

# Quantitative assessment of fish larvae community composition in spawning areas using metabarcoding of bulk samples

FRANCES C. RATCLIFFE, TAMSYN M. UREN WEBSTER, DEIENE RODRIGUEZ-BARRETO, RICHARD O'RORKE, CARLOS GARCIA DE LEANIZ, AND SOFIA CONSUEGRA<sup>1</sup>

*Department of Biosciences, College of Science, Swansea University, Swansea SA2 8PP UK*

*Citation:* Ratcliffe, F. C., Uren Webster, T. M., Rodriguez-Barreto, D., O'Rorke, R., Garcia de Leaniz, C., and Consuegra, S. 2021. Quantitative assessment of fish larvae community composition in spawning areas using metabarcoding of bulk samples. *Ecological Applications* 00(00):e02284. 10.1002/eap.2284

**Abstract.** Accurate assessment of larval community composition in spawning areas is essential for fisheries management and conservation but is often hampered by the cryptic nature of many larvae, which renders them difficult to identify morphologically. Metabarcoding is a rapid and cost-effective method to monitor early life stages for management and environmental impact assessment purposes but its quantitative capability is under discussion. We compared metabarcoding with traditional morphological identification to evaluate taxonomic precision and reliability of abundance estimates, using 332 fish larvae from multinet hauls (0–50 m depth) collected at 14 offshore sampling sites in the Irish and Celtic seas. To improve quantification accuracy (relative abundance estimates), the amount of tissue for each specimen was standardized and mitochondrial primers (12S gene) with conserved binding sites were used. Relative family abundance estimated from metabarcoding reads and morphological assessment were positively correlated, as well as taxon richness ( $R_S = 0.81$ ,  $P = 0.007$ ) and diversity ( $R_S = 0.90$ ,  $P = 0.002$ ). Spatial patterns of community composition did not differ significantly between metabarcoding and morphological assessments. Our results show that DNA metabarcoding of bulk tissue samples can be used to monitor changes in fish larvae abundance and community composition. This represents a feasible, efficient, and faster alternative to morphological methods that can be applied to terrestrial and aquatic habitats.

**Key words:** 12S; bulk samples; Celtic sea; fish larvae; Irish sea; metabarcoding; quantification.

## INTRODUCTION

Assessing larval community composition is needed to provide accurate information about spawning areas for fisheries management and conservation, but the location and dispersal of larval stages are largely unknown aspects of many fish life cycles (Legrand et al. 2019). Early life stages of organisms are particularly sensitive to abiotic stressors (Radchuk et al. 2013) and, for fish, understanding the quantitative relationship between environmental quality and population dynamics remains challenging (Rose 2000). Thus, larval monitoring provides critical information about population changes over time (Asch 2015) to inform conservation and policy (Ellis et al. 2012, Borja et al. 2017), but its application is often hampered by the cryptic morphology of early life-stage organisms (Brechon et al. 2013, Sigut et al. 2017, Kimmerling et al. 2018).

Traditional fish larvae monitoring involves identifying each individual using a light microscope, counting

myotomes, assessing pigmentation patterns and jaw morphology (Russell 1976). Yet, identification keys are incomplete for many parts of the world (Becker et al. 2015) and, where descriptions are available, morphological assessment is time consuming and requires specialist training (Brechon et al. 2013). Morphological taxonomy also relies on the identifying features remaining intact for species level assignment (Russell 1976), but damage is common during sampling (e.g., when using continuous plankton recorders), leading to misidentification and loss of valuable information (Richardson et al. 2006).

In cases where morphological identification is unfeasible, DNA sequencing technologies may be used to identify organisms, as long as their sequences are in the databases (Taberlet et al. 2012). The development of high-throughput sequencing technology allows amplicon-based sequencing (metabarcoding) of multiple individuals of various species concurrently (i.e., bulk samples), providing a relatively quick method of processing many samples to obtain taxonomical information (Taberlet et al. 2012) and estimate biodiversity (Dopheide et al. 2019). However, obtaining accurate absolute abundance (number of individuals) estimates through relative read abundance (RRA) from amplicon sequence data has remained challenging (Deagle et al.

Manuscript received 5 May 2020; revised 10 August 2020; accepted 5 October 2020. Corresponding Editor: Éva E. Plaganyi.

<sup>1</sup>Corresponding Author. E-mail: s.consuegra@swansea.ac.uk

2019, Lamb et al. 2019). This is because biases in RRA estimations can be introduced at different stages of the metabarcoding protocol, for example, cell and DNA quantity, mitochondrial copy number, extraction success and PCR amplification rates can vary between tissue type and species (Lamb et al. 2019, Piñol et al. 2019), leading to inaccurate estimates. Another source of bias can arise from unequal body size of individuals pooled within a bulk sample, which can be mitigated by size fractioning of organisms prior to extraction, increasing the reliability of RRA estimates (Elbrecht et al. 2017). The choice of primers and target region may introduce further bias (Deagle et al. 2014). These biases have led to designing costly and bioinformatically challenging metagenomic approaches (Tang et al. 2015, Kimmerling et al. 2018) or to the use of multiple loci (Richardson et al. 2015) to identify particular species and estimate their abundance.

Improving the reliability of abundance estimates is thus needed to make metabarcoding more useful for biodiversity monitoring, calculation of metrics such as diversity indices, as well as detection of natural shifts in multispecies community composition (Bohmann et al. 2014). Different approaches have been proposed to improve abundance estimates based on RRA, while still using a cost-effective, single-marker, PCR approach (Thomas et al. 2016, Elbrecht et al. 2017). For example, using primers with widely conserved priming sites may reduce taxa specific biases (Krehenwinkel et al. 2017), although taxonomic resolution can be reduced due to highly similar sequences within a family (Thomsen et al. 2016).

Here, using a single mitochondrial marker (12S ribosomal RNA, considered highly specific in fish), we have refined the reliability of DNA metabarcoding abundance estimates by standardizing input material and choosing conserved primer binding sites. Using bulk fish larvae samples from the Irish and Celtic Seas, we compared the sensitivity and accuracy of this approach with traditional morphological identification, to assess whether metabarcoding can be a feasible and rapid alternative to traditional assessment for estimating fish larvae richness, diversity, and community composition metrics.

## MATERIALS AND METHODS

### *Field sampling*

Sampling was carried out onboard the RV Celtic Voyager between 17 and 26 May 2018. Fish larvae (3–30 mm) from 14 hauls (one per site) were sampled using a MultiNet plankton sampler (Hydro-Bios, Kiel, Germany). Sites 1–8 and 12 were sampled with one oblique haul to 50 m depth per site, filtering a mean volume of  $215 \pm 55 \text{ m}^3$  of water. Hauls 9–14 (with the exception of haul 12) consisted of two vertical hauls from the surface to 50 m, filtering a mean volume of  $38 \pm 6 \text{ m}^3$ , which were pooled for each site. Fish larvae from each

haul were separated from other zooplankton species and preserved in RNAlater (Qiagen, Hilden, Germany) at room temperature for 24 h, then refrigerated at 4°C until morphological identification.

### *Morphological identification*

Fish larvae ranged from 2 to 30 mm total length. For morphological identification, all larvae were first separated into major groupings based on body shape following the classification by Russell (1976) and subsequently assigned to family level. Assignment to genus and species where possible, was then carried out. Assignments were checked against the species descriptions first in Russell (1976), and, where possible, double checked against the description by Rodriguez et al. (2017). For taxa that could not be confidently morphologically identified, DNA was extracted from one or more representative individuals (34 individuals of 16 taxa across the survey, Appendix S1: Table S1) using the Qiagen DNeasy Blood and Tissue kit (Qiagen GmbH) following the manufacturer's instructions. Extracted DNA was then amplified using the 12S V5 primers (Riaz et al. 2011a), cleaned using a sodium acetate/EtOH solution, resuspended in 10  $\mu\text{L}$  HiDi Formamide (Applied Biosystems, Foster City, CA, USA) and analysed using Sanger Sequencing on an ABI 3730 DNA Analyzer (Applied Biosystems). Resulting sequences were aligned in BioEdit v 7.2.5 (Hall et al. 2011). When 12S barcoding did not resolve taxonomic identification to species level, due to database limitations or synonymous sequences, the barcoding region of ~650 bp of the CO1 gene (F1, R1; Ward et al. 2005) was used to update taxonomic assignment to the lowest possible taxonomic level, resulting in six additional 12S reference sequences not present in the NCBI nucleotide database (Appendix S1: Table S1; Genbank accession numbers: MN539950, MN539961, MN539952, MN539964, MN539965, MN539966). Taxonomy of Sanger sequenced individuals was assigned to the lowest possible level using the MegaBLAST algorithm (Morgulis et al. 2008) against the National Center for Biotechnology Information (NCBI) GenBank nucleotide database (accessed November 2018) and double checked against the BOLD database (<https://www.boldsystems.org/>). To estimate accuracy and repeatability of taxonomic assignments, a group of 15 specimens were also sent to an experienced taxonomist and verified by CO1 barcoding (morphological taxonomic assignment concordance test).

### *DNA extraction*

After taxonomic identification, bulk tissue samples from all larvae of each haul were prepared for DNA extraction as follows: 2–8 mg of tissue were cut from the area anterior to the tail of each juvenile fish (for individuals <5 mg, the entire larva was used,  $n = 88$ ) and placed in a Falcon tube on ice. Buffer ATL and

proteinase K (Qiagen DNeasy Blood and Tissue kit) were then added to the pooled tissue sample in a ratio of 180  $\mu\text{L}$  of ATL and 20  $\mu\text{L}$  proteinase K for 15 mg of tissue. Each falcon tube (representing one haul) was vortexed thoroughly and incubated overnight to digest at 56°C, shaking at 65 rpm. Samples were visually inspected for tissue remnants, vortexed, and re-incubated until all tissue dissolved. Digestions from each haul were then vortexed for 45 s to ensure thorough mixing of digested products and divided in three sub-samples of 200  $\mu\text{L}$  that were extracted using the Qiagen DNeasy Blood and Tissue kit, following the manufacturer's instructions. Extraction blanks were carried through each step of the process.

#### *Library preparation and sequencing*

A 106-bp fragment of the 12S mitochondrial was amplified with the 12S V5 primers (Riaz et al. 2011b) using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Loughborough, UK), with an annealing temperature of 52°C, in three extraction replicates per haul. Libraries were prepared using a two-step PCR approach, based on the Illumina 16S Metagenomic Sequencing Library preparation guidelines (Illumina, San Diego, California, USA), with following adaptations: in the first PCR step, each extraction replicate was amplified in triplicate in order to increase detection of rare species (Alberdi et al. 2018). Subsequently, 10  $\mu\text{L}$  from each triplicate were pooled prior to first cleanup. Cleanups were performed using Agencourt AMPure XP beads (Beckman Coulter, Brea, California, USA), using a  $1.2 \times$  volume of beads to PCR product. Amplicons were indexed using Nextera XT Index Kit v2 Set C (Illumina), and DNA concentration of each reaction was quantified via Qubit dsDNA HS Assay (Invitrogen, Carlsbad, California, USA) and pooled in equal molar concentrations. PCR and extraction blanks (using molecular grade water instead of template) were subjected to all steps of the library preparation process. In addition, a sequencing/tag jumping blank, where no sample was added prior to sequencing, was used. Pair-end sequencing was carried out at Swansea University using an Illumina MiSeq platform (Illumina) ( $2 \times 300$  bp reads), including 5% PhiX.

#### *Bioinformatics: sequence processing*

De-multiplexed samples containing raw pair-end sequences were processed using Qiime2 (version 2019.1; Bolyen et al. 2019). Initially, raw sequences were quality checked using interactive quality plots, in order to obtain values for sequence trimming and truncation. De-noising was carried out using DADA2 (Callahan et al. 2017) where the first 10 bp of each sequence were trimmed to remove adaptors and all sequences truncated to 100 bp in length based on quality scores. Default DADA2 settings within Qiime2 were used to detect and,

where possible, correct sequencing errors and filter out phiX reads and chimeric sequences, join pair-end reads, and de-replicated sequences. The amplicon sequence variant (ASV) approach was chosen because it provides a higher resolution than a traditional OTU approach, enabling detection of single nucleotide differences (Callahan et al. 2017). After de-noising, the ASV and BIOME tables were exported for taxonomic assignment.

#### *Database construction and taxonomic assignment*

A custom database was constructed using in silico PCR against the NCBI database (downloaded February 2019): 12S V5 primers were allowed to have three base mismatches in silico (search\_PCR command; Edgar 2010) and a corresponding taxonomy file was constructed using the obiannotate tool (OBITools; Boyer et al. 2016). All sequences were trimmed to the target region. A list of all marine fish species encountered in the British Isles, including nonnative fish (366 species; Fish Base: accessed 31 March 2019) was then used to filter the main database to fish species present in the study region, of which 207 were available. The six 12S Sanger sequences (generated with the 12S V5 primers) missing from NCBI database and verified using CO1 barcoding from this study were added to the database (Appendix S1: Table S1), which also included marine mammals, bacteria, and other contaminants (such as *Homo sapiens*) that might be amplified by the primers.

Initially, ASVs were classified using the KNN method in Mothur (Schloss et al. 2009) using the parameter numwanted = 1 (Findley et al. 2013), against the custom database. Because this parameter may lead to false positive assignments, KNN assignments were then verified using NCBI megaBLAST, with max-target sequences = 10. The top 10 assignments were screened for UK species (Fish Base) on a case by case basis. Where the percentage of UK species match fell below 98%, or where multiple UK species matched above a 98% match, MEGAN (6.15.1) was used to assign species to the lowest common ancestor (Huson et al. 2007). ASVs for which there were no vertebrate matches were discarded from downstream analysis.

Tag jumping/cross-contamination (Schnell et al. 2015) was removed on the following basis: a taxon was removed from a haul if it had fewer than 115 reads (maximum reads for a single species in tag jumping control sample) or did not appear in all three replicates.

For spatial analysis, numbers of individuals of each taxon in a haul were estimated from the proportion of reads in the corresponding sample, as follows:

$$A_i = N \times P_i \quad (1)$$

where  $A_i$  is the abundance (number of individuals) of the taxon of interest ( $i$ ) in a given haul,  $N$  is total number of individuals in the haul, and  $P_i$  is the proportion of that taxon in the haul amplicon pool.

### Statistical analysis

The accuracy of estimates of RRA and diversity indices derived from metabarcoding was assessed against results from morphological taxonomy using Spearman's rank correlation analysis performed in R version 3.5.2 (R Core Team 2020). Diversity indices (Shannon Weiner Index, Simpson's Diversity) and richness were estimated based on RRA and morphological relative abundances using the Vegan package (R version 3.5.2) for both lowest possible taxonomic and family level taxon identifications. For spatial analysis, the survey area was divided into three locations along a temperature gradient: Loc 1 (above the Celtic/Irish sea front, 9°–10.99°C), Loc 2 (channel spawning grounds, 11°–12.99°C), and Loc 3 (western Celtic Sea, 13–14°C; Fig. 1). The number of individuals (assessed morphologically) and estimated from reads (Eq. 1), of a given taxon (mean of the three technical replicates per site) were divided by the volume of water filtered in the corresponding haul (Canfield and Jones 1996) to obtain catch per unit filtered (CPUF) or estimated number of individuals from reads per unit filtered (RPUF) values, respectively. This analysis was carried out at both lowest possible taxonomic level and family level. All 14 hauls surveyed were included in this analysis, where only one individual was present in a haul this was divided by the volume of water filtered and included in the both the CPUF and RPUF data sets. The family Ammodytidae was excluded from this analysis, because not all individuals were retained in haul 4. CPUF and RPUF values were square-root transformed, and composition similarity calculated by hierarchical clustering using a Bray-Curtis resemblance matrix. Subsequently, pairwise analysis of similarities (ANOSIM) was used to test whether there was a significant difference in community composition between locations (Clarke 1993), using both the CPUF and RPUF methods. Where significant differences were detected, SIMPER analysis (Clarke 1993) was used to ascertain which taxa accounted for the differences observed. Diversity indices calculations and multivariate spatial analyses were performed using Primer-v7 (Clarke and Gorley 2015).

## RESULTS

### Morphological assessment

A total of 332 fish larvae were caught in 11 of the 14 hauls in the survey. No larvae were encountered in hauls 10, 11, and 14 and only one in hauls 1 and 6, therefore 9 of the 14 hauls were used in metabarcoding. The maximum number of individuals per haul was 63 (haul 2) (Appendix S1: Table S2). Morphological identification assigned 324 (98%) of individuals to family level. It was not possible to assign the families of the remaining eight larvae, due to damaged identifying features. Of those specimens assigned to family level, 255 (77%) were assigned to a genus and 100 (30%) to a species. Sanger

sequencing to check morphological assignment comprising of 34 individuals across nine hauls), contained 15 taxa (Appendix S1: Table S3). In the morphological taxonomic assignment test of the 15 individuals identified by two independent observers and subsequently checked by Sanger sequencing, 100% and 93% were correctly assigned to family level by the first and second observer respectively, 86.3% and 53.3% to genus and 40% to species level in both cases.

Based on morphology alone, before verification with CO1/12S barcoding, taxa within Ammodytidae and Clupeidae could not be assigned further than family level. Most clupeids did not amplify with the CO1 primers and those that did were assigned to *S. sprattus*; for Callionymus there was no *C. reticulatus* sequence to compare with. Incorrect morphological assignments occurred in the cases of *Micromesistius poutassou* (Sanger seq: *M. merlangius*), *Aphia minuta* (Sanger seq: *C. harengus/S. sprattus*) and Mugilidae (Sanger seq: *L. bergylta* and *C. mustela*) (Fig. 2, Appendix S1: Table S3). In addition, *Sardina pilchardus*, *Labrus mixtus/bergylta*, *Molva molva*, and a taxon belonging to the Gobiidae family were only detected using sequencing. In contrast, *M. merlangus* and *Pollachius virens/pollachius*, were identifiable through morphology, but not resolved to species level by 12S metabarcoding due to lack of variability of the 12S fragment. *C. harengus* and *S. sprattus* could not be separated by morphology or 12S metabarcoding.

### Metabarcoding assessment

A total of 3,398,391 raw 300 bp pair-end reads were generated for this study. After Qiime2 DADA de-noising, a total of 2,675,140 reads remained for downstream analysis. Once the taxonomic assignment was complete, reads likely present due to tag jumping from concurrent sample sequencing (Solidae 274 reads, *Scomber scombrus*, 10 reads, *Salmo salar*, 3 reads), and human reads (2,338) were removed from downstream analysis. A total of 49 fish ASVs remained for downstream analysis. Samples contained a mean of 93,223 reads (standard deviation = 31,866) post-filtering, the tag jumping blank contained 146 reads, the PCR blank 64 reads, and extraction blanks 116 and 71 reads, respectively. Tag jumping read removal resulted in 0.046% of reads being excluded from downstream analysis across the samples in the study. Post-filtering, taxa distribution was concordant among the three haul replicates in all nine hauls (Fig. 3).

### Comparison of abundance estimates by morphology and metabarcoding

The relative abundance (percentage) of individuals identified morphologically in a sample and the corresponding RRAs were positively correlated for all families assessed (Spearman's rank: Ammodytidae  $R_S = 0.93$ ,  $P < 0.001$ ; Callionymidae  $R_S = 0.99$ ,  $P < 0.001$ ; Clupeidae  $R_S = 0.97$ ,  $P < 0.001$ ; Gadidae  $R_S = 0.95$ ,

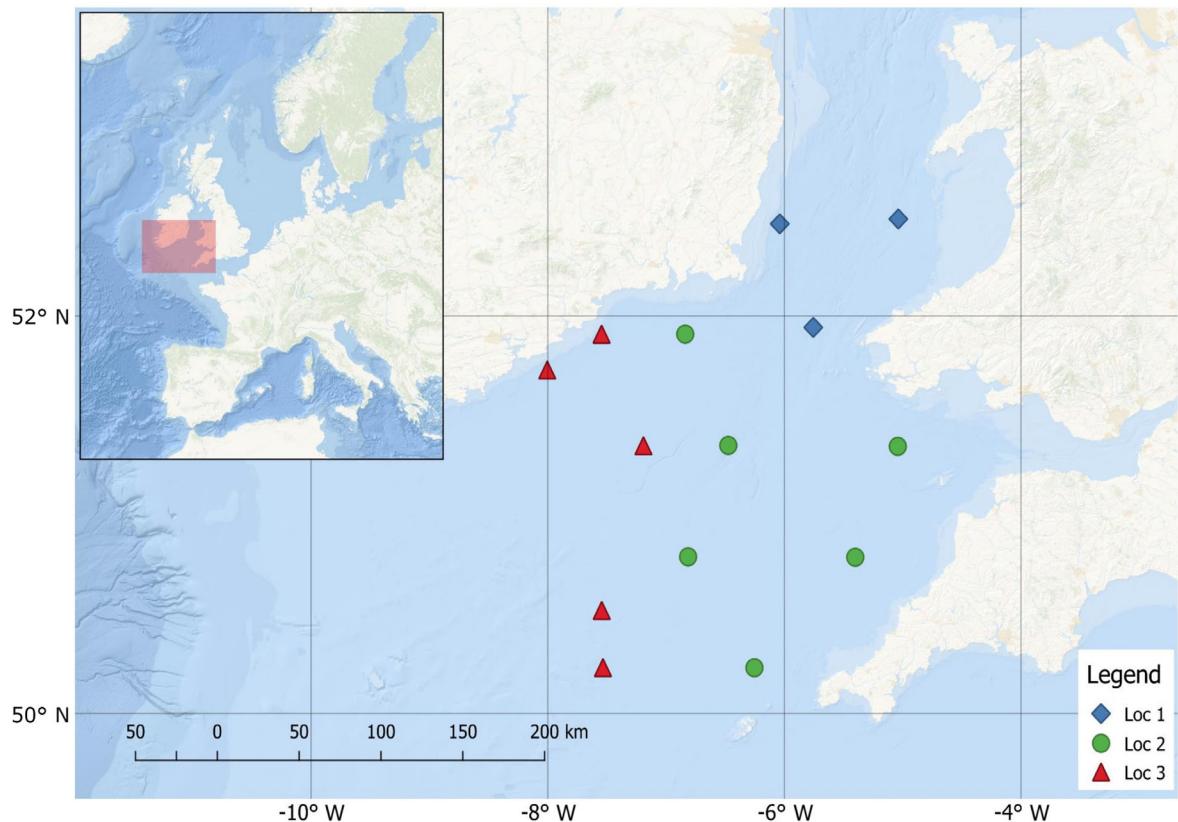


FIG. 1. Multinet haul locations in the Irish and Celtic seas. Locations for spatial analysis, based on SST, are indicated as Loc 1 (above the Celtic/Irish sea front: 9°–10.99°C), Loc 2 (channel spawning grounds: 11°–12.99°C), and Loc 3 (western Celtic Sea: 13°–14°C).

$P < 0.001$ ; Pleuronectidae  $R_S = 0.68$ ,  $P = 0.05$ ; Triglidae  $R_S = 0.88$ ,  $P = 0.002$ ; Appendix S1: Fig. S1). In addition, no difference in diversity and taxon richness were detected between the relative abundance of morphological assignments and RRA assignments at either lowest possible taxonomic level or family level, across hauls (lowest possible taxonomic level; Spearman's rank, richness,  $R_S = 0.84$ ,  $P = 0.005$ ; Shannon index,  $R_S = 0.90$ ,  $P = 0.002$ ; Simpson's diversity,  $R_S = 0.90$ ,  $P = 0.002$ ; family level, Spearman's rank, richness,  $R_S = 0.93$ ,  $P < 0.001$ ; Shannon index,  $R_S = 0.91$ ,  $P = 0.001$ ; Simpson's diversity,  $R_S = 0.80$ ,  $P = 0.01$ ; Fig. 4).

#### *Spatial distribution of larvae assessed by both methods*

Assessment of patterns in community composition yielded comparable results from morphological and metabarcoding assessment at both lowest possible taxonomic level and family level. Catch per unit filtered (CPUF) and back-estimated reads per unit filtered (RPUF), were no different between locations 1 and 2 and 1 and 3, although locations 2 and 3 differed in composition (lowest possible taxonomic level ANOSIM, CPUF  $R = 0.233$ ,  $P = 0.039$ ; RPUF  $R = 0.209$ ,  $P = 0.045$ ; family-level ANOSIM, CPUF  $R = 0.22$ ,

$P = 0.041$ ; RPUF  $R = 0.205$ ,  $P = 0.048$ ; Table 1). SIMPER analysis (percent cumulative dissimilarity contribution) attributed 48.39% (CPUF) and 42.82% (RPUF) of the difference in composition between locations 2 and 3 to three taxa: *C. harengus*/*S. sprattus* (CPUF 21.42%, RPUF 15.74%), Triglidae (CPUF 14.37%, RPUF 14.43%), and *Callionymus* (CPUF 12.61%, RPUF 12.65%) (Appendix S1: Table S4, Fig. S2). The greatest difference observed in dissimilarity contributions for the remaining, less abundant taxa was 2.7% (*C. mustela*). This pattern was repeated at the family level (Appendix S1: Table S4, Fig. S3).

#### DISCUSSION

Here, we demonstrate that metabarcoding is a reliable and practical alternative to traditional morphological assessment. We show that RRA estimates can be achieved by standardizing the amount of tissue analysed per specimen and choosing primers with conserved binding sites. These estimates can then be used to successfully calculate diversity and community composition metrics needed to monitor changes over time. Although more costly in terms of consumables and sequencing, metabarcoding involved considerably less time than

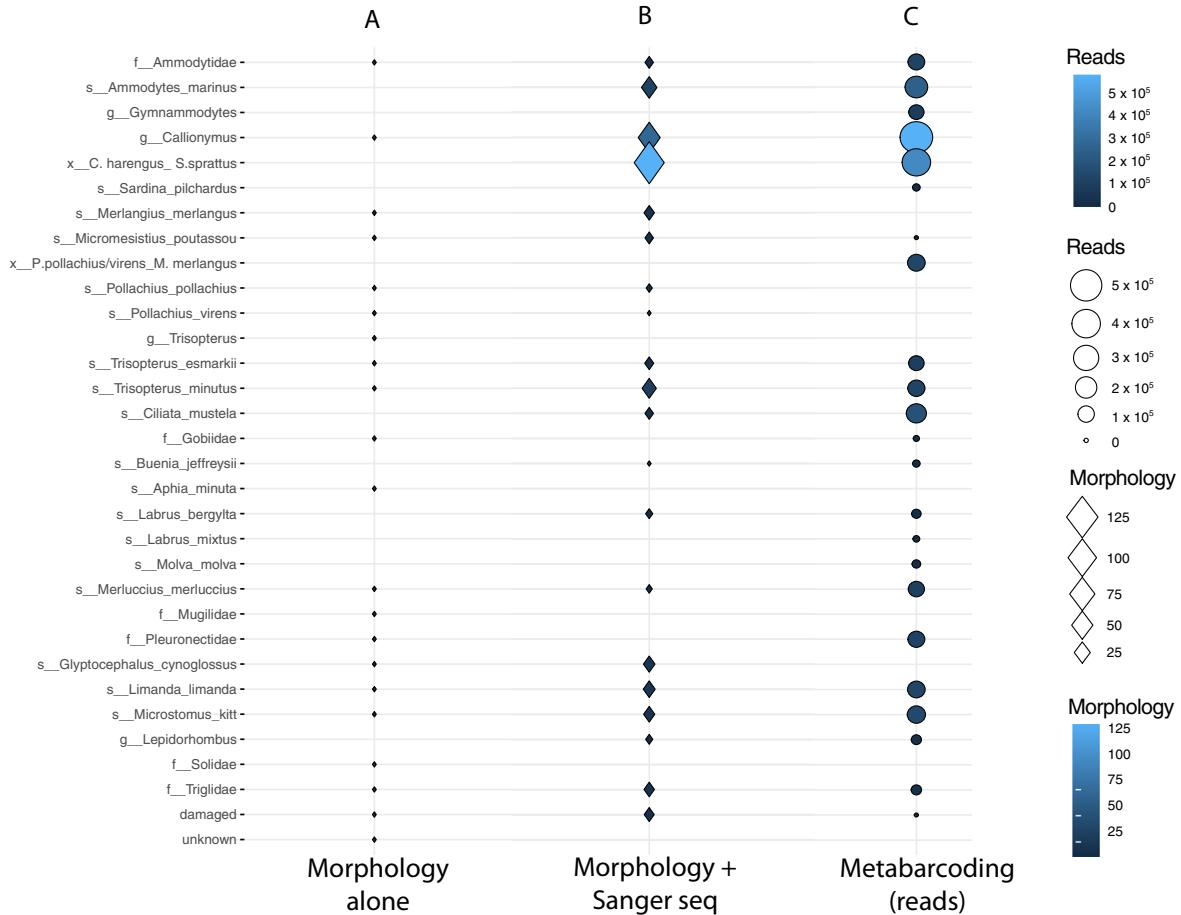


FIG. 2. Overview of larval detections during the survey. (A) Taxonomic assignments using morphology alone (presence/absence). (B) Morphological taxonomic assignments updated with Sanger sequencing, diamonds represent total number of larvae of a taxa observed during the survey. (C) Metabarcoding taxonomic assignments, circles represent total number of reads obtained for each taxa, post-filtering.

morphological identification, particularly for those cases that required additional barcoding to refine the morphological identification. We could use all the individuals collected for metabarcoding, irrespective of their preservation state, while the presence of damaged or poorly preserved specimens made difficult or even impossible their morphological identification, particularly at the species level.

There is considerable debate over whether amplicon sequencing can deliver reliable quantitative data (Deagle et al. 2019). The reliability of abundance estimates from metabarcoding varies considerably between studies, with some showing only a weak correlation between RRA and abundance (Lamb et al. 2019, Piñol et al. 2019). Still, information from RRA tends to be more informative than presence/absence assessments (Deagle et al. 2019). In contrast, metagenomic approaches that do not require PCR amplification can successfully estimate abundance (Kimmerling et al. 2018), although the costs and bioinformatic complexity of this approach may be prohibitive in many contexts (Porter and Hajibabaei

2018). As shown here, amplicon sequencing can be used to estimate abundance and we suggest that further refinements in metabarcoding abundance estimates will enable wider application of amplicon sequencing.

We have shown that the use of approximately equal weights of tissue per individual can improve RRA and diversity estimates. Approaches based on photographically assessing the surface area of taxa and modeling biomass might also eliminate the need for weighing tissue (Kimmerling et al. 2018), although not necessarily reducing time and costs. There are, however, several factors that can bias RRA estimates. For instance, mitochondrial copy number can vary, not only between different species (Piñol et al. 2015), but between different tissue types (Wiesner et al. 1992). We mainly used the region anterior to the tail of each larva to account for one of these biases (tissue type) as much as possible but interspecific biases remain a challenge to the quantitative capabilities of metabarcoding techniques (Deagle et al. 2019). Where they are consistent for a given taxon across all samples within a study, correction factors may

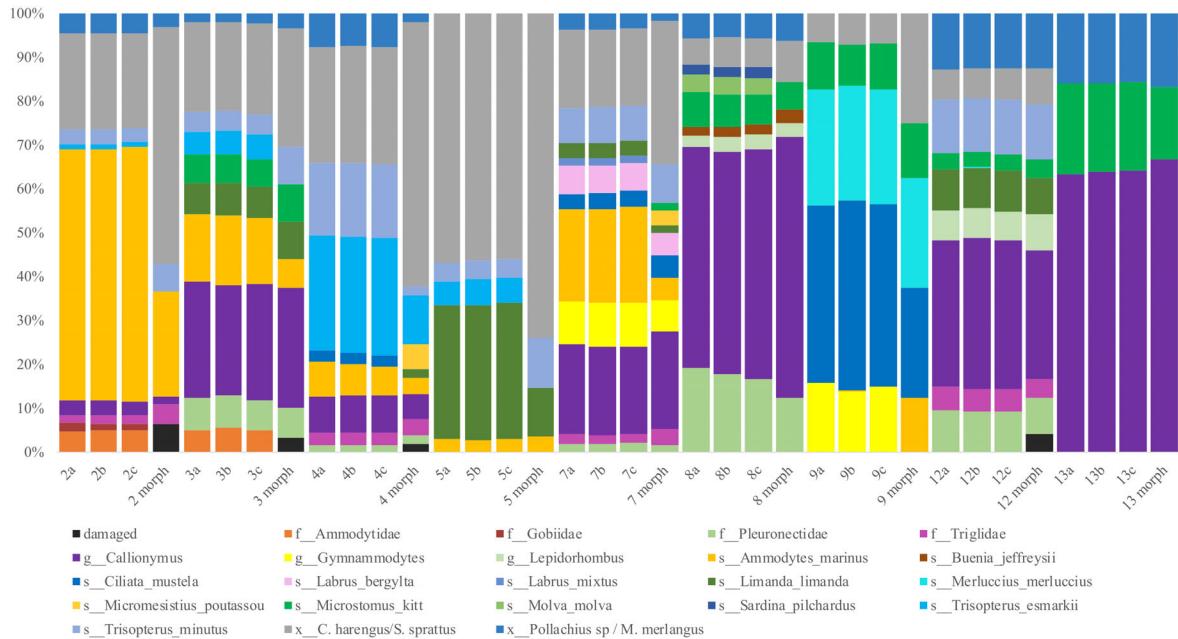


FIG. 3. Comparison of relative read abundances (three replicates per haul, a, b, c samples) and morphological taxonomic assignments, corrected by Sanger sequencing (one per haul, morph samples). Taxa beginning with “f\_” indicates family-level assignment, “s\_” indicates species-level, and “x\_” indicates two to three possible species assignments. Morphological assignments of *P. pollachius/viens*, *M. merlangus* were grouped and morphologically assigned *Glyptocephaus cynoglossus* has been reassigned to Pleuronectidae to match metabarcoding assignments to aid visual interpretation of abundances.

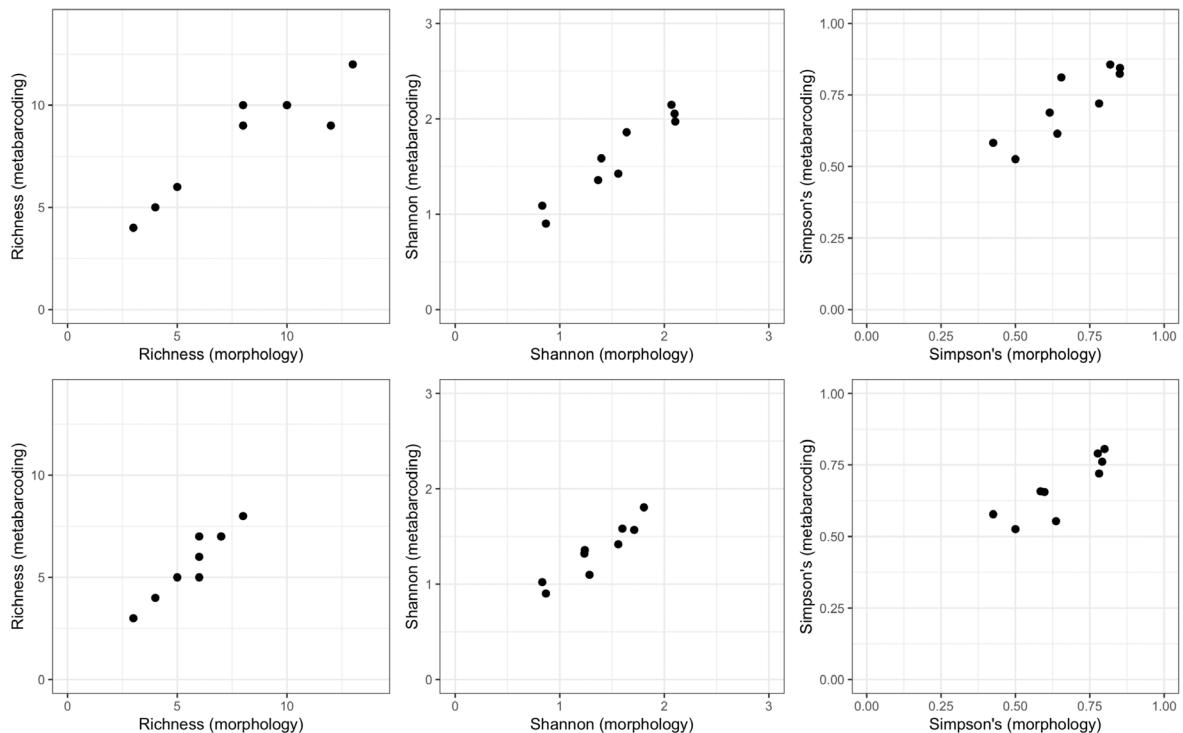


FIG. 4. Consistency of diversity metrics between relative abundances of morphological assignments and relative read abundance assignments, post bioinformatic filtering (mean of three technical replicates per site, for nine sites in the study where >1 larvae was found), for (a) species richness, (b) Shannon Wiener diversity index, and (c) Simpson's diversity (1 -  $\lambda$ ).  $R_s$ , Spearman's rank  $\rho$  values.

TABLE 1. ANOSIM matrix, showing  $R$  values of pairwise comparisons of community composition between three locations in the Irish/Celtic seas, using morphological taxonomic assignments and abundances (CPUF) and metabarcoding taxonomic assignments and back-estimated abundances (RPUF).

CPUF	RPUF		
	Loc 1	Loc 2	Loc 3
Lowest possible taxonomic level			
Loc 1		-0.123	-0.013
Loc 2	-0.111		0.209*
Loc 3	-0.053	0.233*	
Family level			
Loc 1		-0.130	-0.073
Loc 2	-0.123		0.205*
Loc 3	-0.087	0.220*	

\*Significant difference ( $P \leq 0.05$ ) in community composition between two locations.

be applied (Thomas et al. 2016, Krehenwinkel et al. 2017). While we found differences in relative abundance between morphological assessment and metabarcoding, they did not impact the calculation of diversity and community metrics. For example, estimates of the number of individuals of Ammodytidae differed by seven individuals (SD 9.32) compared to those assessed morphologically but this was not sufficient to influence community composition. However, for applications where exact numbers of individuals are needed (e.g., census of particular species) these differences may require consideration.

While there is no perfect marker for all studies (Deagle et al. 2014), we have shown here the benefits of using primers with well conserved binding sites, particularly for RRA estimates. Whereas the CO1 marker has extensive sequence databases as well as a strong capability to discriminate between species, it also carries an increased risk of amplification bias due to the lack of conserved binding sites across a broad range of taxa (Deagle et al. 2014). This can result in false negatives where taxa known to be present in a sample do not amplify (Collins et al. 2019, Nobile et al. 2019). Using more conserved priming sites, such as the 12S marker, may reduce taxa specific biases (Krehenwinkel et al. 2017), although it has been argued that taxonomic resolution may be reduced due to lack of sequence variability within families (Thomsen et al. 2016), and the completeness of reference databases also influences the resolution to species level (Miya et al. 2015). Here, using 12S primers, 40% of the taxa identified with metabarcoding could be assigned to species level, with the rest being assigned to family or genus level. In comparison to morphological identification without the assistance of CO1 Sanger sequencing, 12S metabarcoding achieved higher taxonomic resolution and more accurate identifications to family level. Morphologically assessed groupings supported by barcoding with Sanger sequencing achieved a similar level of assignment at the family level to metabarcoding across the study. Yet, while short reads can struggle to

resolve some families to species level (Thomsen et al. 2016), hindering species level data interpretation, we found that the use of metabarcoding improved taxonomic assignment overall. Morphology only performed better than sequencing in the case of *Glyptocephalus cynoglossus*, due to distinct morphological characteristics, and in a few cases due to lack of information or sequence variation at the 12S region. In general, synonymous sequences at the target region resulted in just two (e.g., *C. harengus/S. sprattus*) or three species (*M. merlangus/P. pollachius/P. virens*) not being distinguished from each other. For studies requiring species-level identification, taxa affected by lack of marker sequence information, or variability, a targeted qPCR approach (Breachon et al. 2013), similar to those carried out to detect particular species using eDNA (Robinson et al. 2018) or a family-specific, multi-primer approach (Riaz et al. 2011b) could be easily used to refine metabarcoding assignments. Combining different markers, as we have done here with the 12S metabarcoding and the CO1 barcoding, can be used to refine the databases by adding novel sequences and by separating species that cannot be identified based solely on small fragments. The completeness of the database used as a reference is critical for the accuracy of the taxonomic assignments and, while databases are continuously increasing in size for the most common metabarcoding makers, given the large diversity of fish and the increasingly lower cost of sequencing, focusing on full mitochondrial genomes may have wider relevance (Collins et al. 2019). We found that in some cases metabarcoding could not resolve identifications to species level, however, for some applications, genus level analysis provides similar diversity and community composition information than species level and would be appropriate, for example to detect responses to environmental change (Hernandez et al. 2013). In some other cases, family-level analysis has been deemed sufficient to detect broadscale changes, e.g., after major environmental disturbance (Hernandez et al. 2013). Therefore, dependent on hypothesis, a single 12S analyses or an additional qPCR can be performed.

Spatial patterns detected in community composition remained the same, independent of whether they were assessed using morphological (CPUF) or metabarcoding (RPUF). The small differences in abundance of rare taxa (Appendix S1: Table S4), were mainly the result of misidentification of *C. mustela* during morphological identification, indicating that metabarcoding of bulk samples may be used as a viable alternative to morphological identification of samples, particularly when the latter proves difficult.

All taxa detected in the survey were known to spawn in the survey area (Acevedo et al. 2002, Ellis et al. 2012) and for the family Ammodytidae, difficult to survey and data limited due to its cryptic morphology (Ellis et al. 2012), metabarcoding identified *A. marinus* and a species of the genus *Gymnammodytes*, further illustrating its potential for detecting cryptic species.

## CONCLUSIONS

We have shown that using a single marker (12S), equal amounts of tissue per sample and estimation of number of individuals from RRA, metabarcoding can provide quantitative abundance estimates for the calculation of alpha and beta diversities. This method could be applied to bulk samples from different terrestrial and marine habitats to improve abundance estimates. Specifically, we recommend the use of markers with highly conserved binding sites and using a small, equally sized pieces of tissue from each specimen to minimize biases and handling steps. This provides a rapid, community level assessment method, that could be used to further understand responses to disturbance and inter-annual or seasonal variability and monitor biodiversity in a changing global climate.

## ACKNOWLEDGMENTS

We are very grateful to those who assisted with sampling: Helen McCormick, Ross O'Neill, Michael Sheridan, Sarah Albuixech-Marti, Katie Costello, and crew of the R.V. Celtic Voyager. This work has been funded by the European Regional Development fund through the Ireland Wales Co-operation Programme 2014–2020 (BlueFish project). Ethics approval: sampling has been conducted following Home Office regulations and approved by Swansea University Ethics Committees under approval No. 181019/1996. S. Consuegra, F. C. Ratcliffe, and C. G. de Leaniz conceived the ideas; F. C. Ratcliffe designed the methodology with help of D. Rodriguez-Barreto; F. C. Rarcliffe and T. M. U. Webster carried out the laboratory work; F. C. Ratcliffe performed bioinformatic analyses with T. M. U. Webster and R. O'Rorke; R. O'Rorke built the species database; F. C. Rarcliffe led the writing of the manuscript with S. Consuegra and all authors contributed critically to the drafts.

## LITERATURE CITED

- Acevedo, S., O. Dwane, and J. M. Fives. 2002. The community structure of larval fish populations in an area of the Celtic Sea in 1998. *Journal of the Marine Biological Association of the United Kingdom* 82:641–648.
- Alberdi, A., O. Aizpurua, M. T. P. Gilbert, and K. Bohmann. 2018. Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution* 9:134–147.
- Asch, R. G. 2015. Climate change and decadal shifts in the phenology of larval fishes in the California Current ecosystem. *Proceedings of the National Academy of Sciences USA* 112: E4065–E4074.
- Becker, R. A., N. G. Sales, G. M. Santos, G. B. Santos, and D. C. Carvalho. 2015. DNA barcoding and morphological identification of neotropical ichthyoplankton from the Upper Paraná and São Francisco. *Journal of Fish Biology* 87:159–168.
- Bohmann, K., A. Evans, M. T. P. Gilbert, G. R. Carvalho, S. Creer, M. Knapp, D. W. Yu, and M. de Bruyn. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution* 29:358–367.
- Bolyen, E., et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* 37:852–857.
- Borja, A., M. Elliott, M. C. Uyarra, J. Carstensen, and M. Mea. 2017. Editorial: Bridging the gap between policy and science in assessing the health status of marine ecosystems. *Frontiers in Marine Science* 4:32. <https://doi.org/10.3389/fmars.2017.00032>
- Boyer, F., C. Mercier, A. Bonin, Y. Le Bras, P. Taberlet, and E. Coissac. 2016. oboitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources* 16:176–182.
- Brechon, A. L., S. H. Coombs, D. W. Sims, and A. M. Griffiths. 2013. Development of a rapid genetic technique for the identification of clupeid larvae in the Western English Channel and investigation of mislabelling in processed fish products. *ICES Journal of Marine Science* 70:399–407.
- Callahan, B. J., P. J. McMurdie, and S. P. Holmes. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal* 11:2639–2643.
- Canfield, T. J., and J. R. Jones. 1996. Zooplankton abundance, biomass, and size-distribution in selected midwestern waterbodies and relation with trophic state. *Journal of Freshwater Ecology* 11:171–181.
- Clarke, K. R. 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* 18:117–143.
- Clarke, K. R., and R. N. Gorley. 2015. Getting started with PRIMER v7. PRIMER-E, Plymouth Marine Laboratory, Plymouth, UK.
- Collins, R. A., J. Bakker, O. S. Wangensteen, A. Z. Soto, L. Corrigan, D. W. Sims, M. J. Genner, and S. Mariani. 2019. Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution* 10:1985–2001.
- Deagle, B. E., S. N. Jarman, E. Coissac, F. Pompanon, and P. Taberlet. 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters* 10:2–5.
- Deagle, B. E., A. C. Thomas, J. C. McInnes, L. J. Clarke, E. J. Vesterinen, E. L. Clare, T. R. Kartzinel, and J. P. Eveson. 2019. Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular Ecology* 28:391–406.
- Dopheide, A., et al. 2019. Estimating the biodiversity of terrestrial invertebrates on a forested island using DNA barcodes and metabarcoding data. *Ecological Applications* 29: e01877.
- Edgar, R. 2010. Usearch. Lawrence Berkeley National Lab (LBNL), Berkeley, California, USA.
- Elbrecht, V., B. Peinert, and F. Leese. 2017. Sorting things out: assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and Evolution* 7:6918–6926.
- Ellis, J. R., S. P. Milligan, L. Readdy, N. Taylor, and M. J. Brown. 2012. Spawning and nursery grounds of selected fish species in UK waters. <https://www.cefas.co.uk/publications/techrep/TechRep147.pdf>
- Findley, K., et al. 2013. Topographic diversity of fungal and bacterial communities in human skin. *Nature* 498:367–370.
- Hall, T., I. Biosciences, and C. Carlsbad. 2011. BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences* 2:60–61.
- Hernandez, F. J., L. Carassou, W. M. Graham, and S. P. Powers. 2013. Evaluation of the taxonomic sufficiency approach for ichthyoplankton community analysis. *Marine Ecology Progress Series* 491:77–90.
- Huson, D. H., A. F. Auch, J. Qi, and S. C. Schuster. 2007. MEGAN analysis of metagenomic data. *Genome Research* 17:377–386.
- Kimmerling, N., et al. 2018. Quantitative species-level ecology of reef fish larvae via metabarcoding. *Nature Ecology & Evolution* 2:306–316.

- Krehenwinkel, H., M. Wolf, J. Y. Lim, A. J. Rominger, W. B. Simison, and R. G. Gillespie. 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports* 7:17668. <https://doi.org/10.1038/s41598-017-17333-x>
- Lamb, P. D., E. Hunter, J. K. Pinnegar, S. Creer, R. G. Davies, and M. I. Taylor. 2019. How quantitative is metabarcoding: a meta-analytical approach. *Molecular Ecology* 28:420–430.
- Legrand, T., A. Di Franco, E. Ser-Giacomi, A. Calo, and V. Rossi. 2019. A multidisciplinary analytical framework to delineate spawning areas and quantify larval dispersal in coastal fish. *Marine Environmental Research* 151:104761.
- Miya, M., et al. 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science* 2:150088.
- Morgulis, A., G. Coulouris, Y. Raytselis, T. L. Madden, R. Agarwala, and A. A. Schäffer. 2008. Database indexing for production MegaBLAST searches. *Bioinformatics* 24:1757–1764.
- Nobile, A. B., D. Freitas-Souza, F. J. Ruiz-Ruano, M. L. M. O. Nobile, G. O. Costa, F. P. de Lima, J. P. M. Camacho, F. Foresti, and C. Oliveira. 2019. DNA metabarcoding of neotropical ichthyoplankton: Enabling high accuracy with lower cost. *Metabarcoding and Metagenomics* 3:69–76.
- Piñol, J., G. Mir, P. Gomez-Polo, and N. Agustí. 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources* 15:819–830.
- Piñol, J., M. A. Senar, and W. O. C. Symondson. 2019. The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology* 28:407–419.
- Porter, T. M., and M. Hajibabaei. 2018. Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. *Molecular Ecology* 27:313–338.
- R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Radchuk, V., C. Turlure, and N. Schtickzelle. 2013. Each life stage matters: the importance of assessing the response to climate change over the complete life cycle in butterflies. *Journal of Animal Ecology* 82:275–285.
- Riaz, T., W. Shehzad, and A. Viari. 2011a. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research* 39:1–11.
- Riaz, T., W. Shehzad, A. Viari, F. Pompanon, P. Taberlet, and E. Coissac. 2011b. EcoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research* 39:1–11.
- Richardson, R. T., C. Lin, J. O. Quijia, N. S. Riusech, K. Goodell, and R. M. Johnson. 2015. Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. *Applications in Plant Sciences* 3:1500043.
- Richardson, A. J., A. W. Walne, A. W. G. John, T. D. Jonas, J. A. Lindley, D. W. Sims, D. Stevens, and M. Witt. 2006. Using continuous plankton recorder data. *Progress in Oceanography* 68:27–74.
- Robinson, C. V., T. M. U. Webster, J. Cable, J. James, and S. Consuegra. 2018. Simultaneous detection of invasive signal crayfish, endangered white-clawed crayfish and the crayfish plague pathogen using environmental DNA. *Biological Conservation* 222:241–252.
- Rodriguez, J. M., F. Alemany, and A. Garcia. 2017. A guide to the eggs and larvae of 100 common Western Mediterranean Sea bony fish species. <https://www.agris.fao.org>
- Rose, K. A. 2000. Why are quantitative relationships between environmental quality and fish populations so elusive? *Ecological Applications* 10:367–385.
- Russell, F. S. 1976. The eggs and planktonic stages of British marine fishes. Academic Press, London, UK.
- Schloss, P. D., et al. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75:7537–7541.
- Schnell, I. B., K. Bohmann, and M. T. P. Gilbert. 2015. Tag jumps illuminated—reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources* 15:1289–1303.
- Sigut, M., M. Kostovčík, J. Hulcr, P. Drozd, and J. Hřeček. 2017. Performance of DNA metabarcoding, standard barcoding, and morphological approach in the identification of host–parasitoid interactions. *PLoS ONE* 12:e0187803–1.
- Taberlet, P., E. Coissac, F. Pompanon, C. Brochmann, and E. Willerslev. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* 21:2045–2050.
- Tang, M., et al. 2015. High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and Evolution* 6:1034–1043.
- Thomas, A. C., B. E. Deagle, J. P. Eveson, C. H. Harsch, and A. W. Trites. 2016. Quantitative DNA metabarcoding: improved estimates of species proportional biomass using correction factors derived from control material. *Molecular Ecology Resources* 16:714–726.
- Thomsen, P. F., P. R. Møller, E. E. Sigsgaard, S. W. Knudsen, O. A. Jørgensen, and E. Willerslev. 2016. Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PLoS ONE* 11:e0165252.
- Ward, R. D., T. S. Zemlak, B. H. Innes, P. R. Last, and P. D. N. Hebert. 2005. DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B* 360:1847–1857.
- Wiesner, R. J., J. C. Rüegg, and I. Morano. 1992. Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissues. *Biochemical and Biophysical Research Communications* 183:553–559.

## SUPPORTING INFORMATION

Additional supporting information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/eap.2284/full>

## DATA AVAILABILITY

Sequences from metabarcoding have been deposited in the NCBI under accession reference BioProject PRJNA576002. Sanger sequences for the reference collection have been deposited in GenBank under accession numbers MN539918–MN539945 (CO-I) and MN539946–MN539976 (12S).