Tees, County Durham (54.631327,-1.164447);
189 two sites within the River Esk estuary, North Yorkshire (54.491633,-0.611833; 54.48975,-0.612617); the
190 River Test, Hampshire (50.901563,-1.440836); and Whitsand Bay, Devon (50.329616,-4.243751). The former
191 four are estuarine sites, while the latter is an inshore coastal area, approximately 1 km from shore. Fish
192 sampling in the River Esk estuary was done by duplicate fyke nets (Esk-fyke) and duplicate beach-seine
193 nets (Esk-seine), in different locations. At the River Tees sampling site, duplicate beach-seine netting and
194 two shallow beam trawls were carried out. The River Test site comprised a 24 h fish impingement survey
195 conducted at Marchwood Power Station. Whitsand Bay was surveyed by four otter trawls, as described in
196 [McHugh et al. \(2011\)](#). The variety of fishing techniques used in the different sampling locations are part of
197 the currently ongoing fish monitoring programmes implemented by local collaborating organisations: the
198 Environment Agency, PISCES Conservation Ltd. and the Marine Biological Association. Further details are
199 presented in Supplementary Information.

200 ***Water processing and DNA extraction***

201 Three 2 L water sample replicates per site were collected immediately prior to the traditional fish survey
202 commencing, using Nalgene HDPE collection bottles pre-sterilised with a 10% bleach solution. Water was
203 pre-strained with a 250 μm nylon mesh filter to remove debris, if required. After collection, the water samples
204 were put into individual sterile plastic bags, and stored in an ice box while being transported back to the
205 laboratory. Within five hours, each 2 L sample was filtered through an 0.22 μm Sterivex-GP PES filter (Merck
206 Millipore) using a 100 mL polypropylene syringe or a peristaltic pump, and cleared of water. When the full 2
207 L could not be passed due to filter clogging, the volume of water was recorded. After filtration, the filters
208 were stored at -20°C . DNA was extracted from the filters using the DNeasy PowerSoil DNA Isolation Kit
209 (MoBio/Qiagen), following the manufacturers' protocol, with the addition of an initial 2 h agitation step to
210 promote the release of DNA from the filter, during which the filter membranes were placed in tubes with lysis
211 buffer C1 and garnet beads from the PowerWater Isolation kit and shaken at 65°C . Filtration blank controls
212 were processed in parallel. All processing was carried out in dedicated eDNA extraction laboratories, and
213 equipment and surfaces were regularly cleaned using a 10% bleach solution. The eDNA extraction, pre-PCR
214 preparations and post-PCR procedures were carried out in separate rooms.

215 ***PCR and library preparation***

216 Four primer sets were selected to go forward for *in vitro* testing: three COI primer sets (Leray-XT, SeaDNA-
217 short, SeaDNA-mid), and one best-performing primer set from the *in silico* analysis (12S MiFish-U). All
218 PCR amplifications were done in duplicate reactions each with a unique 7/8-mer oligo-tag barcode, differing
219 by at least three bases ([Guardiola et al., 2015](#)). In order to increase variability of the amplicon sequences,
220 a variable number (two, three or four) of fully degenerate positions (Ns) were added at the 5' end of the
221 oligo tags ([Wangensteen et al., 2018](#)). For PCR amplification with the newly designed SeaDNA-short and
222 SeaDNA-mid primers, a two-step protocol was used, first using untagged primers, then tagged primers in
223 a second PCR round. The reaction for the first PCR step included 10 μL AmpliTaq Gold 360 Master Mix

224 (ThermoFisher), with 1 μL of each 5 μM forward and reverse primer, 0.16 μL of bovine serum albumin
225 and 10 ng of purified DNA in a total volume of 20 μL per sample. Thermocycling profile for the first step
226 included an initial denaturation at 95°C for 10 minutes, then 40 cycles of 94°C for 30 sec, 47°C for 45 sec and
227 72°C for 30 sec, and then a final extension of 72°C for 5 minutes. The profile for the second PCR step was
228 identical, except for the annealing temperature being 50°C instead of 47°C. Amplifications were assessed by
229 electrophoresis on a 1.5% agarose gel, and the field and laboratory controls were checked for the presence of
230 amplicons. Between the first and second PCR step, amplicons were purified using MinElute PCR purification
231 columns (QIAGEN) and diluted by a factor of ten prior to being used as a template for the second PCR. After
232 the second PCR, all tagged amplicons were pooled by marker, purified again using MinElute columns and
233 eluted into a total volume of 45 μL , in order to concentrate the amplicons approximately 15 times. For 12S
234 MiFish and Leray-XT we used a one-step procedure with tagged PCR primers, with PCR cycling conditions
235 following [Miya et al. \(2015\)](#) and [Wangensteen et al. \(2018\)](#), respectively. Reagents and volumes were the
236 same as for the two-step protocol.

237 Libraries (one for each primer set) were built using the PCR-free NEXTFlex library preparation kit (BIOO
238 Scientific). The libraries were quantified using the NEBNext qPCR quantification kit (New England Biolabs)
239 and spiked with with 1% PhiX (Illumina). The libraries were sequenced on an Illumina MiSeq platform,
240 using V3 chemistry (2 \times 75 bp paired-end) for the SeaDNA-short library, which was run along with two other
241 libraries from unrelated projects. For the MiFish-U and SeaDNA-mid libraries, V2 chemistry (2 \times 150 bp
242 paired-end) was used, and these were sequenced in the same run. The Leray-XT library was run using V2
243 chemistry (2 \times 250 bp paired-end) along with another library from an unrelated project.

244 **Bioinformatic processing**

245 Raw sequencing data were converted to fastq format using *bc12fastq* v2.20 (<https://support.illumina.com/sequencing/sequencing-conversion-software.html>). The remaining bioinformatic steps were carried out using *cutadapt* v2.3 ([Martin, 2011](#)) and *dada2* v1.10.1 ([Callahan et al., 2016](#)). Because a PCR-free library preparation kit was used,
248 adapters could have been ligated to either the 5' or the 3' end of the amplicon, and in order to take advantage
249 of the Illumina error profiling in the *dada2* denoising step, the sense- and antisense-orientated sequences were
250 first isolated and processed independently. This was achieved using *cutadapt* by filtering the R1 fastq files for
251 reads with the forward PCR primer, and then for those with the reverse PCR primer. The reads were then
252 demultiplexed by tag, followed by primer and adapter trimming. Quality trimming was carried out in *dada2*
253 using default settings, but with read truncation length “truncLen” determined to give an approximate 30 bp
254 overlap between forward and reverse reads. The reads were then denoised, dereplicated, merged, cleaned
255 of chimaeras and reorientated, using the *dada2* workflow. Our reference library sequences for each primer
256 set were used as priors to avoid low abundance but valid sequences being discarded during denoising. A
257 homology filter was then implemented by aligning the ASVs against a hidden Markov model of the expected
258 fragment using *HMMER hmmsearch*, and the non-homologous reads discarded.

259 Taxonomy assignment of the amplicon sequence variants (ASVs) produced by *dada2* was carried out
260 using a multi-step procedure, incorporating distance-based and phylogenetic methods. First, a preformatted
261 “nt” blast database was downloaded from NCBI (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/v5>; 21 March 2019). Each
262 ASV sequence was then locally blasted against this database using *blastn* v2.9.0 (‘-task blastn -evalue 1000
263 -word_size 11 -max_target_seqs 500’), and the results filtered to obtain a rough taxonomic classification based

264 on the best-scoring blast hit. Next, a more stringent procedure was carried out, with the putative fish sequences
265 extracted from this initial blast result subjected to a second *blastn* search, this time using our curated reference
266 library of British Isles fishes as the blast database (same settings as the “nt” search but with ‘-word_size 7’).
267 The same reads were then run through the Evolutionary Placement Algorithm (*EPA-ng* v0.3.5, *gappa* v0.2.0;
268 [Barbera et al., 2018](#); [Czech and Stamatakis, 2018](#)). Species name(s) were assigned based on either of the
269 following rules: (i) species-level EPA placement same as the best scoring blast hit, with an aligned match
270 length of $\geq 90\%$ of the modal length of the fragment, and an identity of $\geq 97\%$; or (ii) highest likelihood EPA
271 placement same as the best scoring blast hit, with an EPA probability $\geq 90\%$ and blast identity $\geq 90\%$. Rule
272 (i) finds assignments that are congruent between both the *EPA-ng* and *blastn* methods, but rejects assignments
273 with low similarity and short match lengths. Rule (ii) allows for dissimilar hits, but only ones that have a
274 high phylogenetic probability, and which are usually indicative of low abundance variants with errors. Our
275 prior knowledge of the expected fish fauna of the sites was used to set these cut-off values, with the aim of
276 conservatively minimising false positive assignments. The fish reads were also summarised by OTU clustering
277 using *Swarm* v2.2.2 ([Mahé et al., 2015](#)), with $d = 1$ and the “fastidious” option enabled. This step permitted
278 an evaluation of possible misassigned and unassigned species.

279 RESULTS

280 *In silico* analyses

281 A total of 531 species were identified as part of the United Kingdom marine and freshwater fish fauna. Of
282 these, 176 names were flagged as “common” species, having been identified as relatively widespread marine or
283 freshwater taxa that are likely to be encountered during survey work of coastal and inland habitats ([Henderson,](#)
284 [2014](#); [Kottelat and Freyhof, 2007](#)). The remainder were mostly highly localised species, deep water offshore
285 species, or rare pelagic migrants. The combined reference library for all primer sets, after cleaning, duplicate
286 removal and quality control, comprised 43,366 sequences from 491 total species, and 25,799 sequences from
287 172 common species.

288 In terms of reference database coverage for individual primer sets ([Table 2](#)), COI primers had the greatest
289 number of reference sequences at 23,911–24,058, covering 91% of species. The “Minamoto-fish” cytochrome
290 *b* set had 15,405 sequences and a species coverage of 65%. Of the ribosomal primer sets, the “Berry-fish”
291 16S set had the greatest number of sequences at 4,089, with species coverage at 77%. Among the 12S
292 sets, the “Riaz-V5” primers had the greatest number of sequences (2,416; species coverage 69%), while
293 the “Valentini-tele01” set had the fewest sequences (1,699; species coverage 51%). The “MiFish” primers
294 and their variants (MiFish-U/E, Taberlet-tele02, Taberlet-elas02) had 1,904 sequences, and a coverage of
295 61%. Per species, the average number of reference sequences was greatest for the COI primer sets (mean
296 49–50; median 24), followed by cytochrome *b* (mean 45; median 7), 16S (mean 9.9; median 4), and then 12S
297 (mean 5.9–6.6; median 2–3). When only the subset of common species was considered, the species coverage
298 increased for all primer sets, as did the average number of sequences per species ([Table 2](#)).

299 In terms of taxonomic discrimination of the fragments obtained from each primer set ([Table 2](#)), the
300 proportion of British Isles fish species where all individuals could be unambiguously identified was greatest
301 for the Leray-XT COI fragment at 95%, while the shorter SeaDNA-mid and SeaDNA-short COI fragments
302 resolved 91% and 87% respectively. The cytochrome *b* fragment discriminated 91%. The MiFish fragment had

Table 2. Statistics for reference library coverage, taxonomic discriminatory power, and primer universality as estimated by *in silico* PCR, for twelve primer sets from COI, cytochrome *b*, 16S and 12S. Library coverage is calculated as the number of species for which at least one sequence was available out of the total ($n = 531$) or common species subset ($n = 176$) of British Isles marine and freshwater fishes (proportion in parentheses). Library sequences per species is the mean (median in parentheses) number of sequences available for each species. Taxonomic discrimination is the proportion of species for which all individuals can be unambiguously identified by a unique DNA sequence, with values in parentheses showing the proportion for the subset of species that are shared over all primer sets ($n = 221$ for all; $n = 88$ for common). Primer universality represents the mean Primer Pair Coverage (PPC) percent statistic from *MFEprimer*, and was calculated using the 184 British Isles fish species for which data were available for all species. The standard DNA barcode marker for fishes (Ward et al., 2005) is presented for reference. The highly degenerate Leray-XT primers were simplified to overcome analytical RAM limitations (see Table 1).

Locus	Primer pair	Species subset	Total number sequences	Library species coverage	Library sequences per species	Fragment taxonomic discrimination	Primer % universality (Actinopterygii)	Primer % universality (Elasmobranchii)	
COI	Leray-XT	All	24,058	481 (0.91)	50 (24)	0.95 (0.96)	27.8	39	
	SeaDNA-mid		24,045	481 (0.91)	50 (24)	0.91 (0.94)	23	22.9	
	SeaDNA-short		23,911	481 (0.91)	49.7 (24)	0.87 (0.9)	34.5	21.5	
	Ward-barcode		23,975	481 (0.91)	49.8 (24)	0.95 (0.97)	6.3	1.2	
CYTB	Minamoto-fish	Common	15,405	344 (0.65)	44.8 (6.5)	0.91 (0.91)	13.5	14.4	
12S	MiFish-U		1,904	322 (0.61)	5.9 (3)	0.93 (0.91)	71.3	2.4	
	Taberlet-tele02		1,904	322 (0.61)	5.9 (3)	0.93 (0.91)	85.3	7.7	
	MiFish-E		1,904	322 (0.61)	5.9 (3)	0.93 (0.91)	0.4	39.3	
	Taberlet-elas02		1,904	322 (0.61)	5.9 (3)	0.93 (0.91)	0.4	68.8	
	Valentini-tele01		1,699	273 (0.51)	6.2 (2)	0.86 (0.85)	68.2	60.4	
	Riaz-V5		2,416	364 (0.69)	6.6 (2)	0.79 (0.78)	92.2	11.2	
	16S		Berry-fish	4,089	411 (0.77)	9.9 (4)	0.89 (0.86)	47.5	0
			COI	Leray-XT	12,698	170 (0.97)	74.7 (38.5)	0.97 (1)	23.3
COI	SeaDNA-mid		12,639	170 (0.97)	74.3 (37.5)	0.93 (1)	17	29	
	SeaDNA-short		12,553	170 (0.97)	73.8 (37.5)	0.93 (1)	32.8	28.9	
	Ward-barcode		12,579	170 (0.97)	74 (37.5)	0.97 (1)	6.3	0	
	CYTB	Minamoto-fish	10,936	143 (0.81)	76.5 (16)	0.94 (1)	13.6	9.1	
12S	MiFish-U	941	109 (0.62)	8.6 (3)	0.94 (0.94)	75.6	0		
	Taberlet-tele02	941	109 (0.62)	8.6 (3)	0.94 (0.94)	89.3	0		
	MiFish-E	941	109 (0.62)	8.6 (3)	0.94 (0.94)	0	52.4		
	Taberlet-elas02	941	109 (0.62)	8.6 (3)	0.94 (0.94)	0	82.3		
	Valentini-tele01	852	99 (0.56)	8.6 (2)	0.93 (0.94)	67.6	60.4		
	Riaz-V5	1,398	143 (0.81)	9.8 (3)	0.85 (0.83)	96.4	0		
	16S	Berry-fish	2,296	167 (0.95)	13.7 (6)	0.87 (0.91)	50.3	0	

303 the greatest discrimination among the ribosomal primer sets at 93%, with the Berry-fish 16S, Valentini-tele01,
 304 and Riaz-V5 pairs having lower rates (89%, 86%, and 79% respectively). When a standardised dataset of
 305 species common to all primer sets ($n = 88$) was used, the overall pattern remained similar (Table 2).

306 In terms of primer universality as estimated by *in silico* PCR for British Isles fish species with comparable
 307 data available for all markers ($n = 184$; Table 2), the 12S primer sets targeting actinopterygians had a higher
 308 mean PPC than all other markers, at between 68.2% (Valentini-tele01) and 92.2% (Riaz-V5), compared to
 309 between 13.5% (cytochrome *b*) and 47.5% (16S). The best performing COI marker for actinopterygians
 310 (SeaDNA-short) had a PPC value of 34.5%. For elasmobranchs, three 12S primer pairs had the highest
 311 mean PPC values, with Taberlet-elas02 at 68.8%, Valentini-tele01 at 60.4%, and MiFish-E at 39.3%. The
 312 12S Riaz-V5 primers, the cytochrome *b* primers, and the 16S primers, had the lowest PPC values (11.2%,
 313 14.4% and 0% respectively), while the COI primers had PPC values between 21.5% (SeaDNA-short) and 39%
 314 (simplified Leray-XT). These patterns remained when only common species were compared (Table 2).

315 *In vitro* analyses

316 Total reads from Illumina sequencing (Table 3) varied between 3.4 million (12S MiFish-U) and 14.3 million
 317 (COI SeaDNA-mid). After bioinformatic processing, the proportions of reads retained were 46% (COI
 318 SeaDNA-short), 54% (COI Leray-XT), 61% (COI SeaDNA-mid) and 63% (12S MiFish-U). Mean cleaned
 319 reads recovered per sampling event (triplicate water samples, duplicate PCR tags; $n = 6$) were: 107,458 (SD =
 320 46,924) for Leray-XT; 290,104 (SD = 118,592) for SeaDNA-mid; 135,804 (SD = 44,993) for SeaDNA-short;
 321 and 71,912 (SD = 13,682) for 12S MiFish-U. Supplementary Figure 1 shows distributions of read depths
 322 per sample for each site and primer set. The 12S MiFish-U primers provided the greatest proportion of
 323 chordate and fish reads (100% and 76% of cleaned reads, respectively), resulting in more than 1.6 million
 324 putative fish reads and 156 fish ASVs. From these fish reads, 96% were assigned to 41 species and 67
 325 *Swarm* OTU clusters. A total of 73,377 fish reads comprising 18 *Swarm* OTUs could not be assigned, and
 326 in addition to PCR and sequencing artefacts, these likely represent at least nine species not present in the
 327 reference library (Supplementary Table 1). For the COI primer sets, chordate reads comprised between 0.2%
 328 (Leray-XT) and 6% (SeaDNA-short) of the total cleaned reads, with between 0.1% and 5% putative fish reads
 329 comprising between 22 (Leray-XT) and 29 (SeaDNA-short) assigned species. Between 42% (Leray-XT) and
 330 85% (SeaDNA-short) of the putative fish reads were unassigned to species. The non-chordate reads were
 331 inferred from the preliminary blast search to consist of DNA from other metazoans (4–10%) and eukaryotes
 332 (41–83%), or bacteria (17–59%).

Table 3. Number of reads remaining after seven bioinformatic steps, as well as the number of estimated reads for taxonomic groups (assignments were carried out on the reads remaining after the homology search step 7). Fish reads (putative) are reads assigned to fishes based on the best scoring *blastn* hit using the NCBI “nt” blast database. Fish reads (assigned) are reads assigned to fish species by the stringent taxonomic identification step using *blastn* and *EPA-ng* on our curated reference library. Fish reads (unassigned) are putative fish reads that could not be assigned to species by the stringent taxonomic identification step.

Filtering step	COI Leray-XT	COI SeaDNA-mid	COI SeaDNA-short	12S MiFish-U
Total passing filter	5,967,313	14,291,168	8,881,088	3,436,278
(1) Detect primers	4,828,799	11,535,904	6,428,030	2,776,073
(2) Demultiplex	4,648,811	10,879,223	5,994,815	2,473,594
(3) Trim primers	4,618,236	10,300,907	5,852,555	2,462,936
(4) Quality filter	4,519,097	10,344,024	5,856,045	2,455,532
(5) Merge	3,395,057	9,658,709	4,804,502	2,383,162
(6) Remove chimera	3,225,240	9,404,746	4,416,647	2,271,541
(7) Homology search	3,223,743	8,703,109	4,074,123	2,157,365
Bacteria	1,476,994	1,388,681	2,242,220	4
Eukaryota	1,745,295	7,294,762	1,815,928	2,157,361
Metazoa	321,590	1,161,769	412,871	2,157,361
Chordata	6,351	337,901	250,650	2,157,361
Fish (putative)	2,371	234,219	193,593	1,637,728
Fish (assigned)	1,368	109,486	30,026	1,564,351
Fish (unassigned)	1,003	124,733	163,567	73,377

333 Per sampling location the 12S MiFish-U primer set detected a consistently greater number of total species
 334 across sites than the COI markers, at between 2.2 (River Test) and 2.6 (Whitsand Bay) fold higher (Figure 1).
 335 The SeaDNA-short primers detected a greater number of species than both the SeaDNA-mid and Leray-XT
 336 primers, except at the River Tees site where SeaDNA-mid detected one more.

337 In terms of reproducibility (Figure 2), the 12S MiFish-U primer set showed a greater proportion of shared

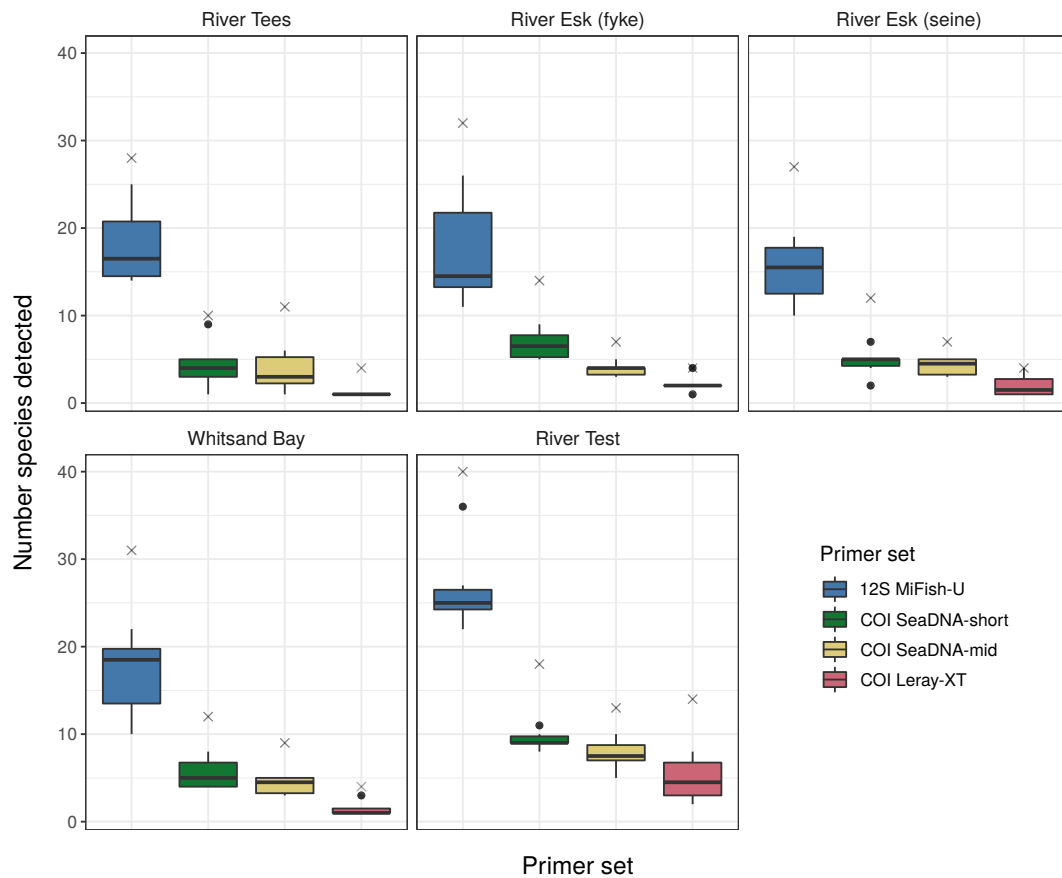


Figure 1. Fish species richness as estimated by four primer pairs at five sampling locations. Per primer-location combination there are three water sample replicates and two uniquely tagged PCR replicates ($n = 6$). The horizontal represents the median value, the boxes represent the 25–75th percentiles, the whiskers represent the values less than 1.5 times the interquartile range, dots represent the outlying data points, and crosses represent the cumulative number of species.

338 species—the top ten species by read abundance at each location—amplified across water sample and PCR
 339 replicates, with a 71% mean reproducibility over all sampling locations. The COI primer sets had mean
 340 reproducibility values of 36% (SeaDNA-short), 29% (SeaDNA-mid) and 12% (Leray-XT).

341 When compared to traditional survey methods—with the freshwater species omitted from the eDNA
 342 results as they were not expected to be found on the traditional fish surveys of the estuarine and coastal
 343 habitats—the 12S MiFish-U primer set showed the greatest congruence (Figure 3), at between 15% (Whitsand
 344 Bay) and 54% (River Test). The COI primers were between 9% (Leray-XT) and 13% (SeaDNA-short)
 345 congruent overall. The MiFish-U primer set also amplified a greater number of marine/estuarine species to
 346 the traditional survey methods at all locations except for Whitsand Bay (26 versus 23 species). The COI
 347 primer sets amplified fewer marine/estuarine species than the traditional surveys in all cases, except for the
 348 SeaDNA-short primer set at the River Tees and River Esk sites. For each site survey, reads per species (eDNA
 349 survey) and individuals per species (traditional survey) are presented in Supplementary Tables 2–6.

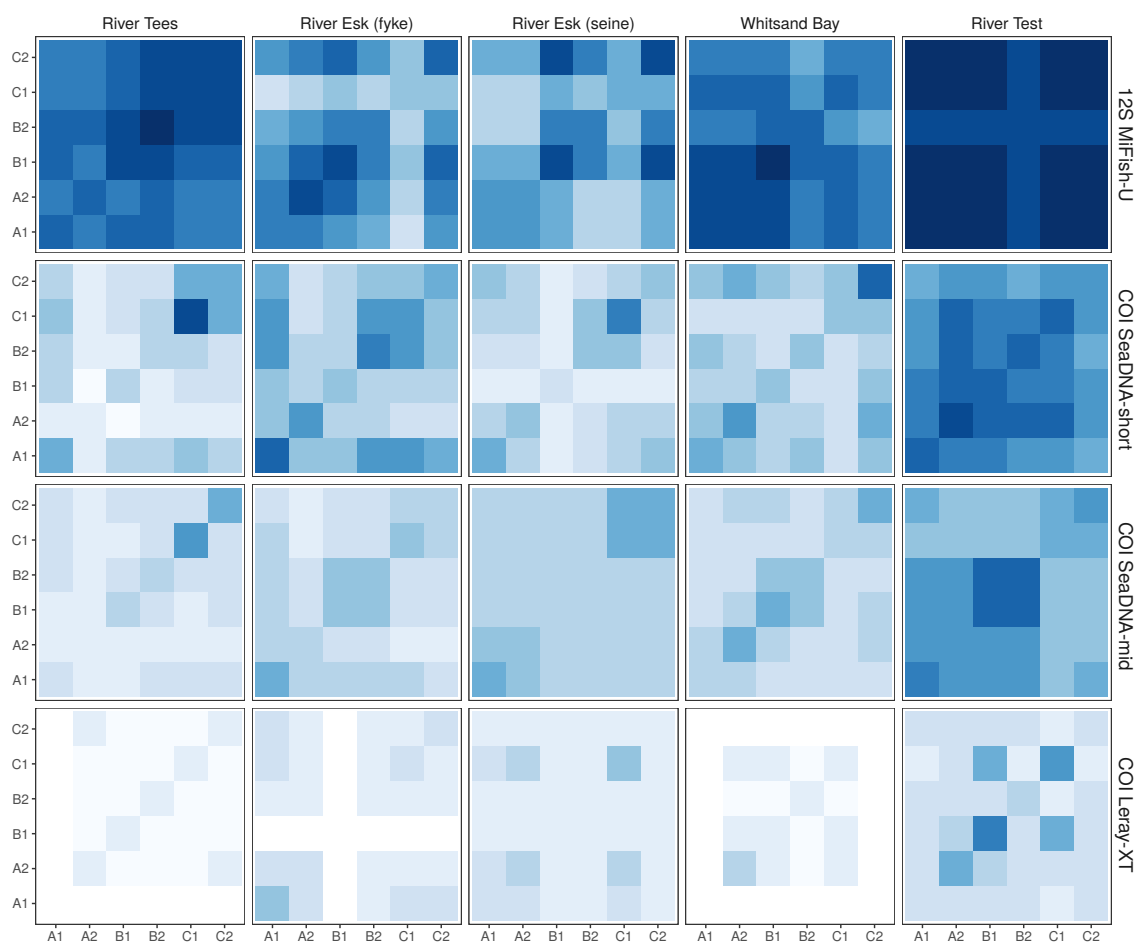


Figure 2. Reproducibility heatmaps of four primer pairs at five sampling locations for the top ten fish species found at each location by read abundance. Letters A, B, and C represent the three water samples taken, while numbers 1 and 2 represent the independent PCR reactions with uniquely tagged primers. There are ten shades showing 10% increments. The darkest shade shows a reproducibility of 100%, i.e. reads from all of the ten species were common to both PCRs. The lightest shade shows 0% reproducibility, i.e. none of the species were present in both of the PCRs. Diagonals show the proportion of the top ten species amplified in that single PCR.

350 DISCUSSION

351 A single metabarcoding marker for fishes?

352 Of arguably the greatest importance in the ability of metabarcoding to answer a particular question, is that
 353 of the choice of marker and primer (Alberdi et al., 2018; Elbrecht and Leese, 2017; Clarke et al., 2017;
 354 Deagle et al., 2014; Valentini et al., 2016). The ideal genetic marker for eDNA metabarcoding marker should
 355 be flexible, allowing different primer sets to target different taxonomic groups, but requiring only a single
 356 reference library. Each individual primer set must also be designed with the following qualities: (i) it must
 357 be universal, i.e. amplifying a large proportion of the target taxonomic group; (ii) it must be specific, i.e. it
 358 must not amplify other taxa at the expense of the target group; (iii) it must be unbiased, i.e. not preferentially
 359 amplifying a subset of the target group; (iv) it must be discriminatory, i.e. the DNA fragment recovered should
 360 differentiate at the appropriate taxonomic level for the question; and (v) it must be replete, i.e. associated



Figure 3. Overlap between fish species found by eDNA metabarcoding (red) and traditional fish surveying (blue). Sizes of circles are proportional only within each primer-location comparison, and not between. Numbers represent number of species in each set. Only marine and estuarine species are shown; freshwater species recorded by the eDNA surveys were removed to allow an equivalent comparison.

361 with a reference library enabling identifications within the target taxonomic group. Here, we assess these
 362 characteristics for COI, cytochrome *b*, 12S, and 16S primer sets using the example of marine and freshwater
 363 fishes from the British Isles.

364 Which primers have the best reference library?

365 In terms of reference libraries, the COI primers were substantially better endowed than all other marker genes,
 366 with between 1.6 times (cytochrome *b*) and 14 times (Valentini-tele01) more public sequence data available
 367 for all species. This was also reflected in the common species coverage, at up to 97% for COI. The 16S (95%),
 368 cytochrome *b* (81%), and 12S Riaz-V5 libraries (81%) were also well developed for common species, but
 369 coverage for other 12S primer sets was lower, at 56–62%. A reference library with broad taxonomic depth
 370 will allow inferences beyond a comparison of anonymous MOTUs, thereby leveraging the wealth of scientific
 371 information that a taxonomic name brings with it (Ward et al., 2009). Deep coverage in the COI reference
 372 library—i.e. the number and geographic distribution of sequences per species—also has advantages in terms
 373 of potential for population level assignments, and for flagging spuriously identified sequences (due to the

374 lesser weight of evidence from the low numbers of sequences, misidentifications were harder to confirm for
375 12S during the quality control step). Furthermore, in terms of voucher specimen and location data etc, much of
376 the ribosomal data on GenBank are not validated to the same standard as COI data on BOLD are (Ward et al.,
377 2009). However, it is important to remember that despite the success of 15 years of the DNA barcode initiative
378 producing COI coverage spanning the majority of northern European fish species, the BOLD database still
379 remains seriously underdeveloped for many other taxonomic groups such as marine invertebrates (Bucklin
380 et al., 2011; Leray and Knowlton, 2016).

381 Which primers best discriminate species?

382 In terms of the discriminatory power for our dataset of British Isles fish species, all primer sets gave a
383 resolution above 90% except for SeaDNA-short (COI), Valentini-tele01 (12S), Riaz-V5 (12S) and Berry-fish
384 (16S). Predictably, the longer COI fragments resolved more species than the shorter ones, at 95% for the 313
385 bp Leray-XT and 87% for the 55 bp SeaDNA-short fragment. The 12S primers did not show this pattern
386 as clearly, with the shorter Valentini-tele01 fragment having a better taxonomic resolution (86%) than the
387 longer Riaz-V5 fragment (79%); the longest, MiFish-U/E and Taberlet-tele02/elas02 primers, had the greatest
388 species resolution at 93%. While discriminatory power may depend on the range of species in that particular
389 library, the observed patterns held up when a dataset of sequences that were shared for all primer sets was
390 used. Discriminatory power also tended to remain the same or increase when only the common species were
391 considered, most likely because rare congeners were excluded.

392 Which primers are most universal?

393 Primer universality as estimated by *in silico* PCR varied greatly. Our results show that the metabarcoding
394 primers targeting protein-coding genes—COI and cytochrome *b*—are likely to exhibit a greater degree of
395 species-level primer bias (i.e. lower universality) than ribosomal 12S and 16S, as indicated by the lower
396 mean PPC values; a mean PPC of 96% was estimated for common actinopterygian species amplified with
397 the Riaz-V5 primers. Previous studies have also reported or predicted less primer bias with rRNA targets
398 than protein coding ones (Clarke et al., 2014; Elbrecht et al., 2016; Deagle et al., 2014; Marquina et al.,
399 2019). It is also important to note again that due to the high level of degeneracy the Leray-XT primers were
400 simplified to overcome RAM limitations of the analysis, and therefore the value presented is likely to be
401 an underestimate of their true potential, as highly degenerate COI primers have been shown to reduce bias
402 substantially (Marquina et al., 2019).

403 Regarding higher level taxonomic bias, for the 12S and 16S primers tested here, no set except Valentini-
404 tele01 appeared suited to amplify actinopterygians and elasmobranchs equally. The COI primers were,
405 however, unbiased in regard to higher taxonomic group. The MiFish primers and the Taberlet et al. (2018)
406 variants of the same sets were both published with actinopterygian (MiFish-U) and elasmobranch (MiFish-E)
407 versions, due to a number of mismatches in the conserved regions (Miya et al., 2015). Unsurprisingly, both of
408 these performed substantially better for their respective taxa. The Taberlet et al. (2018) primers were also
409 predicted here to exhibit substantially less species-level primer bias than the original MiFish versions, for
410 both elasmobranchs and actinopterygians.

411 Many studies computationally predict primer amplification by the number of mismatches between primer
412 and template (e.g. Riaz et al., 2011), or by the number of mismatches and their type and position (e.g. Elbrecht

413 [et al., 2017](#)), but often do not fully consider the thermodynamics of a primer-template reaction. We used
414 the thermodynamics-based PCR simulation implemented in MFEprimer ([Qu et al., 2012](#)), but regardless of
415 whether this method is more realistic or accurate than alternative methods, it is important to remember that
416 these are predicted amplifications, and were used here to compare relative performances between primer sets.
417 Therefore, the lower values estimated do not represent amplification failure *per se*, but rather are indicative
418 of increased bias associated with that primer set ([Deagle et al., 2014](#)). For example, the standard COI DNA
419 barcode primers for fishes (Ward-barcode) had a very low PPC, but these are tried-and-tested primers for
420 amplifying a wide range of fish taxa in standard PCR for Sanger sequencing ([Ward et al., 2005](#)). The use
421 of mock communities is an important step in quality controlling an assay if primer bias is suspected ([Piñol
422 et al., 2015](#); [Elbrecht and Leese, 2017](#); [Bista et al., 2018](#)), but *in silico* PCR has been demonstrated to be an
423 effective proxy in its absence ([Clarke et al., 2014](#)).

424 We used the results of our *in silico* analyses to inform our choices for the *in vitro* experiments. All COI
425 primer sets were selected for testing *in vitro* because of the advantages in terms of reference library and
426 taxonomic discrimination. We chose only one 12S set for comparison, and here we chose the MiFish-U primer
427 pair because this pair had better predicted universality for actinopterygians and more reference sequences
428 available than the Valentini-tele01 primers, and greater taxonomic discrimination than the Riaz-V5 primers.
429 Due to the better predicted universality of the Taberlet-tele02 primer set compared to MiFish-U, these would
430 have been chosen had they been publicly available at the time the experiment was implemented. Despite the
431 well developed reference libraries and good taxonomic discrimination, we did not select cytochrome *b* or 16S
432 because of the lower predicted universality of these primers in comparison to 12S.

433 **Which primers are the most specific?**

434 Despite having the fewest total raw reads, the MiFish-U primer set produced the greatest number and
435 proportion of usable fish reads (76% of processed reads, 48% of raw reads), the greatest overall species
436 richness (41 species), and the greatest proportion of fish reads that were assigned to species (96%). The COI
437 primers amplified a very low proportion of chordate and fish reads compared to the overall sequencing depth
438 (maximum 5% of cleaned reads were fishes). The majority of the SeaDNA-short and SeaDNA-mid reads were
439 estimated by preliminary blast search to have come from bacteria or non-metazoan eukaryotes (86–90%).

440 That the highly degenerate Leray-XT primers produced a low proportion of fish reads is unsurprising
441 given that previous studies on environmental samples using degenerate COI primers have demonstrated that
442 they can amplify widely beyond their target taxa, and can produce large proportions of unassigned reads
443 ([Macher et al., 2018](#); [Stat et al., 2017](#); [Lim et al., 2016](#); [Singer et al., 2019](#)). The proportion of bacterial reads
444 are generally lower when metabarcoding bulk organismal samples, however, with most reads belonging to
445 metazoans ([Wangenstein et al., 2018](#); [Leray and Knowlton, 2015](#); [Macher et al., 2018](#)). More surprising was
446 the poor specificity of the SeaDNA-short and SeaDNA-mid primers, which were designed to target fishes, and
447 with minimal degeneracy. These data are, however, consistent with those of an analysis of shark diversity by
448 [Bakker et al. \(2017\)](#), who used COI mini-barcode primers designed on sharks, and reported a similar level of
449 non-specific amplification.

450 The cause of this non-specific amplification is likely to be the extensive homoplasy (nucleotide con-
451 vergence) apparent in the mutationally saturated COI gene and its homologs. [Siddall et al. \(2009\)](#) demon-
452 strated that metazoan-targeted COI primers are likely to co-amplify many marine prokaryote groups—

453 gammaproteobacteria being a particularly diverse and abundant lineage (Sunagawa et al., 2015)—thereby
454 compromising the specificity of these primer sets. Optimisation of PCR protocols or library preparation
455 methods may increase specificity of the assay (Siddall et al., 2009), but it is probably unlikely that it can
456 increase to a level that makes the proportion of usable reads viable for eDNA metabarcoding of targeted
457 taxonomic groups. While this phenomenon was first observed in marine prokaryotes, studies on freshwater
458 and soil faunas have shown a similar pattern, also with large numbers of unassigned reads (Lim et al., 2016;
459 Yang et al., 2014).

460 Which primers give the most reproducible results?

461 The low number of usable fish reads for the COI primers is reflected in the reproducibility of the assays across
462 water sample and PCR replicates. For the most frequently amplified species at each site, the COI primers were
463 less consistent than 12S MiFish-U overall. Low quantities of template DNA and stochasticity in early PCR
464 cycles is a known factor in causing poor reproducibility (Leray and Knowlton, 2017; Alberdi et al., 2018;
465 Collins et al., 2018), and can be ameliorated by performing multiple PCR technical replicates (Ficetola et al.,
466 2015). We show that this effect is exacerbated when primer specificity is low and non-target organisms are
467 abundant, as is the case in highly diverse environmental samples such as seawater. For many applications
468 repeatability between assays or sampling sites is a requirement, such as the detection of an endangered or
469 invasive species (Grey et al., 2018). Our results, even considering only the top ten common species, show that
470 detectability can vary between sites with the same genetic marker, and that many more than two PCRs will be
471 required if the rare species are to be detected across multiple PCR and water sample replicates (Dopheide
472 et al., 2018).

473 Species richness estimates at all sampling sites were greatest with 12S MiFish-U, and this was despite
474 the deficiencies in the reference library, at only 61% species coverage. For example, species including the
475 European plaice (*Pleuronectes platessa*) and European flounder (*Platichthys flesus*)—both common fishes
476 present at all sampling locations—were missing from the reference library and therefore not represented when
477 comparing with the traditional fish surveys. Most of the large number of reads that were assigned to American
478 plaice, *Hippoglossoides platessoides* ($n = 198,445$), were likely misassigned and actually belong to European
479 plaice and flounder (Supplementary Table 1). The Swarm OTU analysis showed a greater number of clusters
480 (67) than assigned species (41), also suggesting that some species missing from the reference library are
481 likely to have been misassigned. While a small number of the 73,377 unassigned 12S fish reads were low
482 abundance sequences derived from artefacts, almost all could be inferred by phylogenetic analysis or
483 by similarity to geographically disjunct congeners, to belong to at least eight species that were known to be
484 missing from the reference library (Supplementary Table 1). Despite this major handicap, the 12S MiFish
485 primers remained superior to COI in terms of congruence with the traditional fish surveys, by recovering a
486 greater overlap of species in all cases. The 12S MiFish primers amplified more species than the traditional
487 surveys at all sites, except Whitsand Bay. This was mainly due to the underrepresentation of the fauna of that
488 site in the 12S reference library, with over half of the surveyed species absent from the library, and a higher
489 proportion of elasmobranchs (five species) than the other sites, which the MiFish-U primers fail to amplify.
490 Overall, no species that were recorded in the traditional surveys were missing from the COI reference libraries,
491 but eighteen species were missing from the 12S MiFish library (37%). The low numbers of species recorded
492 by the traditional surveys at the Esk and Tees sites in comparison to the Whitsand Bay and River Test sites, is

493 partly due to the inherently less diverse fauna of these northerly estuaries, as well as a reflection of the survey
494 techniques, with fyke and seine netting likely to detect fewer species than otter trawling (Whitsand Bay) or
495 a 24 h power station impingement (River Test). It should also be noted that there is no *a priori* assumption
496 that the eDNA and traditional survey data will be completely congruent, as most fish survey methods are
497 imperfect, sampling a moving target of diversity and abundance over difficult-to-define spatio-temporal points.
498 For example, eDNA can be transported in or out by tides, while some species are difficult to sample using
499 particular fishing gears, due to effects of size, behaviour or abundance. Therefore, overlap between eDNA
500 and traditional survey data is best interpreted as a relative measure between the primer sets.

501 CONCLUSIONS

502 While PCR-free methods are being actively investigated, it is clear that despite the limitations in quantification,
503 the majority of environmental metabarcoding will be based around amplicon sequencing, at least for the
504 medium term (Wilcox et al., 2018; Stat et al., 2017; Bista et al., 2018; Creer et al., 2016). Particularly
505 important for regulatory applications, or where researchers wish to compare results over time or between
506 studies, some degree of standardisation is desirable (Hering et al., 2018). Our results—and those of previous
507 studies using similar primer sets (Macher et al., 2018; Stat et al., 2017; Lim et al., 2016; Bakker et al.,
508 2017; Yang et al., 2014; Jeunen et al., 2018; Singer et al., 2019)—show that environmental metabarcoding
509 for restricted taxonomic groups using degenerate COI primers results in excessive volumes of “wasted”
510 sequencing effort. This co-amplification of prokaryotic and non-target eukaryotic DNAs and subsequent lack
511 of specificity is due to the nature of mutation patterns in COI (Siddall et al., 2009). Therefore, while we
512 fully support the arguments presented by Andújar et al. (2018) regarding the overall advantages of COI as
513 a bulk-sample metabarcoding marker, we find it difficult to recommend for metabarcoding environmental
514 samples with low target template concentrations and high microbial and plankton diversity, such as natural
515 water bodies.

516 While the use of multiple primer sets and markers are probably required for a comprehensive view of total
517 biodiversity (Stat et al., 2017; Drummond et al., 2015), for specific taxonomic groups such as fishes a single
518 assay should be a feasible proposition. Unfortunately, no single 12S primer set was shown to be optimal for
519 eDNA fish surveys. The MiFish-U primer set—and *in silico*, the Taberlet et al. (2018) modified versions—
520 performed well in terms of specificity, discriminatory power, and reproducibility. Despite this, MiFish-U is
521 not universal for all fishes, because a separate MiFish-E assay is required to amplify elasmobranchs. The
522 MiFish reference library was also inadequate in this case, missing large numbers of common taxa. The
523 Valentini-tele01 primer set amplifies actinopterygians and elasmobranchs in a single assay, but suffers from
524 an even more poorly populated reference library than MiFish-U, and weaker taxonomic resolution. The
525 Riaz-V5 primers had the most complete reference library of the 12S primer pairs, but also do not amplify
526 elasmobranchs and have the poorest discriminatory power.

527 Because no single alternative primer set to COI will be optimal for all applications, it is clear that the
528 current DNA barcode reference libraries will need to be augmented with data from multiple mitochondrial
529 regions to enable their wider utility for vertebrate metabarcoding. However, rather than sequencing individual
530 12S regions on an ad hoc basis, a better solution is to generate whole mitochondrial genomes which can act as
531 an extended or linking barcode if sequenced from the same collection material (Coissac et al., 2016; Collins

532 and Cruickshank, 2014). Low coverage genome skimming techniques now produce high quality mitogenomes,
533 and are compatible with existing—frequently ethanol-based—tissue collections, and therefore will not require
534 the recollection of specimens (Linard et al., 2016; Gillett et al., 2014). Environmental DNA techniques could
535 potentially be the default survey methodology for aquatic ecosystems, but the existing gap between recovered
536 genotypes and their corresponding phenotypic and historical data can only be filled with substantially more
537 comprehensive reference libraries.

538 ACKNOWLEDGEMENTS

539 For assistance with logistics, eDNA sampling and providing the traditional survey data, we thank: Peter
540 Henderson and Robin Somes from Pisces Conservation Ltd.; Tony Gray, Hannah Westoby and the staff at
541 the Environment Agency in Newcastle Upon Tyne; and Aisling Smith and Sophie Rainbird from the Marine
542 Biological Association in Plymouth. This work was co-funded by the Natural Environment Research Council
543 grants NE/N005759/1 and NE/N005937/1 (project SeaDNA), and the University of Salford R&E strategy
544 funding.

545 DECLARATION OF INTEREST

546 The authors declare that they have no competing interests.

547 DATA ACCESSIBILITY

548 The full reference library and code to reproduce it can be found at <https://doi.org/10.6084/m9.figshare.7464521>.
549 Code to reproduce all other analyses in this study can be found at <https://doi.org/10.6084/m9.figshare.8291660>.

550 REFERENCES

- 551 Alberdi, A., Aizpurua, O., Gilbert, M. T. P., and Bohmann, K. (2018). Scrutinizing key steps for reli-
552 able metabarcoding of environmental samples. *Methods in Ecology and Evolution*, 9:134–147, DOI:
553 [10.1111/2041-210X.12849](https://doi.org/10.1111/2041-210X.12849).
- 554 Andruszkiewicz, E. A., Starks, H. A., Chavez, F. P., Sassoubre, L. M., Block, B. A., and Boehm, A. B.
555 (2017). Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *PLoS ONE*,
556 12:e0176343, DOI: [10.1371/journal.pone.0176343](https://doi.org/10.1371/journal.pone.0176343).
- 557 Andújar, C., Arribas, P., Yu, D. W., Vogler, A. P., and Emerson, B. C. (2018). Why the COI barcode
558 should be the community DNA metabarcode for the Metazoa. *Molecular Ecology*, 27:3968–3975, DOI:
559 [10.1111/mec.14844](https://doi.org/10.1111/mec.14844).
- 560 Bakker, J., Wangensteen, O. S., Chapman, D. D., Boussarie, G., Buddo, D., Guttridge, T. L., Hertler,
561 H., Mouillot, D., Vigliola, L., and Mariani, S. (2017). Environmental DNA reveals tropical
562 shark diversity in contrasting levels of anthropogenic impact. *Scientific Reports*, 7:16886, DOI:
563 [10.1038/s41598-017-17150-2](https://doi.org/10.1038/s41598-017-17150-2).
- 564 Barbera, P., Kozlov, A. M., Czech, L., Morel, B., Darriba, D., Flouri, T., and Stamatakis, A. (2018).
565 EPA-ng: massively parallel evolutionary placement of genetic sequences. *Systematic Biology*, DOI:
566 [10.1101/291658](https://doi.org/10.1101/291658).

- 567 Berry, T. E., Osterrieder, S. K., Murray, D. C., Coghlan, M. L., Richardson, A. J., Grealy, A. K., Stat,
568 M., Bejder, L., and Bunce, M. (2017). Metabarcoding for diet analysis and biodiversity: A case study
569 using the endangered Australian sea lion (*Neophoca cinerea*). *Ecology and Evolution*, pages 1–19, DOI:
570 [10.1002/ece3.3123](https://doi.org/10.1002/ece3.3123).
- 571 Bista, I., Carvalho, G. R., Tang, M., Walsh, K., Zhou, X., Hajibabaei, M., Shokralla, S., Seymour, M., Bradley,
572 D., Liu, S., Christmas, M., and Creer, S. (2018). Performance of amplicon and shotgun sequencing for
573 accurate biomass estimation in invertebrate community samples. *Molecular Ecology Resources*, 18:1020–
574 1034, DOI: [10.1111/1755-0998.12888](https://doi.org/10.1111/1755-0998.12888).
- 575 Boettiger, C., Lang, D. T., and Wainwright, P. C. (2012). rfishbase: exploring , manipulatin
576 g and visualizing FishBase data from R. *Journal of Fish Biology*, 81:2030–2039, DOI:
577 [10.1111/j.1095-8649.2012.03464.x](https://doi.org/10.1111/j.1095-8649.2012.03464.x).
- 578 Boyer, S., Brown, S. D. J., Collins, R. A., Cruickshank, R. H., Lefort, M.-C., Malumbres-Olarte, J., and
579 Wratten, S. D. (2012). Sliding window analyses for optimal selection of mini-barcodes, and application
580 to 454-pyrosequencing for specimen identification from degraded DNA. *PLoS ONE*, 7:e38215, DOI:
581 [10.1371/journal.pone.0038215](https://doi.org/10.1371/journal.pone.0038215).
- 582 Brown, S. D. J., Collins, R. A., Boyer, S., Lefort, M.-C., Malumbres-Olarte, J., Vink, C. J., and
583 Cruickshank, R. H. (2012). Spider: an R package for the analysis of species identity and evolu
584 tion, with particular reference to DNA barcoding. *Molecular Ecology Resources*, 12:562–565, DOI:
585 [10.1111/j.1755-0998.2011.03108.x](https://doi.org/10.1111/j.1755-0998.2011.03108.x).
- 586 Bucklin, A., Steinke, D., and Blanco-Bercial, L. (2011). DNA Barcoding of Marine Metazoa. *Annual Review*
587 *of Marine Science*, 3:471–508, DOI: [10.1146/annurev-marine-120308-080950](https://doi.org/10.1146/annurev-marine-120308-080950).
- 588 Bylemans, J., Gleeson, D. M., Hardy, C. M., and Furlan, E. (2018). Toward an ecoregion scale evaluation
589 of eDNA metabarcoding primers: A case study for the freshwater fish biodiversity of the Murray-Darling
590 Basin (Australia). *Ecology and Evolution*, 8:8697–8712, DOI: [10.1002/ece3.4387](https://doi.org/10.1002/ece3.4387).
- 591 Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016).
592 DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13:581–583,
593 DOI: [10.1038/nmeth.3869](https://doi.org/10.1038/nmeth.3869).
- 594 Chamberlain, S. (2018). bold: Interface to Bold Systems API. [https://cran.r-project.org/
595 package=bold](https://cran.r-project.org/package=bold).
- 596 Chamberlain, S. and Boettiger, C. (2017). R Python, and Ruby clients for GBIF species occurrence data.
597 *PeerJ PrePrints*, DOI: [10.7287/peerj.preprints.3304v1](https://doi.org/10.7287/peerj.preprints.3304v1).
- 598 Chamberlain, S., Foster, Z., Bartomeus, I., LeBauer, D., Black, C., and Harris, D. (2018). traits: species trait
599 data from around the web. <https://github.com/ropensci/traits>.
- 600 Clarke, L. J., Beard, J. M., Swadling, K. M., and Deagle, B. E. (2017). Effect of marker choice and thermal
601 cycling protocol on zooplankton DNA metabarcoding studies. *Ecology and Evolution*, 7:873–883, DOI:
602 [10.1002/ece3.2667](https://doi.org/10.1002/ece3.2667).
- 603 Clarke, L. J., Soubrier, J., Weyrich, L. S., and Cooper, A. (2014). Environmental metabarcodes for insects:
604 In silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, 14:1160–1170, DOI:
605 [10.1111/1755-0998.12265](https://doi.org/10.1111/1755-0998.12265).
- 606 Coissac, E., Hollingsworth, P. M., Lavergne, S., and Taberlet, P. (2016). From barcodes to genomes: Extending

- 607 the concept of DNA barcoding. *Molecular Ecology*, 25:1423–1428, DOI: [10.1111/mec.13549](https://doi.org/10.1111/mec.13549).
- 608 Collins, R. A. and Cruickshank, R. H. (2014). Known knowns, known unknowns, unknown unknowns and
609 unknown knowns in DNA barcoding: A comment on Downton et al. *Systematic Biology*, 63:1005–1009,
610 DOI: [10.1093/sysbio/syu060](https://doi.org/10.1093/sysbio/syu060).
- 611 Collins, R. A., Wangensteen, O. S., Sims, D. W., Genner, M. J., and Mariani, S. (2018). Per-
612 sistence of environmental DNA in marine systems. *Communications Biology*, 1:185, DOI:
613 [10.1038/s42003-018-0192-6](https://doi.org/10.1038/s42003-018-0192-6).
- 614 Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, K., Potter, C., and Bik, H. (2016). The
615 ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*,
616 56:68–74, DOI: [10.1111/2041-210X.12574](https://doi.org/10.1111/2041-210X.12574).
- 617 Czech, L. and Stamatakis, A. (2018). Scalable methods for post-processing , visualizing , and analyzing
618 phylogenetic placements. *bioRxiv*, pages 1–36, DOI: [10.1101/346353](https://doi.org/10.1101/346353).
- 619 Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., and Taberlet, P. (2014). DNA metabarcoding and
620 the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, 10:20140562, DOI:
621 [10.1098/rsbl.2014.0562](https://doi.org/10.1098/rsbl.2014.0562).
- 622 Dopheide, A., Xie, D., Buckley, T. R., Drummond, A. J., and Newcomb, R. D. (2018). Impacts of DNA
623 extraction and PCR on DNA metabarcoding estimates of soil biodiversity. *Methods in Ecology and*
624 *Evolution*, DOI: [10.1111/2041-210X.13086](https://doi.org/10.1111/2041-210X.13086).
- 625 Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C. M., Heled, J., Ross,
626 H. A., Tooman, L., Grosser, S., Park, D., Demetras, N. J., Stevens, M. I., Russell, J. C., Anderson, S. H.,
627 Carter, A., and Nelson, N. (2015). Evaluating a multigene environmental DNA approach for biodiversity
628 assessment. *GigaScience*, 4:46, DOI: [10.1186/s13742-015-0086-1](https://doi.org/10.1186/s13742-015-0086-1).
- 629 Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, 14:755–763, DOI:
630 [10.1093/bioinformatics/14.9.755](https://doi.org/10.1093/bioinformatics/14.9.755).
- 631 Elbrecht, V. and Leese, F. (2017). Validation and Development of COI Metabarcoding Primers for
632 Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science*, 5:1–11, DOI:
633 [10.3389/fenvs.2017.00011](https://doi.org/10.3389/fenvs.2017.00011).
- 634 Elbrecht, V., Taberlet, P., Dejean, T., Valentini, A., Usseglio-Polatera, P., Beisel, J.-N., Coissac, E., Boyer, F.,
635 and Leese, F. (2016). Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects.
636 *PeerJ*, 4:e1966, DOI: [10.7717/peerj.1966](https://doi.org/10.7717/peerj.1966).
- 637 Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., and Leese, F. (2017). Assessing strengths and weaknesses
638 of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in*
639 *Ecology and Evolution*, 8:1265–1275, DOI: [10.1111/2041-210X.12789](https://doi.org/10.1111/2041-210X.12789).
- 640 Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., and Shehzad, W. (2010). An in silico approach for the
641 evaluation of DNA barcodes. *BMC Genomics*, 11:434, DOI: [10.1186/1471-2164-11-434](https://doi.org/10.1186/1471-2164-11-434).
- 642 Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguët-Covex, C., De Barba, M., Gielly, L., Lopes, C. M.,
643 Boyer, F., Pompanon, F., Rayé, G., and Taberlet, P. (2015). Replication levels, false presences and
644 the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*,
645 15:543–556, DOI: [10.1111/1755-0998.12338](https://doi.org/10.1111/1755-0998.12338).
- 646 Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of

- 647 mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine*
648 *Biology and Biotechnology*, 3:294–299.
- 649 Gillett, C. P. D. T., Crampton-Platt, A., Timmermans, M. J. T. N., Jordal, B., Emerson, B. C., and
650 Vogler, A. P. (2014). Bulk de novo mitogenome assembly from pooled total DNA elucidates the phy-
651 logeny of weevils (Coleoptera: Curculionidae). *Molecular Biology and Evolution*, 31:2223–2237, DOI:
652 [10.1093/molbev/msu154](https://doi.org/10.1093/molbev/msu154).
- 653 Grey, E. K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K. L., Lacoursière-Roussel, A.,
654 Leong, S. C. Y., Li, Y., Olds, B., Pfrender, M. E., Prowse, T. A., Renshaw, M. A., and Lodge, D. M. (2018).
655 Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding
656 surveys. *Scientific Reports*, 8:2–11, DOI: [10.1038/s41598-018-27048-2](https://doi.org/10.1038/s41598-018-27048-2).
- 657 Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangenstein, O. S., and Turon, X. (2015). Deep-sea,
658 deep-sequencing: Metabarcoding extracellular DNA from sediments of marine canyons. *PLoS ONE*,
659 10:e0139633, DOI: [10.1371/journal.pone.0139633](https://doi.org/10.1371/journal.pone.0139633).
- 660 Hänfling, B., Lawson Handley, L., Read, D. S., Hahn, C., Li, J., Nichols, P., Blackman, R. C., Oliver, A., and
661 Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data
662 from established survey methods. *Molecular Ecology*, 25:3101–3119, DOI: [10.1111/mec.13660](https://doi.org/10.1111/mec.13660).
- 663 Hebert, P. D. N., Cywinska, A., Ball, S. L., and DeWaard, J. R. (2003). Biological identifications
664 through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270:313–321, DOI:
665 [10.1098/rspb.2002.2218](https://doi.org/10.1098/rspb.2002.2218).
- 666 Henderson, P. A. (2014). *Identification Guide to the Inshore Fish of the British Isles*. Pisces Conservation
667 Ltd., Pennington, UK.
- 668 Hering, D., Borja, A., Jones, J. I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakare, S., Hänfling, B., Kahlert,
669 M., Leese, F., Meissner, K., Mergen, P., Reyjol, Y., Segurado, P., Vogler, A., and Kelly, M. (2018). Imple-
670 mentation options for DNA-based identification into ecological status assessment under the European Water
671 Framework Directive. *Water Research*, 138:192–205, DOI: [10.1016/j.watres.2018.03.003](https://doi.org/10.1016/j.watres.2018.03.003).
- 672 Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., Sado, T., Mabuchi, K.,
673 Takeshima, H., Miya, M., and Nishida, M. (2013). Mitofish and mitoannotator: A mitochondrial genome
674 database of fish with an accurate and automatic annotation pipeline. *Molecular Biology and Evolution*,
675 30:2531–2540, DOI: [10.1093/molbev/mst141](https://doi.org/10.1093/molbev/mst141).
- 676 Jeunen, G.-J., Knapp, M., Spencer, H. G., Lamare, M. D., Taylor, H. R., Stat, M., Bunce, M., and
677 Gemmell, N. J. (2018). Environmental DNA (eDNA) metabarcoding reveals strong discrimination
678 among diverse marine habitats connected by water movement. *Molecular Ecology Resources*, DOI:
679 [10.1111/1755-0998.12982](https://doi.org/10.1111/1755-0998.12982).
- 680 Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Wood-
681 cock, P., Edwards, F. A., Larsen, T. H., Hsu, W. W., Benedick, S., Hamer, K. C., Wilcove, D. S., Bruce, C.,
682 Wang, X., Levi, T., Lott, M., Emerson, B. C., and Yu, D. W. (2013). Reliable, verifiable and efficient moni-
683 toring of biodiversity via metabarcoding. *Ecology Letters*, 16:1245–1257, DOI: [10.1111/ele.12162](https://doi.org/10.1111/ele.12162).
- 684 Kartzinel, T. R., Chen, P. A., Coverdale, T. C., Erickson, D. L., Kress, W. J., Kuzmina, M. L., Rubenstein,
685 D. I., Wang, W., and Pringle, R. M. (2015). DNA metabarcoding illuminates dietary niche partitioning
686 by African large herbivores. *Proceedings of the National Academy of Sciences*, 112:8019–8024, DOI:

687 [10.1073/pnas.1503283112](https://doi.org/10.1073/pnas.1503283112).

688 Katoh, K. and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7:
689 improvements in performance and usability. *Molecular Biology and Evolution*, 30:772–780, DOI:
690 [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).

691 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A.,
692 Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012). Geneious Basic:
693 an integrated and extendable desktop software platform for the organization and analysis of sequence data.
694 *Bioinformatics*, 28:1647–1649, DOI: [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199).

695 Kelly, R. P., Closek, C. J., O'Donnell, J. L., Kralj, J. E., Shelton, A. O., and Samhuri, J. F. (2017). Genetic
696 and manual survey methods yield different and complementary views of an ecosystem. *Frontiers in Marine
697 Science*, 3:1–11, DOI: [10.3389/fmars.2016.00283](https://doi.org/10.3389/fmars.2016.00283).

698 Kelly, R. P., Port, J. A., Yamahara, K. M., and Crowder, L. B. (2014a). Using environmental DNA to census
699 marine fishes in a large mesocosm. *PLoS ONE*, 9:e86175, DOI: [10.1371/journal.pone.0086175](https://doi.org/10.1371/journal.pone.0086175).

700 Kelly, R. P., Port, J. A., Yamahara, K. M., Martone, R. G., Lowell, N., Thomsen, P. F., Mach, M. E., Bennett,
701 M., Prahler, E., Caldwell, M. R., and Crowder, L. B. (2014b). Harnessing DNA to improve environmental
702 management. *Science*, 344:1455–1456, DOI: [10.1126/science.1251156](https://doi.org/10.1126/science.1251156).

703 Kottelat, M. and Freyhof, J. (2007). *Handbook of European Freshwater Fishes*. Publications Kottelat, Cornol,
704 Switzerland.

705 Leray, M. and Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal
706 patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, 112:2076–2081,
707 DOI: [10.1073/pnas.1424997112](https://doi.org/10.1073/pnas.1424997112).

708 Leray, M. and Knowlton, N. (2016). Censusing marine eukaryotic diversity in the twenty-first cen-
709 tury. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371:20150331, DOI:
710 [10.1098/rstb.2015.0331](https://doi.org/10.1098/rstb.2015.0331).

711 Leray, M. and Knowlton, N. (2017). Random sampling causes the low reproducibility of rare eukaryotic
712 OTUs in Illumina COI metabarcoding. *PeerJ*, 5:e3006, DOI: [10.7717/peerj.3006](https://doi.org/10.7717/peerj.3006).

713 Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., and Machida,
714 R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for
715 metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in
716 Zoology*, 10:34, DOI: [10.1186/1742-9994-10-34](https://doi.org/10.1186/1742-9994-10-34).

717 Lim, N. K. M., Tay, Y. C., Tan, J. W. T., Kwik, J. T. B., Baloglu, B., Meier, R., and Yeo, D. C. J. (2016).
718 Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals
719 species-rich and communities. *Royal Society Open Science*, 3:160635, DOI: [10.1098/rsos.160635](https://doi.org/10.1098/rsos.160635).

720 Linard, B., Arribas, P., Andújar, C., Crampton-Platt, A., and Vogler, A. P. (2016). Lessons from
721 genome skimming of arthropod-preserving ethanol. *Molecular Ecology Resources*, 16:1365–1377, DOI:
722 [10.1111/1755-0998.12539](https://doi.org/10.1111/1755-0998.12539).

723 Macher, J. N., Vivancos, A., Piggott, J. J., Centeno, F. C., Matthaei, C. D., and Leese, F. (2018). Comparison
724 of environmental DNA and bulk-sample metabarcoding using highly degenerate cytochrome c oxidase I
725 primers. *Molecular Ecology Resources*, 18:1456–1468, DOI: [10.1111/1755-0998.12940](https://doi.org/10.1111/1755-0998.12940).

726 Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. (2015). Swarm v2: highly-scalable and

- 727 high-resolution amplicon clustering. *PeerJ*, 3:e1420, DOI: [10.7717/peerj.1420](https://doi.org/10.7717/peerj.1420).
- 728 Mariani, S., Griffiths, A. M., Velasco, A., Kappel, K., Jerome, M., Perez-Martin, R. I., Schroder, U., Verrez-
729 Bagnis, V., Silva, H., Vandamme, S. G., Boufana, B., Mendes, R., Shorten, M., Smith, C., Hankard, E.,
730 Hook, S. A., Weymer, A. S., Gunning, D., and Sotelo, C. G. (2015). Low mislabeling rates indicate
731 marked improvements in European seafood market operations. *Frontiers in Ecology and the Environment*,
732 13:536–540, DOI: [10.1890/150119](https://doi.org/10.1890/150119).
- 733 Marquina, D., Andersson, A. F., and Ronquist, F. (2019). New mitochondrial primers for metabarcoding of
734 insects, designed and evaluated using in silico methods. *Molecular Ecology Resources*, 19:90–104, DOI:
735 [10.1111/1755-0998.12942](https://doi.org/10.1111/1755-0998.12942).
- 736 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMB-
737 net.journal*, 17:10–12, DOI: [10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200).
- 738 McHugh, M., Sims, D. W., Partridge, J. C., and Genner, M. J. (2011). A century later: Long-term
739 change of an inshore temperate marine fish assemblage. *Journal of Sea Research*, 65:187–194, DOI:
740 [10.1016/j.seares.2010.09.006](https://doi.org/10.1016/j.seares.2010.09.006).
- 741 Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M. N., and Kawabata, Z. (2012). Surveil-
742 lance of fish species composition using environmental DNA. *Limnology*, 13:193–197, DOI:
743 [10.1007/s10201-011-0362-4](https://doi.org/10.1007/s10201-011-0362-4).
- 744 Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka,
745 H., Araki, H., Kondoh, M., and Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabar-
746 coding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal
747 Society Open Science*, 2:150088, DOI: [10.1098/rsos.150088](https://doi.org/10.1098/rsos.150088).
- 748 Pedersen, M. W., Overballe-Petersen, S., Ermini, L., Der Sarkissian, C., Haile, J., Hellstrom, M., Spens,
749 J., Thomsen, P. F., Bohmann, K., Cappellini, E., Schnell, I. B., Wales, N. A., Carøe, C., Campos, F.,
750 Schmidt, A. M. Z., Gilbert, M. T. P., Hansen, A. J., Orlando, L., and Willerslev, E. (2015). Ancient and
751 modern environmental DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences*,
752 370:20130383, DOI: [10.1098/rstb.2013.0383](https://doi.org/10.1098/rstb.2013.0383).
- 753 Piñol, J., Mir, G., Gomez-Polo, P., and Agustí, N. (2015). Universal and blocking primer mismatches limit
754 the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular
755 Ecology Resources*, 15:819–830, DOI: [10.1111/1755-0998.12355](https://doi.org/10.1111/1755-0998.12355).
- 756 Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., Yamahara,
757 K. M., and Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using
758 environmental DNA. *Molecular Ecology*, 25:527–541, DOI: [10.1111/mec.13481](https://doi.org/10.1111/mec.13481).
- 759 Qu, W., Zhou, Y., Zhang, Y., Lu, Y., Wang, X., Zhao, D., Yang, Y., and Zhang, C. (2012). MFEprimer-2.0:
760 A fast thermodynamics-based program for checking PCR primer specificity. *Nucleic Acids Research*,
761 40:205–208, DOI: [10.1093/nar/gks552](https://doi.org/10.1093/nar/gks552).
- 762 Ratnasingham, S. and Hebert, P. D. N. (2007). BOLD: The Barcode of Life
763 Data System (www.barcodinglife.org). *Molecular Ecology Notes*, 7:355–364, DOI:
764 [10.1111/j.1471-8286.2006.01678.x](https://doi.org/10.1111/j.1471-8286.2006.01678.x).
- 765 Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., and Gough, K. C. (2014). The detection
766 of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology.

- 767 *Journal of Applied Ecology*, 51:1450–1459, DOI: [10.1111/1365-2664.12306](https://doi.org/10.1111/1365-2664.12306).
- 768 Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., and Coissac, E. (2011). EcoPrimers: Inference of
769 new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, 39:e145, DOI:
770 [10.1093/nar/gkr732](https://doi.org/10.1093/nar/gkr732).
- 771 Shaw, J. L. A., Clarke, L. J., Wedderburn, S. D., Barnes, T. C., Weyrich, L. S., and Cooper, A. (2016).
772 Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system.
773 *Biological Conservation*, 197:131–138, DOI: [10.1016/j.biocon.2016.03.010](https://doi.org/10.1016/j.biocon.2016.03.010).
- 774 Shokralla, S., Hellberg, R. S., Handy, S. M., King, I., and Hajibabaei, M. (2015). A DNA mini-
775 barcoding system for authentication of processed fish products. *Scientific Reports*, 5:15894, DOI:
776 [10.1038/srep15894](https://doi.org/10.1038/srep15894).
- 777 Siddall, M. E., Fontanella, F. M., Watson, S. C., Kvist, S., and Erséus, C. (2009). Barcoding bamboozled by
778 bacteria: convergence to metazoan mitochondrial primer targets by marine microbes. *Systematic Biology*,
779 58:445–451, DOI: [10.1093/sysbio/syp033](https://doi.org/10.1093/sysbio/syp033).
- 780 Singer, G. A. C., Fahner, N. A., Barnes, J. G., Mccarthy, A., and Hajibabaei, M. (2019). Comprehensive
781 biodiversity analysis via ultra-deep patterned flow cell technology: a case study of eDNA metabarcoding
782 seawater. *Scientific Reports*, 9:5991, DOI: [10.1038/s41598-019-42455-9](https://doi.org/10.1038/s41598-019-42455-9).
- 783 Spens, J., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S. T., Sigsgaard, E. E.,
784 and Hellström, M. (2017). Comparison of capture and storage methods for aqueous microbial eDNA
785 using an optimized extraction protocol: advantage of enclosed filter. *Methods in Ecology and Evolution*,
786 8:635–645, DOI: [10.1111/2041-210X.12683](https://doi.org/10.1111/2041-210X.12683).
- 787 Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web
788 servers. *Systematic Biology*, 57:758–771, DOI: [10.1080/10635150802429642](https://doi.org/10.1080/10635150802429642).
- 789 Stat, M., Huggett, M. J., Bernasconi, R., Dibattista, J. D., Berry, T. E., Newman, S. J., Harvey, E. S., and
790 Bunce, M. (2017). Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical
791 marine environment. *Scientific Reports*, 7:0–22, DOI: [10.1038/s41598-017-12501-5](https://doi.org/10.1038/s41598-017-12501-5).
- 792 Stat, M., John, J., DiBattista, J. D., Newman, S. J., Bunce, M., and Harvey, E. S. (2018). Combined use of
793 eDNA metabarcoding and video surveillance for the assessment of fish biodiversity. *Conservation Biology*,
794 33:196–205, DOI: [10.1111/cobi.13183](https://doi.org/10.1111/cobi.13183).
- 795 Stoeckle, M. Y., Soboleva, L., and Charlop-Powers, Z. (2017). Aquatic environmental DNA detects
796 seasonal fish abundance and habitat preference in an urban estuary. *PLoS ONE*, 12:e0175186, DOI:
797 [10.1371/journal.pone.0175186](https://doi.org/10.1371/journal.pone.0175186).
- 798 Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller,
799 G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., D’Ovidio, F., Engelen, S.,
800 Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., Kokoszka, F., Lepoivre, C., Lima-Mendez, G., Poulain,
801 J., Poulos, B. T., Royo-Llonch, M., Sarmiento, H., Vieira-Silva, S., Dimier, C., Picheral, M., Searson, S.,
802 Kandels-Lewis, S., Boss, E., Follows, M., Karp-Boss, L., Krzic, U., Reynaud, E. G., Sardet, C., Sieracki,
803 M., Velayoudon, D., Bowler, C., De Vargas, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D.,
804 Jaillon, O., Not, F., Ogata, H., Pesant, S., Speich, S., Stemmann, L., Sullivan, M. B., Weissenbach, J.,
805 Wincker, P., Karsenti, E., Raes, J., Acinas, S. G., and Bork, P. (2015). Structure and function of the global
806 ocean microbiome. *Science*, 348:1–10, DOI: [10.1126/science.1261359](https://doi.org/10.1126/science.1261359).

- 807 Taberlet, P., Bonin, A., Zinger, L., and Coissac, E. (2018). *Environmental DNA: For Biodiversity Research and*
808 *Monitoring*. Oxford University Press, Oxford, DOI: [10.1093/oso/9780198767220.001.0001](https://doi.org/10.1093/oso/9780198767220.001.0001).
- 809 Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., and Willerslev, E. (2012). Towards next-
810 generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21:2045–2050, DOI:
811 [10.1111/j.1365-294X.2012.05470.x](https://doi.org/10.1111/j.1365-294X.2012.05470.x).
- 812 Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., and Willerslev, E. (2012a). Detection
813 of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE*, 7:e41732,
814 DOI: [10.1371/journal.pone.0041732](https://doi.org/10.1371/journal.pone.0041732).
- 815 Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., Orlando, L.,
816 and Willerslev, E. (2012b). Monitoring endangered freshwater biodiversity using environmental DNA.
817 *Molecular Ecology*, 21:2565–2573, DOI: [10.1111/j.1365-294X.2011.05418.x](https://doi.org/10.1111/j.1365-294X.2011.05418.x).
- 818 Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., and Willerslev, E. (2016).
819 Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes.
820 *PLoS ONE*, 11:e0165252, DOI: [10.1371/journal.pone.0165252](https://doi.org/10.1371/journal.pone.0165252).
- 821 Thomsen, P. F. and Willerslev, E. (2015). Environmental DNA - An emerging tool in conser-
822 vation for monitoring past and present biodiversity. *Biological Conservation*, 183:4–18, DOI:
823 [10.1016/j.biocon.2014.11.019](https://doi.org/10.1016/j.biocon.2014.11.019).
- 824 Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen,
825 S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, 40:e115, DOI:
826 [10.1093/nar/gks596](https://doi.org/10.1093/nar/gks596).
- 827 Ushio, M., Murakami, H., Masuda, R., Sado, T., Miya, M., Sakurai, S., Yamanaka, H., Minamoto, T., and
828 Kondoh, M. (2018). Quantitative monitoring of multispecies fish environmental DNA using high-throughput
829 sequencing. *Metabarcoding and Metagenomics*, 2:1–15, DOI: [10.3897/mbmg.2.23297](https://doi.org/10.3897/mbmg.2.23297).
- 830 Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., Bellemain, E., Besnard, A.,
831 Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G. H., Geniez, P., Pont, D.,
832 Argillier, C., Baudoin, J. M., Peroux, T., Crivelli, A. J., Olivier, A., Acqueberge, M., Le Brun, M., Møller,
833 P. R., Willerslev, E., and Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using
834 environmental DNA metabarcoding. *Molecular Ecology*, 25:929–942, DOI: [10.1111/mec.13428](https://doi.org/10.1111/mec.13428).
- 835 Vandamme, S. G., Griffiths, A. M., Taylor, S.-A., Di Muri, C., Hankard, E. A., Towne, J. A., Watson,
836 M., and Mariani, S. (2016). Sushi barcoding in the UK: another kettle of fish. *PeerJ*, 4:e1891, DOI:
837 [10.7717/peerj.1891](https://doi.org/10.7717/peerj.1891).
- 838 Wangenstein, O. S., Palacín, C., Guardiola, M., and Turon, X. (2018). DNA metabarcoding of littoral
839 hard-bottom communities: high diversity and database gaps revealed by two molecular markers. *PeerJ*,
840 6:e4705, DOI: [10.7717/peerj.4705](https://doi.org/10.7717/peerj.4705).
- 841 Ward, R. D. (2009). DNA barcode divergence among species and genera of birds and fishes. *Molecular*
842 *Ecology Resources*, 9:1077–1085, DOI: [10.1111/j.1755-0998.2009.02541.x](https://doi.org/10.1111/j.1755-0998.2009.02541.x).
- 843 Ward, R. D., Hanner, R., and Hebert, P. D. N. (2009). The campaign to DNA barcode all fishes, FISH-BOL.
844 *Journal of Fish Biology*, 74:329–56, DOI: [10.1111/j.1095-8649.2008.02080.x](https://doi.org/10.1111/j.1095-8649.2008.02080.x).
- 845 Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., and Hebert, P. D. N. (2005). DNA barcoding Australia's
846 fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360:1847–1857, DOI:

847 [10.1098/rstb.2005.1716](https://doi.org/10.1098/rstb.2005.1716).

848 Wilcox, T. M., Zarn, K. E., Piggott, M. P., Young, M. K., McKelvey, K. S., and Schwartz, M. K. (2018).
849 Capture enrichment of aquatic environmental DNA: A first proof of concept. *Molecular Ecology Resources*,
850 18:1392–1401, DOI: [10.1111/1755-0998.12928](https://doi.org/10.1111/1755-0998.12928).

851 Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., Bunce, M., Wiuf, C.,
852 Gilichinsky, D. A., and Cooper, A. (2003). Diverse plant and animal genetic records from Holocene and
853 Pleistocene sediments. *Science*, 300:791–795, DOI: [10.1126/science.1084114](https://doi.org/10.1126/science.1084114).

854 Winter, D. J. (2017). rentrez: an R package for the NCBI eUtils API. *The R Journal*, 9:520–526.

855 Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., Minamoto, T., and Miya, M. (2017).
856 Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific*
857 *Reports*, 7:40368, DOI: [10.1038/srep40368](https://doi.org/10.1038/srep40368).

858 Yang, C., Wang, X., Miller, J. A., De Blécourt, M., Ji, Y., Yang, C., Harrison, R. D., and Yu, D. W. (2014). Us-
859 ing metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity
860 indicator. *Ecological Indicators*, 46:379–389, DOI: [10.1016/j.ecolind.2014.06.028](https://doi.org/10.1016/j.ecolind.2014.06.028).

861 Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., and Ding, Z. (2012). Biodiversity soup:
862 metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and*
863 *Evolution*, 3:613–623, DOI: [10.1111/j.2041-210X.2012.00198.x](https://doi.org/10.1111/j.2041-210X.2012.00198.x).

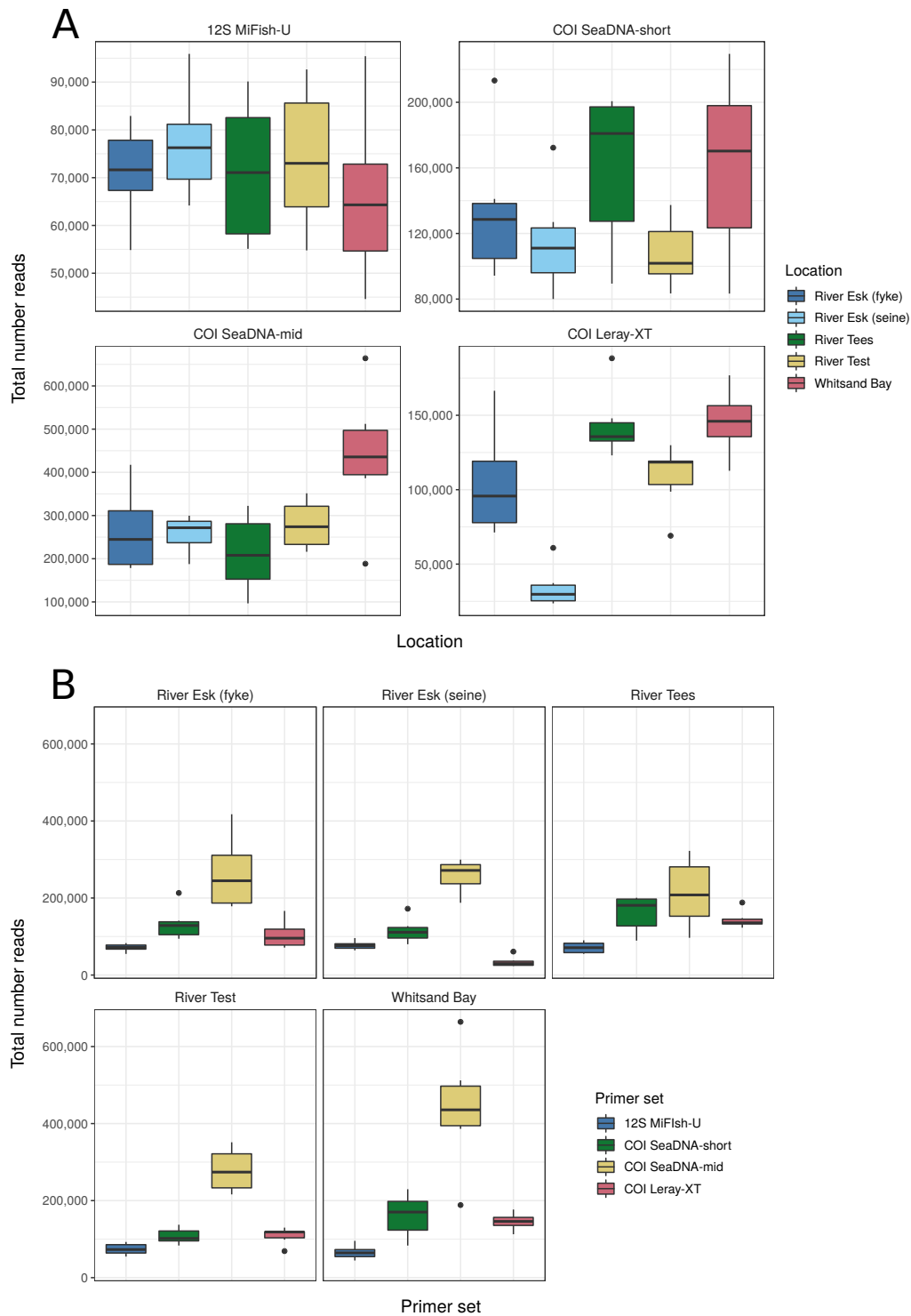
864

Supplementary information for:
**Non-specific amplification compromises
environmental DNA metabarcoding with COI**

Rupert A. Collins, Judith Bakker, Owen S. Wangensteen, Ana Z. Soto,
Laura Corrigan, David W. Sims, Martin J. Genner, Stefano Mariani

June 17, 2019

865



Supplementary Figure 1: (A) Read depth (after bioinformatic processing) per location by primer set. (B) Read depth per primer set by location. Per primer-location combination there are three water sample replicates and for each of these, two uniquely tagged PCR replicates ($n = 6$). The horizontal represents the median value, the boxes represent the 25–75th percentiles, the whiskers represent the values less than 1.5 times the interquartile range, and dots represent the outlying data points.

866

Supplementary Table 1: Unassigned 12S fish reads ($n = 73,377$) obtained from the *in vitro* analyses of water samples taken from five sites. These were reads that were not assigned to species using our curated reference database of UK fishes, under the criteria of: (i) species-level EPA placement same as the best scoring blast hit, with an aligned match length of $\geq 90\%$ of the modal length of the fragment, and an identity of $\geq 97\%$; or (ii) highest likelihood EPA placement same as the best scoring blast hit, with an EPA probability $\geq 90\%$ and blast identity $\geq 90\%$. The assumed identification is reported after conducting additional phylogenetic analyses, additional BLAST searches, and considering the most common species in the area and the species missing from the reference library. There was also a total of 198,445 reads assigned to *Hippoglossoides platessoides*, which were most probably mis-assigned, and actually belong to the more common pleuronectiform species, such as *Pleuronectes platessa* and *Platichthys flesus*, that were absent from the reference library. Likewise 18,746 reads were assigned to *Syngnathus typhle*, but are more likely to belong to *Syngnathus acus* or *Syngnathus rostellatus*. OTUs (operational taxonomic units) were clustered using *Swarm* from the ASVs (amplicon sequence variants) produced by *dada2*.

OTU	Number ASVs	Total reads	GenBank BLAST match	EPA identification	Assumed species	Comment	Locations
otu11	2	33,143	<i>Chelidonichthys spinosus</i>	Actinopterygii	<i>Chelidonichthys lucerna</i>	Reference not in library	Test, Tees, Esk-Seine, Whitsand Bay, Esk-Fyke
otu23	3	14,512	<i>Gaidropsarus argentatus</i>	Lotidae	<i>Gaidropsarus vulgaris</i>	Reference not in library	Esk-Seine, Tees, Whitsand Bay
otu26	1	10,086	<i>Labrus merula</i>	<i>Labrus mixtus</i>	<i>Labrus bergylla</i>	Reference not in library	Tees, Whitsand Bay, Esk-Fyke, Test
otu27	2	9,179	<i>Ammodytes personatus</i>	<i>Ammodytes americanus</i>	<i>Ammodytes tobianus</i>	Reference not in library	Tees, Whitsand Bay, Esk-Fyke, Test
otu35	1	2,677	<i>Symphodus ocellatus</i>	<i>Symphodus melops</i>	<i>Symphodus melops</i>	Misidentified by BLAST due to short reference	Test
otu46	2	1,851	<i>Parablennius yatabei</i>	Vertebrata, Gobiidae	<i>Parablennius gatrongine</i>	Reference not in library	Tees, Test, Esk-Seine, Esk-Fyke
otu45	1	1,119	<i>Eleutherochir maccaddeni</i>	Lophiiformes	<i>Callionymus lyra</i>	Reference not in library	Whitsand Bay, Esk-Seine, Test
otu50	2	746	<i>Psettina iijimae</i>	Stomiiformes	<i>Pleuronectiformes sp.</i>	Possibly <i>Arnoglossus</i>	Test, Whitsand Bay
otu56	1	12	<i>Clupea harengus</i>	<i>Clupea harengus</i>	<i>Clupea harengus</i>	Sequencing/PCR error	Tees
otu57	1	9	<i>Sprattus sprattus</i>	<i>Clupea harengus</i>	<i>Clupea ot Sprattus</i>	Sequencing/PCR error	Tees
otu58	1	7	<i>Clupea harengus</i>	<i>Clupea harengus</i>	<i>Clupea ot Sprattus</i>	Sequencing/PCR error	Tees
otu59	1	7	<i>Lepidopsetta bilineata</i>	Actinopterygii	<i>Pleuronectiformes sp.</i>	Possible sequencing/PCR error	Esk-Fyke
otu60	1	7	<i>Chelidonichthys kumu</i>	<i>Alepisaurus ferox</i>	<i>Clupeidae</i>	Sequencing/PCR error	Whitsand Bay
otu61	1	7	<i>Conger erebennus</i>	<i>Nessorhamphus ingolfianus</i>	<i>Conger conger</i>	Reference not in library	Whitsand Bay
otu62	1	7	<i>Salmo trutta</i>	<i>Salmo trutta</i>	<i>Salmo trutta</i>	Sequencing/PCR error	Whitsand Bay, Esk-Seine
otu64	1	3	<i>Chelidonichthys spinosus</i>	Perciformes	<i>Chelidonichthys lucerna</i>	Reference not in library	Esk-Fyke, Esk-Seine
otu66	1	3	<i>Enchelyopus cimbrius</i>	Lotidae	<i>Gaidropsarus vulgaris</i>	Reference not in library	Tees
otu67	1	2	<i>Clupea pallasii</i>	<i>Clupea harengus</i>	<i>Clupea harengus</i>	Sequencing/PCR error	Tees

867

Supplementary Table 2: Metabarcoding and traditional fish survey results for the River Tees site survey. Values correspond to the number of reads identified to species for the molecular markers, and the number of individuals caught on the traditional surveys. Species separated by semicolon are those for which matches were ambiguous. Predominantly freshwater species that are generally not caught on the traditional surveys, are highlighted with an asterisk.

Species	Traditional	12S MiFish-U	COI Leray-XT	COI SeaDNA-mid	COI SeaDNA-short
<i>Ammodytes tobianus</i>	1				
<i>Anguilla anguilla</i>		85			
<i>Aphia minuta</i>		5		2	211
<i>Atherina boyeri</i>		34	3		
<i>Barbatula barbatula*</i>		42		2	
<i>Chelon labrosus; Liza ramada</i>		43			
<i>Clupea harengus</i>	29	98,907			10
<i>Clupea harengus; Sprattus sprattus</i>		137,123			
<i>Cottus gobio*</i>		25			
<i>Cyclopterus lumpus</i>		8			
<i>Dicentrarchus labrax</i>		265			
<i>Gadus morhua</i>		41,495			
<i>Gasterosteus aculeatus*</i>		30		4	
<i>Gobio gobio*</i>		22		2	
<i>Gobius paganellus</i>		33			3
<i>Hippoglossoides platessoides</i>		13,968			
<i>Limanda limanda</i>	1				
<i>Melanogrammus aeglefinus; Merlangius merlangus</i>		71			
<i>Merlangius merlangus</i>				14	38
<i>Molva molva</i>					31
<i>Oncorhynchus mykiss*</i>		139		83	159
<i>Perca fluviatilis*</i>		19			
<i>Phoxinus phoxinus*</i>		38			
<i>Platichthys flesus</i>	1				
<i>Pleuronectes platessa</i>	12				
<i>Pomatoschistus microps</i>					6
<i>Pomatoschistus minutus</i>	3	24,247			
<i>Salmo salar*</i>				7	
<i>Salmo trutta*</i>		13,086		713	158
<i>Sardina pilchardus</i>		307			
<i>Scomber scombrus</i>		101	3		
<i>Sprattus sprattus</i>	233		3	4	
<i>Squalius cephalus*</i>		8			
<i>Syngnathus acus</i>					47
<i>Syngnathus rostellatus</i>				4	
<i>Syngnathus typhle</i>		16			
<i>Taurulus bubalis</i>		29,189			
<i>Trachurus trachurus</i>		198		10	
<i>Trisopterus luscus</i>					3
<i>Trisopterus minutus</i>		28			
<i>Zeugopterus punctatus</i>			7		

868

Supplementary Table 3: Metabarcoding and traditional fish survey results for the River Esk (fyke) site survey. Values correspond to the number of reads identified to species for the molecular markers, and the number of individuals caught on the traditional surveys. Species separated by semicolon are those for which matches were ambiguous. Predominantly freshwater species that are generally not caught on the traditional surveys, are highlighted with an asterisk.

Species	Traditional	12S MiFish-U	COI Leray-XT	COI SeaDNA-mid	COI SeaDNA-short
<i>Anguilla anguilla</i>	3	364			
<i>Aphia minuta</i>		25			1,105
<i>Atherina boyeri</i>		50			
<i>Barbatula barbatula*</i>		183		4	
<i>Chelidonichthys lucerna</i>				179	
<i>Chelon labrosus; Liza ramada</i>		18			
<i>Ciliata mustela</i>	11				
<i>Clupea harengus</i>		9,258			65
<i>Clupea harengus; Sprattus sprattus</i>		367			
<i>Cottus gobio*</i>		26			
<i>Cyclopterus lumpus</i>		8			
<i>Dicentrarchus labrax</i>		165			
<i>Eutrigla gurnardus</i>				27	5
<i>Gadus morhua</i>	16	23,958		53	690
<i>Gasterosteus aculeatus*</i>		51			
<i>Gobio gobio*</i>		85			10
<i>Gobius paganellus</i>		97			7
<i>Hippoglossoides platessoides</i>		45,006			
<i>Lampetra fluviatilis; Lampetra planeri*</i>		1,562			81
<i>Melanogrammus aeglefinus; Merlangius merlangus</i>		13			
<i>Merlangius merlangus</i>					159
<i>Molva molva</i>		16,319	12		4,443
<i>Oncorhynchus mykiss*</i>		271		32	92
<i>Perca fluviatilis*</i>		14			
<i>Phoxinus phoxinus*</i>		171			
<i>Platichthys flesus</i>	12				
<i>Pleuronectes platessa</i>	2				
<i>Pollachius pollachius</i>	2	1,704			
<i>Pollachius virens</i>	11				
<i>Pomatoschistus minutus</i>		10,794			42
<i>Salmo salar*</i>		13	22	415	
<i>Salmo trutta*</i>		73,142	172	15,963	1,871
<i>Sardina pilchardus</i>		81			
<i>Scomber scombrus</i>		15,844			
<i>Squalius cephalus*</i>		10			
<i>Syngnathus acus</i>					340
<i>Taurulus bubalis</i>	19	17			
<i>Trachurus trachurus</i>		38			14
<i>Trisopterus luscus</i>		5			
<i>Trisopterus minutus</i>		12			
<i>Zeugopterus punctatus</i>			16		
<i>Zoarces viviparus</i>	2				

869

Supplementary Table 4: Metabarcoding and traditional fish survey results for the River Esk (seine) site survey. Values correspond to the number of reads identified to species for the molecular markers, and the number of individuals caught on the traditional surveys. Species separated by semicolon are those for which matches were ambiguous. Predominantly freshwater species that are generally not caught on the traditional surveys, are highlighted with an asterisk.

Species	Traditional	12S MiFish-U	COI Leray-XT	COI SeaDNA-mid	COI SeaDNA-short
<i>Anguilla anguilla</i>		20,004			31
<i>Aphia minuta</i>		4			191
<i>Atherina boyeri</i>		17			
<i>Barbatula barbatula</i> *		11,688		558	8
<i>Chelon labrosus; Liza ramada</i>		83			
<i>Clupea harengus</i>		225			309
<i>Clupea harengus; Sprattus sprattus</i>		278			
<i>Cottus gobio</i> *		7			
<i>Dicentrarchus labrax</i>		184			
<i>Gadus morhua</i>		224			
<i>Gasterosteus aculeatus</i> *		6,633			
<i>Gobio gobio</i> *		4,349		331	697
<i>Gobius paganellus</i>		21			
<i>Hippoglossoides platessoides</i>		45,936			
<i>Lampetra fluviatilis; Lampetra planeri</i> *					10
<i>Melanogrammus aeglefinus; Merlangius merlangus</i>		14			
<i>Merlangius merlangus</i>					165
<i>Molva molva</i>		9			
<i>Oncorhynchus mykiss</i> *		114		32	94
<i>Phoxinus phoxinus</i> *		43,149	31		
<i>Platichthys flesus</i>	2				
<i>Pleuronectes platessa</i>	1				
<i>Pomatoschistus microps</i>					89
<i>Pomatoschistus minutus</i>		79			
<i>Salmo salar</i> *		3,424	71	3,770	290
<i>Salmo trutta</i> *	2	177,271	703	74,004	6,108
<i>Sardina pilchardus</i>		260			
<i>Scomber scombrus</i>		220			
<i>Spondyliosoma cantharus</i>			4		
<i>Sprattus sprattus</i>	1				
<i>Syngnathus acus</i>					50
<i>Syngnathus typhle</i>		97			
<i>Taurulus bubalis</i>		33			
<i>Trachurus trachurus</i>		99		4	
<i>Trisopterus minutus</i>		47			
<i>Zoarces viviparus</i>				53	

870

Supplementary Table 5: Metabarcoding and traditional fish survey results for the Whitsand Bay site survey. Values correspond to the number of reads identified to species for the molecular markers, and the number of individuals caught on the traditional surveys. Species separated by semicolon are those for which matches were ambiguous. Predominantly freshwater species that are generally not caught on the traditional surveys, are highlighted with an asterisk.

Species	Traditional	12S MiFish-U	COI Leray-XT	COI SeaDNA-mid	COI SeaDNA-short
<i>Ammodytes tobianus</i>	52				
<i>Anguilla anguilla</i>		90			
<i>Aphia minuta</i>		11			1,322
<i>Arnoglossus laterna</i>	13		2		
<i>Atherina boyeri</i>		26			
<i>Barbatula barbatula*</i>		31			
<i>Buglossidium luteum</i>	16				
<i>Callionymus lyra</i>	21		3		
<i>Centrolabrus exoletus</i>			2		
<i>Chelidonichthys lucerna</i>	1				
<i>Chelon labrosus; Liza ramada</i>		11,013			
<i>Clupea harengus</i>		7,016			
<i>Clupea harengus; Sprattus sprattus</i>		177			
<i>Conger conger</i>				2	
<i>Cottus gobio*</i>		19			
<i>Ctenolabrus rupestris</i>				44	
<i>Cyclopterus lumpus</i>		4			
<i>Dicentrarchus labrax</i>		30,746			
<i>Echiichthys vipera</i>	7				
<i>Eutrigla gurnardus</i>	26				
<i>Gadus morhua</i>		167			39
<i>Gasterosteus aculeatus*</i>		8			
<i>Gobio gobio*</i>		24			
<i>Gobius paganellus</i>		67			
<i>Hippoglossoides platessoides</i>		90,664			
<i>Hyperoplus immaculatus</i>	24				
<i>Hyperoplus lanceolatus</i>	1				
<i>Limanda limanda</i>	8				
<i>Lophius piscatorius</i>	3				
<i>Melanogrammus aeglefinus; Merlangius merlangus</i>		13,398			
<i>Merlangius merlangus</i>	6			16	87
<i>Molva molva</i>		10			
<i>Mullus surmuletus</i>	7				
<i>Oncorhynchus mykiss*</i>		191		6	32
<i>Pagrus pagrus</i>	10				
<i>Pegusa lascaris</i>	4				
<i>Perca fluviatilis*</i>		5			
<i>Phoxinus phoxinus*</i>		49			
<i>Pleuronectes platessa</i>	71				
<i>Pomatoschistus microps</i>					5
<i>Pomatoschistus minutus</i>	192	55			
<i>Raja brachyura</i>	1				
<i>Raja clavata</i>	3				
<i>Raja microocellata</i>	2				
<i>Raja montagui</i>	6				
<i>Salmo trutta*</i>		427		237	149
<i>Sardina pilchardus</i>		89,488			150
<i>Scomber scombrus</i>		15,546			
<i>Scophthalmus maximus</i>	8				
<i>Scophthalmus rhombus</i>	3				3
<i>Scyliorhinus canicula</i>	1				
<i>Solea solea</i>	3	4		4	
<i>Squalius cephalus*</i>		6			
<i>Syngnathus acus</i>					287
<i>Syngnathus rostellatus</i>				122	
<i>Syngnathus typhle</i>		18,597			
<i>Taurulus bubalis</i>		19			4
<i>Trachurus trachurus</i>	4	49,801		274	209
<i>Trisopterus luscus</i>		7			
<i>Trisopterus minutus</i>		12,953	7		
<i>Zeus faber</i>				4	5

871

Supplementary Table 6: Metabarcoding and traditional fish survey results for the River Test site survey. Values correspond to the number of reads identified to species for the molecular markers, and the number of individuals caught on the traditional surveys. Species separated by semicolon are those for which matches were ambiguous. Predominantly freshwater species that are generally not caught on the traditional surveys, are highlighted with an asterisk.

Species	Traditional	12S MiFish-U	COI Leray-XT	COI SeaDNA-mid	COI SeaDNA-short
<i>Abramis brama</i> *					8
<i>Anguilla anguilla</i>	2	1,704			
<i>Aphia minuta</i>	111	6,493	7	242	220
<i>Atherina boyeri</i>	240	6,154	5	159	
<i>Barbatula barbatula</i> *		2,470	7	11	
<i>Belone belone</i>			6		
<i>Chelidonichthys lucerna</i>				6	
<i>Chelon labrosus</i> ; <i>Liza ramada</i>		17,658			
<i>Ciliata mustela</i>	3		3		
<i>Clupea harengus</i>	24	30,097			288
<i>Clupea harengus</i> ; <i>Sprattus sprattus</i>		21,893			
<i>Cottus gobio</i> *		11,718			
<i>Cyclopterus lumpus</i>		1,609			
<i>Cyprinus carpio</i> *		597			
<i>Dicentrarchus labrax</i>	4	39,417	8		
<i>Gadus morhua</i>		2,841			14
<i>Gasterosteus aculeatus</i> *		13,671		306	14
<i>Gobio gobio</i> *		390	2		
<i>Gobius niger</i>	18				
<i>Gobius paganellus</i>	170	21,225			727
<i>Hippoglossoides platessoides</i>		2,871			
<i>Lampetra fluviatilis</i> ; <i>Lampetra planeri</i> *		215			30
<i>Leuciscus leuciscus</i> *		2,151	2		
<i>Limanda limanda</i>		1,399			
<i>Liparis liparis</i>	1				
<i>Liza aurata</i>		875			
<i>Liza ramada</i>	1				17
<i>Melanogrammus aeglefinus</i> ; <i>Merlangius merlangus</i>		5,364			
<i>Merlangius merlangus</i>	11			56	500
<i>Molva molva</i>		640			46
<i>Oncorhynchus mykiss</i> *		94,231	139	6,263	7,275
<i>Perca fluviatilis</i> *		2,196			
<i>Phoxinus phoxinus</i> *		22,812			
<i>Platichthys flesus</i>	1				
<i>Pleuronectes platessa</i>	1				
<i>Pollachius pollachius</i>		92			
<i>Pomatoschistus microps</i>				14	83
<i>Pomatoschistus minutus</i>	114	12,195			228
<i>Pomatoschistus pictus</i>	3				
<i>Pseudorasbora parva</i> *					35
<i>Raja clavata</i>	1				
<i>Rutilus rutilus</i> *		888		16	13
<i>Salmo salar</i> *			10	46	
<i>Salmo trutta</i> *		12,049	87	5,362	545
<i>Sardina pilchardus</i>		293			
<i>Scardinius erythrophthalmus</i> *		1,361			
<i>Scomber scombrus</i>		505	4		
<i>Scyliorhinus canicula</i>	1				
<i>Solea solea</i>	3	784			
<i>Sprattus sprattus</i>	241		7	24	
<i>Squalius cephalus</i> *		2,646			
<i>Symphodus bailloni</i>	1				
<i>Symphodus melops</i>	2				
<i>Syngnathus rostellatus</i>	1				
<i>Syngnathus typhle</i>		36			
<i>Taurulus bubalis</i>	2	3,171			
<i>Thymallus thymallus</i> *		1,626			7
<i>Trachurus trachurus</i>		216			
<i>Trisopterus luscus</i>	27	3,046	20	2	52
<i>Trisopterus minutus</i>		461			

872 **Supporting Information: Traditional fish survey protocols**

Marchwood Power Station, River Test, Hampshire, Pisces Conservation Ltd.

Outline. Fish entering the station can have four possible fates. They may be returned to sea via the fish return system, they may be washed into the trash basket, captured on the coarse trash screens, or if they are small, they may pass through the station and back to the sea. To estimate the total impingement/entrainment of the station, all possible fates must be quantified. The condition of fish returned to sea is also assessed.

Fish return system monitoring. The fish, invertebrates and weed passing through the fish return system are collected by diverting the flow into a net mounted in the tank built within the system. The water is diverted for a period of 18 hours, usually from 15:15 until 09:15 the following day. A further 6 one-hour samples are then undertaken to complete the full 24-hour monitoring period. The nets used to collect the samples are 1 cm mesh.

From each sample, the debris is sorted and the fish and invertebrates present identified to species. For each fish species present, up to 5 individuals are selected from each size or age class, and their lengths and weights recorded. For fish with no distinct size-classes, individual lengths and weights are recorded for the first 50 individuals. Individual lengths and a combined weight are then recorded for the next 100 individuals of each species. Any further individuals of each species are counted and a combined weight recorded.

Trash basket monitoring. The trash basket is lined with a net, and a 24-hour sample collected and sorted. Fish and invertebrates are measured as described above for the fish return system. The net used to collect the sample is 1 cm mesh.

Trash rake monitoring. A net is placed into the trash skip which receives the rakings from the coarse trash screen. The screens are raked just before the sample is started, and the 24-hr catch is recorded. Mostly the rakings consist of weed and woody debris. The occasional large fish is caught. These data are added to the data on the number of organisms not entering the return system.

⁸⁷³ *Rivers Esk (North Yorkshire) and Tees (County Durham), Environment Agency*

Outline. The Water Framework Directive (WFD) monitoring programme consists of two survey approaches: (i) a suite of methods that include fyke nets, seine nets and small (1.5 metre) beam trawl in the shallower, intertidal parts of each water body. These methods are undertaken twice a year during spring and autumn, in combination per site or per water body, depending upon conditions; (ii) a coastal survey vessel to deploy otter trawls in deeper waters. This method is undertaken once a year during autumn where appropriate.

The combination of results from the above methods provides an assessment of the fish communities present throughout the water body.

Seine netting. Two hauls at least within site area, ideally at low slack (high slack may be needed at shallow upstream sites).

Fyke netting. One deployment per sample station. Use two pairs of nets over a full 12 hour tidal cycle.

1.5 metre beam trawl. One tow of 200 metres.

Data. The transitional fish monitoring programme requires the following mandatory data to be collected at each location for each sample: (i) date, time, trawl duration and tide state; (ii) method used; (iii) equipment used, including net dimensions; (iv) sampler names; (v) fish species present; (vi) abundance of each species; (vii) individual length measurements (freshwater and migratory species record fork length, marine species record total length); (ix) water chemistry data (dissolved oxygen, salinity, temperature; and (x) GPS position.

⁸⁷⁴ *Whitsand Bay, Devon, Marine Biological Association*

For the otter trawl methodology, refer to:

McHugh, M., Sims, D. W., Partridge, J. C., and Genner, M. J. (2011). A century later: Long-term change of an inshore temperate marine fish assemblage. *Journal of Sea Research*, 65:187–194, DOI: 10.1016/j.seares.2010.09.006.