GoFish: a streamlined environmental DNA presence/absence assay for marine vertebrates

Mark Y. Stoeckle^{1*}, Mithun Das Mishu², and Zachary Charlop-Powers³

¹Program for the Human Environment, The Rockefeller University, New York, New York, United States of America

²Hunter College, New York, NY, United States of America

³Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, New York, NY, United States of America

*Corresponding author [MYS]

Email: mark.stoeckle@rockefeller.edu

Abstract

Here we describe GoFish, a streamlined environmental DNA (eDNA) presence/absence assay. The assay amplifies a 12S segment with broad-range vertebrate primers, followed by nested PCR with M13-tailed, species-specific primers. Sanger sequencing confirms positives detected by gel electrophoresis. We first obtained 12S sequences from 77 fish specimens representing 36 northwestern Atlantic taxa not well documented in GenBank. Using the newly obtained and published 12S records, we designed GoFish assays for 11 bony fish species common in the lower Hudson River estuary and tested seasonal abundance and habitat preference at two sites. Additional assays detected nine cartilaginous fish species and a marine mammal, bottlenose dolphin, in southern New York Bight. GoFish sensitivity was equivalent to Illumina MiSeq metabarcoding. Unlike quantitative PCR (qPCR), GoFish does not require tissues of target and related species for assay development and a basic thermal cycler is sufficient. Unlike Illumina metabarcoding, indexing and batching samples are unnecessary and advanced bioinformatics expertise is not needed. The assay can be carried out from water collection to result in three days. The main limitations so far are species with shared target sequences and inconsistent amplification of rarer eDNAs. We think this approach will be a useful addition to current eDNA methods when analyzing presence/absence of known species, when turnaround time is important, and in educational settings.

Introduction

DNA profiling of ecological communities was first applied to terrestrial microbes [1]. DNA extracted from soil samples—amplified with ribosomal RNA gene primers, cloned, and analyzed by Sanger sequencing—revealed an enormous diversity of uncultured organisms. Whole genome shotgun sequencing provided an alternative culture-independent approach [2]. Combining targeted amplification with high-throughput sequencing, first 454 then Illumina, eliminated cloning and Sanger sequencing, greatly facilitating microbiome study [3-5].

Around the same time, ancient DNA techniques began to be applied to environmental samples, with recovery of 10,000 years-old to 400,000 years-old plant and animal DNA from fecal samples and sediments [6,7]. The earliest reports examining contemporary materials include differentiating human and domestic sources in sewage-contaminated water [8] and recovery of Arctic fox DNA from snow footprints [9]. Taberlet and colleagues were the first to apply an environmental DNA approach to present-day ecology, demonstrating pond water eDNA accurately surveys an invasive frog species [10]. Subsequent work revealed aquatic eDNA detects diverse vertebrates and invertebrates in multiple habitats [11-15]. Aquatic eDNA assays now routinely monitor rare and invasive freshwater species [16-19].

Beginning in 2003, the DNA barcoding initiative firmly demonstrated that most animal species are distinguished by a short stretch of mitochondrial (mt) COI gene [20-22]. This led researchers to assess animal communities by “metabarcoding”, i.e., high-throughput sequencing of mtDNA segments amplified from environmental samples [23-26]. The sequence variability that makes COI an excellent identifier of species hobbles broad-range primer design [27]. Primers targeting highly conserved regions in vertebrate 12S or 16S mt genes [28-30]

successfully profile aquatic vertebrate communities [31-38]. Multi-gene metabarcoding promises kingdom-wide surveys of eukaryotic diversity [39-43].

Aquatic vertebrate eDNA assays challenge in design and execution. Developing a single-species qPCR test typically necessitates obtaining tissue samples of the target organism and potential confounding species [e.g., 44], and running assays requires a dedicated thermal cycler. High-throughput metabarcoding involves indexing and batching a large number of samples for each sequencing run, and advanced bioinformatics expertise to decode output files.

To facilitate wider use, we aimed for an eDNA assay that did not require tissue samples for validation and could be completed in less than a week. One straightforward technique is species-specific amplification followed by gel electrophoresis and Sanger sequencing, as in early eDNA reports [10]. However, our preliminary experiments generated visible products only in samples with a high number of MiSeq reads, indicating low sensitivity. In addition, multiple bands were frequent, likely to interfere with Sanger sequencing.

Nested PCR is a highly sensitive and specific approach to identifying genetic variants [e.g., 45]. Nested PCR improves detection of earthworm eDNA from soil samples archived for more than 30 years [46] and enables highly sensitive eDNA assays for salmonid fishes and a fresh water mussel [47,48].

We reasoned that nested PCR applied to a metabarcoding target could be a more general approach. Amplification with broad-range metabarcoding primers could provide a “universal” first round, followed by nested PCR with species-specific primers, taking advantage of sequence differences within the amplified segment.

Here we test whether a metabarcoding/nested PCR protocol detects aquatic vertebrate eDNA in time-series water samples from lower Hudson River estuary and southern New York

Bight and compare results to those obtained with MiSeq metabarcoding. Because this assay involves querying amplified material one species at a time, we name it after the children's card game Go Fish in which a player might ask "do you have any Jacks?"

Results

New 12S reference sequences

Regional checklist species [49,50] with absent or incomplete GenBank 12S records were flagged. For bony fish, we focused on common species; for the much smaller number of cartilaginous fish we sought specimens from all not in GenBank. Seventy-seven specimens representing 36 target species were obtained from NOAA Northeast Spring Trawl Survey, Monmouth University, fish markets, or beach wrack (Tables S1,S2).

Specimen DNAs were sequenced for a 750 base pair (bp) 12S segment [51] encompassing three commonly used vertebrate eDNA target sites [29,30,32] (Fig 1), and for 648 bp COI barcode region. COI sequences confirmed taxonomic identifications, showing 99.4% to 100% identity to GenBank reference accessions, excepting that of Northern stargazer [*Astroscopus guttatus*], which at the time of this study had no GenBank COI records (Tables S1,S2).

Fig 1. GoFish eDNA assay. Top, schematic of GoFish and MiSeq metabarcoding protocols. Bottom, diagram of 12S and flanking tRNA genes, with locations and sizes of vertebrate metabarcoding targets (MiFish, ECO, Teleo) and the Li segment sequenced from reference specimens as indicated.

GoFish nested PCR assay

Of three commonly used vertebrate 12S metabarcoding targets (Fig 1), the MiFish segment is longer than the other two and has hypervariable regions near the ends, features facilitating species-specific nested PCR. In addition, by targeting a different segment than what our laboratory uses for MiSeq metabarcoding (12S ECO V5), we aimed to minimize potential cross-contamination between GoFish and MiSeq assays. First-round PCR for bony fish was done with MiFish primer set [30] (Fig 1, Table 1). The resultant reaction mix, diluted 1:20 in Elution Buffer (10mM Tris pH 8.3, Qiagen) served as input DNA for species-specific PCRs.

GoFish primers were designed for 11 fish species that together account for most (92%) lower Hudson River estuary fish eDNA reads (Table 2) [52]. The nested primers generated unambiguous presence/absence bands on gel electrophoresis (Fig 2). In all samples analyzed so far, Sanger sequencing confirmed GoFish amplified only targeted species.

Fig 2. GoFish amplifications visualized on 2.5% agarose gel with SYBER Safe. Lanes bracketed by dates are time series samples; the last three lanes in each panel are negative controls detailed in Materials and Methods. Marker indicates dye front at approximately 150 bp. Gel positives were sent for Sanger sequencing to confirm target species amplification.

Table 1. Broad-range vertebrate 12S primers. PCR parameters and expected amplicon sizes are shown. M13 and Illumina tails in Li primers and ECO V5 primers, respectively, are highlighted in bold.

	Name	Tm (C) (not incl M13 or Illumina tail)	Sequence
12S Li segment (Reference specimen sequencing)			
Bony fish	M13-12S229F	56.6	TGT AAA ACG ACG GCC AGT GYC GGT AAA AYT CGT G
	M13-12S954R	60.6	CAG GAA ACA GCT ATG AC YCC AAG YGC ACC TTC CG
Cartilaginous fish	M13-12S229SF	54.6	TGT AAA ACG ACG GCC AGT GTT GGT HAA TCT CGT G
	M13-12S954SR	52.1	CAG GAA ACA GCT ATG AC TCC AAG TRC ACT TTC CA
12S MiFish segment (GoFish first-round PCR)			
Bony fish	MiFish-U-F	58.7	GTC GGT AAA ACT CGT GCC AGC
	MiFish-U-R	56.7	CAT AGT GGG GTA TCT AAT CCC AGT TTG
Cartilaginous fish	MiFish-E-F	56.5	GTT GGT AAA TCT CGT GCC AGC
	MiFish-E2-R	52.2	CAT AGT AGG GTA TCT AAT CCT AGT TTG
Mammals	MiFish-W-F	55.1	GTT GGT AAA TTT CGT GCC AGC
	MiFish-U-R		same as for bony fish
12S ECO V5 segment (MiSeq metabarcoding)			
Bony fish, mammals	ECO-V5-F	50.1	TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG A
	ECO-V5-R	49.6	GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G
Cartilaginous fish	ECO-V5-SF	48.7	TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG A
	ECO-V5-R		same as for bony fish

Table 2. Species-specific GoFish primers. Amplicon sizes, PCR parameters, and target specificity as shown. M13 tails are highlighted in bold.

	Primer name	Tm species-specific segment (C)	Primer sequence
Bony fish			
American eel (<i>Anguilla rostrata</i>)	M13ameeF M13ameeR	47.7 49.5	TGT AAA ACG ACG GCC AGT GGG CTC AAA TTG ATA TTA C CAG GAA ACA GCT ATG AC C GTG AGT TCA AAG GTG T
Atlantic menhaden (<i>Brevoortia tyrannus</i>)	M13atmeF M13atmeR	48.2 48.2	TGT AAA ACG ACG GCC AGT GAG TGG TTA TGG AGA ACT CAG GAA ACA GCT ATG AC ATC CCA GTT TGT GTC CCG
Bay anchovy (<i>Anchoa mitchilli</i>)	M13baanF M13baanR	48.3 50.3	TGT AAA ACG ACG GCC AGT GTG GTT ATG GAA TTC TTT T CAG GAA ACA GCT ATG AC GAT AAA GTC ACT TTC GTG TC
Black sea bass (<i>Centropristis striata</i>)	M13blsbF M13blsbR	51.2 51.2	TGT AAA ACG ACG GCC AGT GGG TGG TTA GGA CAT ACT CAG GAA ACA GCT ATG AC CTT TCG TGG GTT CAG AAT AA
Bluefish (<i>Pomatomus saltatrix</i>)	M13blfiF M13blfiR	54.6 57.1	TGT AAA ACG ACG GCC AGT AGA GTG GTT AAG GAA AGC CAG GAA ACA GCT ATG AC TCG TGG GGT CAG GAA TGG
Cunner (<i>Tautoglabrus adspersus</i>)	M13cunnF M13cunnR	54.6 57.5	TGT AAA ACG ACG GCC AGT GTA AAG AGT GGT TAG GGC CAG GAA ACA GCT ATG AC CTC TCG TGG GGT CAG GTG
Oyster toadfish (<i>Opsanus tau</i>)	M13oytoF M13oytoR	52.6 50.3	TGT AAA ACG ACG GCC AGT CGC GGT TAC ACG AAT GA CAG GAA ACA GCT ATG AC ATA GTT TAC GTG GTG TCA AA
Scup (<i>Stenotomus chrysops</i>)	M13scupF M13scupR	48.7 50.1	TGT AAA ACG ACG GCC AGT GGG TGG TTA AGA ATA AAC CAG GAA ACA GCT ATG AC AAT CCC AGT TTG TGT CTC
Seaboard goby (<i>Gobiosoma ginsburgii</i>)	M13segoF M13segoR	52.8 51.7	TGT AAA ACG ACG GCC AGT GCC CAA GTT GAC AAC TCA CAG GAA ACA GCT ATG AC CTT TCG TGG GGT CAT ATG TA
Striped bass (<i>Morone saxatilis</i>)	M13stbaF M13stbaR	53.0 57.4	TGT AAA ACG ACG GCC AGT GGT TAA GGG CCC AAC TTT CAG GAA ACA GCT ATG AC TTT CGT GGG GTC AGG TTT G
Tautog (<i>Tautoga onitis</i>)	M13tautF M13tautR	50.4 55.7	TGT AAA ACG ACG GCC AGT GTA AAG AGT GGT TAG GAT CAG GAA ACA GCT ATG AC CTC TCG TGG GGT CAG GTA
Sharks, rays, skates			
Sand tiger shark (<i>Carcharias taurus</i>)	M13stshF M13stshR	50.8 51.6	TGT AAA ACG ACG GCC AGT CGA GTA ACT TAT ATT AAT A CAG GAA ACA GCT ATG AC TGA CAT CAA GAT TTC TAG TA
Sandbar shark (<i>Carcharhinus plumbeus</i>)	M13sbshF M13sbshR	51.3 50.4	TGT AAA ACG ACG GCC AGT CGA GTA ACT CAC ATT AAC A CAG GAA ACA GCT ATG AC GTG ACA TCA AGG TTC CTT AC
Smooth dogfish shark (<i>Mustelus canis</i>)	M13smdoF M13smdoR	50.9 52.2	TGT AAA ACG ACG GCC AGT CGA GTG ACT CAT ATT AAC A CAG GAA ACA GCT ATG AC GCA TCA AGG CTC CTT GA
Bullnose ray (<i>Myliobatis freminvillei</i>)	M13buraF M13buraR	51.4 50.1	TGT AAA ACG ACG GCC AGT AGG GTG ATT AGA ATT AAT C CAG GAA ACA GCT ATG AC TGT CGT GAG GTC AAA AAC

Cownose ray (<i>Rhinoptera bonasus</i>)	M13coraF M13coraR	50.2 51.3	TGT AAA ACG ACG GCC AGT GGT GAT TAG AAA TAA TCT C CAG GAA ACA GCT ATG AC CGT GAG GTC AAA AAT TCT GT
Roughtail stingray (<i>Dasyatis centroura</i>)	M13rostF M13rostR	50.8 50.8	TGT AAA ACG ACG GCC AGT ACG AGT GAC ACA AAT TAA T CAG GAA ACA GCT ATG AC GTG AGG TCA AAA ACT CTG TT
Spiny butterfly ray (<i>Gymnura altavela</i>)	M13sbraF M13sbra-R	50.9 50.6	TGT AAA ACG ACG GCC AGT TAA GGG TGA TTA GAA AAA T CAG GAA ACA GCT ATG AC AGG TCA AAA ATT CTG TTG TG
Clearence skate (<i>Raja eglanteria</i>)	M13clskF M13clskR	49.2 52.6	TGT AAA ACG ACG GCC AGT CGA GTA ACT CAT ATT AAT A CAG GAA ACA GCT ATG AC GTC GTG AAT TCA AAA GCT CT
Little skate (<i>Leucoraja erinacea</i>)	M13liskF M13liskR	51.2 52.7	TGT AAA ACG ACG GCC AGT CGA GTA ACT CAC ATT AAT A CAG GAA ACA GCT ATG AC TGT CGT GAG GTC AAA AGC
Marine mammals			
Bottlenose dolphin (<i>Tursiops truncatus</i>)	M13bodoF M13bodoR	49.2 49.6	TGT AAA ACG ACG GCC AGT TGA CCC AAA CTA ATA GAC A CAG GAA ACA GCT ATG AC TCT TAG TTG TCG TGT ATT CA

We applied these GoFish assays to a four-month time series of water samples collected weekly at two contrasting lower Hudson River estuary locations—a high flow, rocky tidal channel on the east side of Manhattan, and a low-flow, sandy bottom site in outer New York harbor (Fig 3]. Species detections increased seasonally at both sites, consistent with historical trawl surveys and a metabarcoding eDNA time series [52]. Despite large tidal flows in the estuary, eDNAs differed by site consistent with habitat preferences, with rocky bottom specialists (cunner, oyster toadfish, seaboard goby) more commonly detected in East River than in outer New York harbor.

Fig 3. GoFish detections at two lower Hudson estuary locations sampled weekly from March to August 2017. At top, black indicates detections, with species arranged by decreasing number of positives; at bottom, number of species present on each date is shown.

Cartilaginous fish, marine mammals

We tested this approach on cartilaginous fish and marine mammals, groups relatively understudied by eDNA so far [12,34,36,53-55]. First-round MiFish primers were modified to favor cartilaginous fish or mammals (Table 1). Species-specific GoFish amplifications successfully detected three shark species, four rays, and two skates (Table 2). In a nine-month time series of water samples from southern New Jersey, cartilaginous fish eDNAs peaked in mid-summer, consistent with seasonal migration patterns (Fig 4) [56]. One exception was little skate (*Leucoraja erinacea*), present largely in fall samples, in line with its greater fall and winter abundance [57]. Sanger sequencing confirmed species ID for all gel-positive amplifications, except that clearnose skate (*Raja englanteria*) and little skate primers amplified non-target sequences in some samples. A GoFish assay for bottlenose dolphin (*Tursiops truncatus*) was positive in most summer and fall samples; sequencing verified all (Table 2, Fig 4).

Fig 4. GoFish detections of cartilaginous fish and bottlenose dolphin eDNA. Water samples were collected at one- to two-week intervals from April to December 2017 in southern New York Bight.

Comparison to Illumina metabarcoding

The New York City time series samples were analyzed by an Illumina MiSeq metabarcoding protocol targeting 12S ECO V5 segment (Fig 1). The apparent sensitivity (method detections/total detections) for both protocols was about 80% (Fig 5). As expected, the proportion of GoFish “drop-outs” differed by MiSeq read number—more abundant eDNAs were amplified more consistently than were rarer eDNAs.

Fig 5. Comparison of GoFish and metabarcoding. A. Method detections for samples and species analyzed in Fig 3. B. GoFish detections stratified by MiSeq read number.

Discussion

Here we report species-specific nested PCR eDNA assays for 20 marine fish and one marine mammal. The GoFish assay can potentially be adapted to detect any vertebrate with a 12S reference sequence; tissue specimens are not necessary. It can be completed in less than a week with standard molecular biology equipment and interpreted with Sanger sequencing-level bioinformatics. A single broad-range amplification suffices for multiple species-specific assays.

To facilitate primer design we sequenced a 12S fragment, covering three commonly analyzed vertebrate eDNA metabarcoding targets, from 77 specimens representing 36 local species, boosting GenBank 12S coverage to 95% of lower Hudson River estuary checklist species [52].

The main limitations so far are species lacking sequence differences in the target segment (assays require species-specific substitutions in both primer sites and in amplified segment) and inconsistent amplification of rarer eDNAs, shortcomings shared with metabarcoding. GoFish can be viewed as an alternate form of metabarcoding, one that uses species-specific amplification rather than high-throughput sequencing to analyze the pool of PCR products generated by broad-range primers.

Several pairs or sets of local species have insufficient MiFish segment sequence differences for GoFish assays. Of note, these include *Alosa* herrings, of commercial and conservation interest: alewife (*A. pseudoharengus*), American shad (*A. sapidissima*), blueback

herring (*A. aestivalis*), and hickory shad (*A. mediocris*). It may be possible to design nested PCR assays for *Alosa* herrings by targeting a different mtDNA region.

GoFish sensitivity was equivalent to that of MiSeq metabarcoding, with drop-outs in both assays (Fig 5). As expected, GoFish drop-outs were mostly those with lower MiSeq read numbers, and presumably represent rarer eDNAs. Inconsistent amplification of low abundance DNAs was recognized as a hazard early on [10]. If desired, replicate amplification or other PCR enhancement strategies [58] could be applied to GoFish. False-negatives may be inherent to broad-range primers, which “rarely detect lineages accounting for less than 0.05% of the total read count, even after 15 PCR replicates” [59] [also 40]. Looked at more generally, all ecological survey methods generate false-negatives; site occupancy modeling can help infer true presence/absence [60-63].

We assumed that sequences matching regional species indicated the presence of that species. This could lead to overlooking extralimital occurrences of taxa with shared sequences. For instance, the locally abundant Atlantic menhaden (*Brevoortia tyrannus*) shares GoFish target sequences with Gulf menhaden (*B. patronus*), found in Gulf of Mexico.

Although likely impractical to apply GoFish to the hundreds of fish species typically resident in any given marine region, it may be possible to characterize communities by targeting the smaller number of species that account for the majority of biomass (e.g., Fig 3).

The non-labor costs for a GoFish assay were about \$15 per sample for one species, and \$8 per sample per additional species. One hard to quantify advantage is constrained cross-contamination. Because it is cost-effective to test small sets of samples, a GoFish assay puts fewer results at risk than does high-throughput sequencing. This feature could be particularly

valuable in educational settings with less expert performers, instead of putting “all your eggs in one basket” in a MiSeq run.

eDNA promises to help better understand and appreciate ocean life. We think GoFish will be a useful addition to eDNA tools when species of interest are known and relatively few in number, when turnaround time is important, and in educational settings.

Materials and Methods

The GoFish protocol was designed for persons familiar with basic molecular biology techniques and access to essential molecular biology laboratory equipment. To facilitate use, we utilized commercial kits and open source software, and standardized PCR and sequencing protocols. Procedures were performed on an open bench following routine molecular biology precautions. Particulars include gloves worn for all laboratory procedures and changed after handling water samples and PCR reactions, filtration equipment scrubbed and rinsed thoroughly after each use with tap water, and pipettors and workspace areas wiped with 10% bleach after use. Unfiltered pipette tips were employed; after each procedure used tips were discarded and collection containers rinsed with 10% bleach. Our aquatic eDNA methods are posted online at protocols.io site (<https://dx.doi.org/10.17504/protocols.io.p9gdr3w>).

Standardized GoFish 12S amplification protocol

Limited customization of cycle number and annealing temperature was applied (Tables 1,2). Materials and conditions were as follows: GE Illustra beads in 0.2 ml tubes (8 tube strips); 25 µl reaction volume; 5 µl input DNA; 250 nM each primer; and thermal cycler program of

95 °C for 5 m, (25 or 35) cycles of [95 °C for 20 s, (60 °C or 65 °C) for 20 s, 72 °C for 20 s], and 72 °C for 1 m. Primers were obtained from Integrated DNA Technologies (IDT).

New 12S, COI reference sequences

DNA was extracted from tissues using PowerSoil kit (MoBio). 12S primer sequences and PCR parameters applied to reference specimen DNAs are shown in Table 1. PCR conditions were otherwise as listed in GoFish protocol above. Amplifications were confirmed by agarose gel electrophoresis with SYBER Safe dye (Thermo Fisher Scientific), and PCR clean-up and bidirectional sequencing with M13 primers were done at GENEWIZ. Consensus sequences were assembled in MEGA, using 4Peaks to assess trace files [64,65]. For COI, COI-3 primer cocktail [66], 35 cycles, and 55 °C annealing were used with GoFish protocol outlined above. Substitute forward primers were employed for specimens that failed to generate high-quality sequences with COI-3 cocktail [hickory shad, American shad (M13alosaCOI-F, 5'-TGT AAA ACG ACG GCC AGT TCA ACT AAT CAT AAA GAT ATT GGT AC-3'); windowpane flounder (M13wiflCOI-F, 5'-TGT AAA ACG ACG GCC AGT CTA CCA ACC ACA AAG ATA TCG G-3')]. 12S and COI reference sequences (Files S1,S2) are deposited in GenBank (Accession nos. MH377759-MH377835 and MH379020-MH379090, respectively).

Water collection, filtration, DNA extraction

Water sampling was done under permit from New York City Department of Parks and Recreation at two locations: East River (40.760443, -73.956354), a rocky, high-flow tidal channel on the east side of Manhattan, and Steeplechase Pier, Coney Island (40.569576, -

73.983297), a sandy bottom, low-flow location in outer New York Harbor (Fig 3). One-liter surface water samples were collected weekly at both sites from March 31, 2017 to August 3, 2017 (34 samples in total).

With authorization from New Jersey Department of Environmental Protection, surface water samples were collected on a barrier island beach (39.741641, -74.112961) about 70 miles south of New York City and halfway to Cape May, the southern border of New York Bight (Fig 3). 22 one-liter samples were collected at one-week to two-week intervals from April 2, 2017 to December 23, 2017.

Samples were filtered within 1 h of collection or stored at 4 °C for up to 48 h beforehand. Water was poured through a paper coffee filter to exclude large particulate matter and then into a filtration apparatus consisting of a 1000 ml side arm flask attached to wall suction, a frittered glass filter holder (Millipore), and a 47 mm, 0.45 µm pore size nylon filter (Millipore). Filters were folded to cover the retained material and stored in 15 ml tubes at -20° C until DNA extraction. As negative controls, one-liter samples of laboratory tap water were filtered and DNA extracted using the same equipment and procedures as for environmental samples.

DNA was extracted with PowerSoil kit with modifications from manufacturer's protocol to accommodate the filter [52]. DNA was eluted with 50 µl Buffer 6 and concentration measured using Qubit (Thermo Fisher Scientific). Typical yield was 1 µg to 5 µg DNA per liter water filtered.

No animals were housed or experimented upon as part of this study. No endangered or protected species were collected.

GoFish assay

Broad-range 12S amplification

Primer sequences, annealing temperatures, and cycle numbers are shown in Table 1. Different MiFish primer sets targeted bony fish, cartilaginous fish, or marine mammals. Tap water eDNA and reagent-grade water were included as negative controls on all amplification sets. After PCR, 5 µl of reaction mixture were run on a 2% agarose gel with SYBER Safe to assess amplification. Rather than affinity bead purification, we diluted the reaction mix 20-fold in Elution Buffer and used 5 µl for nested amplifications, effecting a 100-fold dilution of first-round reaction products. With this protocol, a single broad-range amplification sufficed for 80 species-specific assays.

Species-specific primer design, amplification, verification

An alignment of 12S MiFish segment sequences from regional fish species including those obtained in this study, marine mammals, and commonly detected non-marine vertebrates (human, pig, chicken, cow, dog, rat), was generated in MEGA using MUSCLE [67], sorted according to a neighbor-joining tree, exported to Excel, and used to generate a matrix showing differences from the consensus [68]. Primers were selected by eye according to desired criteria: two or more nucleotide mismatches against other species at or near the 3' end, a T_m not including M13 tail of 50.0 °C to 52.0 °C according to IDT website, and diagnostic differences within the amplified segment that confirmed target species detection. G-T or T-G primer-template mismatches were considered relatively permissible and so less useful in conferring specificity [30]. M13 tails enabled a single primer set to sequence all detections and improved 5' end reads; the latter was particularly helpful given the short amplicons generated by GoFish primers (Table 2).

The standardized amplification protocol described above was employed. Default annealing temperature was 60 °C; if non-target amplification occurred, primers were tested at 65 °C. Three negative controls were included in all runs: the two negative controls from the broad-range PCR, and a reagent-grade water blank. A 5 µl aliquot of each PCR reaction was run on an agarose gel with SYBER Safe (Fig 2); positives were sent to GENEWIZ for cleanup and bidirectional sequencing with M13 primers. Sequences were matched to a local file of 12S reference sequences.

Metabarcoding

As a comparison, eDNA samples were also analyzed by MiSeq metabarcoding protocol using broad-range primers that target 12S ECO V5 segment as previously described (Fig 1) [52]. Briefly, DNA samples from PowerSoil extraction were further purified with AMPure XP (Beckman Coulter) and resuspended in 50 µl of Elution Buffer. Primer sequences and amplification parameters are given in Table 1. Each sample was amplified with ECO V5 primer set as previously described, which effectively targets bony fish and mammals, and with a modified ECO V5 primer set that favors cartilaginous fish (Table 1). As for MiFish amplifications, tap water eDNA and reagent-grade water negative controls were included in all sets. 5 µl of each reaction were run on a 2% agarose gel with SYBR Safe dye. Some negative controls gave faint bands; with MiSeq, these turned out to be human or domestic animal DNA, commonly observed in eDNA work [69].

PCR products were diluted 1:20 in Elution Buffer and Nextera index primers [Illumina] were added following the standardized amplification protocol with 13 cycles and annealing temperature 55 °C. 5 µl of each reaction were run on a 2% agarose gel with SYBR Safe dye to

confirm amplification. Indexed PCR libraries were pooled, treated with AMPure XP, and adjusted to 5.4 ng/μl (30 nM assuming 270 bp amplicon) according to Qubit. Sequencing was done at GENEWIZ on an Illumina MiSeq (2 x 150 bp). 56 experimental and 11 control libraries, plus other samples not reported here, were analyzed in three runs with 86-96 libraries per run. Original fastq files with metadata are deposited in NCBI Sequence Read Archive (NCBI BioProject ID PRJNA358446).

Bioinformatic analysis was performed as previously described [52], using DADA2, which identifies all unique sequences rather than lumping according to threshold criteria [70]. DADA2-generated OTU tables and FASTA files of unique sequences are in Supporting Information (Table S3, Files S3-S5). As in prior report, detections representing less than 0.1% of total reads for that sequence were excluded to minimize misassigned reads. Tap water eDNA and reagent-grade water controls were negative for fish eDNA after filtering.

Acknowledgments

We thank Jesse Ausubel for encouragement and editorial comment, Jeanne Garbarino, Nica Rabinowitz, Doug Heigl, and Odaelys Walwyn for laboratory space and assistance, John Galbraith, Jakub Kircun, and Keith Dunton for fish specimens, and Victor Shahov and Thomas Poku for trialing protocols.

References

1. Head IM, Saunders JR, Pickup RW. Microbial evolution, diversity, and ecology: a decade of ribosomal RNA analysis of uncultivated microorganisms. *Micro Ecol.* 1998;35: 1-21.

2. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*. 2004;304: 66-74.
3. Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AK, Kent AD, Daroub SH, Camargo FA, Farmerie WB, Triplett EW. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J*. 2007;4: 283-290.
4. Acosta-Martínez V, Dowd S, Sun Y, Allen V. Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. *Soil Biol Biochem*. 2008;40: 2762-2770.
5. Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Appl Environ Microbiol*. 2011;77: 3848-3852.
6. Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz, Bland H, Evershed RP, Possnert G, Pääbo S. Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science*. 1998;281: 402-406.
7. Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, Shapiro B, et al. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*. 2003;300: 791-795.
8. Martellini A, Payment P, Villemur R. Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine, and ovine sources in fecally contaminated surface water. *Water Res*. 2005;39: 541-548.
9. Dalén L, Götherström, Meijer T, Shapiro B. Recovery of DNA from footprints in the snow. *Canadian Field-Naturalist*. 2007;121: 321-324.
10. Ficetola GF, Miaud C, Pompanon F, Taberlet P. Species detection using environment DNA from water samples. *Biol Lett*. 2008;4: 423-425.
11. Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, Gilbert MTP, et al. Monitoring endangered freshwater biodiversity using environmental DNA. *Mol Ecol*. 2012;21: 2565-2573.
12. Foote AD, Thomsen PF, Sveegaard S, Wahlberg M, Kielgast J, Kyhn LA, et al. Investigating the potential use of environmental DNA [eDNA] for genetic monitoring of marine mammals. *PLOS ONE*. 2012;7: e41781.
13. Deiner K, Fronhofer EA, Mächler E, Walser J-C, Altermatt F. Environmental DNA reveals that rivers are conveyor belts of biodiversity information. *Nature Comm*. 2016;7: e12544.
14. Minamoto T, Yamanaka H, Takahra T, Honjo MN, Kawabata Z. Surveillance of fish species composition using environmental DNA. *Limnology*. 2012;13: 193-197.

- 382 15. Tréguier A, Paillisson J-M, Dejean T, Valentini A, Schlaepfer MA, Roussel J-M.
383 Environmental DNA surveillance for invertebrate species: advantages and technical limitations
384 to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. J Applied Ecol. 2014;51:
385 871-879.
- 386 16. Dejean T, Valentini A, Miquel C, Taberlet P, Bellemain E, Miaud C. Improved detection of
387 an alien invasive species through environmental DNA barcoding: the example of the American
388 bullfrog *Lithobates catesbeianus*. J Applied Ecol. 2012;49: 953-959.
- 389 17. Rees HC, Bishop K, Middleditch DJ, Patmore JRM, Maddison BC, Gough KC. The
390 application of eDNA for monitoring of the Great Crested Newt in the UK. Ecol Evol. 2014;4:
391 4023-4032.
- 392 18. Nathan LM, Simmons M, Wegleitner BJ, Jerde CL, Mahon AR. Quantifying environmental
393 DNA signals for aquatic invasive species across multiple detection platforms. Environ Sci
394 Technol. 2014;48: 12800-12806.
- 395 19. Wilson C, Wright E, Bronnenhuber J, MacDonald F, Belore M, Locke B. Tracking ghosts:
396 combined electrofishing and environmental DNA surveillance efforts for Asian carps in Ontario
397 waters of Lake Erie. Management Biol Invasions. 2014;5: 225-231.
- 398 20. Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM. Identification of birds through DNA
399 barcodes. PLOS Biol. 2004;2: e312.
- 400 21. Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN. DNA barcodes distinguish
401 species of tropical Lepidoptera. Proc Natl Acad Sci U S A. 2006;103: 968-971.
- 402 22. Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, Burrige M, Watkinson D, Dumont P,
403 Curry A, Bentzen P, Zhand J, April J, Bernatchez L. Identifying Canadian freshwater fishes
404 through DNA barcodes. PLOS ONE. 2008;3: e2490.
- 405 23. Pompanon F, Coissac, Taberlet P. Metabarcoding, une nouvelle façon d'analyser la
406 biodiversité. Biofutur. 2011;319: 30-32.
- 407 24. Yoccoz NG. The future of environmental DNA in ecology. Mol Ecol. 2012;21: 2031-2038.
- 408 25. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation
409 biodiversity assessment using DNA metabarcoding. Mol Ecol. 2012;21: 2045-2050.
- 410 26. Lodge DM, Turner CR, Jerde CL, Barnes MA, Chadderton L, Egan SP, Feder JL, Mahon AR,
411 Pfreder ME. Conservation in a cup of water: estimated biodiversity and population abundance
412 from environmental DNA. Mol Ecol. 2012;21: 2555-2558.
- 413 27. Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P. DNA metabarcoding and the
414 cytochrome c oxidase subunit I marker: not a perfect match. Biol Lett. 2014;10: 20140562.

- 415 28. Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC.
416 Dynamics of mitochondrial evolution in animals: amplification and sequencing with conserved
417 primers. *Proc Natl Acad Sci U S A*. 1989;86: 6196-6200.
- 418 29. Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E. ecoPrimers: inference of
419 new DNA barcode markers from whole genome sequence analysis. *Nucl Acids Res*. 2011;39:
420 e145.
- 421 30. Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, et al. MiFish, a set of universal
422 PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230
423 subtropical marine species. *R Soc Open Sci*. 2015;2: 150088.
- 424 31. Thomsen PF, Kielgast J, Iversen LL, Moller PR, Rasmussen M, Willerslev E. Detection of
425 diverse marine fish fauna using environmental DNA from seawater samples. *PLOS ONE*.
426 2012;7: e41732.
- 427 32. Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF et al. Next-generation
428 monitoring of aquatic biodiversity using environmental DNA barcoding. *Mol Ecol*. 2016;25:
429 929-942.
- 430 33. Kelly RP, Port JA, Yamahara KM, Crowder LB. Using environmental DNA to census
431 marine fish in a large mesocosm. *PLOS ONE*. 2014;9: e86175
- 432 34. Thomsen PF, Moller PR, Sigsgaard EE, Knudsen W, Jorgensen OA, Willerslev E.
433 Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater
434 fishes. *PLOS ONE* 2016;11: e0165252.
- 435 35. Port JA, O'Donnell JL, Romero-Maraccini OC, Leary PR, Litvin SY, Nickols KJ, et al.
436 Assessing vertebrate biodiversity in a kelp forest using environmental DNA. *Mol Ecol*. 2016;25:
437 527-541.
- 438 36. Andruszkiewicz EA, Starks HA, Chavez FP, Sassoubre LM, Block BA, Boehm AB.
439 Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *PLOS ONE*
440 2017;12: e0176343.
- 441 37. Yamamoto S, Masuda R, Sato Y, Sado T, Araki H, Kondoh M, et al. Environmental DNA
442 metabarcoding reveals local fish communities in a species-rich coastal sea. *Sci Rep*. 2017;7:
443 40368.
- 444 38. Sigsgaard EE, Nielsen IB, Carl H, Krag MA, Knudsen SW, Xing Y, Holm-Hansen TH,
445 Møller PR, Thomsen PF. Seawater environmental DNA reflects seasonality of a coastal fish
446 community. *Mar Biol*. 2017;164: 128.
- 447 39. Kelly RP, O'Donnell JL, Lowell NC, Shelton AO, Samhoury JE, Hennessey SM, Feist BE,
448 Williams GD. Genetic signatures of ecological diversity along an urbanization gradient. *Peer J*.
449 2016;4: e2444.

450 40. Kelly RP, Closek CJ, O'Donnell JL, Kralj JE, Shelton AO, Samhouri JF. Genetic and manual
451 survey methods yield different and complementary views of an ecosystem. *Front Marine Sci.*
452 2017;3: 283.

453 41. O'Donnell JL, Kelly RP, Shelton AO, Samhouri JF, Lowell NC, Williams GD. Spatial
454 distribution of environmental DNA in a nearshore marine habitat. *Peer J.* 2017; 5: e3044.

455 42. Stat M, Huggett MJ, Bernasconi R, DiBattista JD, Berry TE, Newman SJ, Harvey ES, Bunce
456 M. Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical
457 marine environment. *Sci Rep.* 2017;7: 12240.

458 43. Djurhuus A, Pitz K, Sawaya NA, Rojas-Márquez J, Michaud B, Montes E, Muller-Karger F,
459 Breitbart M. Evaluation of marine zooplankton community structure through environmental
460 DNA metabarcoding. *Limnol Oceanography: Methods.* 2018;10: 1-13

461 44. Laramie MB, Pilliod DS, Goldberg CS. Characterizing the distribution of an endangered
462 salmonid using environmental DNA analysis. *Biol Conserv.* 2015;183: 29-37.

463 45. Snounou G, Viriyakosol S, Zhu XP, Jarra W, Pinheiro L, do Rosario VE, Thaithong S,
464 Brown KN. High sensitivity of detection of human malaria parasites by the use of nested
465 polymerase chain reaction. *Mol Biochem Parasitol.* 1993;61: 315-320.

466 46. Jackson M, Myrholm C, Shaw C, Ramsfield T. Using nested PCR to improve detection of
467 earthworm eDNA in Canada. *Soil Biol Biochem.* 2017;113: 215-218.

468 47. Clusa L, Ardura A, Fernández S, Roca AA, Garcia-Vázquez E. An extremely sensitive
469 nested PCR-RFLP mitochondrial marker for detection and identification of salmonids in eDNA
470 from water samples. *Peer J.* 2017;5: e3045.

471 48. Stoeckle BC, Kuehn R, Geist J. Environmental DNA as a monitoring tool for the endangered
472 freshwater pearl mussel [*Margaritifera margaritifera* L.]: a substitute for classical monitoring
473 approaches? *Aquatic Conserv: Mar Freshwater Ecosyst.* 2015;26: 1120-1129.

474 49. Able KW. Checklist of New Jersey saltwater fishes. *Bull NJ Acad Sci.* 1992;37: 1-11

475 50. Briggs PT, Waldman JR. Annotated list of fishes reported from the marine waters of New
476 York. *Northeastern Naturalist.* 2002;9: 47-80.

477 51. Li C, Orti G. Molecular phylogeny of Clupeiformes (Actinopterygii) inferred from nuclear
478 and mitochondrial DNA sequences. *Mol Phylogenet Evol.* 2007;44: 386-398.

479 52. Stoeckle MY, Soboleva L, Charlop-Powers Z. Aquatic environmental DNA detects seasonal
480 fish abundance and habitat preference in an urban estuary. *PLOS ONE.* 2017;12: e0175186.

481 53. Sigsgaard EE, Nielsen IB, Bach SS, Lorenzen ED, Robinson DP, Knudsen SW, et al.
482 Population characteristics of a large whale shark aggregation inferred from seawater
483 environmental DNA. *Ecol Evol.* 2016;1: 14.

484 54. Bakker J, Wangenstein OS, Chapman DD, Boussarie G, Buddo D, Guttridge TL, Hertler H,
485 Mouillot D, Vigliola L, Mariani S. Environmental DNA reveals tropical shark diversity in
486 contrasting levels of anthropogenic impact. *Sci Rep.* 2017;7: 16886.

487 55. Baker CS, Steel D, Niekirk S, Klinck H. Environmental DNA (eDNA) from the wake of
488 whales: droplet digital PCR for detection and species identification. *Frontiers Marine Sci.* 2018;9:
489 968-980.

490 56. Kells V, Carpenter K. A Field Guide to Coastal Fishes: from Maine to Texas. Baltimore: The
491 Johns Hopkins University Press; 2011.

492 57. Bigelow HB, Schroeder WC. Fishes of the Gulf of Maine, First Revision. Fishery Bull Fish
493 Wildlife Service. 1953;53: 1577.

494 58. Pechgit P, Intarapuk A, Pinyoowong D, Bhumiratana A. Touchdown-touchup nested PCR for
495 low-copy gene detection of benzimidazole-susceptible *Wuchereria bancrofti* with a *Wolbachia*
496 endosymbiont imported by migrant carriers. *Experimental Parasitol.* 2011; 127: 559-568.

497 59. Sato H, Sogo Y, Doi H, Yamanaka H. Usefulness and limitations of sample pooling for
498 environmental DNA metabarcoding of freshwater fish communities. *Sci Rep.* 2017;7: 14860.

499 60. MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA Estimating
500 site occupancy rates when detection probabilities are less than one. *Ecol.* 2002;83: 2248-2255.

501 61. Ficetola GF, Pansu J, Bonin A, Coissac E, GiguetCovex C, DeBarba M, et al. Replication
502 levels, false presences and the estimation of the presence/absence from eDNA metabarcoding
503 data. *Mol Ecol Resources.* 2015;15: 543556.

504 62. Furlan EM, Gleeson D, Hardy CM, Duncan RP. A framework for estimating the sensitivity
505 of eDNA surveys. *Mol Ecol Resources.* 2016;16:641-654.

506 63. Lahoz-Monfort JJ, GuilleraArroita G, Tingley R. Statistical approaches to account for false-
507 positive errors in environmental DNA samples. *Mol Ecol Res.* 2016;16: 673-683.

508 64. Kumar S, Tamura K, Nei M [2004] MEGA3: integrated software for molecular evolutionary
509 genetics analysis and sequence alignment. *Brief Bioinform.* 5: 150–163.

510 65. Griekspoor A, Groothuis T. 4Peaks software, available at <https://nucleobytes.com>. Accessed
511 January 1, 2017.

512 66. Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN. Universal primer cocktails for fish DNA
513 barcoding. *Mol Ecol Notes.* 2007;7: 544-548.

514 67. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
515 *Nucl Acids Res* 2004;32: 1792-1797.

- 516 68. Stoeckle MY, Kerr KCR. Frequency matrix approach demonstrates high sequence quality in
517 avian BARCODEs and highlights cryptic pseudogenes. PLOS ONE. 2012;7: e43992.
- 518 69. Leonard JA, Shanks O, Hofreiter M, Kreuz E, Hodges L, Ream W, et al. Animal DNA in
519 PCR reagents plagues ancient DNA research. J Anthropol Sci. 2007;34: 1361-1366.
- 520 70. Callahan BJ, McMurdie PJ, Rosen NJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-
521 resolution sample inference from Illumina amplicon data. Nature Methods. 2016;13: 581-583.

522

523 **Supporting Information**

524 **Table S1. Bony fish specimens analyzed for 12S, COI.**

525 **Table S2. Cartilaginous fish specimens analyzed for 12S, COI.**

526 **Table S3. DADA2 OTU tables.**

527 **File S1. New 12S reference sequences.**

528 **File S2. New COI reference sequences.**

529 **File S3. DADA2 FASTA file for MiSeq run jun2017.**

530 **File S4. DADA2 FASTA file for MiSeq run oct2017.**

531 **File S5. DADA2 FASTA file for MiSeq run jan2018.**

532

MiSeq

GoFish

Scup
(*Stenotomus chrysops*)

American eel
(*Anguilla rostrata*)

Coney Island

Mar-----Aug

Atlantic menhaden
Tautog
Bluefish
Striped bass
Bay anchovy
Cunner
Scup
Oyster toadfish
Seaboard goby
American eel
Black sea bass

East River

Mar-----Aug

A. GoFish

MiSeq

Detections

B.

		GoFish detection
MiSeq reads	>10K	18/18 (100%)
	>1K	67/77 (87%)
	>0.1K	35/53 (66%)
	overall	120/148 (81%)