Received: 16 November 2017 | Revised: 14 June 2018 | Accepted: 25 June 2018

DOI: 10.1111/1755-0998.12928

RESOURCE ARTICLE

WILEY MOLECULAR ECOLOGY RESOURCES

Capture enrichment of aquatic environmental DNA: A first proof of concept

Taylor M. Wilcox[1,2,3] | Katherine E. Zarn[1,2] | Maxine P. Piggott[3] | Michael K. Young[1] | Kevin S. McKelvey[1] | Michael K. Schwartz[1]

[1]Rocky Mountain Research Station, National Genomics Center for Wildlife and Fish Conservation, U.S. Department of Agriculture, Forest Service, Missoula, Montana

[2]University of Montana, W.A. Franke College of Forestry & Conservation, Missoula, Montana

[3]Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, Acton, Australian Capital Territory, Australia

Correspondence
Taylor M. Wilcox, U.S. Department of Agriculture, Forest Service, National Genomics Center for Wildlife and Fish Conservation, Rocky Mountain Research Station, 800 E. Beckwith Ave., Missoula, MT 59801.
Email: taylor.m.wilcox@gmail.com

Funding information
Australian Research Council, Grant/Award Number: DE130100777; Directorate for Biological Sciences, Grant/Award Number: DGE-1313190; University of Montana; ANU-ActewAGL Endowment Fund, Grant/Award Number: 58 2014

Abstract

Environmental DNA (eDNA) sampling—the detection of genetic material in the environment to infer species presence—has rapidly grown as a tool for sampling aquatic animal communities. A potentially powerful feature of environmental sampling is that all taxa within the habitat shed DNA and so may be detectable, creating opportunity for whole-community assessments. However, animal DNA in the environment tends to be comparatively rare, making it necessary to enrich for genetic targets from focal taxa prior to sequencing. Current metabarcoding approaches for enrichment rely on bulk amplification using conserved primer annealing sites, which can result in skewed relative sequence abundance and failure to detect some taxa because of PCR bias. Here, we test capture enrichment via hybridization as an alternative strategy for target enrichment using a series of experiments on environmental samples and laboratory-generated, known-composition DNA mixtures. Capture enrichment resulted in detecting multiple species in both kinds of samples, and post-capture relative sequence abundance accurately reflected initial relative template abundance. However, further optimization is needed to permit reliable species detection at the very low-DNA quantities typical of environmental samples (<0.1 ng DNA). We estimate that our capture protocols are comparable to, but less sensitive than, current PCR-based eDNA analyses.

KEYWORDS
capture, eDNA, environmental DNA, metabarcoding, metagenetics

1 | INTRODUCTION

Environmental DNA (eDNA) sampling uses genetic material in environmental samples to infer the presence of species (Jerde, Mahon, Chadderton, & Lodge, 2011). This sampling approach has been widely adopted for detecting aquatic animals because it has been found to be more sensitive and less expensive than traditional sampling approaches (Smart, Tingley, Weeks, Rooyen, & McCarthy, 2015; Wilcox et al., 2016). Increasingly, work has focused on using high-throughput sequencing (HTS) to detect multiple taxa simultaneously (Deiner et al., 2017). All animals slough DNA into the environment, a process that potentially leaves traces of all species within a habitat. The exciting implication is that all members of a biotic community are potentially detectable within a single environmental sample (Lodge et al., 2012).

There are, however, a number of obstacles to using eDNA sampling for whole-community assessment. First, DNA in the water is a heterogeneous mixture, often dominated by plant, bacterial and fungal DNA. For example, Turner, Barnes, et al. (2014) found that DNA from an established population of common carp (*Cyprinus carpio*) made up ≤0.0004% of total DNA in water samples (<0.01 ng mitochondrial DNA/litre). Second, reference sequences for taxonomic assignment are only available for one or a few genes of the entire genome of most animal species. Because animal eDNA is rare and

only a small portion of the genome has adequate reference data, eDNA sampling for most species relies on using PCR to enrich for short "barcoding" regions of the genome. This approach, called "metabarcoding," uses "universal" primers to amplify small regions (up to several hundred base pairs), massively increasing their relative abundance so that amplicons dominate the post-PCR DNA sequence population (Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012). This generates HTS libraries composed primarily of barcode reads that can be assigned back to known taxa.

Although metabarcoding has proven successful in some systems (e.g., Thomsen et al., 2012; Valentini et al., 2016), inferring community composition from PCR-based enrichment of DNA from multiple taxa within environmental samples remains problematic. Although P CR primers for metabarcoding are highly conserved among target taxa, they still amplify some templates (DNA from some species) with greater efficiency than others. Small differences in amplification efficiency are compounded during PCR so that templates with high-amplification efficiency are overenriched relative to other templates. These primer biases result in skewed relative sequence abundances and failure to detect some taxa known to be present, particularly those that are rare (e.g., Elbrecht & Leese, 2015; Evans et al., 2016; Kelly, Port, Yamahara, & Crowder, 2014). This is problematic because detection of these rare taxa is often a key motivation for adopting eDNA sampling (e.g., Mckelvey et al., 2016).

To facilitate metabarcoding, much work has focused on optimizing amplification protocols, primer design, and accounting for primer bias within bioinformatic analyses (Elbrecht & Leese, 2016; O'Donnell, Kelly, Lowell, & Port, 2016; Thomas, Deagle, Eveson, Harsch, & Trites, 2016). An optimal solution, however, might be to avoid bulk amplification of relatively generic DNA sequences altogether. In some cases, such as mixed invertebrate samples, mitochondrial enrichment via centrifugation prior to sequencing may be sufficient to characterize community composition (Zhou et al., 2013). For eDNA analysis, several authors have suggested that capture enrichment could offer a primer-bias-free alternative to metabarcoding (Creer et al., 2016; Jones & Good, 2016; Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012). With this approach, rather than amplifying a region of interest to increase its relative abundance, the targeted region is hybridized to a DNA or RNA probe bound to a magnetic bead, and then, all other sequences are washed away. Capture-based approaches have already been used for heavily contaminated ancient DNA (Briggs et al., 2009; Carpenter et al., 2013) and samples containing DNA from multiple invertebrate species (Liu et al., 2016; Shockralla et al., 2016). Capture enrichment may circumvent the issues of primer bias described above and allow any genetic target to be enriched and sequenced (Taberlet, Coissac, Hajibabaei, et al., 2012). However, little is known about the sensitivity and limitations of capture enrichment for aquatic community eDNA analysis.

Here, we test the efficacy of capture enrichment using environmental samples and known-composition genomic DNA mixtures. To our knowledge, this is the first demonstration of capture enrichment for animal DNA detection from water samples. Our results highlight both the potential for capture enrichment of eDNA and some of the technical challenges associated with enrichment from low-concentration, mixed DNA samples.

2 | MATERIALS AND METHODS

We conducted two capture enrichment experiments (Experiments 1 and 2). Our goals were to (a) determine the sensitivity of our capture enrichment protocols for low-concentration DNA samples (*sensitivity assessment*), (b) determine whether capture enrichment preserved relative sequence abundance information (*relative abundance assessment*), (c) test capture enrichment on environmental samples as a proof of concept (*Environmental samples*) and (d) evaluate the effects of two different indexing approaches to assess whether index tag-jumping affected inferences about species composition (*tag-jump assessment*).

2.1 | Capture work-flow

Our capture enrichments targeted commonly sequenced mitochondrial genes of 40 aquatic, semiaquatic and riparian taxa (Table 1). We synthesized probes from biotinylated PCR products as described by Maricic, Whitten, and Pääbo (2010) and Peñalba et al. (2014). In Experiment 1, we used one *COI* (655–749 bp), *cytb* (421–540 bp) and *16S* (550 bp; primers 16SA-L/16SB-H in Vences, Thomas, Meijden, Chiari, & Vietes, 2005) probe for each taxon, plus two *ND2* (550–555 bp) probes for *Pandion haliaetus* only (probe synthesis details in Supporting Information Tables S1–S3 in Appendix S1). In Experiment 2, we only included probes for *COI* and *cytb* for each taxon because preliminary results from Experiment 1 suggested that these were most taxonomically informative and that the *16S* probes appear to have led to the capture of many nontarget chloroplast and bacterial sequences (data not shown). We constrained analyses of both experiments to *COI* and *cytb* to permit direct comparison between experiments.

In each experiment, there were (a) replicate, known-composition mixtures of tissue-extracted DNA, (b) environmental DNA samples and (c) one no-template control sample (n = 80 and 40 libraries for Experiments 1 and 2, respectively; sample descriptions below). We extracted DNA from tissue samples using DNeasy Blood and Tissue Kits (QIAGEN). For invertebrate DNA extractions, we used legs when possible to limit gut microbe contamination of DNA samples. Invertebrates were frozen using liquid nitrogen, crushed using 1.5-ml tube pestles, and RNase A was added according to the kit protocol. We prepared libraries following the protocol described in Meyer and Kircher (2010), shearing DNA to a mean length of 300 bp using an E220 Focused-Ultrasonicator (Covaris). In Experiment 1, libraries were double-indexed with 80 unique combinations of 24 indices (20 P7 and 4 P5 adaptors; Supporting Information Table S1). In Experiment 2, libraries were double-indexed with 40 unique combinations of 80 indices so that no index was used for more than one library (40 P7 and 40 P5 adaptors; Supporting Information Table S4). HPLC-purified indexing primers were

TABLE 1 List of taxa targeted in capture enrichment experiments

Taxonomic group	Taxa
Amphibians	Boreal Toad (*Bufo boreas boreas*)
	Columbia Spotted Frog (*Rana luteiventris*)
	Idaho Giant Salamander (*Dicamptodon aterrimus*)
	Rocky Mountain Tailed Frog (*Ascaphus montanus*)
Birds	Mallard (*Anas platyrhynchos*)
	Osprey (*Pandion haliaetus*)
Fishes	Arctic Grayling (*Thymallus arcticus*)
	Bonneville Cutthroat Trout (*Oncorhynchus clarkii utah*)
	Brook Trout (*Salvelinus fontinalis*)
	Brown Trout (*Salmo trutta*)
	Bull Trout (*Salvelinus confluentus*)
	Cedar Sculpin (*Cottus schitsuumsh*)
	Chinook Salmon (*Oncorhynchus tshawytscha*)
	Coho Salmon (*Oncorhynchus kisutch*)
	Kokanee Salmon (*Oncorhynchus nerka*)
	Lake Trout (*Salvelinus namaycush*)
	Lake Whitefish (*Coregonus clupeaformis*)
	Mottled Sculpin (*Cottus bairdii*)
	Mountain Whitefish (*Prosopium williamsoni*)
	Northern Pike (*Esox lucius*)
	Rainbow Trout (*Oncorhynchus mykiss*)
	Slimy Sculpin (*Cottus cognatus*)
	Sockeye Salmon (*Oncorhynchus nerka*)
	Torrent Sculpin (*Cottus rhotheus*)
	Westslope Cutthroat Trout (*Oncorhynchus clarkii lewisi*)
	Yellowstone Cutthroat Trout (*Oncorhynchus clarkii bouveri*)
Invertebrates	Caddisfly (*Hydropsyche instabilis group*)
	Caddisfly (*Lepidostoma cinereum*)
	Mayfly—Pale Morning Dun (*Ephemerella excrucians*)
	Mayfly—Western Green Drake (*Drunella grandis*)
	Mayfly (*Baetis tricaudatus* complex)
	Signal Crayfish (*Pacifastacus leniusculus*)
	Stonefly—Giant Salmonfly (*Pteronarcys californica*)
	Stonefly—Golden Stone (*Hesperoperla pacifica*)
	Stonefly—Western Stone (*Calineuria californica*)
	Stonefly—Yellow Sally (*Isoperla fulva*)
Mammals	Mink (*Neovison vison*)
	North American Beaver (*Castor canadensis*)
	North American River Otter (*Lontra canadensis*)
	Nutria (*Myocastor coypus*)

purchased from Eurofins Scientific. Libraries were indexed using 9–22 cycles of PCR, based on the number of cycles needed to reach PCR plateau (estimated via qPCR as described in Meyer & Kircher, 2010), then pooled in equimolar concentrations to a total quantity of 2 μg for capture.

We used capture enrichment as described in Maricic et al. (2010) with the addition of 10 μg mouse *COT-1* DNA to minimize capture of repetitive genomic DNA (Peñalba et al., 2014). The postcapture pools were amplified with 12 cycles in four PCR replicates (Experiment 1) or 15 cycles in two PCR replicates (Experiment 2; based on postcapture library quantification). Replicate final pool amplifications were carried out to reduce library bias induced by PCR stochasticity. Each experiment was sequenced using 150-bp paired-end reads on a single MiSeq lane at the University of Montana/USFS Genomics Core (Missoula, MT, USA).

2.2 | Taxonomic assignment

Initial bioinformatic analysis was carried out in Galaxy (release 17.05; Afgan et al., 2016). Demultiplexed forward and reverse reads underwent adapter trimming (*Trimmomatic*; Bolger, Lohse, & Usade, 2014) and were filtered by quality (≥90% bases with PHRED quality scores ≥30). Duplicate reads were removed using *FastqUniq* (Xu et al., 2012), and then, reads were compared with a reference database (78 of 80 targeted *COI* and *cytb* genes from GenBank and our own sequencing; two missing sequences were *cytb* for *Pteronarcys californica* and *Ephemerella excrucians* for which Sanger sequencing was difficult and template DNA was limiting; Supporting Information Table S5) via *BLAST* (Altschul, Gish, Miller, Myers, & Lipman, 1990). We retained all sequences for which both the forward and reverse reads had ≥95% identity and 100% coverage with at least one reference sequence. Using a custom script in *R* (version 3.4.3; R Core Team 2017), we filtered *BLAST* output to retain only sequences where the forward or reverse reads were ≥100 bp. We then assigned sequences to genus if identity with a reference was ≥97% for *COI* or *cytb* and assigned sequences to family if identity with a reference was ≥95%. We removed any sequences that could not be unambiguously assigned to a genus or family. We assessed proportion on-target reads by dividing the total number of reads assigned to genus per sample by the number of reads passing quality filtering.

We determined appropriate thresholds for taxonomic assignment by conducting a sliding window analysis for each of the targeted genes. Unlike metabarcoding, where the loci of amplified sequences are anchored by the primer annealing sites, capture enrichment may result in sequences across and extending beyond the targeted gene. As a result, taxonomic assignment must account for variation in interspecific divergence within and adjacent to the targeted region. We aligned reference sequences for each gene (*CLUSTAL* in *MEGA5*; Tamura et al., 2011; Thompson, Higgins, & Gibson, 1994), then used the *R* package *SPIDER* (Brown et al., 2012) to quantify minimum divergence between genera and families at every possible 100-bp window across each gene. We set thresholds for taxonomic assignment so that there were no windows with less interspecific divergence than the threshold (e.g., minimum raw *COI* window divergence between genera was >3%).

2.3 | Sensitivity assessment

To assess the minimum DNA concentration at which capture enrichment was effective, we created a titration series of a known-composition DNA mixture with masses of 0.01, 0.1, 1, 10 and 100 ng genomic DNA (gDNA). These dilutions were made from a single mixture of tissue-extracted gDNA from each of the 40 focal taxa (equal gDNA mass from each taxon) with 10 replicates (Experiment 1) or three replicates (Experiment 2) at each dilution level (DNA quantified via fluorometry with the Qubit 2.0 dsDNA Broad Range Assay, ThermoFisher Scientific). Because all samples contained the same set of taxa and initial relative abundance of DNA templates, capture enrichment should produce comparable results among replicates. We used Bray–Curtis dissimilarity to assess variation in taxon detection and relative sequence abundances among replicates at each titration level and visualized this via nonmetric multidimensional scaling (NMDS) using the *R* package *vegan* (Oksanen et al., 2017).

2.4 | Relative abundance assessment

To assess our ability to retrieve relative abundances, we created known-composition, two-species mixtures of tissue-extracted DNA (100 ng DNA *total* for each mixture): *Pandion haliaetus* × *Salvelinus confluentus* (*P. haliaetus* at 50%, 90% and 99% of mixture; triplicates in Experiment 1), *Oncorhynchus clarkii lewisi* × *S. confluentus* (*O. c. lewisi* at 50%, 90%, and 99% of mixture; triplicates in Experiment 1), *S. fontinalis* × *O. c. lewisi* (*S. fontinalis* at 50% and 90% of mixture; triplicates in Experiment 2) and *P. haliaetus* × *O. c. lewisi* (*P. haliaetus* at 50% and 90% of mixture; triplicates in Experiment 2).

We created known-composition mixtures based on total genomic DNA mass. However, mtDNA content varies among taxa and tissue types (e.g., Rooney et al., 2015; Thomas et al., 2016). To account for this, we calculated an ad hoc adjustment to the two-species mixtures based on mtDNA copies per ng of total genomic DNA per sample. This was estimated via qPCR analysis of 1 ng DNA from each individual (n = 4) using *12S* primers designed for the amplification of vertebrate DNA (*12S-V* described in Riaz et al., 2011; suitability of primers was checked in silico via comparison with mitochondrial genome sequences available for *P. haliaetus*, *O. c. lewisi* and *S. fontinalis*). Reactions consisted of 10 μl SYBR Green Mastermix (Life Technologies), 150 nM each primer, 1 ng template DNA and molecular grade H_2O for a 20 μl total reaction volume. Cycling conditions were 95°C 5 min, [95°C 15 s, 60°C 1 min] × 40. Duplicate reactions were compared with a standard curve constructed from PCR products (10–10^6 copies/reaction; quantified via Qubit 2.0 dsDNA Broad Range Assay and converted to copy number based on estimated molecular weight). Standard curve r^2 = 0.98, efficiency = 80%, and there was <0.5 C_t spread among replicate reactions. For our analysis, we compared relative sequence abundance (experimental observation) with initial relative template abundance (expectation based on correction above) across samples.

2.5 | Environmental samples

As a proof of concept, we tested capture enrichment using environmental samples from Rattlesnake Creek, Montana, USA (n = 11 in August 2015 for Experiment 1 and 9 samples in March 2016 for Experiment 2; n = 20 total, sampling locations in Supporting Information Table S4). We collected 5-L water samples via in situ filtration according to the protocol described in Carim, McKelvey, Young, Wilcox, and Schwartz (2016), stored filters in silica desiccant until arrival at the laboratory and then extracted DNA from one half of each sample filter using the DNeasy Blood and Tissue Kit with QIAshredder (QIAGEN) as described by Goldberg, Pilliod, Arkle, and Waits (2011). The other filter half was stored for future analyses. Extractions were carried out in dedicated low-DNA laboratory spaces housed at the National Genomics Center for Wildlife and Fish Conservation (Missoula, MT, USA).

2.6 | Tag-jump assessment

We assessed tag-jump events—where indices used to assign sequences to samples are associated with the incorrect sequence—by assessing index combinations (P5 and P7 index combinations that were *not* used for libraries) in Experiment 2. For this experiment, 40 samples were tagged with pairs of unique P7 and P5 indices so that there were 40 *used* and 1,560 *unused* index combinations of 1,600 total possible combinations. As in other studies, we assumed that *used* combinations represent predominantly correctly tagged DNA fragments and that *unused* combinations represent errors (tag-jump events).

To consider all possible index combinations, we demultiplexed the raw MiSeq base call data from Experiment 2 using *bcl2fastq* (version 2.17; Illumina) for all 1,600 possible index combinations. We allowed zero index base pair mismatches and exported the P7 and P5 index reads (I1 and I2) to separate FASTQ files for each index combination. Mixed clusters on the sequencing flow cell can result in misidentification of samples, but this is detectable using double indexing with indexes that are not reused between samples (Kircher, Sawyer, & Meyer, 2012). Wright and Vetsigian (2016) found that bleeding of index sequences due to mixed clusters could be almost entirely removed by filtering sequences based on the quality score of the index sequences. To test this, we filtered used and unused index combinations by the quality scores of the P7 and P5 index sequences at 14 different thresholds (minimum PHRED quality score for both index sequences 20–33). Sequencing error or substitution during PCR amplification could result in returning an incorrect index sequence (Potapov & Ong, 2017; Schirmer et al., 2015). We expect tag-jump events due to sequencing and PCR error to be more common when there are fewer substitution errors necessary for an error to occur. To assess this, we compared the number of raw sequences per unused index combination as a function of that index combination's minimum Hamming distance (number of base pair differences) compared to the other 40 used index combinations. We assumed that remaining unused index combination rates were due to PCR-mediated tag-jumps (including chimera formation) and contamination across indexing primers.

3 | RESULTS

Libraries composed of tissue-derived DNA resulted in 752–863,768 filtered reads each (mean = 10,109 and 400,426 for Experiments 1 and 2, respectively), of which 24–57,283 reads (range = 2.0%–20.8%, mean = 12.7% and 4.2% for Experiments 1 and 2, respectively) assigned to genus and 24–57,410 sequences assigned to family (mean = 1,606 and 12,893 to genus and 1,676 and 12,962 to family for Experiments 1 and 2, respectively). Environmental samples resulted in 568–6,957 high-quality sequences per library (mean = 1,366 and 4,904 for Experiments 1 and 2, respectively), of which 0–20 sequences (range = 0%–1.4%, mean = 0.4% and 0.1% for Experiments 1 and 2, respectively) assigned to a genus and 0–22 sequences assigned to family (mean = 6 and 4 to both genus and family for Experiments 1 and 2, respectively).

In the sensitivity assessment, capture enrichment resulted in similar taxon lists and relative sequence abundance across replicates when there was at least 0.1 ng of total gDNA in a sample. However, similarity among replicates declined sharply at the lowest DNA concentration tested (0.01 ng total; Figure 1).

In the relative abundance assessment, after accounting for variance in mtDNA content among samples, there was a strong, linear relationship between initial template relative abundance and relative sequence abundance in both experiments and for all pairwise species mixtures (r^2 = 0.98; Figure 2).

In the tag-jump analysis, unused index combinations in Experiment 2 made up 1,035,089 of all 16,996,207 reads across all possible index combinations (6.09%). Most unused index combinations had few sequences, but the distribution was right-skewed by several combinations with many sequences (mean = 663, median = 120, range = 0–21,961 raw sequences per unused index combination). Of the eight most abundant unused index combinations (top 0.5%), six were found in a regular series (Figure 3a). This pattern may suggest direct contamination of indexing primers between libraries, although the pattern does not correspond with the order of indexing primer synthesis or orientation of plate columns during library preparation. Filtering by index sequence quality score decreased the proportion of sequences assigning to unused index combinations to 5.90%–6.07%, but also reduced the proportion of remaining sequences assigning to a used index combination to 64.02%–99.66% of the original sequence number (minimum mean quality score 20–33; Figure 3b). The number of sequences per unused index combination was not related to Hamming distance from the most similar used index combination either before (F = 0.226 df = 1,558, p = 0.635; Figure 3c) or after filtering by index sequence quality score (F = 0.1661, df = 1,558, p = 0.684; minimum mean base quality score >33). Experiment 1 environmental samples included detections for 15 different genera, including four which are known to be absent from the sampling area (16.2% of sequences from genera known to be absent from the area). Experiment 2 environmental samples included detections for only eight genera, all of which were listed as recently detected in a county-level website (Montana Field Guide http://fieldguide.mt.gov/ accessed 30 May 2017; Figure 4). The no-template control libraries resulted in 3 (*Hesperoperla*, *Neovision*, *Oncorhynchus*) and 0 sequences assigned to genus and family for Experiments 1 and 2, respectively.

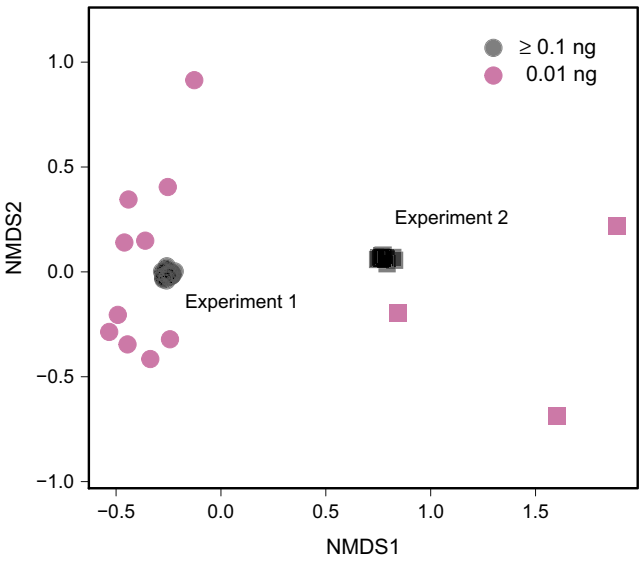

FIGURE 1 Titration replicate similarity. Nonmetric dimensional scaling plot showing Bray–Curtis similarity of titration replicates. Shape of points indicate experiment (circle and square for Experiments 1 and 2, respectively); colour indicates total genomic DNA mass per replicate (n = 10 and 3 replicates at each titration level for Experiments 1 and 2, respectively). Within each experiment, replicate similarity is high for DNA masses ≥0.1 ng, but is low at the lowest titration level (0.01 ng DNA; pink)

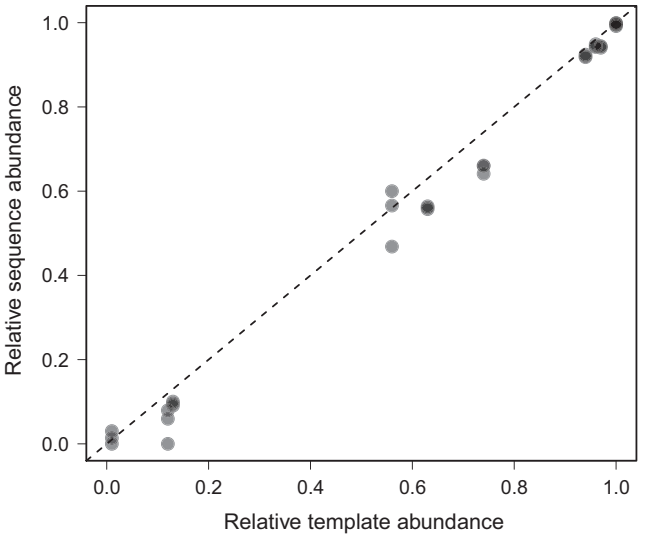

FIGURE 2 Two-taxa relative abundance. Observed relative sequence abundance versus initial relative template abundance in two-taxa mixtures (triplicates for each taxa combination and relative abundance level). Dotted line indicates expected 1:1 relationship

4 | DISCUSSION

We found that capture enrichment was effective for targeting mitochondrial loci from both known-composition DNA mixtures and

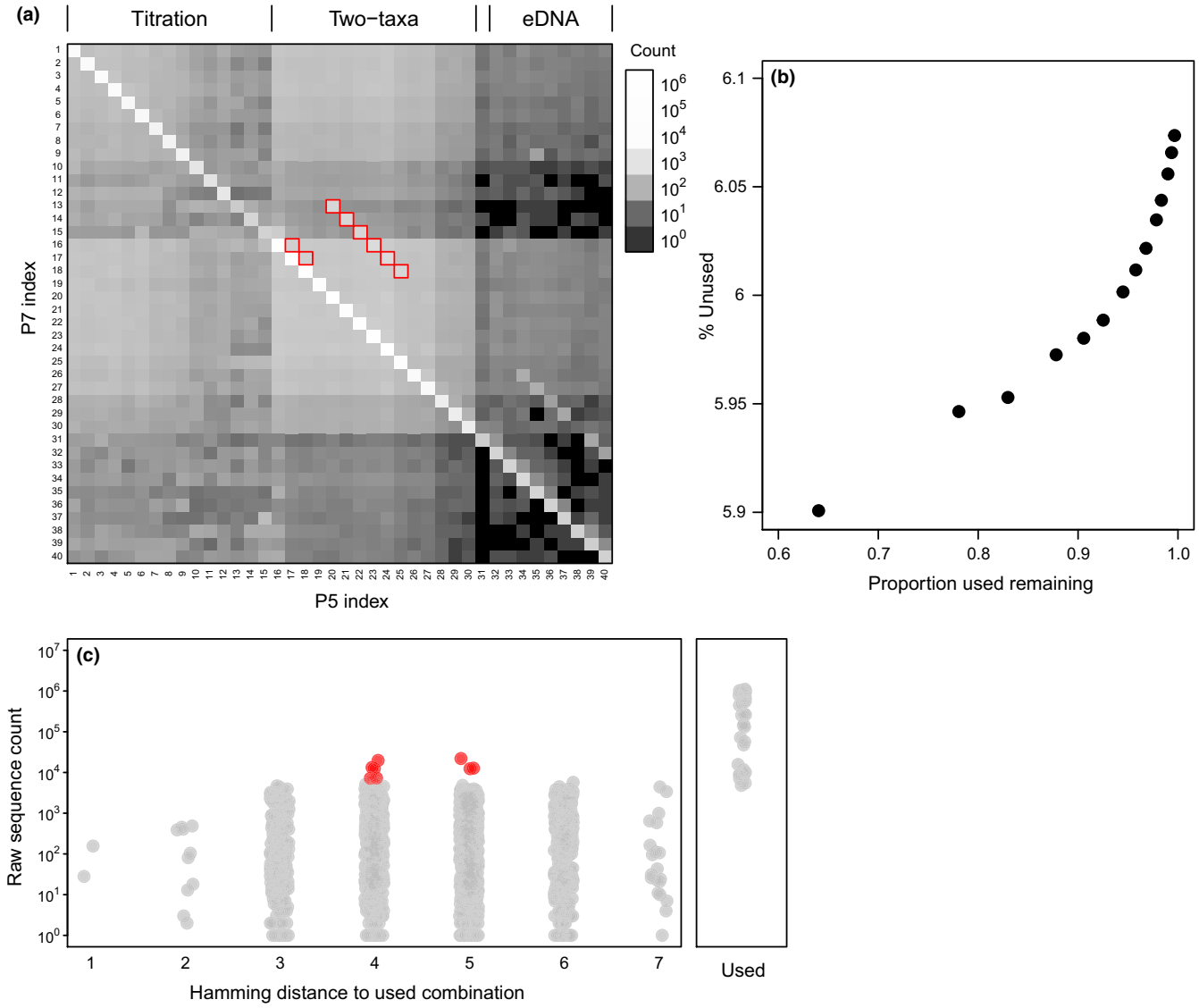

FIGURE 3 Patterns of tag-jump events in Experiment 2. (a) heat map showing the distribution of sequences with all possible combinations of P7 and P5 indices. The 40 samples along the diagonal (white) are the real eDNA and laboratory-concocted samples, and all others are tag-jump events. Highlighted are the eight most abundant tag-jump combinations (top 0.5% of tag-jumps). The regular series for six of these may be suggestive of a contamination event. (b) The proportion of used tag combinations (*x*-axis) that remain versus the per cent of unused tag combinations (*y*-axis) as index sequences are filtered more stringently for quality (mean quality score threshold 22–33). Filtering index sequences for quality reduces the rate of tag-jump events, but only explains a small proportion of the total unused index combinations observed in Experiment 2. (c) Number of sequences per unused index combination (*y*-axis) versus the Hamming distance (base pair differences; x-axis) to the most similar used index combination (distribution of used combination sequence counts on far right). Highlighted in red are the 8 most abundant tag-jump combinations. There is no significant relationship between number of sequences and Hamming distance, suggesting that index sequencing and PCR error are not primary drivers of the observed tag-jump events

environmental samples. To our knowledge, these experiments represent the first demonstration of capture enrichment for animal eDNA analysis from water samples. Although we found that the method requires optimization to improve sensitivity when DNA quantities are low (*Sensitivity assessment*), captured sequences reflected initial template DNA relative abundance (*Relative abundance assessment*) and detected a wide range of aquatic animal taxa (eight genera across six families of both vertebrates and invertebrates) from river water samples with no false-positive detections when libraries were double-indexed (*Environmental samples and tag-jump assessment*).

4.1 | Sensitivity assessment

Future applications of capture enrichment for eDNA will require protocol optimization for very low-concentration DNA samples. In Experiments 1 and 2, capture enrichment resulted in reproducible results among titration mixture replicates, but this reproducibility

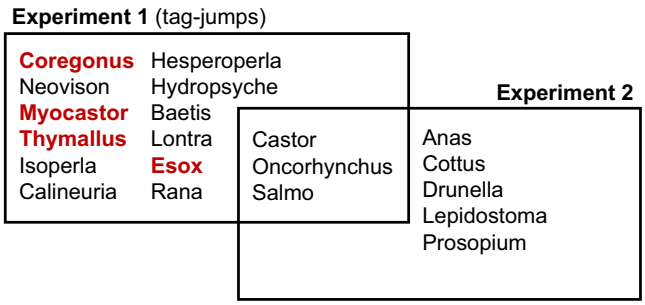

FIGURE 4 Venn diagram showing genera detected in environmental samples collected from Rattlesnake Creek (MT, USA) and analysed in Experiments 1 and 2. Genera not expected to be detected in the sample area indicated in red (Experiment 1 only). Detection of unexpected taxa in Experiment 1 samples was likely driven by tag-jump errors that the Experiment 2 indexing approach was robust to

rapidly declined when there was <0.1 ng of total gDNA (Figure 1). Based on quantification of mtDNA content in the four samples used for two-species mixtures, our 0.01–100 ng titrations are approximately equivalent to 10^3–10^7 total mtDNA copies per library (mean = 2.5×10^1–10^5 copies per taxon). Thus, our capture protocols appear to be less sensitive than PCR-based eDNA analyses (e.g., typical qPCR limit of detection <10 copies per reaction). However, there is also the opportunity to analyse a larger number of target copies for a given eDNA sample via capture versus a single PCR replicate. Unlike PCR-based approaches that typically use 1–4 μl of the DNA extract per analysis, our protocol uses >50 μl of template DNA solution per library, and additional template DNA could be used to increase sensitivity of this method.

Capture enrichment of low-abundance templates is sensitive to small protocol differences (e.g., Bragg, Potter, Bi, & Moritz, 2015; Paijmans, Fickel, Courtiol, Hofreiter, & Förster, 2015). Our capture enrichment protocol might be improved by the use of RNA-based probes (e.g., myBaits; Arbor Biosciences) and optimization of probe length, tiling density and hybridization conditions (e.g., Avila-Arcos et al., 2011; Cruz-Dávalos et al., 2017). The on-target capture rates for tissue-derived DNA (per cent of filtered sequences assigned to genus) observed in our protocol were low relative to similar protocols using tissue-derived DNA (mean = 12.7% and 4.2% for Experiments 1 and 2, respectively, compared to a mean of 15% and 56% for Maricic et al., 2010 and Peñalba et al., 2014; respectively), which may be related to lower initial library mass (0.01–100 ng in this study vs. 200 ng and >1,000 ng for Maricic et al., 2010 and Peñalba et al., 2014; respectively). Environmental DNA sample on-target capture rates were low (0.4% and 0.1% for Experiments 1 and 2, respectively), but within the range of rates observed in previous studies of ancient DNA with low levels of endogenous DNA (e.g., <0.01%–7.6% in 13 studies reviewed in Enk et al., 2014). Although we did not compare our results with shotgun sequencing of unenriched eDNA libraries, we assume that target species' mtDNA made up ≤0.001% of the total DNA, as observed in Turner, Barnes, et al. (2014). Besides optimizing the capture protocol, it may be possible to increase the initial DNA in captured samples by further optimizing field sampling and DNA extraction protocols (e.g., Turner, Miller, Coyne, & Corush, 2014).

4.2 | Relative abundance assessment

We found that relative sequence abundance was strongly correlated with initial relative template abundance in two-species mixtures. Accurately reflecting relative template abundance has been a key motivation in proposed research on capture enrichment for eDNA analysis (Creer et al., 2016; Jones & Good, 2016; Taberlet, Coissac, Hajibabaei, et al., 2012) and our findings support this potential. Often there exists a trade-off in metabarcoding approaches between taxonomic breadth (the level of evolutionary divergence over which a primer set will amplify taxa with low bias) and taxonomic resolution (the ability to distinguish closely related taxa from the amplicon). Here, we targeted entire genes commonly used in barcoding/metabarcoding and made taxonomic assignments at the genus level, but capture enrichment is potentially much more flexible. While metabarcoding is constrained to loci with a short, taxonomically informative region, flanked by conserved primer-binding regions, capture probes can be designed to target more divergent loci to differentiated closely related taxa or more conserved loci to facilitate cross-taxon enrichment. Thus, capture enrichment may facilitate assessment of relative template abundance with good taxonomic resolution across a broader taxonomic range than is generally the case for PCR-based approaches.

An interesting twist to this analysis is that relative sequence abundance reflected initial relative mtDNA copy abundance, *not* initial relative *total* DNA abundance. We found that our *O. c. lewisi* sample had a much higher abundance of mtDNA copies per ng of total DNA than the other samples in our analysis (~20×). We initially interpreted high numbers of *O. c. lewisi* sequences as overenrichment of this DNA due to an abundance of probes for closely related taxa in the capture enrichment pool. However, this apparent discrepancy was resolved by adjusting for differences in mitochondrial content among samples. Variance in mtDNA content among taxa may be an important driver of taxon-specific eDNA concentration–organism abundance relationships along with variance in body size (Elbrecht, Peinert, & Leese, 2017), developmental stage and breeding readiness (Maruyama, Nakamura, Yamanaka, Kondoh, & Minamoto, 2014; Spear, Groves, Williams, & Waits, 2015). Although deriving species abundance from sequence abundance is a common goal in eDNA sampling (e.g., Lacoursière-Roussel, Rosabal, & Bernatchez, 2016; Takahara, Minamoto, Yamanaka, Doi, & Kawabata, 2012), doing so using capture enrichment or other metabarcoding approaches will likely require species- and life stage-specific estimates of organism–eDNA concentration relationships.

4.3 | Environmental samples and tag-jump assessment

Our eDNA results confirm the potential to use hybridization to enrich animal eDNA from water samples and also further highlight the need to account for tag-jumping in inferences about species

composition in environmental samples (Schnell, Bohmann, & Gilbert, 2015). In Experiment 1, multiple sequences were assigned to environmental samples from taxonomic groups known to be absent from the sampling area. In contrast, Experiment 2 used a robust indexing design and there were only detections of taxonomic groups known to inhabit the sampled system. Different sets of samples were analysed in Experiments 1 and 2, which could result in different taxa DNA being present, but the difference in detection of known-absent taxa could only be caused by differences in either contamination (unlikely, because samples were collected and analysed in the same way between experiments) or in index assignment errors. Although there are multiple drivers of index assignment errors, we regard PCR-mediated tag-jumping as the primary culprit. As suggested by Wright and Vetsigian (2016), we filtered by minimum index sequence quality score to remove tag-jumps caused by bleeding between clusters on the sequencing flow cell. Doing this reduced the tag-jump rate by up to 0.19% (minimum mean quality score = 33), which is comparable to rates of error reported in other studies without the potential for PCR-mediated tag-jump events (range = 0.06%–0.21% in Ref. (Kircher et al., 2012; Nelson, Morrison, Benjamino, Grim, & Graf, 2014; Wright & Vetsigian, 2016)). However, this was a small proportion of the total tag-jump rate in Experiment 2, which was higher (6.09%) than tag-jump rates estimated from metabarcoding (1.9%–4% in Schnell et al., 2015). This may be because our postcapture libraries required more cycles of PCR prior to sequencing (e.g., 15 cycles for Experiment 2 vs. 10–12 in Schnell et al., 2015). During pooled PCR, tags may move across templates when chimeric sequences are formed as incompletely extended products from previous cycles anneal to similar templates. Because postcapture libraries tend to have low concentrations of DNA, PCR amplification of the entire pooled library will generally be necessary, which may make capture enrichment particularly sensitive to tag-jump errors. Together, our findings suggest that addressing PCR-mediated tag-jumps is particularly important for accurate inference from eDNA when using capture-based enrichment.

5 | CONCLUSION

Our laboratory and field experiments suggest that capture enrichment is useful for noninvasive biodiversity assessments of aquatic animal communities if sufficient DNA can be collected. Despite the challenges identified in this study, capture enrichment detected eDNA from animals of known presence at our field sites and accurately reflected initial relative template abundance in known-composition samples. With further refinement, capture enrichment shows high potential to facilitate parallel eDNA analysis across diverse taxonomic groups, which has been challenging with metabarcoding and other PCR-based approaches (Elbrecht & Leese, 2015; Kelly et al., 2017).

ACKNOWLEDGEMENTS

We thank Jason Bragg, Colin Callahan, Kellie Carim, Rich Cronn, Jeffrey Good, Matthew R. Jones, Josh Peñalba, Kristy Pilgrim, Sally Potter, Renae Pratt and Dan Vanderpool for advice and assistance in study design and laboratory and bioinformatic analysis. Tiago Antão helped develop the Galaxy bioinformatics pipeline. Cory Engkjer provided invertebrate samples, Erick Greene provided osprey tissue samples, and Craig Stafford provided fish tissue samples. Sean Sullivan and Jeffrey W. Webb assisted with morphological insect identification. This work was in part funded by NSF GRFP and NSF GROW programs (Grant No. DGE-1313190; Taylor Wilcox), the Wesley M. Dixon Graduate Fellowship (University of Montana; Taylor Wilcox), an ARC DECRA (DE130100777; Maxine Piggott) and the ANU-ActewAGL Endowment Fund (Project 58 2014; Maxine Piggott).

AUTHOR CONTRIBUTIONS

Environmental samples were collected by K.E.Z. and M.K.S. Laboratory analyses were conducted by K.E.Z. and T.M.W. Data analysis was conducted by T.M.W. All authors contributed to study design and writing the manuscript.

DATA ACCESSIBILITY

Probe synthesis details are described in Supporting Information Tables S1–S3. Detailed sample information is listed in Supporting Information Table S4. Reference database sequences and corresponding GenBank accession numbers are listed in Supporting Information Table S5. FASTQ files from the two capture enrichment experiments, and scripts for analysis are available on DRYAD (https://doi.org/10.5061/dryad.8j88573).

REFERENCES

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., … Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, *44*, W3–W10. https://doi.org/10.1093/nar/gkw343

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Avila-Arcos, M. C., Cappellini, E., Romero-Navarro, J. A., Wales, N., Moreno-Mayar, J. V., Rasmussen, M., … Gilbert, M. T. (2011). Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Scientific Reports*, *1*, 74. https://doi.org/10.1038/srep00074

Bolger, A. M., Lohse, M., & Usade, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bragg, J. G., Potter, S., Bi, K., & Moritz, C. (2015). Exon capture phylogenomics: Efficacy across scales of divergence. *Molecular Ecology Resources*, *16*, 1059–1068.

Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., … Pääbo, S. (2009). Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science*, *325*, 318–321. https://doi.org/10.1126/science.1174462

Brown, S. D. J., Collins, R. A., Boyer, S., Lefort, M. C., Malumbres-Olarte, J., Vink, C. J., & Cruickshank, R. H. (2012). Spider: An R package for the analysis of species identity and evolution, with particular

reference to DNA barcoding. *Molecular Ecology Resources*, *12*, 562–565. https://doi.org/10.1111/j.1755-0998.2011.03108.x

Carim, K. J., McKelvey, K. S., Young, M. K., Wilcox, T. M., & Schwartz, M. K. (2016). *A protocol for collecting environmental DNA samples from streams* (pp. 1–8). Fort Collins, CO: USFS Rocky Mountain Research Station GTR.

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., … Bustamante, C. D. (2013). Pulling out the 1%: Whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *American Journal of Human Genetics*, *93*, 852–864. https://doi.org/10.1016/j.ajhg.2013.10.002

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., … Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, *56*, 68–74.

Cruz-Dávalos, D. I., Llamas, B., Gaunitz, C., Fages, A., Gamba, C., Soubrier, J., … Orlando, L. (2017). Experimental conditions improving in-solution target enrichment for ancient DNA. *Molecular Ecology Resources*, *17*(3), 508–522. https://doi.org/10.1111/1755-0998.12595

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., … Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*, 5872–5895. https://doi.org/10.1111/mec.14350

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS ONE*, *10*, 1–16.

Elbrecht, V., & Leese, F. (2016). PrimerMiner: An r package for development and in silico validation of DNA metabarcoding primers. *Methods in Ecology and Evolution*, *5*, 622–626.

Elbrecht, V., Peinert, B., & Leese, F. (2017). Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and Evolution*, *7*, 6918–6926. https://doi.org/10.1002/ece3.3192

Enk, J. M., Devault, A. M., Kuch, M., Murgha, Y. S., Rouillard, J., & Poinar, H. N. (2014). Ancient whole genome enrichment using baits built from modern DNA. *Molecular Biology and Evolution*, *31*(5), 1292–1294.

Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., … Lodge, D. M. (2016). Quantification of mesocosm fish and amphibian species diversity via eDNA metabarcoding. *Molecular Ecology Resources*, *16*, 29–41. https://doi.org/10.1111/1755-0998.12433

Goldberg, C. S., Pilliod, D. S., Arkle, R. S., & Waits, L. P. (2011). Molecular detection of vertebrates in stream water: A demonstration using rocky mountain tailed frogs and Idaho giant salamanders. *PLOS ONE*, *6*(7), e22746.

Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, *4*, 150–157. https://doi.org/10.1111/j.1755-263X.2010.00158.x

Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, *25*, 185–202. https://doi.org/10.1111/mec.13304

Kelly, R. P., Closek, C. J., O'Donnell, J. L., Kralj, J. E., Shelton, A. O., & Samhouri, J. F. (2017). Genetic and manual survey methods yield different and complementary views of an ecosystem. *Frontiers in Marine Science*, *3*, 1–11.

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

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*, 1–8.

Lacoursière-Roussel, A., Rosabal, M., & Bernatchez, L. (2016). Estimating fish abundance and biomass from eDNA concentrations: Variability among capture methods and environmental conditions. *Molecular Ecology Resources*, *16*, 1401–1414. https://doi.org/10.1111/1755-0998.12522

Liu, S., Wang, X., Xie, L., Tan, M., Li, Z., Su, X., … Zhou, X. (2016). Mitochondrial capture enriches mito-DNA 100 folds enabling PCR-free mitogenomics biodiversity analysis. *Molecular Ecology Resources*, *16*, 470–479. https://doi.org/10.1111/1755-0998.12472

Lodge, D. M., Turner, C. R., Jerde, C. L., Barnes, M. A., Chadderton, L., Egan, S. P., … Pfrender, M. E. (2012). Conservation in a cup of water: Estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology*, *21*, 2555–2558. https://doi.org/10.1111/j.1365-294X.2012.05600.x

Maricic, T., Whitten, M., & Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE*, *5*(11), e14004. https://doi.org/10.1371/journal.pone.0014004

Maruyama, A., Nakamura, K., Yamanaka, H., Kondoh, M., & Minamoto, T. (2014). The release rate of environmental DNA from Juvenile and adult fish. *PLoS ONE*, *9*, e114639. https://doi.org/10.1371/journal.pone.0114639

Mckelvey, K. S., Young, M. K., Knotek, W. L., Carim, K. J., Wilcox, T. M., Padgett-Stewart, T. M., & Schwartz, M. K. (2016). Sampling large geographic areas for rare species using environmental DNA (eDNA): A study of bull trout occupancy in western Montana. *Journal of Fish Biology*, *88*, 1215–1222. https://doi.org/10.1111/jfb.12863

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Springs Harbor Protocols*, *6*, pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Nelson, M. C., Morrison, H. G., Benjamino, J., Grim, S. L., & Graf, J. (2014). Analysis, optimization and verification of illumina-generated 16s rRNA gene amplicon surveys. *PLoS ONE*, *9*, e94249. https://doi.org/10.1371/journal.pone.0094249

O'Donnell, J. L., Kelly, R. P., Lowell, N. C., & Port, J. A. (2016). Indexed PCR primers induce template-specific bias in large-scale DNA sequencing studies. *PLoS ONE*, *11*, e0148698. https://doi.org/10.1371/journal.pone.0148698

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2017). *vegan: Community ecology package*. Retrieved from https://CRAN.R-project.org/package=vegan

Paijmans, J. L. A., Fickel, J., Courtiol, A., Hofreiter, M., & Förster, D. W. (2015). Impact of enrichment conditions on cross-species capture of fresh and degraded DNA. *Molecular Ecology Resources*, *16*, 42–55.

Peñalba, J. V., Smith, L. L., Tonione, M. A., Sass, C., Hykin, S. M., Skipwith, P. L., … Moritz, C. (2014). Sequence capture using PCR-generated probes: A cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Molecular Ecology Resources*, *14*, 1000–1010.

Potapov, V., & Ong, J. L. (2017). Examining sources of error in PCR by single-molecule sequencing. *PLOS ONE*, *12*(1), e0169774.

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Riaz, T., Schehzad, W., Viara, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). ecoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, *39*(21), e145.

Rooney, J. P., Ryde, I. T., Sanders, L. H., Howlett, E. H., Colton, M. D., Germ, K. E., … Meyer, J. N. (2015). PCR based determination of mitochondrial DNA copy number in multiple species. *Methods in Molecular Biology*, *1241*, 23–38. https://doi.org/10.1007/978-1-4939-1875-1

Schirmer, M., Ijaz, U. Z., D'Amoe, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research*, *43*(6), e37.

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, *15*(6), 1289–1303.

Shockralla, S., Gibson, J., King, I., Baird, D., Janzen, D., Hallwachs, W., & Hajibabaei, M. (2016). Environmental DNA barcode sequence capture: Targeted, PCR-free sequence capture for biodiversity analysis from Bulk environmental samples. *bioRxiv*. https://doi.org/10.1101/087437

Smart, A., Tingley, R., Weeks, A., Rooyen, A., & McCarthy, M. (2015). Environmental DNA sampling is more sensitive than a traditional survey technique for detecting an aquatic invader. *Ecological Applications*, *25*, 1944–1952. https://doi.org/10.1890/14-1751.1

Spear, S. F., Groves, J. D., Williams, L. A., & Waits, L. P. (2015). Using environmental DNA methods to improve detectability in a hellbender (*Cryptobranchus alleganiensis*) monitoring program. *Biological Conservation*, *183*, 38–45. https://doi.org/10.1016/j.biocon.2014.11.016

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, *21*, 1789–1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*, 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., & Kawabata, Z. (2012). Estimation of fish biomass using environmental DNA. *PLoS ONE*, *7*, e35868. https://doi.org/10.1371/journal.pone.0035868

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*, 2731–2739. https://doi.org/10.1093/molbev/msr121

Thomas, A. C., Deagle, B. E., Eveson, J. P., Harsch, C. H., & Trites, A. W. (2016). Quantitative DNA metabarcoding: Improved estimates of species proportional biomass using correction factors derived from control material. *Molecular Ecology Resources*, *16*(3), 714–726.

Thompson, J., Higgins, D., & Gibson, T. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*, 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T., … Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, *21*, 2565–2573. https://doi.org/10.1111/j.1365-294X.2011.05418.x

Turner, C. F., Barnes, M. A., Xu, C. C. Y., Jones, S. E., Jerde, C. L., & Lodge, D. M. (2014). Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, *5*, 676–684. https://doi.org/10.1111/2041-210X.12206

Turner, C. R., Miller, D. J., Coyne, K. J., & Corush, J. (2014). Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian bigheaded carp (*Hypophthalmichthys spp*.). *PLoS ONE*, *9*, e114329. https://doi.org/10.1371/journal.pone.0114329

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., … Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, *25*, 929–942. https://doi.org/10.1111/mec.13428

Vences, M., Thomas, M., Meijden, A., Chiari, Y., & Vietes, D. R. (2005). Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. *Frontiers in Zoology*, *2*, 5.

Wilcox, T. M., McKelvey, K. S., Young, M. K., Sepulveda, A. J., Shepard, B. B., Jane, S. F., … Schwartz, M. K. (2016). Understanding environmental DNA detection probabilities: A case study using a stream-dwelling char (*Salvelinus fontinalis*). *Biological Conservation*, *194*, 209–216. https://doi.org/10.1016/j.biocon.2015.12.023

Wright, E. S., & Vetsigian, K. H. (2016). Quality filtering of Illumina index reads mitigates sample cross-talk. *BMC Genomics*, *17*, 1–7.

Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., … Chen, S. (2012). FastUniq: A fast de novo duplicates removal tool for paired short reads. *PLoS ONE*, *7*, e52249. https://doi.org/10.1371/journal.pone.0052249

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., … Huang, Q. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, *2*, 4. https://doi.org/10.1186/2047-217X-2-4

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Wilcox TM, Zarn KE, Piggott MP, Young MK, McKelvey KS, Schwartz MK. Capture enrichment of aquatic environmental DNA: A first proof of concept. *Mol Ecol Resour*. 2018;18:1392–1401. https://doi.org/10.1111/1755-0998.12928