Diversity and Distributions   WILEY

Environmental DNA metabarcoding as a tool for biodiversity assessment and monitoring: reconstructing established fish communities of north-temperate lakes and rivers

Peter T. Euclide[1] | Yer Lor[2] | Michael J. Spear[3] | Tariq Tajjioui[2] |
Jake Vander Zanden[3] | Wesley A. Larson[1,4] | Jon J. Amberg[2]

[1]Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, Stevens Point, WI, USA

[2]U.S. Geological Survey, Upper Midwest Environmental Sciences Center, WI, USA

[3]Center for Limnology, University of Wisconsin-Madison, Madison, WI, USA

[4]U.S. Geological Survey, Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, Stevens Point, WI, USA

Correspondence
Peter Euclide, Wisconsin Cooperative Fishery Research Unit, College of Natural Resources, University of Wisconsin-Stevens Point, 800 Reserve St., Stevens Point, WI 54481, USA.
Email: euclidp@uwm.edu

Yer Lor, U.S. Geological Survey, Upper Midwest Environmental Sciences Center, La Crosse, WI 54603, USA.
Email: ylor@usgs.gov

Present address
Wesley A. Larson, National Oceanic and Atmospheric Administration, Ted Stevens Marine Research Institute, Juneau, AK, USA

Funding information
U.S. Geological Survey Invasive Species Program; Research Computing Clusters at Old Dominion University

Editor: Yoan Fourcade

Abstract

Aim: To evaluate the ability of precipitation-based environmental DNA (eDNA) sample collection and mitochondrial 12S metabarcoding sequencing to reconstruct well-studied fish communities in lakes and rivers. Specific objectives were to 1) determine correlations between eDNA species detections and known community composition based on conventional field sampling, 2) compare efficiency of eDNA to detect fish biodiversity among systems with variable morphologies and trophic states, and 3) determine if species habitat preferences predict eDNA detection.

Location: Upper Great Lakes Region, North America.

Methods: Fish community composition was estimated for seven lakes and two Mississippi River navigation pools using sequence data from the mitochondrial 12S gene amplified from 10 to 50 water samples per waterbody collected in 50-mL centrifuge tubes at a single time point. Environmental DNA (eDNA) was concentrated without filtration by centrifuging samples to reduce per-sample handling time. Taxonomic detections from eDNA were compared to established community monitoring databases containing up to 40 years of sampling and a detailed habitat/substrate preference matrix to identify patterns of bias.

Results: Mitochondrial 12S gene metabarcoding detected 15%–47% of the known species at each waterbody and 30%–76% of known genera. Non-metric multidimensional scaling (NMDS) assessment of the community structure indicated that eDNA-detected communities grouped in a similar pattern as known communities. Discriminant analysis of principal components indicated that there was a high degree of overlap in habitat/substrate preference of eDNA-detected and eDNA-undetected species suggesting limited habitat bias for eDNA sampling.

Main conclusions: Large numbers of small volume samples sequenced at the mitochondrial 12S gene can describe the coarse community structure of freshwater systems. However, additional conventional sampling and environmental DNA sampling may be necessary for a complete diversity census.

1 | INTRODUCTION

Obtaining basic community structure and diversity estimates is critical for effective biomonitoring and conservation of biodiversity (Dudgeon et al., 2006; Maxwell & Jennings, 2005; Tilman, 1999). In aquatic systems, community composition can be used as an indicator of system health (Branton & Richardson, 2011), ecosystem function (Humbert & Dorigo, 2005) and even physical structure (Dodson et al., 2000). Therefore, describing and monitoring the native community of a system prior to significant change can help inform future mitigation of anthropogenic stressors and the loss of endemic species (Banks et al., 2010; Dudgeon et al., 2006). Unfortunately, freshwater streams and lakes around the world often lack basic biodiversity data despite their critical importance for local communities (Fluet-Chouinard et al., 2018; Lynch et al., 2016). Environmental DNA (eDNA) metabarcoding has the potential to offer ecologists and conservation biologists the ability to efficiently describe community composition in systems that may be impractical to sample with conventional techniques (Bohmann et al., 2014; Evans et al., 2016; Olds et al., 2016).

The promise of eDNA metabarcoding to describe and monitor freshwater communities has led to a large increase in eDNA studies over a short time. In particular, eDNA has been increasingly used to study fish communities which were conventionally surveyed with nets and traps that target different subsets of the community (e.g. Aylagas, Borja, Irigoien, Rodríguez-Ezpeleta, 2016; Yamamoto et al., 2017). Many of these studies have directly assessed how community estimates based on eDNA compare with conventional sampling techniques (Hänfling et al., 2016; Sard et al., 2019). Comparisons between eDNA metabarcoding and conventional sampling indicate that eDNA metabarcoding often identifies similar, but slightly different community estimates as conventional sampling (Fujii et al., 2019; Shaw et al., 2016). In some cases, eDNA metabarcoding requires less sampling effort (Evans et al., 2017; Sard et al., 2019) and identifies novel species that were undetected with conventional sampling (Olds et al., 2016; Valentini et al., 2016). One potential reason for novel detections with eDNA could be life history sampling bias of conventional sampling techniques (Clark et al., 2007; Murphy & Willis, 1996). The growing evidence that eDNA metabarcoding can complement conventional sampling of fish communities has encouraged many agencies to begin considering how to implement eDNA sampling into local monitoring and management efforts (Bohmann et al., 2014; Valentini et al., 2016). Understanding how and why eDNA metabarcoding detections differ from conventional methods is critical for the design and interpretation of eDNA sampling.

One region where eDNA metabarcoding assessments have the potential to expand conservation biology is the Great Lakes region of North America. This region was glaciated until 12,000 YBP and contains thousands of natural lakes with complex fish communities (Downing & Duarte, 2006). Natural resource managers have conducted intensive sampling and monitoring efforts throughout the region. However, the majority of lakes have never been formally surveyed or have no reported biodiversity data. Additionally, where surveys have been conducted, the focus has been on game species, rather than total biodiversity. In Wisconsin, USA, there are 14,973 lakes reported in the Wisconsin Department of Natural Resources database, of which 3,426 (23%) have reported fisheries data from between 1990 and 2020 (Zachary Feiner, Wisconsin Department of Natural Resources, personal communication, July 21, 2020). The ability to quickly survey the community composition of novel systems could help to identify the presence of rare or endangered species (Ficetola et al., 2008), estimate diversity (Olds et al., 2016), or identify dominant fish communities to help predict broad scale patterns of community diversity (Sard et al., 2019).

Sampling protocols can have a large influence on eDNA metabarcoding results (Creer et al., 2016; Dickie et al., 2018). Recent studies have commonly used filtration protocols to sample eDNA from water whereby a predetermined volume of water (usually 1–2 litres) is collected per waterbody and pumped through sterilized filters. Filters can then be stored in preservative or frozen until extraction (Li et al., 2018). While filtration techniques have been used with success, excessive particulate matter can clog filters and greatly increase the processing time per sample. One method that can eliminate filtration and therefore expedite per-sample processing time is the use of a centrifuge to pellet organic matter and eDNA (Ficetola et al., 2008). A version of this protocol has been used successfully in single species detection eDNA studies of invasive bighead and silver carp (*Hypophthalmichthys nobilis and H. molitrix*) populations in the Mississippi drainage which often inhabit areas with a high degree of suspended particulate matter (Erickson et al., 2016; Merkes et al., 2014) as well as for the detection of endangered species (Lor et al., 2020). Because large volumes can be difficult to centrifuge, the centrifugal method relies on a greater number of smaller volume water samples (50 ml), the eDNA of which can be directly extracted from centrifuge-generated pellets without filtration. Because lakes in Wisconsin range from extremely eutrophic with a high density of suspended particulate matter to oligotrophic with a low-density suspended particulate matter (Birge & Juday, 1934), a single sampling technique that can successfully sample both extremes is desirable.

Here, we apply small volume/high number eDNA sampling with centrifugal sample processing and metabarcoding to describe the species communities of seven different reference waterbodies in Wisconsin and two in Illinois and Iowa that have varying degrees of productivity and community compositions well characterized by conventional sampling methods. We incorporate more than 25 years of data collected on fish community structure from the Mississippi

TABLE 1 Waterbody structure and detection thresholds. Known richness was calculated based on all available data for a waterbody between 2014 and 2020 with years of data shown in parentheses. Read counts for Threshold 1 were calculated based on laboratory blanks (n = 18) and read counts for Threshold 2 were calculated from field blanks samples where the number of samples (n) was 10% of total samples collected at a given waterbody. Information not available are denoted with na

Waterbody	Latitude	Longitude	Trophic state	Max depth (m)	Mean depth (m)	Surface area (ha)	Known richness	No. water samples	Threshold 1	Threshold 2
Crystal	46.00	−89.612	Oligotrophic	20.4	10.4	37.5	14 (5)	30	52	1569
McDermott	46.00	−90.161	Mesotrophic	5.8	3.0	33.2	19 (2)	10	52	0
Mendota	43.10	−89.405	Eutrophic	25.3	12.8	3,961.0	31 (5)	50	52	17
Sparkling	46.01	−89.701	Oligotrophic	20	10.9	63.7	19 (5)	25	52	0
Trout	46.03	−89.665	Oligotrophic	35.7	14.6	1565.1	32 (5)	47	52	165
Trout Bog	46.04	−89.686	Eutrophic	7.9	5.6	1.0	1 (5)	30	52	0
UMR_13	41.90	−90.155	na	na	na	na	76 (2)	24	52	25
UMR_19	40.40	−91.374	na	na	na	na	46 (1)	49	52	21
Wingra	43.05	−89.425	Eutrophic	6.7	2.7	140	20 (5)	50	52	0

River of the U.S. Army Corps of Engineers Upper Mississippi River Restoration Program, specifically the Long-Term Resource Monitoring (LTRM) element which is funded by the U.S. Army Corps of Engineers and implemented by the U.S. Geological Survey (USGS), and 39 years of fish community data collected by the University of Wisconsin-Madison which records fish capture data from several Wisconsin lakes as part of their North Temperate Lakes Long-Term Ecological Research (NTL-LTER) Program (https://lter.limnology.wisc.edu/). We evaluated the efficacy of centrifugal eDNA sampling to describe lake and river fish communities in these nine systems using these monitoring data and a detailed habitat/substrate preference database for detected fish species (Berker, 1983; Frimpong & Angermeier, 2009; Froese & Pauly, 2019; USGS, 2005). The objectives of our study were threefold: a) determine correlations between eDNA species detection and current and historical catch data, b) compare efficiency of eDNA to detect fish biodiversity among systems with different known community compositions, and c) identify habitat/substrate preferences that predict if a species can be detected using eDNA.

2 | METHODS

2.1 | Waterbody selection

We selected five inland lakes (McDermott Lake, Lake Mendota, Sparkling Lake, Trout Lake and Lake Wingra) and one bog (Trout Bog) located in Wisconsin, as well as two navigation pools of the Upper Mississippi River System (UMR-13 and UMR-19) located between Illinois and Iowa for this study (Table 1). We selected waterbodies from this region based on the amount of available fish community data and to represent a wide range of latitudes (40.398–46.041 N), mean depth (3–14 m), surface area (1–1565 ha), trophic state (oligotrophic–eutrophic) and species richness (see Table 1). The fish communities at each waterbody were previously well characterized by long-term routine sampling with conventional methods (nets, traps, electrofishing boat, *etc*.) by the NTL-LTER (University of Wisconsin-Madison; Magnuson et al., 2019), the LTRMP; (USGS), and the Long-Term Survey and Assessment of Large River Fishes in Illinois (LTEF; Illinois Department of Natural Resources, Illinois Natural History Survey, and Western Illinois University) or by intensive short-term sampling for fish removal projects by the Illinois Natural History Survey and the University of Wisconsin-Madison (H.S. Embke, University of Wisconsin-Madison, personal communication, July 21, 2018).

2.2 | eDNA sampling

We collected 50-ml water samples at 10–50 sites from each waterbody. Samples were collected at LTER sites over the course of three days between July 23rd and July 25th 2018 (Table S1), and UMR-19 and UMR-13 were sampled on June 19th and June 20th

2018, respectively. The number of samples was approximately proportional to the surface area of each water body, resulting in a total of 327 water samples processed for eDNA metabarcoding across all nine systems. Within each waterbody, sample collection was divided 1:1 between nearshore and offshore sites (defined as less than or greater than 10 m from shore, respectively). Samples were spread approximately evenly across its surface area. Nearshore sites were selected to be representative of the waterbody's major nearshore habitat types (e.g. rock, sand, macrophytes, etc.). For thermally unstratified waterbodies ($n = 5$), all samples were collected from the surface by hand-dipping a 50-mL centrifuge tube just below the surface of the water. Sterile gloves were worn to prevent contamination and were changed between each sample. For thermally stratified waterbodies ($n = 4$), all nearshore samples were collected at the surface and offshore samples were divided 2:1 between surface water and at-depth water.

At-depth samples were taken from haphazardly selected depths within the hypolimnion using a Van Dorn sampler. The Van Dorn sampler was soaked in a 20% bleach solution for 30 min (min.) and rinsed with tap water between waterbodies to prevent DNA cross-contamination. Negative field blanks containing 50 ml of ultrapure water were filled in the laboratory and opened to the air for ~ 5 s and had their lower ends submerged in the water during the sampling of each waterbody. The number of negative field blanks differed by waterbody (1 to 5 samples) but was approximately 10% of the total number of samples for each waterbody. Samples were immediately placed on ice and remained refrigerated for no more than 48 hr before being delivered to the molecular laboratory at the U.S. Geological Survey Upper Midwest Environmental Sciences Center (UMESC) and frozen at −80 degrees C (°C) until sample centrifugation and DNA extraction.

2.3 | DNA extraction and library preparation

All water samples including field blanks were processed according to Merkes et al. (2014). In short, each 50-mL sample was centrifuged at 5,000 g for 30 min. at 4°C. The supernatant was decanted, and DNA was extracted from the remaining pellets with a final elution volume of 100 μl using the gMax Mini Genomic DNA Extraction Kit (IBI Scientific, Peosta, IA, USA) according to the manufacturer's instructions. Two extraction blanks were processed alongside all water samples all the way through sequencing and data analysis to account for potential contamination during extraction.

Illumina MiSeq libraries were prepared using vertebrate-specific primers that amplified a region of the mitochondrial 12S gene (Riaz et al., 2011). These primers have been used successfully in similar metabarcoding studies to amplify this particular region for fish (Sard et al., 2019; Gehri et al., 2020). To ensure that sufficient DNA was sequenced, the samples were amplified using a two-step PCR protocol. We first amplified each sample in 50 μl PCR reactions using Bio-Rad C1000 Touch Thermal Cycler CFX96 Real-Time System (BioRad Laboratories; Hercules, California). Each reaction contained

25 μl Quantabio sparQ HiFi PCR Master Mix (Quantabio), 0.4 μM 12S forward and reverse primers each, and 4 μl of eDNA template. The reactions had an initial denaturation at 98°C for 2 min followed by 18 cycles of denaturation at 98°C for 20 s, annealing at 60°C for 30 s and extension at 72°C for 30 s, followed by a final extension at 72°C for 5 min. Sixteen no-template-controls (NTCs) were processed along with water samples and sequenced to account for potential contamination during PCR amplification.

We used the products of the first step PCR reactions as DNA template in a second 50 μl PCR to affix the Illumina Nextera overhang adapter sequences (OAS) (Illumina, Inc.) to the target amplicons. Each reaction contained 25 μl sparQ HiFi PCR Master Mix, 0.4 μM 12S OAS forward and reverse primers and 4 μl of previously amplified template. The reactions had an initial denaturation at 95°C for 3 min followed by 32 cycles of denaturation at 95°C for 5 s, annealing at 60°C for 15 s and extension at 72°C for 15 s, followed by a final extension at 72°C for 5 min.

Each OAS-tagged PCR product was purified using 1.8X Agencourt Ampure XP magnetic beads (Beckman Coulter Life Sciences, Indianapolis, IN, USA) according to the manufacturer's instructions. The samples were then analysed using a D1000 ScreenTape Assay on an Agilent 4200 TapeStation System (Agilent Technologies, Inc) to check for sufficient presence of the target amplicons and for sample concentration. Samples were normalized and indexed using the Illumina Nextera XT Index Kit v2 (Illumina) according to the manufacturer's instructions.

Each indexed sample was purified again using 1.8X Agencourt Ampure XP magnetic beads according to the manufacturer's instructions. After bead-based purification, the samples were analysed with the D1000 ScreenTape Assay again to verify that the samples were properly indexed before being quantified in triplicate with the KAPA Library Quantification Kit (Roche Sequencing) according to the manufacturer's instructions. The samples were normalized to 2 nM and equimolar pooled then denatured and diluted to 6 pM loading concentration and sequenced on an Illumina MiSeq System using the MiSeq Reagent Kit v2 Nano (300 cycles) and 30% PhiX spike-in according to the manufacturer's instructions (Illumina).

2.4 | Bioinformatics and sequence processing

Sequence data were processed using protocols outlined in Gehri et al. (2020). Raw reads were demultiplexed then primer sequences and putative adapter contamination were trimmed from each sequence using cutadapt version 2.01 (Martin, 2011). Cleaned reads were then processed in DADA2 version 1.16 (Callahan et al., 2016) to remove all sequences < 100 bp or longer than 125 bp and reads with a quality score (truncQ) ≤ 2 or expected error (maxEE) > 2. Next, putative chimeras were identified and removed also using DADA2. As the final step in the DADA2 pipeline, forward and reversed reads were merged and a sequence table containing counts of every remaining unique sequence was generated. A local BLASTn (https://blast.ncbi.nlm.nih.gov) search was used to compare sequences against a

custom reference database containing 2,426 12S sequences from 505 species either known to inhabit the study area or closely related to species in the study area. The reference database was constructed by combining 12S sequences available from the National Center for Biotechnology Information (NCBI) for fish species identified through fishery surveys in inland Michigan lakes by the Michigan Department of Natural Resources or with known distribution in the Upper Mississippi and Great Lakes region with existing data developed during two other metabarcoding studies in the region (Sard et al. 2019; Gehri et al. 2020). Fifteen species with known observations in the study were not included in the reference database 10 did not contain any 12S sequences in GenBank and five contained only partial sequences (Table S2). Using custom R scripts, results were filtered to retain only matches with > 98% sequence identity and alignment lengths > 100 bp. Any sequence that matched a sequence in the reference database was assigned by manually blasting each sequence to the National Center for Biotechnology Information (NCBI) nucleotide database. Reference sequences for all assignments with > 98% similarity in NCBI were added to the custom reference database. The local nucleotide BLAST search (BLASTn) was then repeated so that only sequences with < 98% identity to any sequence in the NCBI database were unassigned. Because many species have identical or similar 12S sequences, we assigned sequences with ambiguous species or genus level assignments to the lowest common taxonomic unit (i.e. family or genus). We used custom R scripts to export feature tables containing all samples and read counts for the lowest taxonomic unit and subsequently filtered to avoid false-positive detections and account for sample contamination.

Field blanks, extraction blanks and NTCs were used to account for contamination following similar procedures as those outlined in Sard et al. (2019) and Evans et al. (2017). First, putative fish detections in each water sample were grouped into two categories based on the number of independent samples with a non-zero read count for a particular waterbody. Any putative detection that occurred in two or more samples (5%–20% of the total samples for a lake depending on sample size) was considered to be a true detection for that system and the taxon was counted as present in the system. Any putative fish detection that occurred in only a single sample was subjected to two additional read count thresholds based on the number of reads in extraction blanks, NTCs, and field blanks. Threshold 1 was determined by using the read counts assigned to fish taxa from NTCs and extraction negatives. Read counts from both extraction blanks and NTCs were combined into a single "laboratory blank" group ($n = 18$) and the average number of reads assigned to any fish taxa was used as the baseline threshold needed to define a species as "detected" if reads were only identified in a single sample at a waterbody. Threshold 2 was defined as the mean number of reads assigned to fish taxa in field blanks from a particular waterbody and then applied to all water samples from that waterbody. Therefore, any read counts found in only a single water sample at a particular waterbody that did not exceed both thresholds 1 and 2 were considered to be putative contamination and removed from further analysis. All downstream data processing and figures were constructed

in R version 3.6 (R Core Team, 2019) using custom scripts and the ggsci, ggpubr and tidyverse packages (Kassambara, 2020; Wickham et al., 2019; Xiao, 2018) and the eDNA species detection matrix (Table S3).

2.5 | Obj 1. Community assessment

We created a matrix of known fish communities based on routine long-term monitoring data or intensive short-term sampling using conventional netting and electrofishing methods (Table S2). For Lakes Mendota, Wingra, Sparkling, Crystal, and Trout, Trout Bog and UMR-13, we considered a species present if it was detected one time in the most recent 5 years of routine long-term monitoring. Although McDermott Lake and UMR-19 are not long-term monitoring waterbodies, fish community compositions in these systems were well characterized by high-intensity sampling events that took place in 2017 and 2018. Community data for McDermott Lake were generated by an intensive netting and electrofishing effort by the University of Wisconsin-Madison which included approximately 40 net set nights and 15 hr of electrofishing in 2017 and 2,500 net set nights and 30 hr electrofishing in 2018 (H.S. Embke, University of Wisconsin-Madison, personal communication, July 21, 2018). Community data for UMR-19 were generated by a netting and electrofishing effort as part of an intensive Asian carp (primarily *H. molitrix* and *H. nobilis*) removal program in 2018. The agreement of the eDNA metabarcoding survey with conventional methods was then determined by calculating the intersection of species, genus, and family level eDNA detections with the known community composition for each waterbody. Because sampling was either exceptionally routine or exceptionally intensive, we treated these data as known fish community compositions.

2.6 | Obj 2. Efficiency of eDNA among systems with different known community compositions

To determine if eDNA metabarcoding could efficiently identify waterbody-level differences in community composition, we conducted non-metric multidimensional scaling (NMDS) using known community compositions and eDNA detection matrices. We conducted the NMDS three times based on family, genus and species-level detections with Bray–Curtis dissimilarity and 2 ordination axes. The similarities between matrices were evaluated using analysis of similarities (ANOSIM) available in the vegan R package (Oksanen et al., 2019).

2.7 | Obj 3. Relationship between habitat/substrate preference traits and species detection

To determine whether habitat/substrate preference influenced the probability of detection with eDNA, we conducted a discriminant

analysis of principal components (DAPC; Jombart, 2008) with five retained principal components using basic habitat/substrate preference matrix as the predictive matrix (Table S4). Habitat preference was a binary variable whereby 1 indicated that species has been frequently found in the habitat and 0 indicated a general absence from the habitat. We included 12 habitat/substrate types in our analysis: bedrock, boulders, clay/silt, cobble, gravel, woody debris, muck, organic debris, pelagic water, lotic environments, sand and vegetation. A DAPC analysis is a multivariate method to describe distinct variable clusters by identifying the largest between-group variance and smallest within-group variance. DAPC first transforms a matrix using principal component analysis (PCA) then performs a discriminant analysis on all the user-retained principal components. Group differences are described as discriminant function loadings whereby species with dissimilar habitat preferences will have dissimilar loadings. By comparing the loadings of each species known to inhabit each system grouped by whether or not the species was detected through eDNA metabarcoding, we were able to determine whether eDNA metabarcoding detected a biased subset of fish communities and, if so, identify traits or sets of traits that may predict eDNA detection. The R package adegenet was used to conduct eight total DAPC's, one for each waterbody, as well as a final test where all waterbodies were combined with each analysis including the first five principal components (Jombart, 2008). Evidence of significant bias in habitat/substrate preference was determined based on reassignment accuracy of eDNA-detected species to the eDNA-detected group of at least 90%.

3 | RESULTS

3.1 | Sequence summary

Twelve samples contained less than the minimum recommended concentration for sequencing (2 nM), so those were removed in the final steps of library preparation. Another sample was removed after sequencing that contained zero sequence reads after sequence filters were applied. This left a total of 356 samples including 16 NTCs, 2 extraction blanks, 23 field blanks and 314 samples. A substantial number of reads were lost during quality control filtering and species assignments (Figure 1). Raw sequence files from water samples contained between 7 and 81,209 total reads (mean = 1,715; median = 826). Each sample lost an average of 779 (46%) of reads following sequence quality filtering, denoising and removal of chimeras. After filtering, 11 samples retained 0 reads including 8 NTCs, 2 field blanks and one sample from Lake Mendota. The remaining 347 sample contained between 3 and 53,003 remaining reads for taxonomic assignment per sample (mean = 913, median = 375). Raw read count in field and laboratory blanks contained between 202 and 21,537 total reads (mean = 2,230, median = 1,221) but lost more reads during filtering (mean loss = 74%).

Of all reads that remained following sequence filtering in DADA2, a total of 267 amplicon sequence variants were identified
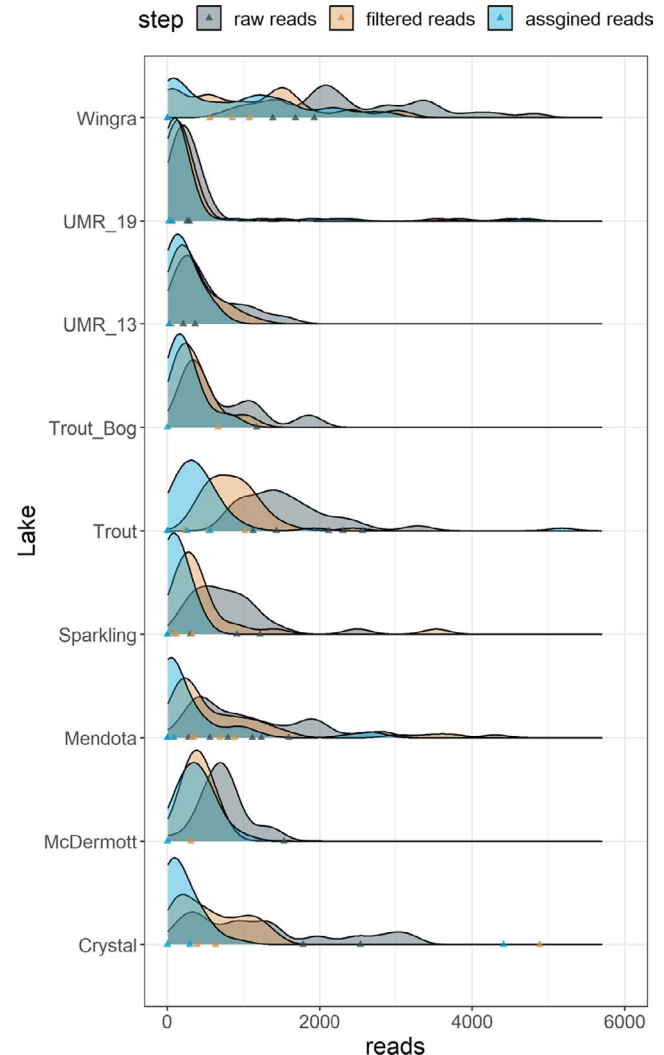
FIGURE 1 Sequential loss of per-sample sequence reads during filtering and assignment for each waterbody. Density distributions show the total number of sequence reads present after each step of the analysis pipeline beginning with the total reads present in demultiplexed fastq files (raw reads), followed by quality filtering and removal of chimeras in DADA2 (filtered reads), and, finally, reads that were successfully assigned to a fish taxa with 98% accuracy (assigned reads). Triangles show the read counts of the 1–5 field blanks at each filtering step. Unfiltered reads in one of the field blanks in Crystal Lake are out of the boundaries of the x-axis because the number of reads exceeded most of the samples

that could be assigned to 74 unique taxa of which 53 fish taxa. Each sample contained between 0 and 17,059 reads that assigned to a fish taxon (mean = 475, median = 172). Field and laboratory blanks contained between 0 and 5,438 reads assigned to fish taxa (mean = 317, median = 0).

To account for false-positive eDNA detections in the water samples (detections of species in waterbodies where there is no resident population), we set thresholds based on the number of samples with positive read counts for a particular taxon and two additional read count thresholds based on the number of reads assigned to each

taxon in field and laboratory blanks. First, taxa detected in two or more samples from the same waterbody were considered true detections and retained (1,008 of 1,054 positive read count observations). Next, the 46 positive read count observations that occurred in only one sample from a waterbody were subjected to the two-part threshold based on read counts from laboratory and field blanks. Of these 46 observations, 39 failed to pass Threshold 1 because they contained fewer than 52 total reads which was the mean read count in laboratory blanks. Another four observations passed Threshold 1 but failed the waterbody-specific Threshold 2 (Table 1). This resulted in 1,011 observations that were considered "detections," of which 1,008 observations were observed in at least two samples at a particular waterbody and an additional three observations that were observed only once at a particular waterbody but passed read count Thresholds 1 and 2. Positive fish taxon detections were identified in 278 of 314 samples and represented 44 unique fish taxa (Table S4).

3.2 | Obj. 1: Intersection between eDNA species detection and known fish communities

We detected 15% of known species at our highest diversity waterbody (UMR-13) and 100% of species at our lowest diversity waterbody (Trout Bog) but detected an average of 39% of the known community across all waterbodies with eDNA metabarcoding data (Table 2). Because many closely related species shared identical or almost identical 12S sequences, species-level designations using a single gene was not possible in many cases. However, there was a marked increase in overlap between known community composition and eDNA detections when taxa were grouped by genus (average = 54%) or family (average = 67%) rather than by species (Table 2). Common species had a higher rate of detection than the total community (Table 3). On average, eDNA detected 52% of the species observed in every year of the last 5 years of conventional sampling surveys.

Species not present in the known community of a particular waterbody were detected with eDNA eight times across four of the nine waterbodies sampled. Three novel species were detected in Trout Lake (*Fundulus diaphanus*, *Pimephales promelas* and *Umbra limi*). Of

the three species, only *P. promelas* has been detected in Trout Lake historically (outside of the most recent five years of monitoring). The species *P. promelas* has been observed eight times and in six different years since 1981, most recently in 2009, suggesting that this could be a true positive detection. Though *F. diaphanus* and *U. limi* have not been observed in Trout Lake since 1981, Trout Bog, which is hydrologically connected to Trout Lake, has a population of *U. limi* which could be the source of *U. limi* individuals and/or *U. limi* eDNA. Therefore, *F. diaphanus* is the only species without a probable explanation for detection. The detection could be a false detection from contamination or misassignment or DNA from populations in surrounding tributaries or wetlands that were unsampled by LTER. Both *U. limi* and *P. promelas* were also detected in McDermott Lake, but neither species was identified in the two years of intensive sampling that was conducted at McDermott Lake. However, both have natural populations throughout Wisconsin, so DNA could have originated from a nearby waterbody. One novel species, *Dorosoma cepedianum*, was detected in Lake Wingra which has not been observed in the lake since 1981; therefore, it is possible that this was a true detection. Finally, two novel species, *Carassius auratus* and *Scaphirhynchus platorynchus*, were identified at UMR-19. *C. auratus* is a common nuisance species introduced from aquarium releases in that region while *S. platorynchus* is native to the region and has been observed every year between 1989 and 2018 at UMR-13 (~240 km upstream) suggesting both detections could be true positive detections.

3.3 | Obj 2. Efficiency of eDNA among systems with different known community compositions

Because Trout Bog had low known species diversity, it was not used for NMDS analysis. The NMDS ordinations differed between eDNA metabarcoding detections and known fish communities when taxa were grouped by species (test statistic = 0.33; $p = .001$) but not genus (test statistic = 0.07; $p = .148$) or family (test statistic = 0.04; $p = .233$; Figure 2). At each taxonomic level, waterbodies were arranged similarly along NMDS axis 1 and 2 for both detection types. For both eDNA-detected communities and known communities, UMR-13 and UMR-19 were clearly differentiated from the lake communities.

TABLE 2 Intersection between the total known community identified through conventional net sampling conducted over the last 5 years in each waterbody (2014 – 2018) and eDNA detections from 12S metabarcoding for species, genus and family level taxonomic assignment

Waterbody	Known species	Detected species	Known genera	Detected genera	Known families	Detected families
Crystal	14	3	9	3	5	4
McDermott	19	8	13	10	6	4
Mendota	31	13	25	14	13	7
Sparkling	19	6	14	5	6	4
Trout	32	15	22	14	9	7
Trout Bog	1	1	1	1	1	1
UMR–13	76	12	44	18	19	11
UMR–19	46	11	31	14	15	11
Wingra	20	6	15	6	11	3

TABLE 3 Intersection between the total known community of common species identified through conventional net sampling conducted over the last 5 years in each waterbody (2014–2018) and eDNA detections from 12S metabarcoding for species, genus, and family level taxonomic assignment. Common taxa are defined as any species identified in every year for the most recent 5 years of survey data (2014–2018) from sampling for long-term monitoring systems or in all years of sampling for short-term intensive sampling systems (McDermott and UMR–19)

Waterbody	Known species	Detected species	Known genera	Detected genera	Known families	Detected families
Crystal	4	2	4	2	4	2
McDermott	16	8	12	9	6	3
Mendota	16	11	15	10	9	6
Sparkling	10	4	8	3	4	2
Trout	12	8	10	8	6	5
Trout Bog	1	1	1	1	1	1
UMR–13	54	11	32	11	14	9
UMR–19	46	11	31	11	15	8
Wingra	11	6	10	6	6	3

Note: Read counts for Threshold 1 were calculated based on laboratory blanks ($n = 18$) and read counts for Threshold 2 were calculated from field blanks samples where the number of samples (n) was 10% of total samples collected at a given waterbody. Information not available is denoted with na

Oligotrophic lakes (Crystal, Trout and Sparkling) appeared more dissimilar from the eutrophic lakes (Mendota and Wingra) than they were from each other, with the mesotrophic lake (McDermott) appearing in between.

3.4 | Obj 3. Relationship between habitat/substrate preference traits and species detection

Trout Bog had low known species diversity, so it was not used for habitat/substrate preference analysis. Discriminant analysis of principal components was used as a method to determine whether there were large differences between the habitat/substrate preferences shared by species that were detected by eDNA metabarcoding and the habitat/substrate preferences shared by species that were not detected by eDNA metabarcoding. We predicted that if species with certain habitat/substrate preferences were more easily detected than others, species loading scores of detected and undetected groups along Discriminant Function 1 should be diverged. However, we observed high overlap in species loading scores Discriminant Function 1 (Figure 3) indicating that detected and undetected species shared many of the same habitat/substrate preferences. The scale of the overlap between groups can be summarized using a reassignment accuracy estimate, whereby if habitat/substrate preferences are strongly diverged between groups, reassignment accuracy should be high. In our analysis, the reassignment accuracy for detected and undetected species was low both overall when detections from all lakes were combined (28%) and within each waterbody (8%–66%; mean = 42%). The explanatory contribution of the habitat variables (i.e. proportion of variance along Discriminant Function 1 explained by preference for a certain habitat/substrate) was inconsistent among waterbodies (Figure 4). Preference for lotic environments appeared to somewhat consistently explain the most variance in eDNA detection but was only higher than 20% in 4 out of 7 sites. Other habitat preferences explained similarly high amounts of variance (>20%) in certain waterbodies. However, because the variable

contribution of same habitat preference was rarely high in more than one waterbody, we concluded that habitat preference was not a good predictor of whether or not a species would be detected with eDNA metabarcoding.

4 | DISCUSSION

Freshwater systems represent a disproportionate amount of global biodiversity by surface area and face substantial anthropogenic stresses including climate change, pollution and species introductions (Collingsworth et al., 2017; Strayer & Dudgeon, 2010). However, biodiversity and fisheries data for freshwater systems can be limited (Bower et al., 2020; Strayer & Dudgeon, 2010). Our study shows that centrifugal DNA isolation followed by 12S eDNA metabarcoding sequencing can provide basic biodiversity information for a broad range of freshwater systems. Using 50-ml water samples collected from waterbodies with a wide range of physical and trophic characteristics, we were able to detect 30%–50% of the known fish community and 50%–60% of common fish species in a single eDNA sampling event using sequence data from a single gene region and publicly available reference sequences. Sampling and sample processing effort for this study represents a fraction of the hundreds of hours of effort involved in establishing the known fish community which is based on 5 years of data and hundreds of hours of sampling with beach seins, electrofishing, gillnetting, fyke netting and minnow traps (NTL-LTER, LTRM). Additional eDNA surveys would almost definitely result in the detection of more species, but these results indicate that even single-pass metabarcoding could provide useful information about the biodiversity in understudied systems.

Using eDNA metabarcoding, we reconstructed the broad patterns of known community structure and successfully differentiated lake and river pool communities as well as communities between lakes of different trophic states. These results indicate that eDNA metabarcoding data could be useful to conduct coarse but consistent biodiversity assessments across a large spatial scale to elucidate
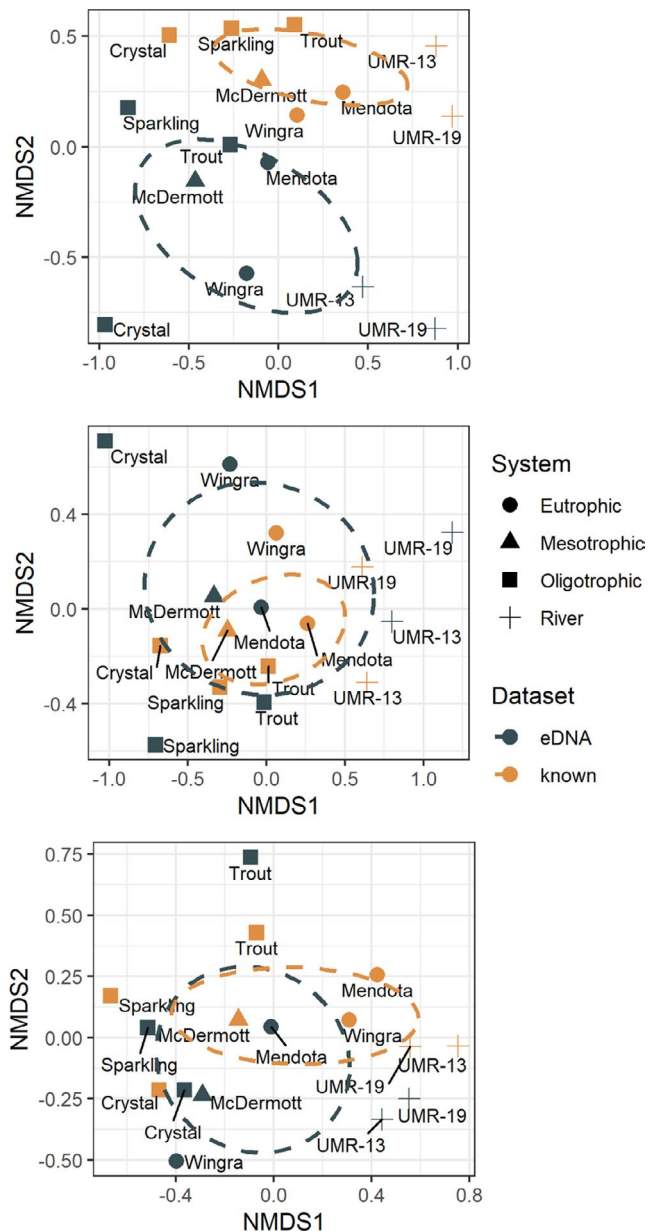
FIGURE 2 Non-metric multidimensional scaling (NMDS) for waterbodies based on eDNA and known community data 0sets with detections for three taxonomic levels: species (top), genera (middle) and family (bottom). Ellipses show the standard deviation around the centroid of each group. Shapes group waterbodies by basic system physical structure

how communities change in space and time. Our results echo those of other eDNA metabarcoding studies that have found that eDNA can provide useful community information for relatively low effort but also highlight that a certain amount of type I and II errors should be expected (Ficetola et al., 2016; Hänfling et al., 2016). Though species-specific assays or high-volume water filtration methods may be more appropriate if detecting rare species is a priority, we show that a single sampling event of low-volume/high-sample number study design combined with centrifugation and metabarcoding can

reconstruct basic fish community structure from a variety of fresh-water systems.

4.1 | Optimal sampling strategy and evaluation of centrifugal methods

Our study is one of the few demonstrations of non-filtration-based water sampling for fish metabarcoding assessments (but see Deiner et al. 2015; Ficetola et al. 2008; Thomsen et al. 2012). While many reviews consider both techniques (e.g. Bohmann et al. 2014; Goldberg et al. 2015), the field of eDNA metabarcoding in aquatic systems has increasingly moved towards the use of filtration approaches. However, precipitation-based extractions can produce similar concentrations of DNA as filtration (Deiner et al., 2015). In our study, we found that not using a filtration method enabled field crews to collect a large number of water samples from many lakes without returning to the laboratory for filtration. Excluding transportation time between lakes, collection of water samples generally took 1–3 hr per lake. Also, minimal field experience was required to conduct sampling which means that sample collectors could be trained quickly which further expedited the time required to collect samples.

While the employed sampling strategy could be completed quickly, the low per-sample volume and sequencing depth both likely contributed to lower rates of species detections than other similar metabarcoding studies that conducted higher coverage sampling (Hänfling et al., 2016; Sard et al., 2019). Although we found low read counts in control samples (often zero), the number of sequence reads in many of the actual water samples was also low; therefore, many taxonomic assignments were based on fewer than 100 reads. The use of a larger sample volume, collection of additional 50 ml samples or inclusion of additional field and laboratory negatives could have boosted total read counts and improved our study's ability to distinguish true detections from contamination. Another major factor that contributed to low total read counts was the use of a MiSeq Reagent v2 Nano kit. When possible, future studies should consider sequencing at a greater depth by using a standard MiSeq Reagent kit which can produce six times the reads of a Nano kit or consider other sequencing platforms, such as NovaSeq, to obtain an even higher depth of coverage. The increased overall read counts should help to discriminate between negative control contamination and true sample detections and increase detection rates of low abundance species.

The number of control samples collected at each waterbody is another important consideration. One to five field blanks were collected for each waterbody which made the identification of outliers difficult. For example, one of the three field control samples from Crystal Lake contained over 4,000 *Lepomis macrochirus* reads while the other two field control samples contained 0 reads. If more field control samples had been available for Crystal Lake, potential anomalies, like the control sample with 4,000 reads, may have been able to be confidently identified and possibly removed as outliers. However, even when read counts were high in some control samples, the high

FIGURE 3 Discriminant analysis of principal components (DAPC) comparing the habitat/substrate preference of species not detected with eDNA to species detected by eDNA for all waterbodies combined and for specific communities in at each waterbody. Density curves are estimated from species loading scores along Discriminant Function 1 which is a measure of group separation. Thus, the more overlap in species habitat/substrate preference between groups results in higher overlap between Not Detected and Detected density curves. The points along the x-axis show the point estimates of individual species loading scores that were summarized to generate density curves. Species with the maximum and minimum loadings are labelled as representatives of the most extreme habitat/substrate preference differences based on DAPC. Note that y-axes are on different scales
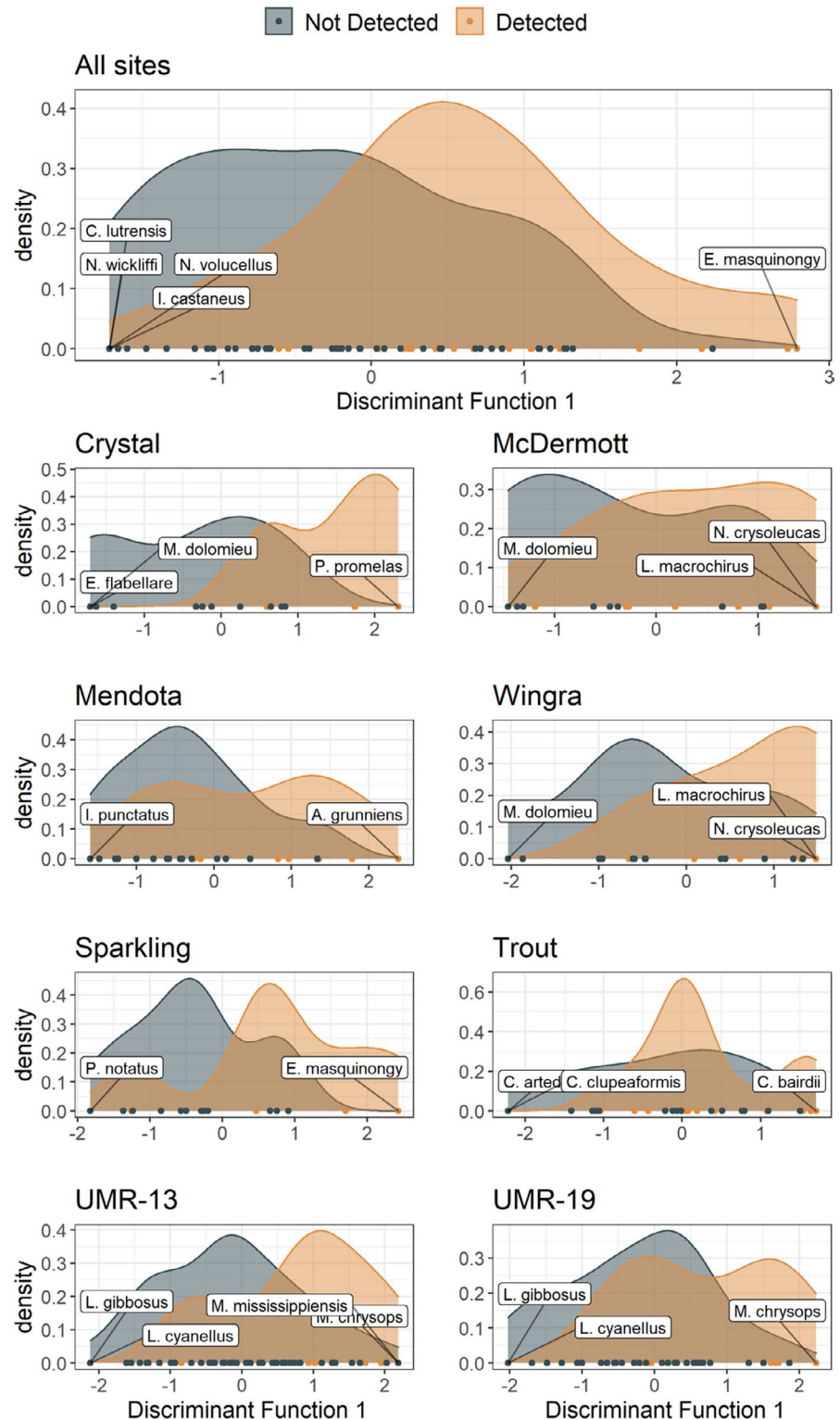
number of water samples collected per waterbody made it possible to still confidently identify taxa in each system. This suggests that our approach could be a successful means of eDNA capture especially when efficiency in the field and high sample size per system are desired. Many filtration protocols collected fewer water samples per waterbody than we did even if a larger total volume was sampled with filtration (Dickie et al., 2018). By collecting many independent

samples per waterbody, it was possible to use the number of observations of a taxon in water samples as a control for false-positive detection in addition to overall read counts. This approach is similar to other studies that collected a large number of independent samples per waterbody and is suggested to be a balanced approach for handling putative false-positive and false-negative errors (Evans, Li, et al., 2017). This additional filter parameter made it possible to
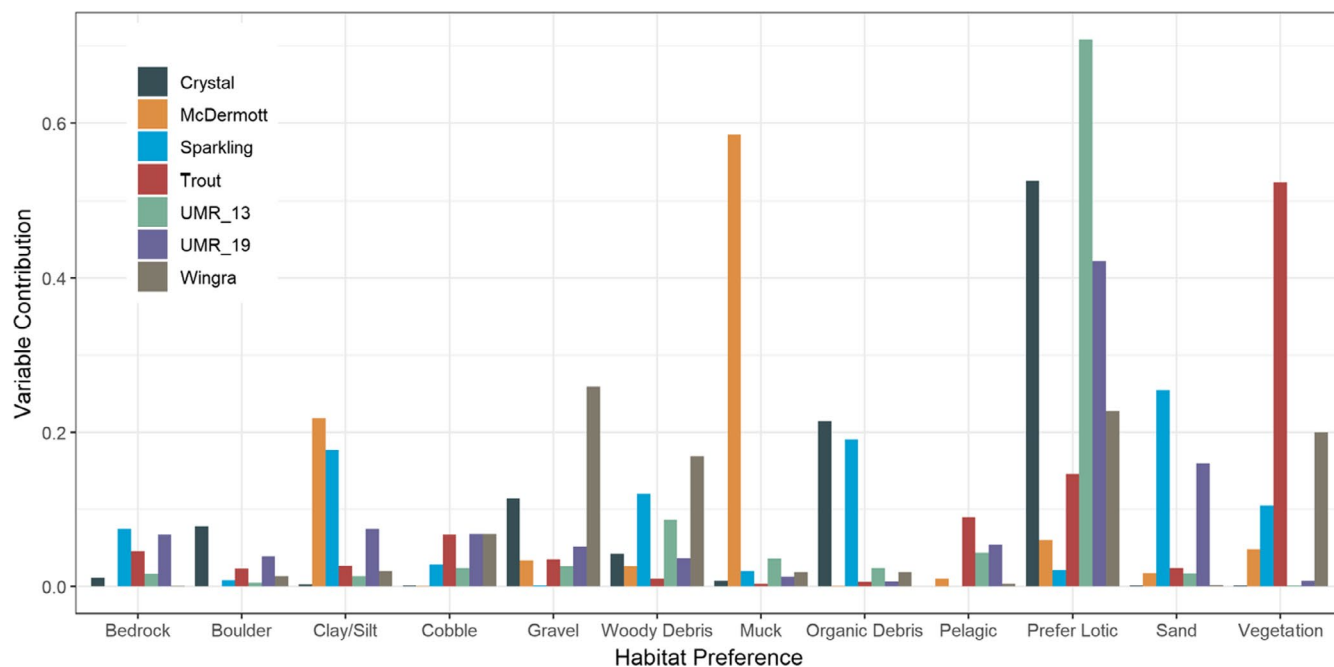
FIGURE 4 The proportional variable contributions of the 12 habitat/substrate variables used to explain the differences between species not detected with eDNA metabarcoding and species detected by eDNA for all waterbodies in Discriminant Analysis of Principal Components (DAPC). Higher variable contribution indicates that there is a higher differentiation in preference for that habitat/substrate type between detected and undetected groups, while lower variable contribution indicates a lack of difference between groups

consider taxa that were poorly amplified, such as *S. vitreus*, to be detected while still protecting against potential false-positives due to contamination.

4.2 | Reconstruction of known community structure

Our results complement much of the existing eDNA metabarcoding literature in finding that eDNA metabarcoding is capable of detecting a subset of the known fish community across a broad range of life histories and habitat preferences (Hänfling et al., 2016; Sard et al., 2019). Also, our data represent a low estimate for species detection because only a single gene was used for taxonomic assignments. Many eDNA metabarcoding studies use multiple genes which can help to increase the number of species detected and increase the number of species-level taxonomic assignments (e.g. Sard et al. 2019; Shaw et al. 2016). We chose to use a single gene because the majority of taxa can often be identified using a single marker and our goal was to evaluate the ability of a very basic eDNA monitoring protocol to reconstruct known fish communities. Future studies could include additional metabarcoding markers such as 16S or COI to increase species-level assignments and detections.

We collected samples from a broad range of habitats within each waterbody including below the thermocline in stratified lakes. This may have helped to ensure the detection of species with diverse habitat preferences. Some studies have found that eDNA is often well mixed throughout a system (Zhang et al. 2020), but others found that DNA can clump together (Furlan et al. 2016; Williams,

Huyvaert, Vercauteren, Davis, Piaggio, 2018); therefore, the optimal sampling strategy should consider both spatial distribution which could be important for stagnant systems where there is low mixing or large waterbodies and total volume which could be important to mitigate effects of DNA clumping. Conventional sampling approaches often have biased sampling efficiency for different species (Murphy & Willis, 1996); therefore, multiple types of sampling are necessary to fully describe local fish communities (Clark et al., 2007). For example, the data sets used to construct our known lake communities were derived from a combination of different conventional sampling techniques: gillnets, fyke nets, minnow traps, electrofishing and beach seines. Our findings are consistent with previous studies that have illustrated the use of eDNA metabarcoding as a complement to conventional sampling efforts (Hanfling et al., 2016; Shaw et al., 2016); however, our study also suggests that eDNA metabarcoding was able to detect species from a diverse background of habitat/substrate preferences. In situations when only a single sampling method can be adopted, eDNA may provide a more representative sample of the overall fish community than certain conventional sampling approaches; therefore, eDNA metabarcoding surveys could provide useful information for waterbodies that lack fish community data. However, more detailed research evaluating life history biases should be conducted to identify the limits of this approach. One important consideration for future studies should be the confound between life history bias and primer bias for related taxa. These sources of bias are confounded because related taxa often share life history traits but will also have similar 12S sequences leading to similar amplification rates (i.e. primer bias). Without careful evaluation,

over- or under-detection of a taxa could be attributed to life history when the actual cause is laboratory bias during PCR enrichment or *vice versa*. Once optimized, eDNA metabarcoding surveys could help improve the sensitivity of conservation tools such as species distribution models which can perform poorly when environments are not well sampled (Wisz et al., 2008).

While we found limited sampling bias, careful interpretation of eDNA sequence data is still crucial for accurate community assessments (Valentini et al., 2016) because substantial laboratory and bioinformatic biases do occur (Kelly et al., 2019; Piggott, 2016). For example, *S. vitreus* was a common species in 6 out of the 9 waterbodies we sampled, but we observed few *S. vitreus* detections across all samples, and where detections did occur, they contained low read counts. This suggests potential primer bias against *S. vitreus*; however, walleye have been successfully detected using these same primers in other studies (Gehri et al. 2020). The primers we used in this study are general vertebrate primers that are known to amplify a broad range of vertebrate taxa including the dominant taxonomic groups of fish common to lakes and rivers in our study area (Riaz et al. 2011; Gehri et al. 2020; Sard et al. 2019). Still, given the low sequencing depth which may compound the effects of primer bias and lack of a second genetic marker which would help to identify variance in sequence abundance we cannot rule out that some taxa are over- or under-represented in our data set. For this reason, primer bias can be especially problematic when attempting to estimate relative abundance based on metabarcoding (Piñol et al., 2019). We chose to use eDNA data as a method of detection instead of quantification of relative abundance which means that primers would need to fail to amplify sequences from a given taxa for it to be excluded from our analysis. However, presence-only data can still provide important information on species occurrence that is necessary to guide conservation action (Hefley et al., 2017; McDonald et al., 2013).

False-positive detections due to contamination or misassignment are a common concern in eDNA studies (Lahoz-Monfort et al., 2016; Wilcox et al., 2013). Here, we identified eight species in three waterbodies that had not been previously reported. Two novel species detections in Trout Lake and one in Lake Wingra have never been observed in either lake in 40 years of routine monitoring. Given the extensive conventional sampling of both lakes, it is unlikely that the detections are from DNA shed from individuals residing in each waterbody. In certain cases, such as the detection of *F. diaphanus* in Lake Wingra, it is possible that *F. diaphanus* recently recolonized the lake and is a true detection (Willink et al., 2018). Unfortunately, it is very difficult conclusively say whether detections are false-positives due to contamination, misassignment, or from DNA that originated from adjacent unsampled streams or wetlands or transferred to the lake by some other means. Additionally, Cyprinidae DNA closely related to *Ctenopharyngodon idella* 12S reference sequences was identified in Trout Bog, which only has one known species of fish (*U. limi*). The molecular laboratory at UMESC where the water samples were processed frequently works with Asian carp, including *C. idella*, indicating contamination as the likely source of this detection, but it could have originated elsewhere. Misassignments could also be the

result of an incomplete reference database; however, this concern was mitigated by developing a reference database with as many species known to inhabit local lakes as possible and by conducting GenBank BLASTn searches on all unknown sequences. Nonetheless, 12S sequences were missing in GenBank for 15 species that have been observed in one of our survey sites previously. If sequences from any of these species were in our samples, it is likely they assigned to a closely related species or assigned to a higher taxonomic level (e.g. genus or family). Until a complete archive of barcoding sequences is available for all known species in a region, there will be the potential for taxa to be misassigned or un-assigned as a result of missing reference data. Extensive knowledge of fish communities in our study systems allowed us to make informed decisions about unexpected detections; however, when study systems lack previous survey data, it will be important to acknowledge type I errors when describing communities and plan for false-positive corrections (Ficetola et al., 2016) since these detections could bias future species distribution studies (Gormley et al., 2011).

We were able to successfully identify established differences and similarities among fish communities using the subset of the total fish community detected with eDNA. A great deal of community ecology research aims to quantify and describe patterns of diversity across time and space (Lomolino & Rosenzweig, 1996). However, this requires consistent methodology and effort to appropriately compare among systems (Gotelli, 2008). It can be difficult to conduct biodiversity surveys consistently at multiple locations because of differences in gear, sampling effort, or scientific expertise (Bried & Hinchliffe, 2019). Our study highlights the capability of eDNA metabarcoding to consistently sample a variety of systems, allowing simple cross-system comparison. eDNA detections described broad community structure similar to that achieved by the most recent 5 years of routine long-term monitoring or 1–2 years of intensive short-term monitoring with conventional methods. Waterbody grouping in NMDS mirrored established differences in fish communities such as strong community differences between river and lake habitats and smaller but still significant differences between eutrophic and oligotrophic lakes. These cross-system differences can help to identify at-risk systems or differences in ecosystem health among systems (Argillier et al., 2013). Our results and others suggest that eDNA metabarcoding could facilitate community surveys at a larger scale (10s to 100s of waterbodies; Hänfling et al., 2016; Li et al., 2019; Valentini et al., 2016).

4.3 | Implications for biodiversity monitoring and conservation management in the Anthropocene

Describing diversity and community structure continues to be a crucial area of research for conservation (Dudgeon et al., 2006). Establishing baseline community composition and consistent monitoring of a diverse range of systems is one step in meeting conservation goals of the future (Strayer & Dudgeon, 2010). Many freshwater systems throughout the world are currently

experiencing increased stress (Fluet-Chouinard et al., 2018; Heino et al., 2009) and lack of fish community data makes the sustainable management of these systems much more difficult (Beard et al., 2011; Bower et al., 2020). Our study indicates that a simple eDNA metabarcoding assessment using a high number of low-volume samples, centrifugation and a single gene for metabarcoding can provide quality information about fish community composition that could be incorporated into monitoring regimes and allow for increased scales of sampling. As methods are adopted, sampling and sample processing can be optimized which will help increase the sensitivity and accuracy of assessments. We liken an eDNA metabarcoding approach such as ours to the fundamental limnological tool the Secchi disc. Though a single measurement with a Secchi disc is coarse and imperfect, its simplicity and low effort allow for easy scaling through time and space that make it a powerful tool for understanding freshwater ecosystems (Tyler 1968). Likewise, eDNA metabarcoding offers the opportunity to collect coarse but consistent biodiversity estimates across large spatial and temporal scales.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

There is no conflict of interest declared in this article.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ddi.13253.

DATA AVAILABILITY STATEMENT

Annotated R Scripts can be found on GitHub at https://github.com/peuclide/eDNA_metabarcoding_pipeline. Sequencing and associated metadata are available via ScienceBase https://doi.org/10.5066/P9CORN8Q.

ORCID

Peter T. Euclide https://orcid.org/0000-0002-1212-0435
Yer Lor https://orcid.org/0000-0002-5738-2412
Michael J. Spear https://orcid.org/0000-0002-1365-3583
Tariq Tajjioui https://orcid.org/0000-0002-0113-0451
Jake Vander Zanden https://orcid.org/0000-0002-8122-1416
Wesley A. Larson https://orcid.org/0000-0003-4473-3401
Jon J. Amberg https://orcid.org/0000-0002-8351-4861

REFERENCES

Argillier, C., Caussé, S., Gevrey, M., Pédron, S., De Bortoli, J., Brucet, S., Emmrich, M., Jeppesen, E., Lauridsen, T., Mehner, T., Olin, M., Rask, M., Volta, P., Winfield, I. J., Kelly, F., Krause, T., Palm, A., & Holmgren, K. (2013). Development of a fish-based index to assess the eutrophication status of European lakes. *Hydrobiologia*, https://doi.org/10.1007/s10750-012-1282-y

Aylagas, E., Borja, Á., Irigoien, X., & Rodríguez-Ezpeleta, N. (2016). Benchmarking DNA metabarcoding for biodiversity-based monitoring and assessment. *Frontiers in Marine Science*, 3. https://doi.org/10.3389/fmars.2016.00096.

Banks, J. E., Ackleh, A. S., & Stark, J. D. (2010). The use of surrogate species in risk assessment: Using life history data to safeguard against false negatives. *Risk Analysis*, 30(2), 175–182. https://doi.org/10.1111/j.1539-6924.2009.01349.x

Beard, T. D., Arlinghaus, R., Cooke, S. J., McIntyre, P. B., De Silva, S., Bartley, D., & Cowx, I. G. (2011). Ecosystem approach to inland fisheries: Research needs and implementation strategies. *Biology Letters*, 7(4), 481–483. https://doi.org/10.1098/rsbl.2011.0046

Berker, G. C. (1983). *Fishes of Wisconsin*. University of Wisconsin Press.

Birge, E. A., & Juday, C. (1934). Particulate and dissolved organic matter in inland lakes. *Ecological Monographs*, 4(4), 440–474. https://doi.org/10.2307/1961650

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, 29(6), 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Bower, S. D., Aas, Ø., Arlinghaus, R., Douglas Beard, T., Cowx, I. G., Danylchuk, A. J., Freire, K. M. F., Potts, W. M., Sutton, S. G., & Cooke, S. J. (2020). Knowledge gaps and management priorities for recreational fisheries in the developing world. *Reviews in Fisheries Science and Aquaculture*, 28, 518–535. https://doi.org/10.1080/23308249.2020.1770689

Branton, M., & Richardson, J. S. (2011). Assessing the value of the umbrella-species concept for conservation planning with meta-analysis. *Conservation Biology*, 25(1), 9–20. https://doi.org/10.1111/j.1523-1739.2010.01606.x

Bried, J. T., & Hinchliffe, R. P. (2019). Improving taxonomic resolution in large-scale freshwater biodiversity monitoring: An example using wetlands and Odonata. *Insect Conservation and Diversity*, 12(1), 9–17. https://doi.org/10.1111/icad.12323

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. https://doi.org/10.1038/nmeth.3869

Clark, S. J., Jackson, J. R., & Lochmann, S. E. (2007). A comparison of shoreline seines with fyke nets for sampling littoral fish communities in floodplain lakes. *North American Journal of Fisheries Management*, 27(2), 676–680. https://doi.org/10.1577/m06-197.1

Collingsworth, P. D., Bunnell, D. B., Murray, M. W., Kao, Y.-C., Feiner, Z. S., Claramunt, R. M., Lofgren, B. M., Höök, T. O., & Ludsin, S. A. (2017). Climate change as a long-term stressor for the fisheries of the Laurentian Great Lakes of North America. *Reviews in Fish Biology and Fisheries*, 27(2), 363–391. https://doi.org/10.1007/s11160-017-9480-3

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., Potter, C., & Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, 7(9), 1008–1018. https://doi.org/10.1111/2041-210X.12574

Deiner, K., Walser, J. C., Mächler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, 183, 53–63. https://doi.org/10.1016/j.biocon.2014.11.018

Dickie, I. A., Boyer, S., Buckley, H. L., Duncan, R. P., Gardner, P. P., Hogg, I. D., Holdaway, R. J., Lear, G., Makiola, A., Morales, S. E., Powell, J. R., &

Weaver, L. (2018). Towards robust and repeatable sampling methods in eDNA-based studies. *Molecular Ecology Resources*, 18(5), 940–952. https://doi.org/10.1111/1755-0998.12907

Dodson, S. I., Arnott, S. E., & Cottingham, K. L. (2000). The relationship in lake communities between primary productivity and species richness. *Ecology*, 81(10), 2662–2679.

Downing, J. A., & Duarte, C. M. (2006). Abundance and size distribution of lakes, ponds and impoundments. *Limnology and Oceanography*, 51(5), 2388–2397. https://doi.org/10.1016/B978-012370626-3.00025-9

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A.-H., Soto, D., Stiassny, M. L. J., & Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, 81(02), 163–182. https://doi.org/10.1017/S1464793105006950

Erickson, R. A., Rees, C. B., Coulter, A. A., Merkes, C. M., McCalla, S. G., Touzinsky, K. F., & Amberg, J. J. (2016). Detecting the movement and spawning activity of bigheaded carps with environmental DNA. *Molecular Ecology Resources*, 16(4), 957–965. https://doi.org/10.1111/1755-0998.12533

Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., Jerde, C. L., Lodge, D. M., Lamberti, G. A., & Pfrender, M. E. (2017). Fish community assessment with eDNA metabarcoding: Effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(9), 1–13. https://doi.org/10.1139/cjfas-2016-0306

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 environmental DNA metabarcoding. *Molecular Ecology Resources*, 16(1), 29–41. https://doi.org/10.1111/1755-0998.12433

Evans, N. T., Shirey, P. D., Wieringa, J. G., Mahon, A. R., & Lamberti, G. A. (2017). Comparative cost and effort of fish distribution detection via environmental DNA analysis and electrofishing. *Fisheries*, 42(2), 90–99. https://doi.org/10.1080/03632415.2017.1276329

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, 4(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Ficetola, G. F., Taberlet, P., & Coissac, E. (2016). May 1). How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources*, 16, 604–607. https://doi.org/10.1111/1755-0998.12508

Fluet-Chouinard, E., Funge-Smith, S., & McIntyre, P. B. (2018). Global hidden harvest of freshwater fish revealed by household surveys. *Proceedings of the National Academy of Sciences of the United States of America*, 115(29), 7623–7628. https://doi.org/10.1073/pnas.1721097115

Frimpong, E. A., & Angermeier, P. L. (2009). A database of ecological and life-history traits of freshwater fishes of the United States. *Fisheries*, 34(10).

Froese, R., & Pauly, D. (2019). FishBase. World Wide Web Electronic Publication. www.fishbase.org, version (12/2019).

Fujii, K., Doi, H., Matsuoka, S., Nagano, M., Sato, H., & Yamanaka, H. (2019). Environmental DNA metabarcoding for fish community analysis in backwater lakes: A comparison of capture methods. *PLoS One*, 14(1), e0210357. https://doi.org/10.1371/journal.pone.0210357

Furlan, E. M., Gleeson, D., Hardy, C. M., & Duncan, R. P. (2016). A framework for estimating the sensitivity of eDNA surveys. *Molecular Ecology Resources*, 16(3), 641–654. http://dx.doi.org/10.1111/1755-0998.12483

Gehri, R. R., Larson, W. A., Gruenthal, K., & Sard, N. (2020). eDNA metabarcoding outperforms traditional fisheries sampling and reveals fine-scale heterogeneity in a temperate freshwater lake. *BioRxiv*, 800, 10.20.347203. https://doi.org/10.1101/2020.10.20.347203

Gotelli, N. (2008). *A Primer of Ecology*, 4th edn. Sunderland, MA: Sinauer Associates.

Goldberg, C. S., Strickler, K. M., & Pilliod, D. S. (2015). Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. *Biological Conservation*, 183, 1–3. https://doi.org/10.1016/j.biocon.2014.11.040

Gormley, A. M., Forsyth, D. M., Griffioen, P., Lindeman, M., Ramsey, D. S. L., Scroggie, M. P., & Woodford, L. (2011). Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species. *Journal of Applied Ecology*, 48(1), 25–34. https://doi.org/10.1111/j.1365-2664.2010.01911.x

Hänfling, B., Lawson Handley, L., Read, D. S., Hahn, C., Li, J., Nichols, P., Blackman, R. C., Oliver, A., & Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology*, 25(13), 3101–3119. https://doi.org/10.1111/mec.13660

Hefley, T. J., Brost, B. M., & Hooten, M. B. (2017). Bias correction of bounded location errors in presence-only data. *Methods in Ecology and Evolution*, 8(11), 1566–1573. https://doi.org/10.1111/2041-210X.12793

Heino, J., Virkkala, R., & Toivonen, H. (2009). Climate change and freshwater biodiversity: Detected patterns, future trends and adaptations in northern regions. *Biological Reviews*, 84(1), 39–54. https://doi.org/10.1111/j.1469-185X.2008.00060.x

Humbert, J.-F., & Dorigo, U. (2005). Biodiversity and aquatic ecosystem functioning: A mini-review. *Aquatic Ecosystem Health & Management*, 8(4), 367–374. https://doi.org/10.1080/14634980500457773

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kassambara, A. (2020). ggpubr: "ggplot2" Based Publication Ready Plots. Retrieved from https://CRAN.R-project.org/package=ggpubr.

Kelly, R. P., Shelton, A. O., & Gallego, R. (2019). Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. *Scientific Reports*, 9(1), 660530. https://doi.org/10.1038/s41598-019-48546-x

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Tingley, R. (2016). Statistical approaches to account for false-positive errors in environmental DNA samples. *Molecular Ecology Resources*, 16(3), 673–685. https://doi.org/10.1111/1755-0998.12486

Li, J., Hatton-Ellis, T. W., Lawson Handley, L. J., Kimbell, H. S., Benucci, M., Peirson, G., & Hänfling, B. (2019). Ground-truthing of a fish-based environmental DNA metabarcoding method for assessing the quality of lakes. *Journal of Applied Ecology*, 56(5), 1232–1244. https://doi.org/10.1111/1365-2664.13352

Li, J., Lawson Handley, L. J., Read, D. S., & Hänfling, B. (2018). The effect of filtration method on the efficiency of environmental DNA capture and quantification via metabarcoding. *Molecular Ecology Resources*, 18(5), 1102–1114. https://doi.org/10.1111/1755-0998.12899

Lomolino, M. V., & Rosenzweig, M. L. (1996). Species diversity in space and time. *The Journal of Wildlife Management*, 60(4), 971. https://doi.org/10.2307/3802400

Lor, Y., Schreier, T. M., Waller, D. L., & Merkes, C. M. (2020). Using environmental DNA (eDNA) to detect the endangered Spectaclecase mussel (*Margaritifera monodonta*). *Freshwater Science*, 711673, 1120–1129. https://doi.org/10.1086/711673

Lynch, A. J., Cooke, S. J., Deines, A. M., Bower, S. D., Bunnell, D. B., Cowx, I. G., Nguyen, V. M., Nohner, J., Phouthavong, K., Riley, B., Rogers, M. W., Taylor, W. W., Woelmer, W., Youn, S.-J., & Beard, T. D. (2016). The social, economic, and environmental importance of inland fish and fisheries. *Environmental Reviews*, 24(2), 115–121. https://doi.org/10.1139/er-2015-0064

Magnuson, J. J., Carpenter, S. R., & Stanley, E. H. (2019). North Temperate Lakes LTER: Fish Abundance 1981 - current ver 39. Environmental Data Initiative, Ver 39. https://doi.org/10.6073/pasta/ffca5d232e1aa84d9508a3370f955885

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17(1), 10. https://doi.org/10.14806/ej.17.1.200.

Maxwell, D., & Jennings, S. (2005). Power of monitoring programmes to detect decline and recovery of rare and vulnerable fish. *Journal of Applied Ecology*, *42*(1), 25–37. https://doi.org/10.1111/j.1365-2664.2005.01000.x

McDonald, L., Manly, B., Huettmann, F., & Thogmartin, W. (2013). Location-only and use-availability data: Analysis methods converge. *Journal of Animal Ecology*, *82*(6), 1120–1124. https://doi.org/10.1111/1365-2656.12145

Merkes, C. M., McCalla, S. G., Jensen, N. R., Gaikowski, M. P., & Amberg, J. J. (2014). Persistence of DNA in carcasses, slime and avian feces may affect interpretation of environmental DNA data. *PLoS One*, *9*(11), e113346. https://doi.org/10.1371/journal.pone.0113346

Murphy, B. R., & Willis, D. W. (1996). *Fisheries Techniques* (2nd ed.; B. R. Murphy & D. W. Willis, Eds.). American Fisheries Society.

Oksanen, J., Kindt, R., Pierre, L., O'Hara, B., Simpson, G. L., Solymos, P., & Wagner, H. (2019). vegan: Community Ecology Package, R package version 2.5-6. R Package Version 2.5-6. Retrieved from http://vegan.r-forge.r-project.org.

Olds, B. P., Jerde, C. L., Renshaw, M. A., Li, Y., Evans, N. T., Turner, C. R., & Lamberti, G. A. (2016). Estimating species richness using environmental DNA. *Ecology and Evolution*, *6*(12), 4214–4226. https://doi.org/10.1002/ece3.2186

Piggott, M. P. (2016). Evaluating the effects of laboratory protocols on eDNA detection probability for an endangered freshwater fish. *Ecology and Evolution*, *6*, 2739–2750.

Piñol, J., Senar, M. A., & Symondson, W. O. C. (2019). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology*, *28*(2), 407–419. https://doi.org/10.1111/mec.14776

R Core Team (2019). *R: A language and environment for statistical computing*. Retrieved from http://www.r-project.org/.

Riaz, T., Shehzad, W., Viari, 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. https://doi.org/10.1093/nar/gkr732

Sard, N. M., Herbst, S. J., Nathan, L., Uhrig, G., Kanefsky, J., Robinson, J. D., & Scribner, K. T. (2019). Comparison of fish detections, community diversity, and relative abundance using environmental DNA metabarcoding and traditional gears. *Environmental DNA*, *1*(4), 368–384. https://doi.org/10.1002/edn3.38

Shaw, J. L. A., Clarke, L. J., Wedderburn, S. D., Barnes, T. C., Weyrich, L. S., & Cooper, A. (2016). Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biological Conservation*, *197*, 131–138. https://doi.org/10.1016/j.biocon.2016.03.010

Strayer, D. L., & Dudgeon, D. (2010). Freshwater biodiversity conservation: Recent progress and future challenges. *Journal of the North American Benthological Society*, *29*(1), 344–358. https://doi.org/10.1899/08-171.1

Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., Orlando, L., & Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, *21*(11), 2565–2573. https://doi.org/10.1111/j.1365-294X.2011.05418.x

Tilman, D. (1999). The ecological consequences of changes in biodiversity: A search for general principles. *Ecology*, *80*(5), 1455–1474. https://doi.org/10.2307/176540.

Tyler, J. E. (1968). The Secchi disc. *Limnology and Oceanography*, *13*(1), 1–6. http://dx.doi.org/10.4319/lo.1968.13.1.0001.

USGS (2005). USGS Long Term Monitoring Program Technical Report.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G. H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.-M., ... Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, *25*(4), 929–942. https://doi.org/10.1111/mec.13428

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686.

Wilcox, T. M., McKelvey, K. S., Young, M. K., Jane, S. F., Lowe, W. H., Whiteley, A. R., & Schwartz, M. K. (2013). Robust detection of rare species using environmental DNA: The importance of primer specificity. *PLoS One*, *8*(3), e59520. https://doi.org/10.1371/journal.pone.0059520

Willink, P. W., Widloe, T. A., Santucci, V. J., Makauskas, D., Tiemann, J. S., Hertel, S. D., Lamer, J. T., & Sherwood, J. L. (2018). Rapid expansion of banded killifish *Fundulus diaphanus* across Northern Illinois: Dramatic recovery or invasive species? *American Midland Naturalist*, *179*, 179–190. https://doi.org/10.1674/0003-0031-179.2.179

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A., & Zimmermann, N. E. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, *14*(5), 763–773. https://doi.org/10.1111/j.1472-4642.2008.00482.x

Xiao, N. (2018). ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for "ggplot2". Retrieved from https://CRAN.R-project.org/package=ggsci.

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., Minamoto, T., & Miya, M. (2017). Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific Reports*, *7*, 10361. https://doi.org/10.1038/srep40368

Zhang, S., Lu, Q., Wang, Y., Wang, X., Zhao, J., & Yao, M. (2020). Assessment of fish communities using environmental DNA: Effect of spatial sampling design in lentic systems of different sizes. *Molecular Ecology Resources*, *20*(1), 242–255. https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.13105

BIOSKETCH

The team of authors includes academic researchers and federal scientists from the Molecular Conservation Genetics Laboratory at University of Wisconsin Stevens, The Center for Limnology, University of Wisconsin-Madison, and the U.S. Geological Survey Upper Midwest Environmental Sciences Center. The team shares an interest in applying molecular tools to inform conservation and management of fish in the Great Lakes regions, USA. In addition to eDNA metabarcoding, the authors' research spans themes of population genetics and evolution (PTE, WAL), aquatic ecology (MJS, JVZ), and environmental DNA methods development and application (YL, TT, JJA).

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.