ICES Journal of Marine Science

ICES Journal of Marine Science (2022), 79(2), 403–412. https://doi.org/10.1093/icesjms/fsab273

Original Article

Current laboratory protocols for detecting fish species with environmental DNA optimize sensitivity and reproducibility, especially for more abundant populations

Mark Y. Stoeckle [1,*], Jason Adolf[2], Jesse H. Ausubel[1], Zach Charlop-Powers[1], Keith J. Dunton[2], and Greg Hinks[3]

[1]*Program for the Human Environment, The Rockefeller University , New York, NY, USA*
[2]*Department of Biology, Monmouth University, West Long Branch, NJ, USA*
[3]*Bureau of Marine Fisheries, Department of Environmental Protection, Port Republic, NJ, USA*

[*]*Corresponding author: tel: +212 327-8630; fax: +212 327-7519; e-mail: mark.stoeckle@rockefeller.edu.*

Stoeckle, M. Y., Adolf, J., Ausubel, J. H., Charlop-Powers, Z., Dunton, K. J., and Hinks, G. Current laboratory protocols for detecting fish species with environmental DNA optimize sensitivity and reproducibility, especially for more abundant populations. – ICES Journal of Marine Science, 79: 403–412.

Received 13 October 2021; revised 23 December 2021; accepted 25 December 2021; advance access publication 11 January 2022.

Analysing environmental DNA (eDNA) in seawater can aid in monitoring marine fish populations. However, the extent to which current methods optimize fish eDNA detection from water samples is unknown. Here, we test modifications to laboratory components of an eDNA metabarcoding protocol targeting marine finfish. As compared to baseline methods, amplifying a smaller proportion of extracted DNA yielded fewer species, and, conversely, amplifying a larger proportion identified more taxa. Higher-read species were amplified more reproducibly and with less variation in read number than were lower-read species. Among pooled samples, 20-fold deeper sequencing recovered one additional fish species out of a total of 63 species. No benefit was observed with additional PCR cycles, alternative primer concentrations, or fish-selective primers. Experiments using an exogenous DNA standard to assess absolute eDNA concentration suggested that, for a given proportion of a DNA sample, current laboratory methods for metabarcoding marine fish eDNA are near to maximally sensitive. Our results support the unofficial standard collection volume of one liter for eDNA assessment of commonly encountered marine fish species. We conclude that eDNA rarity poses the main challenge to current methods.

Keywords: bottom trawl survey, coastal ecosystem, eDNA, environmental DNA, marine fish, metabarcoding, neritic zone, ocean management, Poisson distribution, stock assessment.

Introduction

Environmental DNA (eDNA) analysis has attracted interest as a tool for monitoring marine animals. As compared to traditional field observation and capture techniques, including trawl, gillnet, seine, and trap, collecting water for eDNA is relatively inexpensive, harmless to marine life and the physical environment, requires modest equipment, and can be performed by a wide variety of personnel (Bourlat *et al.*, 2013; Hansen *et al.*, 2018). However, more work is needed to establish reproducible, replicable, and generalizable laboratory procedures (Joskow 2015), and to benchmark eDNA against traditional methods (Gilbey *et al.*, 2021). Here, we focus on laboratory procedures used for metabarcoding fish eDNA extracted from a water sample. The goal was to maximize fish species detection from a given water sample and so to reduce the need for additional water collection in the field.

First, we briefly review eDNA biology and analytic methods. Seawater contains DNA from nearby animals, in the form of cells and

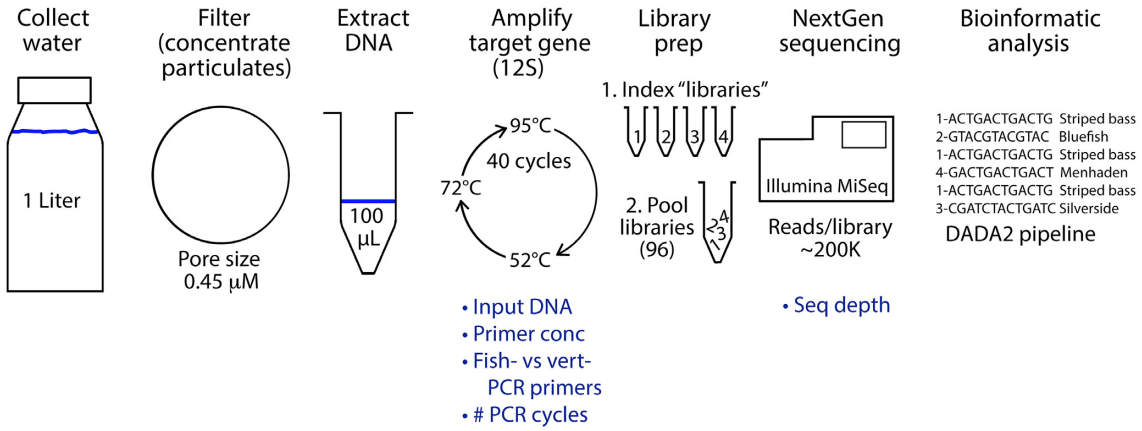

Figure 1. Schematic of eDNA metabarcoding protocol for marine fish. Major components and baseline parameters are shown. Aspects modified in this study are listed in blue text.

cell fragments, either shed from internal or external body surfaces, or released from tissues following injury or mortality. In the ocean, eDNA is lost through degradation and dispersal relatively quickly, such that detection usually indicates animals were nearby in the last few days (Goldberg *et al.*, 2016; Allan *et al.*, 2020). There is increasing evidence that the concentration of a fish species eDNA in a water sample can serve as an index to the local abundance of that species (Yates *et al.*, 2020; Stoeckle *et al.*, 2021). eDNA analysis starts with collecting water, typically 1 l, then filtration to capture particulates, and extraction of DNA from the retained material. Most (> 99%) of the DNA obtained by this process is from bacteria, algae, and other planktonic organisms (Turner *et al.*, 2014). PCR is then used to amplify the DNA of the species of interest (Figure 1). eDNA metabarcoding uses PCR primers that target multiple species in a taxonomic group, e.g. vertebrates (Pompanon *et al.*, 2011; Kelly *et al.*, 2014; Valentini *et al.*, 2015; Andruszkiewicz *et al.*, 2017). This strategy takes advantage of the fact that short segments of mitochondrial genes are identical among related species, and primers that bind at the conserved segments will amplify the DNA of any of those species. When applied to an eDNA sample, multiple species are amplified, and the resulting mixture is sequenced using next generation technology. Sequences are identified by matching to a genetic reference library derived from named specimens. Metabarcoding methods are sufficiently robust to be codified in handbooks (Laramie *et al.*, 2015; Taberlet *et al.*, 2018; Pawlowski *et al.*, 2020). Regarding optimization, considerable work has been done comparing results according to filter material, filter pore size, DNA extraction method, PCR primers, and presence of DNA inhibitors (Djurhuus *et al.*, 2017; Hunter *et al.*, 2019; Jeunen *et al.*, 2019; Muha *et al.*, 2019; Sanches and Schreier, 2020; Valsecchi *et al.*, 2020; Zhang *et al.*, 2020). Marine eDNA metabarcoding surveys may fail to detect some fish species likely or known to be present (Port *et al.*, 2016; Stoeckle *et al.*, 2017; Bessey *et al.*, 2020; Fernandez *et al.*, 2021).

In this study, we tested modifications to our laboratory eDNA protocol to better understand determinants of species detection. The goal was to improve finfish detection without increasing water collection volume. We focused on laboratory aspects that have attracted relatively little attention, including the amount of DNA used for PCR, number of PCR cycles, primer concentration, sequencing depth, and fish- vs. vertebrate-selective primers (Figure 1). For each modification, the null hypothesis was that it would not improve detection of fish species.

Methods

eDNA sources

Archived DNA extracts from New Jersey Trawl-eDNA Project (NJTrawl-eDNA) water samples collected during 2019 (Stoeckle *et al.*, 2021), and from additional trawl collections in January 2020 that were not part of the original report, were employed to study protocol modifications. Water samples were collected during normal trawl operations, prior to net deployment, at surface and near bottom, and at about one quarter of trawl sites each survey month. Processing details include 1 l water sample volume, vacuum filtration with wall suction using a 47 mm diameter, 0.45 μM pore size nitrocellulose filter (Whatman), DNA extraction with DNeasy PowerSoil Kit (Qiagen), and re-suspension of extracted DNA in 100 μl Buffer C6. In addition to NJTrawl-eDNA samples, we made use of aquatic eDNA samples from an ongoing regional habitat survey conducted at shoreline sites in Massachusetts, Rhode Island, New York, and New Jersey during 2019–2020. Processing followed similar procedures as for NJTrawl–eDNA samples. Briefly, vacuum filtration of 500–1000 ml water was done out of laboratory with a hand pump (Nalgene) or in laboratory with wall suction and was performed on same day as collection, or water was stored at 4°C and filtered within 24 h. Filters were 47 mm, 0.45 μM pore size, nylon (Millipore), or nitrocellulose. In our experience, nitrocellulose vs. nylon filters and hand pump vs. wall suction produced equivalent results. We note that the lower volume habitat survey samples may have contained eDNA of fewer fish species. All experiments involved comparing aliquots of the same DNA sample, so differences in filter type, filtration method, or water volume are unlikely to have affected findings. For negative controls, 1 l of laboratory tap water was processed using the same equipment and procedures as for field samples. Filters were stored at −20°C until extraction. DNA was isolated from filters with DNeasy PowerSoil Kit, with modifications from the manufacturer's protocol as previously described (Stoeckle *et al.*, 2020). At completion, DNA was eluted with 100 μl Buffer C6 and stored at −20°C. Collection date, location, processing intervals, and DNA yield (average 1 μg per liter filtered) for samples analysed in this report are in Supplementary Table 1. No animals were housed or ex-

perimented upon as part of this study. No endangered or protected species were collected.

DNA amplification

PCR reactions were carried out in 25 μl total volume with TaKaRa High Yield PCR EcoDry™ Premix containing TaqStart® Antibody for hot-start PCR. Standard conditions were 5 μl of DNA (representing 1/20th of DNA extracted from a water sample, containing on average 55 ng DNA) or 5 μl of molecular biology grade water, and 200 nM Illumina-tailed "ecoPrimers" (IDT) that target an approximately 110 bp segment of the vertebrate mitochondrial gene for 12S rRNA (12S gene; Riaz *et al.*, 2011). Primer sequences are shown in Supplementary Table 2. Thermal cycling conditions were 95°C × 5 m, 40 cycles of (95°C 20 s, 52°C 20 s, and 72°C 20 s), extension at 72°C for 1 m, and hold at 4°C. Negative control reactions were included in all amplification sets. A total of 5 μl of each reaction mix were run on a 2.5% agarose gel with SYBR Safe (Invitrogen) to assess amplification. The remaining 20 μl were diluted 1:20 in 10 mM Tris-HCl, pH 8.5 (Buffer EB, Qiagen) to be used as template for indexing, described below. Where noted, DNA input, primer concentration, and cycle number were altered.

Indexing

Indexing tags individual libraries with unique DNA sequences so that multiple libraries can be analysed the same sequencing run. Indexing was done in 25 μl reaction volume with Illustra PuReTaq Ready-To-Go PCR beads. PCR cocktails included 5 μl diluted reaction mix from initial PCR (see above) and 2.5 μl N7xx and S5xx indexing primers (Nextera XT Index Kit v2 set A, 96 indices; Illumina). Parameters were as above, except 10 cycles and 55°C annealing temperature were used. A total of 5 μl of each reaction mix were run on a 2.5% agarose gel with SYBR Safe to assess amplification. The remaining 20 μl from each of 96 libraries were pooled. A total of 400 μl of pooled libraries were cleaned with AMPure beads (Beckman Coulter) at 1:1 and eluted with equivalent volume of EB. Concentration was checked by Qubit (Thermo Fisher Scientific; typical yield 10 ng μl^{-1}), and pooled sample was sent for sequencing.

Next-generation sequencing, bioinformatic analysis

Sequencing was performed at GENEWIZ on an Illumina MiSeq, 2×150 bp. Findings in this study are based on 139 eDNA and 28 negative control libraries, which were analysed together with other samples not reported here, in seven MiSeq runs with 96 libraries per run. PhiX was not routinely employed. Bioinformatic analysis was performed on Illumina FASTQ files using DADA2, which identifies all amplicon sequence variants (ASVs) without a similarity threshold (Callahan *et al.*, 2016, 2017). Our DADA2 pipeline (Stoeckle *et al.*, 2017) generated taxon assignments by comparison to an internal 12S gene reference library of regional fishes and other commonly amplified vertebrate ASVs (Supplementary File 1). In addition, all ASVs were manually submitted to GenBank to recover overlooked matches to 12S gene sequences not included in internal library. All identifications were based on 100% match to a reference sequence. Tap water eDNA and reagent grade water libraries were negative for fish ASVs after filtering DADA2 output tables as previously described (Stoeckle *et al.*, 2020). Filtering consisted of excluding detections comprising less than 1/1000th of the total for that taxon among all libraries in the run. Except where noted, filtered DADA2 reads are presented as obtained without normalization. All sequence data analysed here are new, except Supplementary Figure 1 re-analysed FASTQ files from an earlier study (Stoeckle *et al.*, 2021). To assess possible benefit of increased sequencing depth, one set of pooled libraries was additionally sequenced at GENEWIZ on Illumina HiSeq 2×150 bp with 10% Phi-X spike-in. HiSeq FASTQ files were analysed with a modified DADA2 pipeline (Supplementary File 2). To assess whether fish species were eliminated by bioinformatic processing, TextEdit search function was applied to MiSeq FASTQ files. Linear regression, Fisher's exact test, *t*-test, and Poisson statistics were made with Prism 8.

Exogenous 12S gene DNA standard

We selected ostrich (*Struthio camelus*) as an exogenous standard, as it was unlikely to be present in regional environmental samples and Riaz 12S gene metabarcoding primer sites are identical to those in bony fish. DNA was extracted from ostrich meat dog treats (American Ostrich Farms) and amplified with M13-tailed primers that targeted a 689 bp segment of ostrich 12S gene covering the Riaz primer target site and flanking regions. Primer sequences are in Supplementary Table 2. PCR parameters were as above, except 45 cycles and 55°C annealing temperature were used. Gel electrophoresis showed a single band of expected size and Sanger sequencing with M13 primers at GENEWIZ confirmed 100% identity to *S. camelus* mitochondrial reference genome NC002785. Ostrich PCR product was purified with AMPure beads at 1:1, concentration checked by Qubit, and a series of 100-fold dilutions was made in EB with resulting concentrations of 10 pg μl^{-1}–0.1 ag μl^{-1}. Where noted, 5 μl, containing 50 pg, 500 fg, 5 fg, 50 ag, or 0.5 ag, were added to four eDNA samples (two from January 2020 NJTrawl–eDNA Project, two from habitat survey) before amplification. Given amplicon size including primers of 765 bp, these aliquots were calculated to contain on average 6×10^7, 6×10^5, 6×10^3, 60, or 0.6 copies of ostrich DNA, respectively. For lowest concentration, 0.6 copies, expected content of a 5 μl aliquot was 0–2 copies, according to Poisson distribution. To assess whether metabarcoding reads were proportional to template copies, for each sample we compared ostrich reads in the 6000-copy library to those in the 60-copy library, using fish reads in the respective libraries to normalize each value. This generated a "6K/60 index" as follows, with expected value of 100:

$$\frac{(\text{Ostrich reads/fish reads in } 6000 - \text{copy} - \text{library})}{(\text{Ostrich reads/fish reads in } 60 - \text{copy} - \text{library})}.$$

To evaluate accuracy of ostrich DNA standard, libraries were prepared from eight replicate PCRs with 0.6 or 6 copies of ostrich DNA. Based on Poisson distribution, about half (45.1%) of 0.6-copy amplifications were expected to contain ostrich DNA, as compared to essentially all (99.8%) of 6-copy inputs.

Fish-selective primers

Riaz 12S gene primers employed in this study commonly recover human and domestic animal DNA from near shore marine samples (Stoeckle *et al.*, 2017). Potential sources include wastewater contamination, laboratory procedures, and laboratory reagents (Leonard *et al.*, 2007; Champlot *et al.*, 2010). At the same time, pilot experiments demonstrated MiFish-U primers (Miya *et al.*, 2015), which are selective for fish vs. other vertebrates, frequently amplified a 256 bp segment of bacterial 16S rRNA gene. With both Riaz and MiFish primers, non-target reads sometimes exceeded fish

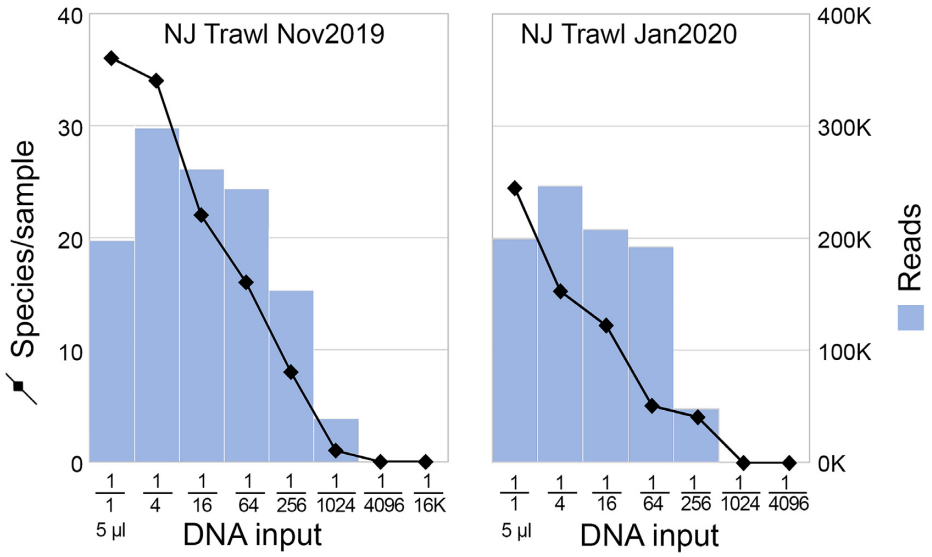

Figure 2. Species decumulation curve: effect of smaller amount of DNA template on fish species detection and total fish reads. Black diamonds with connecting lines represent number of fish species per library, and blue columns depict fish reads per library. Each black diamond and blue column represent one library prepared by amplifying indicated fraction of NJTrawl–eDNA sample collected during month shown. Libraries in each panel were analysed in a single MiSeq run. Source data are in Supplementary Tables 3A and B.

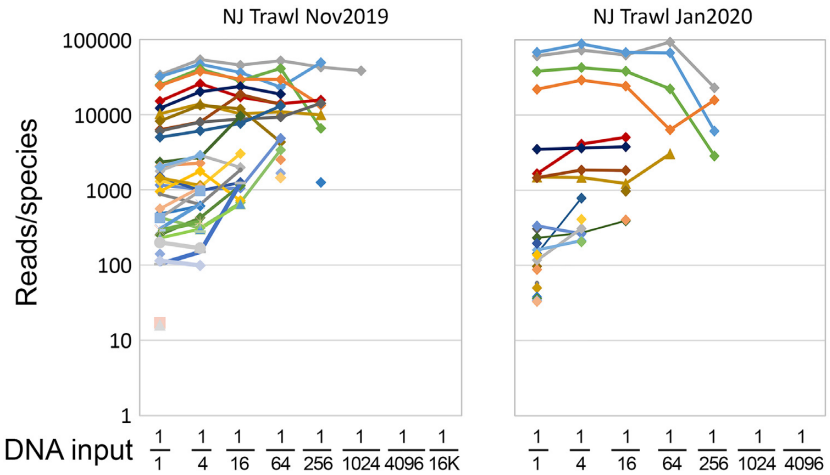

Figure 3. Species decumulation: effect of smaller amount of DNA template on reads per fish species. Reads per species from libraries depicted in Figure 2 are shown. Coloured symbols with connecting lines represent individual species, and designated species differ between charts. As noted in Figure 1 legend, libraries in each panel were analysed in a single MiSeq run. Source data are in Supplementary Tables 3A and B.

reads. To assess the possible benefit of fish-selective primers, we modified the MiFish-U reverse primer by adding a T at 3' end. This change was based on observation that most (98%) regional fish species had a T in this position, whereas off-target bacterial 16S DNA amplicons had a C. PCR with the modified MiFish-U primer set was done as described above except that annealing temperature was 60°C. Primer sequences are listed in Supplementary Table 2. PCR products were cleaned with AMPure at 1:1, and indexing and sequencing were performed at GENEWIZ on Illumina MiSeq 2 × 250 bp.

Results

Protocol modifications with a negative effect on species detection

Smaller amount of DNA template

Libraries prepared from progressively smaller amounts of a DNA sample yielded progressively fewer species (Figure 2, lines). No fish were recovered with less than 1/1024th and 1/256th of November 2019 and January 2020 NJTrawl–eDNA samples, respectively, which implied 5 μl of these undiluted samples contained at least

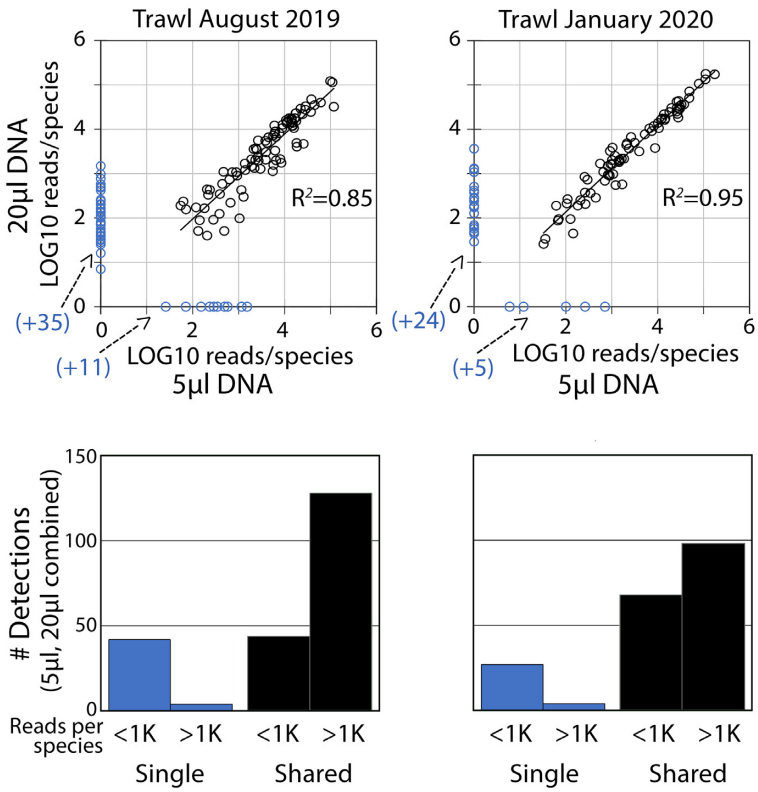

Figure 4. Effect of larger amount of DNA template on fish species detection. In total, five NJTrawl–eDNA samples collected during months shown were analysed. For each sample, separate libraries were prepared from amplifications of 5- and 20-μl extracted DNA. Libraries in each panel were analysed in a single MiSeq run. Each circle represents one species from one water sample. Black circles denote "shared" detections, i.e. present in both 5 and 20 μl libraries, and blue circles represent "single" detections, i.e. in 5 or 20 μl library only. Charts at bottom compare single and shared detections according to reads per species. A larger proportion of singles had fewer than 1000 reads (91% vs. 26%, and 87% vs. 41%, for 2019 and 2020 samples, respectively; $p < 0.0001$ for both months, Fisher's exact test). In addition, shared positives with fewer than 1000 reads were more variable than were those with more than 1000 reads (August 2019, mean absolute arithmetic fold-difference reads, 2.3 vs. 1.6, $p = 0.0003$; January 2020, 1.7 vs. 1.4, $p = 0.035$; unpaired *t*-test). Source data are in Supplementary Table 4.

1024 and 256 eDNA copies of fish eDNA. In contrast to reduction in species number, total fish reads (Figure 2, columns) were relatively unchanged with as little as 1/256th and 1/64th of standard input for these samples, respectively. For both samples, there was a step-up in total reads with $\frac{1}{4}$ of standard amount of DNA, which might reflect reduced PCR inhibition. However, this was not accompanied by improved species detection. Species decumulation curves in Figure 2 reflected progressive drop-out of those species with the fewest reads (Figure 3). This suggested that, at each step down, the rarer templates were lost by dilution. Reads for the remaining species were largely maintained with decreased DNA, illustrating that metabarcoding reports relative, not absolute, eDNA copy number.

Protocol modifications with a positive effect

Larger amount of DNA template

Replicate metabarcoding amplification typically reveals additional species and fails to detect others (e.g. Sato *et al.*, 2017). Here, we refer to these as "singles," i.e. species found in only one replicate. Metabarcoding libraries generated with a fourfold larger amount of template DNA yielded about fourfold more singles than did libraries generated with standard amount (Figure 4). Most singles had fewer than 1000 reads, and shared positives with fewer than 1000 reads were more variable (see Figure 4 legend). Replicate amplification with 5 μl followed the same pattern, namely, most single detections had fewer than 1000 reads (single vs. shared, 99% vs. 47%, respectively, $p < 0.0001$, Fisher's exact test), and shared positives with fewer than 1000 reads were more variable (Supplementary Figure 1).

Increased sequencing depth

Repeat sequencing of a 96-library pool with Illumina HiSeq yielded about 2 million fish reads per library, about 20-fold more than with MiSeq. Among pool of 96 libraries, DADA2 processing of HiSeq and MiSeq files recovered 63 and 62 bony fish species, respectively, with highly correlated read counts (Figure 5). The one HiSeq-only species was present in MiSeq FASTQ files, but was screened out by DADA2 bioinformatic pipeline. At the level of individual libraries, HiSeq recovered on average about 10% more species than MiSeq. As with pooled results, these novelties were present in corresponding MiSeq FASTQ files but were screened out by DADA2 pipeline (Supplementary Figure 2). In addition, HiSeq ASV table contained multiple very low-read-number detections not represented in corresponding MiSeq FASTQ files (Supplementary Figure 3). These were

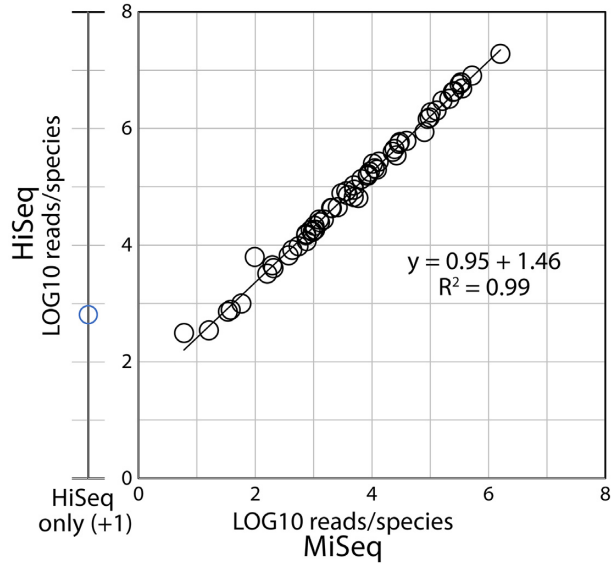

Figure 5. Increased sequencing depth. A 96-library pool was sequenced on HiSeq and MiSeq platforms, and files were analysed by DADA2 pipeline. Among all libraries, 62 fish species were detected by both platforms, represented by black circles in grid. HiSeq yielded one additional species, shown as blue circle. Source data are in Supplementary Table 9.

exclusive to libraries that shared an index with a high-read number, presumably accurate, positive. We, therefore, concluded that very low-read-number positives reflected library mis-assignment due to PCR, sequencing, or bioinformatic error.

Protocol modifications with no effect

Increased PCR cycles

Duplicate 5 μl aliquots from four habitat survey DNA samples were amplified for 40 or 50 cycles and used to prepare separate libraries. No gain in species counts was observed (Supplementary Figure 4).

Alternative primer concentration

Replicate 5 μl aliquots from two habitat survey DNA samples were amplified with primer concentrations ranging in twofold steps from 50 to 800 nM. As compared to standard 200 nM, higher concentrations yielded more PCR product, and lower yielded less. No difference in species number was observed (Supplementary Figure 5).

Fish-selective primers

A total of six habitat survey samples that generated a large number of non-fish reads (mostly human) with Riaz primers were selected for re-analysis. The modified MiFish-U primer set (see methods) successfully improved target specificity; however, species recovery was unchanged (Supplementary Figure 6). As an aside, this amended primer set may be useful in other settings to reduce non-target reads, given that gel purification of MiFish-U amplicons is commonly required (e.g. Miya *et al.*, 2015).

Assessing dynamic range and absolute copy number using an exogenous standard

High-copy-number ostrich DNA (6×10^7 and 6×10^5) completely suppressed fish amplification, indicating a limit to dynamic range of assay (Figure 6). With lower amounts, 6000 and 60 copies, fish eDNA was amplified. The relative number of ostrich vs fish reads was closely proportional to relative number of ostrich vs fish DNA template copies (Figure 6, 6K/60 index). In 6000-copy-ostrich-DNA libraries, even though ostrich DNA occupied 90% or more of reads, eDNA from all, or most all, fish species was amplified, albeit at lower read numbers. This demonstrated that the dynamic range of the metabarcoding protocol was greater than the range of eDNA abundance in these samples. Based on comparison to ostrich standard, the apparent fish 12S eDNA template copy number was 170–700 per 5 μl extract, in the same range as minimums extrapolated from decumulation analysis (Figures 2 and 3; albeit with different DNA samples), consistent with detection of single-copy template. To assess accuracy of ostrich DNA concentration, replicate libraries were prepared with 0.6 or 6 copies of ostrich DNA, and no fish eDNA. In total, five of eight (62.5%) 0.6-copy libraries and eight of eight (100%) 6-copy libraries contained ostrich reads (Supplementary Table 6), consistent with expected Poisson distribution outcomes of 45% and 100%, respectively. Given observed outcome with 0.6-copy libraries, 95% confidence interval of actual concentration was 0.3–2.2 copies per aliquot.

Discussion

In this study, we tested modifications to laboratory components of an eDNA metabarcoding protocol targeting marine finfish. The baseline methods were representative of those in other fish metabarcoding applications. The goal was to enhance species recovery from a DNA sample, thereby improving fish eDNA detection without the need to filter larger volumes of water. The baseline protocol, which used 1/20th of DNA obtained from one liter of seawater, gave reproducible results for most fish species in our neritic zone samples. However, the proportion of the DNA sample used for PCR was critical—a smaller proportion produced fewer species and a larger proportion yielded more. Higher-read species were amplified more reproducibly and with less variation in read number than were lower-read species. We hypothesize that variable amplification reflects Poisson distribution of low-copy-number templates among aliquots subject to PCR (Figure 7). Non-detection due to Poisson distribution implies an average concentration of 3 or fewer template copies per aliquot and becomes the norm ($p > 0.5$) only when average concentration is less than one copy per aliquot, implying protocol detected single copy eDNA. Experiments with ostrich DNA as an exogenous standard supported this inference.

Our findings help demonstrate that metabarcoding reports relative, not absolute abundance—reads per species were largely maintained over 1000-fold range of DNA template concentrations (Figure 3). As others have noted, measuring absolute eDNA concentration with metabarcoding requires an exogenous standard (Ushio *et al.*, 2018). We infer a threshold for reproducible detection of about 100 copies l^{-1} seawater (Figure 7). Most species detections in our neritic zone samples were reproducible and likely represented by eDNA concentration above this threshold. Detections were largely maintained in the presence of non-fish vertebrate DNA up to 10-fold excess (Figure 6) and, consistent with this result, there was no benefit of fish-specific primers (Supplementary

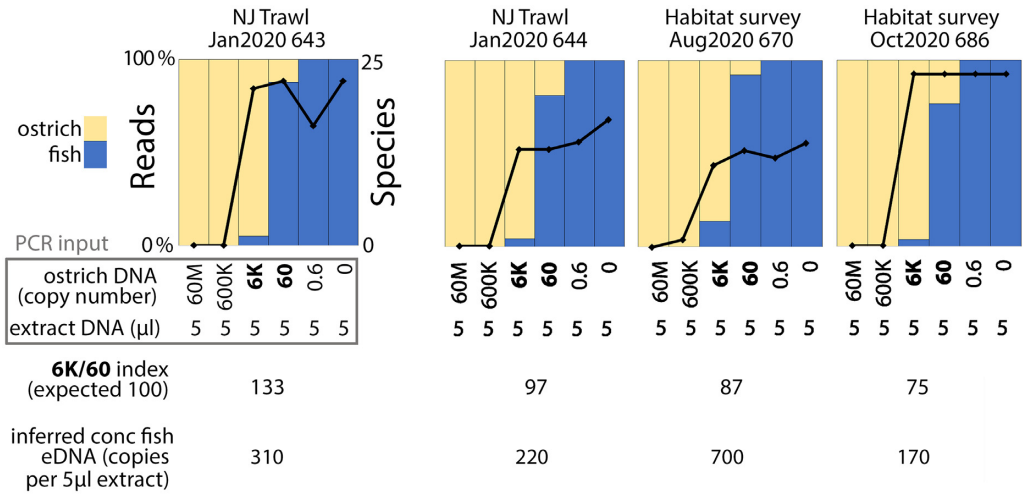

Figure 6. Amplification of DNA samples spiked with exogenous 12S gene DNA standard. Libraries were prepared by amplifying 5 μl of eDNA sample mixed with indicated number of copies of ostrich 12S gene amplicon. Yellow and blue columns represent relative proportion of ostrich and fish reads, respectively, and black lines indicate number fish species detected. The 6K/60 index compares ostrich reads in 6000- and 60-copy libraries (see methods for details), with expected value of 100. Libraries in each panel were analysed in a single MiSeq run. Source data are in Supplementary Tables 5A and B.

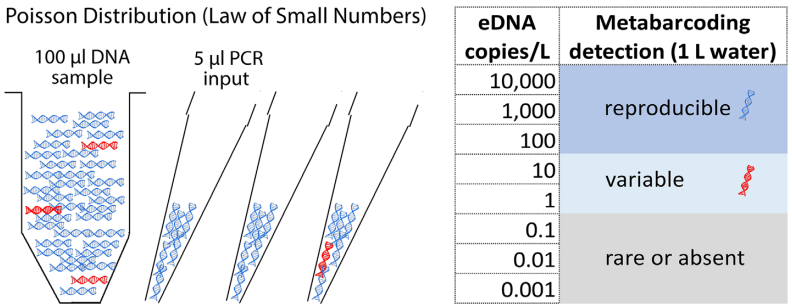

Figure 7. Model of metabarcoding detection, eDNA abundance, and law of small numbers. Illustration assumes 1 l water collection, 100 μl eDNA elution volume, 5 μl eDNA sample in first round PCR, and single amplification. The illustrated cutoffs are approximate—with given parameters, non-detection due to Poisson distribution is expected to become significant ($p \geq 0.05$) when eDNA concentration is below 60 copies l^{-1}—and assume 100% capture of eDNA from water sample.

Figure 6). These findings predict that blocking primers designed to prevent amplification of unwanted templates from aquatic eDNA, such as human DNA (e.g. Thomsen *et al.*, 2016), may increase reads per species, but will not improve the number of species detected.

We conclude that eDNA rarity poses the main challenge to current methods. This has several practical implications. First, reducing the proportion of a DNA sample subject to PCR, as commonly done to minimize PCR inhibition (e.g. Ushio *et al.*, 2018), can be expected to reduce recovery of rare eDNA. The absence of a plateau with increased template concentration (Figure 4) implies that identifying all the species in an eDNA sample may require analysing the entire sample. Beyond detection, amplifying a larger proportion of an eDNA sample can be expected to improve reproducibility in read counts via lessened Poisson effects. Second, advances in sequencing technology alone are unlikely to dramatically improve fish species recovery over current protocols. There was some benefit to 20-fold deeper sequencing, about 10% more species per library, however, this gain was obviated by pooling multiple libraries. Given that HiSeq-only species were present in MiSeq files prior to bioinformatic processing, it may be fruitful to investigate whether software modification could improve findings. Third, reliable monitoring of scarce eDNA may require analysing much larger water volumes (e.g. $\geq$ 10X), either by collecting multiple water samples or via alternative filtration devices. Scarcity of eDNA may arise from organism rarity, reduced eDNA shedding (Harper *et al.*, 2020), or enhanced degradation and dispersal (Allan *et al.*, 2020). Other approaches that may aid in the recovery of rare eDNA include collection near to organisms (Baker *et al.*, 2018; Dugal *et al.*, 2021), analysis of invertebrate filter feeders such as sponges that trap suspended eDNA (Mariani *et al.*, 2019), and passive collection methods (Bessy *et al.*, 2021). Rarity challenges all survey technologies, including nets, traps, cameras, and sound. Given the enormous gain in sensitivity (fish species detected per unit water volume) with eDNA compared to a bottom trawl—seven orders of magnitude (Stoeckle *et al.*, 2021)—eDNA may be more practical than other methods.

The general concept of "more water, more fish" supported by the experiments performed here is intuitive and can be logically extended to a recommendation to sample more water in field, as noted above. Filtering a larger volume, however, presents challenges in practice because of the tendency of near shore seawater samples to clog the small-pore-size filters typically used to recover eDNA, due to sediment, phytoplankton, zooplankton, and other biological

and physical debris. In the present experiments, eDNA was recovered from 1 l of seawater collected on a 0.45 μm-pore-size cellulose nitrate filter, and in most cases filtering more water would have been difficult due to clogging. Capsule filtration devices, which have higher surface areas and are less prone to clogging, have been used to process $\geq$ 20 l volumes for eDNA analysis of freshwater systems (Vences *et al.*, 2016; Cantera *et al.*, 2019). Giovannoni *et al.* (1990), employed tangential flow filtration (TFF) to concentrate picoplankton from thousands of liters of seawater against a 0.1 μm pore size filter. TFF was used for decades as the standard for filtering large volumes of seawater for viruses, although the excessive time and expense associated with TFF has led the field toward alternative collection methods (reviewed in Duhaime and Sullivan, 2012). Other strategies such as filtering multiple 1 l samples and combining filters before extraction may be feasible.

Our results add evidence that filtering larger water volumes recovers more aquatic taxa (Mächler *et al.*, 2016; Lopes *et al.*, 2017; Cantera *et al.*, 2019; McClenaghan *et al.*, 2020). Bessey *et al.* (2020), demonstrated improved marine fish eDNA recovery by increasing filtered water volume from 25 to 2000 mL, and further gains when results from ten 2000 mL samples were combined. The calculated fish eDNA copy number in this study (total for all species, 3400–14000 copies l^{-1}, Figure 6), is consistent with reported concentrations in other environments (Takahara *et al.*, 2012; Ushio *et al.*, 2018; Collins *et al.*, 2018; Salter *et al.*, 2019; Ramón-Laca *et al.*, 2021). Our findings are compatible with equivalent sensitivity of qPCR and metabarcoding (Harper *et al.*, 2018; Ushio *et al.*, 2018).

Limitations of this study include potential primer bias (Krehenwinkel *et al.*, 2017; Kelly *et al.*, 2019). Although the 12S gene primer binding sites are highly conserved among vertebrates, we cannot exclude the contribution of primer bias to inconsistent detection and variable read number. In addition, even with no primer bias, PCR efficiency may differ among templates. In particular, the ostrich DNA amplicon used as a standard may amplify with a different efficiency than do native mitochondrial DNA templates which are likely present in diverse fragments. Calculated template copy numbers may be inaccurate due to pipetting errors or a sink effect, such as binding of DNA to plastic surfaces (Ellison *et al.*, 2006), or to non-vertebrate DNA, which is orders of magnitude more abundant (Turner *et al.*, 2014). Evidence against this speculation is that low concentration ostrich DNA (0.6 copies) gave expected success, although this estimate has a substantial confidence interval. The hallmark of PCR inhibition, namely, a larger number of species detected with reduced amount of DNA template (Goldberg *et al.*, 2016), was not observed. However, PCR inhibition may have contributed to some of the findings, such as increased total reads noted with fractional input (Figure 1). A limited dynamic range of PCR may have prevented amplification of some low-copy-number templates, although near complete species recovery with excess ostrich template suggests this aspect was not a constraint (Figure 6). Finally, the sequencing technology used in this study, Illumina MiSeq, may have limitations. Singer *et al.* (2019), demonstrated improved detection of rare ASVs with Illumina Novaseq, which uses a different type of flow cell for sequencing and generates more reads per sample, as compared to results with Illumina MiSeq.

Coastal oceans are subject to increasing human impacts—offshore wind development, shipping, resource extraction, runoff pollution, commercial and recreational fishing, aquaculture—and to long-term changes in weather and climate. Together with traditional surveys, eDNA promises to enable more timely, accurate assessments of marine animal populations. Current eDNA methods appear sufficient for assessment of commonly encountered marine fish species, which includes most managed fish stocks.

Data availability statement

Illumina FASTQ files underlying this article are deposited in NCBI Bioproject ID PRJNA793893..

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

Author contributions statement

Designed study: MS, JA, KJD, and GH.
Designed bioinformatics: ZC.
Collected samples: MS, JHA, and GH.
Analysed samples: MS, and JA.
Analysed data: MS, JA, JHA, and KD.
Prepared figures: MS.
Wrote first draft of manuscript: MS.
Revised manuscript: MS, JA, JHA, and KJD.

Funding statement

This work was supported by NOAA Grant NA20OAR01104XX. Sampling during the New Jersey Ocean Trawl Survey was supported by US Fish and Wildlife Service Wildlife and Sport Fish Restoration Program. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

At New Jersey Ocean Trawl Survey, we thank Stony Brook University, Captain Chris Harter and the crew of the R/V Seawolf. At The Rockefeller University, we thank Jeanne Garbarino, Disan Davis, and Odaelys Walwyn for laboratory resources and assistance.

References

Andruszkiewicz, E. A., Starks, H. A., Chavez, F. P., Sassoubre, L. M., Block, B. A., and Boehm, A. B. 2017. Biomonitoring of marine vertebrates in monterey bay using eDNA metabarcoding. Plos ONE, 12: e0176343.

Allan, E. A., Zhang, W. G., Lavery, A. C., and Govindarajan, A. 2020. Environmental DNA shedding and decay rates from diverse animal forms and thermal regimes. Environmental DNA, 3: 1–23.

Baker, C. S., Steel, D., Nieukirk, S., and Klinck, H. 2018. Environmental DNA (eDNA) from the wake of the whales: droplet digital PCR for detection and species identification. Frontiers in Marine Science, 5: 133.

Bessey, C., Jarman, S. N., Berry, O., Olsen, Y. S., Bunce, M., Simpson, T., Power, M. *et al.* 2020. Maximizing fish detection with eDNA metabarcoding. Environmental DNA, 2: 493–504.

Bessy, C., Jarman, S. N., Simpson, T., Miller, H., Stewart, T., Keesing, J. K., and Berry, O. 2021. Passive eDNA collection enhances aquatic biodiversity analysis. Communications Biology, 4: 236.

Bourlat, S. J., Borfa, A., Gilbert, J., Taylor, M. I., Davies, N., Weisberg, S. B., Griffith, J. F. *et al.* 2013. Genomics in marine monitoring: new opportunities for assessing marine health status. Marine Pollution Bulletin, 74: 19–31.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, J. A., and Holmes, S. P. 2016. DADA2: high-resolution sample inference from illumina amplicon data. Nature Methods, 13: 581–583.

Callahan, B. J., McMurdie, P. J., and Holmes, S. P. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME Journal, 11: 2639–2643.

Cantera, I., Cilleros, K., Valentini, A., Cerdan, A., Dejean, T., Iribar, A., Taberlet, P. *et al.* 2019. Optimizing environmental DNA sampling effort for fish inventories in tropical streams and rivers. Scientific Reports, 9: 3085.

Champlot, S., Berthelot, C., Pruvost, M., Bennett, E. A., Grange, T., and Geigl, E.-M. 2010. An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. Plos ONE, 5: e13042.

Collins, R. A., Wangensteen, O. S., O'Gorman, E. J., Mariani, S., Sims, D. W., and Genner, M. J. 2018. Persistence of environmental DNA in marine systems. Communications Biology, 1: 185.

Djurhuus, A., Port, J., Closek, C. J., Yamahara, K. M., Romero-Maraccini, O., Walz, K. R., Goldsmith, D. B. *et al.* 2017. Evaluation of filtration and DNA extraction methods for environmental DNA biodiversity assessments across multiple trophic levels. Frontiers in Marine Science, 4: 314.

Dugal, L., Thomas, L., Jensen, M. R., Sigsgaard, E. El, Simpson, T., Jarman, S., Thomsen, P. F. *et al.* 2021. Individual haplotyping of whale sharks from seawater environmental DNA. Molecular Ecology Resources, 00: 1–10.

Duhaime, M. B., and Sullivan, M. B. 2012. Ocean viruses: rigorously evaluating the metagenomic sample-to-sequence pipeline. Virology, 434:181–186.

Ellison, S. L. R., English, C. A., Burns, M. J., and Keer, J. T. 2006. Routes to improving the reliability of low-level DNA analysis using real-time PCR. BMC Biotechnology, 6: 33.

Fernandez, A. P., Marques, V., Fopp, F., Juhel, J.-B., Borrero-Pérez, G. H., Cheutin, M.-C., Dejean, T. *et al.* 2021. Comparing environmental DNA metabarcoding and underwater visual census to monitor tropical reef fishes. Environmental DNA, 3: 142–156.

Gilbey, J., Carvalho, G., Castilho, R., Coscia, I., Coulson, M. W., Dahle, G., Derycke, S. *et al.* 2021. Life in a drop: sampling environmental DNA for marine fishery management and ecosystem monitoring. Marine Policy, 124: 104331.

Giovannoni, S., DeLong, E. F., Schmidt, T. M., and Pace, N. R. 1990. Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. Applied Environmental Microbiology, 56: 2572–2575.

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P.F., Murphy, M. A., Spear, S. F. *et al.* 2016. Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods in Ecology and Evolution, 7: 1299–1307.

Hansen, B. K., Bekkevold, D., Clausen, L. W., and Nielsen, E. E. 2018. The skeptical optimist: challenges and perspectives for the application of environmental DNA in marine fisheries. Fish and Fisheries, 19: 751–768.

Harper, L. R., Handley, L. L., Hahn, C., Boonham, N., Rees, H. C., Gough, K. C., Lewis, E. *et al.* 2018. Needle in a haystack? Comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (*Triturus cristatus*). Ecology and Evolution, 8: 6330–6341.

Harper, K. J., Goodwin, K. D., Harper, L. R., LaCasella, E. L., Frey, A., and Dutton, P. H. 2020. Finding crush: environmental DNA analysis as a tool for tracking the green sea turtle *Chelonia mydas* in a marine estuary. Frontiers in Marine Science, 6: 810.

Hunter, M. E., Ferrante, J. A., Meigs-Friend, G., and Ulmer, A. 2019. Improving eDNA yield and inhibitor reduction through increased water volumes and multi-filter isolation techniques. Scientific Reports, 9: 5259.

Jeunen, G.-J., Knapp, M., Spencer, H. G., Taylor, H. R., Lamare, M. D., Stat, M., Bunce, M. *et al.* 2019. Species-level biodiversity assessment using marine environmental DNA metabarcoding requires protocol optimization and standardization. Ecology and Evolution, 2019: 1–13.

Joskow, P. L. 2015. Research integrity and reproducibility. Alfred P. Sloan Foundation, 2014 Annual Report, p. XIII–XXIII. Alfred P. Sloan Foundation.

Kelly, R. P., Port, J. A., Yamahara, K. M., and Crowder, L. B. 2014. Using environmental DNA to census marine fishes in a large mesocosm. Plos ONE, 9: e86175.

Kelly, R. P., Shelton, A. O., and Gallego, R. 2019. Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. Scientific Reports, 9: 12133.

Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W.B., and Gillespie, R. G. 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific Reports, 7: 17668.

Laramie, M.B., Pilliod, D. S., Goldberg, C. S., and Strickler, K.M. 2015. Environmental DNA Sampling Protocol— Filtering Water to Capture DNA from Aquatic Organisms: Techniques and Methods, Book 2, chapter A13. U.S. Geological Survey, 15p. doi: 10.3133/tm2A13.

Leonard, J. A., Shanks, O., Hofreiter, M., Kreuz, E., Hodges, L., Ream, W., Wayne, R. K. *et al.* (2007). Animal DNA in PCR reagents plagues ancient DNA research. Journal of Archaeological Science, 34: 1361–1366.

Lopes, C. M., Sasso, T., Valentini, A., Dejean, T., Martins, M., Zamudio, K. R., and Haddad, F. B. 2017. eDNA metabarcoding: a promising method for anuran surveys in highly diverse tropical forests. Molecular Ecology Resources, 17: 904–919.

Mächler, E., Deiner, K., Spahn, F., and Altermatt, F. 2016. Fishing in the water: effect of sampled water volume on environmental DNA-based detection of macroinvertebrates. Environmental Science and Technology, 50: 305–312.

Mariani, S., Baillie, C., Colosimo, G., and Riesgo, A. 2019. Sponges as natural environmental samplers. Current Biology, 29: R395–R402.

McClenaghan, B., Fahner, N., Cote, D., Chawarski, J., McCarthy, A., Rajabi, H., Singer, G. *et al.* 2020. Harnessing the power of eDNA metabarcoding for the detection of deep-sea fishes. Plos ONE, 15: e0236540.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T. *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.

Muha, T. P., Robinson, C. V., de Leaniz, C. G., and Consuegra, S. 2019. An optimised eDNA protocol for detecting fish in lentic and lotic freshwaters using a small water volume. Plos ONE, 14: e0219218.

Pawlowski, J., Apothéloz-Perret-Gentil, L., Mächler, E., and Altermatt, F. 2020. Environmental DNA applications in biomonitoring and bio-assessment of aquatic ecosystems. Guidelines. Environmental Studies No 2010. Federal Office for the Environment, Bern. 71pp.

Pompanon, F., Coissac, E., and Taberlet, P. (2011). Metabarcoding: a new way to analyze biodiversity. Biofutur, 319: 30–32.

Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., and Yamahara, K. M. 2016. Assessing vertebrate biodiversity in a kelp forest using environmental DNA. Molecular Ecology, 25: 527–541.

Ramón-Laca, A., Wells, A., and Park, L. 2021. A workflow for the relative quantification of multiple fish species from oceanic water samples using environmental DNA (eDNA) to support large-scale fisheries surveys. Plos ONE, 16: e0257773.

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., and Coissac, E. 2011. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Research, 39: e145.

Salter, I., Joensen, M., Kristiansen, R., Steingrund, P., and Vestergaard, P. 2019. Environmental DNA concentrations are correlated with regional biomass of Atlantic cod in oceanic waters. Communications Biology, 2: 461.

Sanches, T. M., and Schreier, A.D. 2020. Optimizing an eDNA protocol for estuarine environments: balancing sensitivity, cost and time. Plos ONE, 15: e0233522.

Sato, H., Sogo, Y., Doi, H., and Yamanaka, H. 2017. Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. Scientific Reports, 7: 14860.

Singer, G. A. C., Fahner, N. A., Barnes, J. G., McCarthy, A., and Hajibabaei, M. 2019. Comprehensive biodiversity analysis via ultra-deep patterned flow cell technology: a case study of eDNA metabarcoding seawater. Scientific Reports, 9: 5991.

Stoeckle, M. Y., Soboleva, L., and Charlop-Powers, Z. 2017. Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. Plos ONE, 12: e0175186.

Stoeckle, M. Y., Das Mishu, M., and Charlop-Powers, Z. 2020. Improved environmental DNA reference library identifies overlooked marine fishes in coastal New Jersey, U.S.A. Frontiers in Marine Science, 7: 226.

Stoeckle, M. Y., Adolf, J., Charlop-Powers, Z, Dunton, K. J., Hinks, G., and VanMorter, S. M. 2021. Trawl and eDNA assessment of marine fish diversity, seasonality, and relative abundance in coastal New Jersey, USA. ICES Journal of Marine Science, 78: 293–304.

Taberlet, P., Bonin, A., Zinger, L., and Coissac, E. 2018. Environmental DNA for Biodiversity Research and Monitoring. New York, NY: Oxford University Press.

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., and Kawabata, Z. 2012. Estimation of fish biomass using environmental DNA. Plos ONE, 7: e35868.

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, W., Jørgensen, O. A., and Willerslev, E. 2016. Environmental DNA from seawater samples correlate with trawl catches of subArctic, deepwater fishes. Plos ONE, 11: e0165252.

Turner, C. R., Barnes, M. A., Xu, C. C. Y., Jones, S. E., Jerde, C. L, and Lodge, D. M. 2014. Particle size distribution and optimal capture of aqueous macrobial DNA. Methods in Ecology and Evolution, 5: 676–684.

Ushio, M., Murakami, H., Masuda, R., Sado, T., Miya, M., Sakurai, S., Yamanaka, H. *et al.* 2018. Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. Metabarcoding and Metagenomics, 2, 1–15.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., Thomsen, PF, Bellemain, E. *et al.* 2015. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology, 25: 929–942.

Valsecchi, E., Bylemans, J., Goodman., S. J., Lombardi, R., Carr, I., Castellano, L., Galimberti, A. *et al.* 2020. Novel universal primers for metabarcoding environmental DNA surveys of marine mammals and other marine vertebrates. Environmental DNA, 00: 1–17.

Vences, M., Lyra, M. L., Perl, R. G. B., Bletz, M. C., Stanković, D., Lopes, C. M., Jarek, M. *et al.* 2016. Freshwater vertebrate metabarcoding on illumina platforms using double-indexed primers of the mitochondrial 16S rRNA gene. Conservation Genetics Resources, 8: 323–327.

Yates, M. C., Glaser, D., Post, J., Cristescu, M. E., Fraser, D. J, and Derry, A.M. 2020. The relationship between eDNA particle concentration and organism abundance in nature is strengthened by allometric scaling. Molecular Ecology, 30: 1–15.

Zhang, S., Zhao, J., and Yao, M. 2020. A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. Methods in Ecology and Evolution, 00: 1–17.

Handling Editor: William Grant