Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering

Journal Article

Author(s):
Evans, Nathan T.; Li, Yiyuan; Renshaw, Mark A.; Olds, Brett P.; Deiner, Kristy (iD); Turner, Cameron R.; Jerde, Christopher L.; Lodge, David M.; Lamberti, Gary A.; Pfrender, Michael E.

ARTICLE

Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering

Nathan T. Evans, Yiyuan Li, Mark A. Renshaw, Brett P. Olds, Kristy Deiner, Cameron R. Turner, Christopher L. Jerde, David M. Lodge, Gary A. Lamberti, and Michael E. Pfrender

Abstract: Species richness is a metric of biodiversity that represents the number of species present in a community. Traditional fisheries assessments that rely on capture of organisms often underestimate true species richness. Environmental DNA (eDNA) metabarcoding is an alternative tool that infers species richness by collecting and sequencing DNA present in the ecosystem. Our objective was to determine how spatial distribution of samples and "bioinformatic stringency" affected eDNA-metabarcoding estimates of species richness compared with capture-based estimates in a 2.2 ha reservoir. When bioinformatic criteria required species to be detected only in a single sample, eDNA metabarcoding detected all species captured with traditional methods plus an additional 11 noncaptured species. However, when we required species to be detected with multiple markers and in multiple samples, eDNA metabarcoding detected only seven of the captured species. Our analysis of the spatial patterns of species detection indicated that eDNA was distributed relatively homogeneously throughout the reservoir, except near the inflowing stream. We suggest that interpretation of eDNA metabarcoding data must consider the potential effects of water body type, spatial resolution, and bioinformatic stringency.

Résumé : La richesse spécifique est une mesure de la biodiversité qui représente le nombre d'espèces présentes dans une communauté. Les évaluations traditionnelles des ressources halieutiques qui reposent sur la capture d'organismes sous-estiment souvent la richesse spécifique réelle. Les méta-codes à barres d'ADN environnemental) (ADNe) constituent un autre outil qui permet d'inférer la richesse spécifique en recueillant et en séquençant l'ADN présent dans l'écosystème. Notre objectif consistait à déterminer comment la répartition spatiale des échantillons et la « rigueur bioinformatique » influent sur les estimations de la richesse spécifique reposant sur les méta-codes à barres d'ADNe par rapport aux estimations reposant sur la capture, dans un réservoir de 2,2 ha. Quand les critères bioinformatiques exigeaient la détection d'une espèce dans un seul échantillon, la méthode des méta-codes à barres d'ADNe a détecté toutes les espèces capturées par les méthodes traditionnelles en plus de 11 autres espèces non capturées. Toutefois, quand il fallait que les espèces soient détectées sur la base de plus d'un marqueur et dans plus d'un échantillon, les méta-codes à barres d'ADNe n'ont détecté que sept des espèces capturées. Notre analyse de la répartition spatiale de la détection d'espèces indique que l'ADNe était réparti de manière assez uniforme dans tout le réservoir, sauf près de l'embouchure du cours d'eau qui l'alimente. Nous proposons que l'interprétation des données obtenues par la méthode des méta-codes à barres d'ADNe doit tenir compte des effets potentiels du type de plan d'eau, de la résolution spatiale et de la rigueur bioinformatique. [Traduit par la Rédaction]

Introduction

Species richness is a biodiversity metric used in community ecology to describe the number of species in a given area at a given time, and has strong underpinnings in ecological theory (William 1964; MacArthur and Wilson 1967; Connell 1978; Hubbell 2001; Holyoak et al. 2005). Further, the effectiveness of human management efforts is commonly assessed using species richness metrics (Bailey et al. 2004a; Hubert and Quist 2010). Traditionally, assessment of fish species richness has relied on capture-based

sampling of organisms via netting, trapping, or electrofishing (Murphy and Willis 1996; Bonar et al. 2009). However, owing to difficulties related to underwater sampling and the mobility of fishes, traditional capture-based sampling often limits the accuracy of species richness estimates (Bayley and Peterson 2001; Gu and Swihart 2004; Mackenzie and Royle 2005).

Theoretically, a progressive increase in sampling effort should eventually detect all of the species present in the community (McDonald 2004). However, increased effort combined with mul-

N.T. Evans. University of Notre Dame, Department of Biological Sciences and Environmental Change Initiative, Notre Dame, IN, USA; Florida International University, Southeast Environmental Research Center, Miami, FL, USA.

Y. Li, G.A. Lamberti, and M.E. Pfrender. University of Notre Dame, Department of Biological Sciences and Environmental Change Initiative, Notre Dame, IN, USA.

M.A. Renshaw and B.P. Olds. University of Notre Dame, Department of Biological Sciences and Environmental Change Initiative, Notre Dame, IN, USA; Oceanic Institute at Hawai'i Pacific University, Shrimp Department, Waimanalo, HI, USA.

K. Deiner and D.M. Lodge. University of Notre Dame, Department of Biological Sciences and Environmental Change Initiative, Notre Dame, IN, USA; Cornell University, Department of Ecology and Evolutionary Biology and Atkinson Center for a Sustainable Future, Ithaca, NY, USA.

C.R. Turner. University of Notre Dame, Department of Biological Sciences and Environmental Change Initiative, Notre Dame, IN, USA; EcoSystem Genetics LLC, South Bend, IN, USA.

C.L. Jerde. University of Notre Dame, Department of Biological Sciences and Environmental Change Initiative, Notre Dame, IN, USA; University of California, Santa Barbara, Marine Science Institute, Santa Barbara, CA, USA.

Corresponding author: Nathan T. Evans (email: nevans@fiu.edu).

Fig. 1. Conceptual diagram illustrating the relationship between bioinformatic stringency and strength of certainty about the presence of species detected with eDNA metabarcoding.
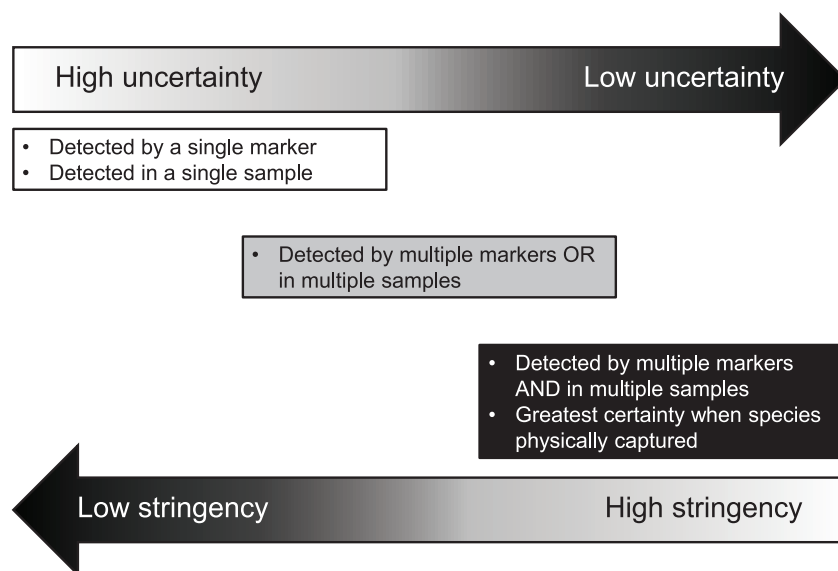
tiple sampling approaches may be needed to accurately measure species richness if all species in the community are not biologically or behaviorally susceptible to a single sampling modality (Peterson and Paukert 2009). For example, both active and passive sample methods are often required to estimate freshwater fish species richness, as passive gears such as fyke nets and gill nets tend to select for mobile species (Hubert 1996), while sedentary species are more susceptible to active gear types such as electrofishing and trawl nets (Hayes et al. 1996). Therefore, as a result of practical limitations in cost and effort, traditional sampling methods in many contexts can be suboptimal in generating estimates of species richness. A potential alternative for estimating species richness is the use of environmental DNA (eDNA) metabarcoding (Lodge et al. 2012).

eDNA metabarcoding infers taxa richness through the identification of taxa-specific DNA fragments collected in relatively small environmental samples (e.g., 250 mL of water). This bioassessment technique is highly sensitive (Ficetola et al. 2008; Bohmann et al. 2014; Rees et al. 2014) and capable of detecting multiple species (Thomsen and Willerslev 2015). Although a relatively recent technological development, eDNA metabarcoding expands eDNA analysis beyond species-specific detection and allows for en masse detection of assemblage-level species richness.

Previous research has shown that eDNA can be effective in determining the identity of fish species in freshwater ecosystems (Dejean et al. 2011; Jerde et al. 2011; Thomsen et al. 2012). Evans et al. (2016) illustrated that eDNA metabarcoding could effectively measure the complete fish and amphibian species richness in experimental mesocosms with varying densities and relative abundances. Olds et al. (2016) used eDNA metabarcoding to measure the complete fish species richness of a natural stream ecosystem and were able to identify DNA from an additional four species not captured via electrofishing but likely present in the ecosystem. Similarly, Valentini et al. (2016) detected at least as many fish as traditional sampling methods in 89% of 23 aquatic sampling sites that included ponds, rivers, mountain lakes, streams, and ditches. Likewise, using eDNA metabarcoding, Hänfling et al. (2016) detected 14 of 16 historically known fish species in a 1480 ha natural lake. Lastly, using eDNA metabarcoding, Shaw et al. (2016) de-

tected all fish species captured with fyke nets in each of two Australian river systems. To date, however, with the exception of post hoc evaluation (Ficetola et al. 2015) and the influence of water column depth and shoreline proximity (Hänfling et al. 2016), studies have provided little guidance on eDNA metabarcoding sampling design or bioinformatic criteria necessary to infer detection.

The ability to use eDNA metabarcoding as an ecological research and conservation tool requires a clear understanding of the data-filtering steps that occur throughout the analysis process. Data filtering takes places at multiple steps in the eDNA metabarcoding process. Initially, the raw sequence data are processed to remove low-quality and nontarget reads (Schloss et al. 2011; Nguyen et al. 2015; Thomsen and Willerslev 2015). There is little consensus, across studies, about what criteria constitute a species detection. This lack of consensus is a result of context dependency (influenced by total species diversity, sequencing depth, marker specificity, etc.) and the trade-off that exists between stringency and uncertainty during the interpretation of eDNA metabarcoding results (Fig. 1). This trade-off in stringency versus uncertainty results from choices about how many and what markers are used to infer detection (Fig. 1). Furthermore, the trade-off occurs when requirements are set on the frequency with which DNA from an organism must be observed in a sample or across samples before it is considered detected.

A full continuum of the stringency–uncertainty trade-off is illustrated in recent eDNA metabarcoding studies on freshwater fish communities, with each study "defining" what constitutes a species detection in a unique way (see online Supplemental Table S1[1]). These studies exhibit diversity in both their filtering steps and in the types and number of markers used. The studies also demonstrate varied ways in which the detection of species can be inferred from postfiltered results. The lack of consensus among these eDNA metabarcoding studies provides little guidance about how to optimize filtering stringency to best define species detections during eDNA metabarcoding.

The overall objective of this study was to test the effectiveness of eDNA metabarcoding to estimate the fish species richness of a small freshwater reservoir by comparing species richness estimates derived from capture-based sampling and eDNA metabar-

---

1364

Can. J. Fish. Aquat. Sci. Vol. 74, 2017

coding. Specifically, we investigated three research questions: (*i*) What species does eDNA metabarcoding detect relative to traditional capture-based sampling? (*ii*) What is the effect of sample size and the spatial distribution of samples on our ability to estimate species richness using eDNA metabarcoding? (*iii*) How does the stringency of bioinformatic criteria applied to species detections, in terms of samples and genetic markers, influence our ability to measure species richness via eDNA metabarcoding?

Materials and methods

Study site

Lawler Pond is a 2.2 ha surface-area reservoir contained within the Fort Custer Training Center (FCTC) of the Michigan Army National Guard located near Battle Creek, Michigan, USA. Lawler Pond is a shallow impoundment (maximum depth ≈3 m) created by a dirt levee and containing a warm-water fish assemblage. Fish habitat within Lawler Pond is relatively homogeneous, with a sand bottom and abundant submerged aquatic vegetation (predominantly *Chara* spp.) throughout the reservoir. A small first-order stream flows into and out of Lawler Pond, which drains a watershed area of approximately 1.4 km². An approximately 2 m wide × 3 m deep channel is located along the northern edge of the reservoir beginning where the stream enters the reservoir from the east and flows out in the northwest corner (Fig. 2). Prior to our sampling, 26 fish species were known to inhabit aquatic ecosystems at FCTC (Michele Richards, FCTC Environmental Biologist, personal communication); however, the fish assemblage of Lawler Pond had not previously been surveyed.
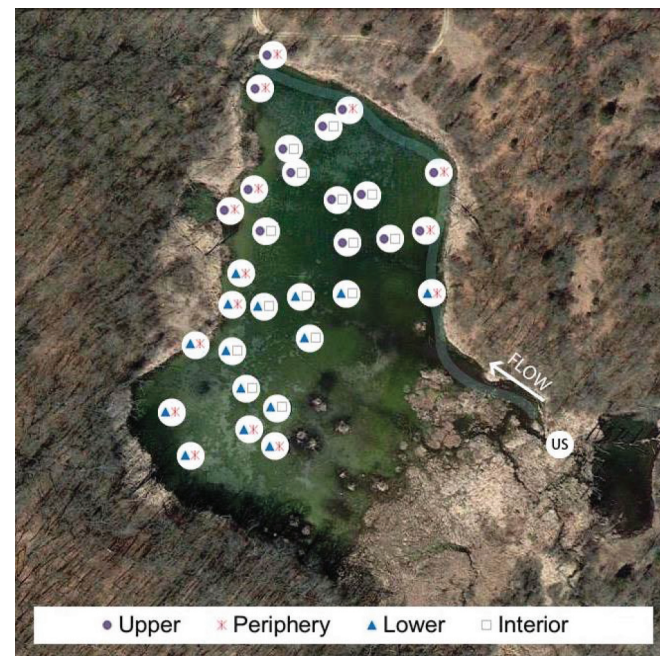
Capture-based sampling

We directly assessed fish species richness in Lawler Pond using a combination of 17 unbaited metal minnow traps and three unbaited modified fyke nets, a 2 m diameter cast net, and handheld dip nets. Modified fyke nets were constructed from two rectangular 91 cm × 183 cm steel frames, four 76 cm diameter steel hoops, and 13 mm knotless nylon bar mesh. From 2 to 6 June 2014, all minnow traps and modified fyke nets were deployed at approximately noon (1200), emptied at approximately 1030 the following morning, then redeployed for a total of four net-nights per net (*n* = 12 total net-nights) and trap (*n* = 68 total trap-nights). Twenty cast net throws were conducted from a boat on the morning of 6 June after the completion of fyke netting. Handheld dip nets were used to target schools of small (<2 cm total length) fishes whenever they were observed. It is important to note that we were not permitted to electrofish in Lawler Pond owing to military regulations and safety concerns (i.e., unexploded munitions). All captured fish were identified to species based on morphological features (and knowledge of local fish fauna), measured for total length and mass, and then returned to the center of the reservoir.

eDNA sampling

On 1 June 2014, 1 day prior to the start of our capture-based sampling, we collected one 250 mL water sample (Evans et al. 2016; Olds et al. 2016) from each of 30 locations distributed throughout Lawler Pond (Fig. 2). In addition, we collected one 250 mL water sample from the stream inflow into Lawler Pond (Fig. 2). Each water sample was collected from the surface of the reservoir by a researcher in a kayak. Prior to sampling, the kayak was decontaminated via a 10 min exposure to 10% bleach solution and then rinsed with reverse osmosis water as recommended by Prince and Andrus (1992) to remove any viable DNA on the surface of the kayak. To minimize the potential for vectoring eDNA among sampling locations within Lawler Pond, samples were collected immediately upon arriving at each sampling location from the bow of the kayak at arm's length (~0.5 m). Additionally, to avoid disturbing future sampling locations, samples were collected starting near the Lawler Pond outflow then proceeded along a single zigzag pattern, ending in the southeast corner of

Fig. 2. Aerial photograph of Lawler Pond (Michigan, USA) illustrating the collection location of each eDNA water sample taken from the impoundment and the inflowing stream (US) as well as the location of the deeper channel (shaded). The 15 samples included in each of the four spatial subsampling designs are indicated by the following symbols: circle (upper samples), asterisk (periphery samples), triangle (lower samples), square (interior samples). Each sample was included in two spatial sampling designs as indicated by the two symbols per sample.

the reservoir. The location of each sample was recorded with a handheld geographic positioning system (GPS) (Garmin Corp, Lenexa, Kansas, USA). Each water sample (250 mL bottle) was wiped with a 10% bleach solution and immediately placed in a cooler containing ice for transport back to the laboratory.

Sample processing and extraction

In the laboratory on that same day, water samples were vacuum-filtered onto 47 mm, 1.2 μm pore size polycarbonate membrane filters (EMD Millipore, Billerica, Massachusetts, USA). Filters containing sample retentate were placed in 2.0 mL microcentrifuge tubes containing 700 μL of CTAB and stored at −20 °C until extraction. DNA was isolated following a modified chloroform – isoamyl alcohol (24:1; Amresco) extraction and an isopropanol precipitation. (Renshaw et al. 2015; see full details in the Supplementary material, Appendix S1[1]). To remove potential inhibitors, resuspended DNA was treated with the OneStepPCR Inhibitor Removal Kit (Zymo Research, Irvine, California, USA).

PCR-based Illumina library preparation and sequencing

We amplified three mitochondrial gene fragments: the cytochrome B gene (CytB; primer set: L14735/H15149c), 12S rRNA (primer set: Am12S), and 16S rRNA (primer set: Ac16S) as described in Evans et al. (2016). Amplified gene fragments were prepped for Illumina sequencing following a two-step PCR-based approach as outlined in the Illumina 16S Metagenomic Sequencing Library preparation guidelines (Illumina, Inc., San Diego, California, USA).

PCR products were electrophoresed through a 2% agarose gel, stained with ethidium bromide, then visualized on a ultraviolet light platform. Each amplified product was manually excised from the gels using single-use razor blades, cleaned with the QIAquick Gel Extraction Kit (Qiagen, Venlo, Netherlands), and eluted from spin columns with 30 μL of buffer EB. We excised a

band from the agarose gel at the expected amplicon size for the extraction and PCR negative controls and, regardless of visual confirmation of amplification, carried each through the remaining library prep for subsequent Illumina sequencing per the recommendation of Nguyen et al. (2015). DNA concentration of each elution was quantified via Qubit dsDNA HS Assay (Life Technologies, Carlsbad, California, USA). Libraries were pooled in equal molar concentrations along with 25% PhiX (v3, Illumina), then paired-end sequenced on an Illumina MiSeq in a single MiSeq flow cell by the University of Notre Dame's Genomics and Bioinformatics Core Facility (http://genomics.nd.edu/) using a MiSeq Reagent Kit v3 (600 cycles; Illumina). To ensure sufficient read depth, libraries were sequenced via two MiSeq runs with 17 libraries per run.

Positive and negative controls

Three types of controls were used to monitor potential contamination during the filtration and laboratory analysis of samples. First, a single mock community sample was constructed (Schloss et al. 2011) and run in parallel with the DNA extraction step. The mock community sample was composed of equal amounts of tissue-derived DNA (measured with Qubit dsDNA HS assay) from six Indo-Pacific marine fishes: ocellaris clownfish (*Amphiprion ocellaris*), jewelled blenny (*Salarias fasciatus*), bicolor blenny (*Ecsenius bicolor*), twospined angelfish (*Centropyge bispinosa*), dispar anthias (*Pseudanthias dispar*), and black leopard wrasse (*Macropharyngodon negrosensis*). Second, a single extraction blank was constructed by using only extraction reagents without a filter, and the blank was subsequently processed alongside the 31 eDNA samples for all laboratory steps. Lastly, a PCR no-template control (NTC) was used for each of the three gene regions amplified and pooled as described above during library preparation. The NTC consisted of sterile water that was added as template during the first round of PCR amplification. A band was then excised from the agarose gel at the anticipated amplicon size, cleaned, and used as template for the second round of PCR amplification, which included the addition of a unique barcode.

Bioinformatics analysis

Raw sequence reads were filtered based on their quality (Q20), merged (Q0.5), and clustered (97%) to species information following the procedure and parameters detailed in Olds et al. (2016). In brief, to detect nontarget (non-vertebrate) operational taxonomic units (OTUs), usually of bacterial origin, we filtered with HMMER (Wheeler and Eddy 2013) using the same parameter values as those used by Olds et al. (2016). Centroid sequences from each OTU were assigned to species with two different approaches. First, we used SAP version 1.9.3 (Munch et al. 2008) to assign species using the NCBI NR database (95% match to reference). Second, we used USEARCH version 8.0.1623 (Edgar 2010) to confirm species assignments using an in-house reference database (Supplemental Table S2[1]) of regional species (97% match to reference). Our in-house database included sequences for additional species, previously identified as present on Fort Custer, not available on GenBank. Sequences for the in-house database were obtained via in-house Sanger sequencing of tissue samples and have since been uploaded to GenBank (accession numbers provided in Supplemental Table S2[1]). We manually checked all OTUs that had a closely related OTU (90%–96.9% similarity) against those in the NCBI GenBank.

Following species assignment, we assessed potential cross-sample contamination, on a per-marker basis, by screening for the presence of any species detected in the 31 Lawler Pond samples in the mock community, extraction blank, and NTC sample libraries. If sequence reads from any species were detected in the three control libraries, we applied a threshold correction (Hänfling et al. 2016; Valentini et al. 2016). For the correction, the cumulative relative frequency of contaminant reads for the detected species in the control libraries functioned as a minimum

detection threshold. For the Lawler Pond samples, any species with a frequency of occurrence (relative proportion of reads) less than that of the detection threshold were discarded (Supplemental Table S3[1]). This correction is similar to the procedure performed by Hänfling et al. (2016), but is based on the false positive reads found in the negative control samples rather than the false positive reads found in their mock community species being detected in field samples.

To determine the effect of bioinformatic decisions on our ability to infer the presence of fishes in Lawler Pond, we then evaluated the effect of three stringency scenarios representing low, moderate, and high stringency (Fig. 1). For the low-stringency scenario, a species was considered detected if its eDNA was found in at least one sample using at least one marker. For the moderate-stringency scenario, species detection required the presence of sequences in at least two samples or by at least two markers from a single sample. For the high-stringency scenario, a species detection required the presence of sequences in both a minimum of two samples and by a minimum of two markers (species were not required to be detected by the same two markers among samples).

### Species accumulation and richness estimation

We estimated species richness based on the Chao II bias-corrected estimator (Chao 2005; Colwell 2013). We calculated all species richness estimates and 95% confidence intervals using EstimateS version 9 (Colwell 2013). The number of samples necessary to measure both the total observed (S_{obs}; detected) and the estimated (Chao II) species richness were calculated via rarefaction analysis with 1000 sample-order randomizations for each of the three bioinformatic criteria scenarios. Sample-based species accumulation curves and 95% confidence intervals were analytically derived using the S_{est} "Mao Tau" estimator in EstimateS version 9 (Colwell 2013). The motivation for including both directly observed species richness (S_{obs}) and an estimator, such as the Chao II bias-corrected, is to evaluate the effects that variable community composition, sampling size, spatial sampling effort, and bioinformatics criteria have on the measured uncertainty in our estimation of species richness, including those not directly observed in the sampling effort.

Sample similarity and spatial analysis

Similarity in the detected species richness of each of the 31 Lawler Pond samples was calculated via the Sørensen coefficient (S_s; Cao et al. 1997). Sørensen dissimilarity (D_s) is calculated as $1 - S_s$. We express both S_s and D_s as percentages by multiplying the index scores by 100. We calculated the Euclidean distance between each of the samples based on GPS coordinates for each of the samples. The effect of spatial separation on species richness similarity was evaluated via a Mantel test of correlation between Euclidean distance and sample similarity using the three bioinformatic stringency criteria used to determine species richness in a sample.

The effect of sample spatial distribution on our ability to estimate species richness was evaluated by subsampling 15 of the 30 available (stream sample omitted) Lawler Pond eDNA samples using four spatial sampling designs: (*i*) subsampling the samples from the periphery of the reservoir, (*ii*) subsampling the samples from the interior of the reservoir, (*iii*) subsampling the upper (north; N) half of the reservoir relative to the inflow, and (*iv*) subsampling the lower (south; S) half of the reservoir relative to the inflow (Fig. 2). The stream sample was excluded from the subsampling as it was located outside of the analysis' scope of inference (Lawler Pond). Chao II species richness estimates were calculated via rarefaction analysis of 1000 sample-order randomizations for each sampling design. The resulting species richness estimates and rarefaction curves were then compared across the four sampling designs and using the three bioinformatic stringency criteria used to determine species richness in a sample.

Fig. 3. Proportional catch of the nine species captured from Lawler Pond, Fort Custer Training Center, Michigan. Number of fishes captured by each method is indicated above each bar. Sampling effort consisted of 12 modified fyke net-nights, 76 minnow trap-nights, 20 cast net throws, and three targeted dip-net dips. Sampling was conducted 3–6 June 2014. In addition to nine species physically captured, common carp were visually observed. [Colour online.]
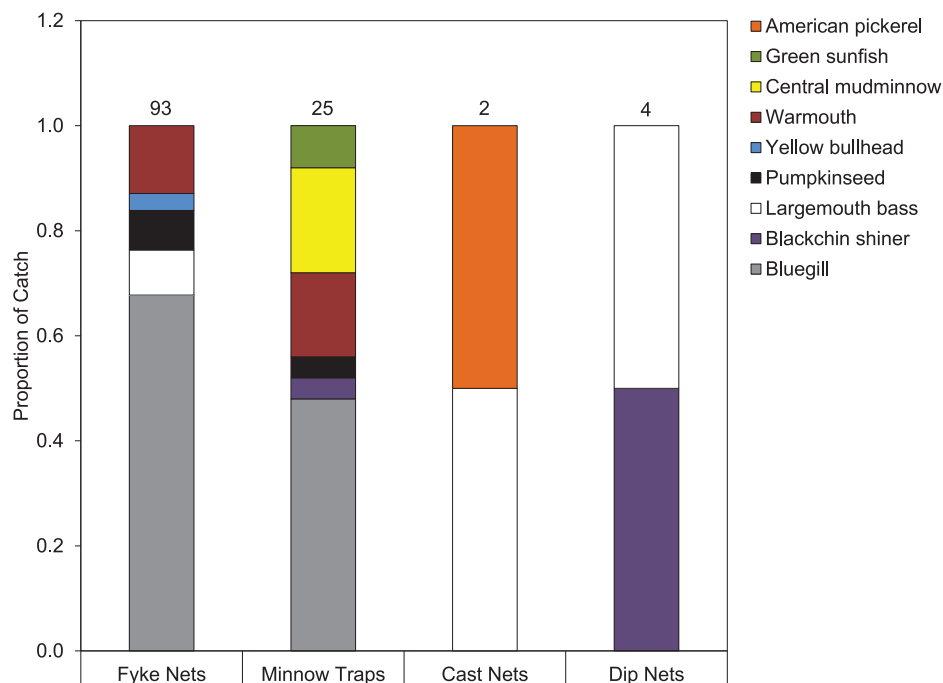
Results

Traditional capture-based sampling

In total, we captured nine species of fishes from Lawler Pond (Fig. 3) in at least one of the four deployed gear types. The majority of the species were captured in the modified fyke nets and minnow traps, with most individuals being captured in the modified fyke nets. In addition to the nine captured species, we visually observed common carp (*Cyprinus carpio*) roaming throughout Lawler Pond but were unable to capture any of the individuals. Because multiple capture-based sampling gears, with differing sampling efficiencies, were deployed over a four-night temporal sampling regime, we were unable to estimate species richness via the Chao II estimator in an equivalent fashion to the estimates derived from the spatially collected eDNA samples.

High-throughput sequencing statistics and effect on species detection

We generated 30.3 million paired-end reads from two Illumina MiSeq runs. After primer demultiplexing, 19.8 million paired-end reads were retained (Supplemental Table S4[1]). The demultiplexing rate was 71.4% for the Lawler Pond samples and 27.5% for the control samples owing to a large proportion of nonspecific amplicons in the PCR negative controls and extraction blanks. In total, 41.3% of the raw reads passed the stringent filtering criteria. USEARCH analysis for OTUs on the combined pools of amplicon-specific sequences and subsequent HMMER modeling (to remove non-vertebrate OTUs) for each of the three markers resulted in detection of 32 OTUs from the 16S fragment, 42 OTUs from the 12S fragment, and 29 OTUs from the CytB fragment (Supplemental Table S4[1]). Several OTUs occurred in low abundance (≤1% of the total number of reads) and matched a reference sequence with only 90%–96% similarity. When manually checked, none of the low-abundance, low-similarity OTUs matched a more similar reference in NCBI GenBank. Therefore, these low-similarity OTUs were excluded from further analysis. Species assignment (see below) further reduced the number of OTUs included in the bioin-

formatic stringency analysis. Prior to subtracting potential cross-library contamination and removing species with only one read per sample, a total of 28 fish species, two turtle species, and humans (all non-fish species were excluded from further analysis) were detected in at least one of the 31 Lawler Pond samples with at least one marker (Table 1).

Comparison of genetic marker species assignments

Based on both the initial species assignment to NCBI NR using SAP and the secondary species assignment to our in-house reference database using USEARCH, we matched 22 OTUs with species-level assignments to the 16S marker (including four mock community species), 19 OTUs with species-level assignments to the 12S marker (including six mock community and human), and 24 OTUs with species-level assignments to the CytB marker (including five mock community, human, and two turtle species) (Table 1). For the 16S and 12S markers, one OTU was assigned to eastern mudminnow (*Umbra pygmaea*), a species that is not believed to occur in Michigan (Bailey et al. 2004b). However, the genetic distance of central mudminnow (*Umbra limi*) and eastern mudminnow is less than 3%. Therefore, we were unable to distinguish between the two species using the three markers employed in this study. We consider all *Umbra* spp. detections to be central mudminnow, a species that is known to occur at Fort Custer. Another species, chain pickerel (*Esox niger*), was detected in multiple samples by both the 16S and 12S markers; however, for these two markers, no reference exists for American pickerel (*Esox americanus*), which was captured via traditional sampling at the time of our sampling. In fact, in 15 of 16 samples where American pickerel were detected via the CytB marker, chain pickerel was detected in the same sample with the 16S or 12S markers. Because chain pickerel is not known to occur in inland Michigan (Bailey et al. 2004b), it is likely that these detections were a misidentification of American pickerel due to a lack of NCBI reference data. We did not consider chain pickerel detections to be accurate iden-

Table 1. Species identified with operational taxonomic unit species assignment for each marker under the low bioinformatic stringency scenario.

Species	Primary habitat	16S	12S	CytB
Fish				
American pickerel	Lentic and lotic (warm water)	×	×	×
Blackchin shiner	Lentic and lotic (warm water)		×	×
Bluegill	Lentic and lotic (warm water)	×	×	×
Bluntnose minnow	Lentic and lotic (warm water)	×		
Brook trout	Lentic and lotic (cold water)			×
Brown trout	Lentic and lotic (cold water)	×		
Central mudminnow	Lentic and lotic (warm water)	×	×	×
Channel catfish	Lentic and lotic (warm water)			×
Common carp	Lentic and lotic (warm water)	×	×	×
Creek chub	Lotic (warm water)	×		
Green sunfish	Lentic and lotic (warm water)	×	×	
Iowa darter	Lentic and lotic (warm water)	×	×	×
Johnny darter	Lotic (cool water)			×
Lake chubsucker	Lentic (warm water)	×	×	×
Largemouth bass	Lentic and lotic (warm water)	×	×	×
Least darter	Lentic and lotic (cool water)			×
Mottled sculpin	Lotic (cool water)	×		
Pumpkinseed	Lentic and lotic (warm water)	×	×	×
Warmouth	Lentic and lotic (warm water)	×	×	×
White sucker	Lentic and lotic (cool water)	×		
Yellow bullhead	Lentic and lotic (warm water)	×		×
Mock community species				
Ocellaris clownfish (*Amphiprion ocellaris*)	Marine	×	×	×
Twospined angelfish (*Centropyge bispinosa*)	Marine	×	×	
Bicolor blenny (*Ecsenius bicolor*)	Marine		×	×
Black leopard wrasse (*Macropharyngodon negrosensis*)	Marine		×	×
Dispar anthias (*Pseudanthias dispar*)	Marine	×	×	×
Jewelled blenny (*Salarias fasciatus*)	Marine	×	×	×
Non-fish vertebrate species				
Human (*Homo sapien*)	Terrestrial		×	
Common snapping turtle (*Chelydra serpentina*)	Lentic and lotic (warm water)			×
Spiny softshell turtle (*Apalone spinifera*)	Lentic and lotic (warm water)			×

Note: Primary habitats for each species were identified based on information available at www.natureserve.org. Refer to Table 2 for scientific names of fishes.

tifications and considered all chain pickerel identifications to be American pickerel detections.

The number of species detected varied among the three genetic markers. No single marker discovered all 21 of the eDNA-detected species under our low-stringency scenario. The highest number of species detected by a single marker was 16 species detected by the 16S marker. Similarly effective was the CytB marker that detected 15 species. The 12S marker was the least effective, detecting just 10 species (Table 1). Of the 16 species detected by the 16S marker, five species were unique to the gene region and not identified by either of the other two markers. Of the 15 species detected by the CytB marker, five species were unique to that gene region and not identified by either of the other two markers. In total, nine species were identified by all three markers, three species were identified by just two markers, and nine species were identified by a single marker (Table 1). Overall, all 21 species could be detected with just the 16S and CytB markers. All species detected with the 12S marker were identified by at least one of the other two markers.

Effects of bioinformatic stringency on species detections and richness estimation

In our low-stringency scenario, eDNA metabarcoding detected 21 species of fishes, including the 10 species observed using traditional sampling (Table 2). eDNA metabarcoding at this stringency level detected an additional 11 fish species. The moderate bioinformatic stringency scenario resulted in the detection of 15 fish species, including the 10 species directly observed. Our high bioinformatic stringency scenario resulted in the detection of eight fish species, including only seven of the 10 species directly observed.

In the low- and moderate-stringency scenarios, sample-specific richness ranged from three to 11 species (Supplemental Figs. S1, S2[1]). Under the high-stringency scenario, sample-specific richness ranged from two to seven species (Supplemental Fig. S3[1]). In the low-stringency scenario, six "singleton" species were detected in only one sample (Supplemental Table S5[1]): brook trout (*Salvelinus fontinalis*), brown trout (*Salmo trutta*), channel catfish (*Ictalurus punctatus*), johnny darter (*Etheostoma nigrum*), mottled sculpin (*Cottus bairdii*), and white sucker (*Catostomus commersonii*). All singleton species were excluded by the high- and moderate-stringency scenarios, as each of the six singleton species were detected by a single marker (three species by 16S and three species by CytB). Moreover, two "doubleton" species, green sunfish (*Lepomis cyanellus*) and creek chub (*Semotilus atromaculatus*), were detected in only two samples (Supplemental Table S5[1]). Despite only being detected in two samples, both green sunfish and creek chub were detected in the moderate-stringency scenario (Supplemental Table S6[1]). However, neither green sunfish, creek chub, lake chubsucker (*Erimyzon sucetta*), bluntnose minnow (*Pimephales notatus*), yellow bullhead (*Ameiurus natalis*), blackchin shiner (*Notropis heterodon*), or least darter (*Etheostoma microperca*) were detected in the high-stringency scenario (Supplemental Table S7[1]).

For the low-stringency scenario, the mean Chao II species richness estimate using all 31 Lawler Pond samples (including the one upstream sample) was 25.8 species present with a 95% confidence interval of 21.8–49.1 species compared with 10 species captured via traditional sampling (Fig. 4a). For the moderate-stringency scenario, the mean Chao II species richness estimate for the metabar-

Table 2. Species observed (capture-based) and detected (eDNA) in Lawler Pond, Fort Custer Training Center, Michigan, under each of the three bioinformatic stringency scenarios: low stringency (low), moderate stringency (moderate), and high stringency (high).

Species	Capture-based	Low	Moderate	High
American pickerel (*Esox americanus*)	×	×	×	×
Blackchin shiner (*Notropis heterodon*)	×	×	×	FN
Bluegill sunfish (*Lepomis macrochirus*)	×	×	×	×
Bluntnose minnow (*Pimephales notatus*)		×	×	
Brook trout (*Salvelinus fontinalis*)		×		
Brown trout (*Salmo trutta*)		×		
Central mudminnow (*Umbra limi*)	×	×	×	×
Channel catfish (*Ictalurus punctatus*)		×		
Common carp (*Cyprinus carpio*)	×	×	×	×
Creek chub (*Semotilus atromaculatus*)		×	×	
Green sunfish (*Lepomis cyanellus*)	×	×	×	FN
Iowa darter (*Etheostoma exile*)		×	×	×
Johnny darter (*Etheostoma nigrum*)		×		
Lake chubsucker (*Erimyzon sucetta*)		×	×	
Largemouth bass (*Micropterus salmoides*)	×	×	×	×
Least darter (*Etheostoma microperca*)		×	×	
Mottled sculpin (*Cottus bairdii*)		×		
Pumpkinseed sunfish (*Lepomis gibbosus*)	×	×	×	×
Warmouth sunfish (*Lepomis gulosus*)	×	×	×	×
White sucker (*Catostomus commersonii*)		×		
Yellow bullhead (*Ameiurus natalis*)	×	×	×	FN
Cumulative species richness	10	21	15	8

Note: An "×" indicates species that were detected via traditional sampling and (or) environmental DNA (eDNA) metabarcoding. "FN" indicates eDNA metabarcoding false negative detections (i.e., species captured via traditional sampling but not detected with eDNA). Blank cells indicate species not detected with either traditional sampling or eDNA metabarcoding.

coding approach was 15 species present with a 95% confidence interval of 15.0–16.2 species (Fig. 4b). For the high-stringency scenario, the mean Chao II species richness estimate for the metabarcoding approach was eight species present with a 95% confidence interval of 8.0–8.3 species (Fig. 4c).

Effects of sample size on estimated species richness

For all three bioinformatic stringency scenarios (low, moderate, and high), the accumulated number of species and the Chao II estimate of species richness varied depending on the number of 250 mL samples included in the analysis. For the low-stringency scenario, the species accumulation curve illustrated that the observed species richness accumulated steadily all the way through inclusion of all 31 eDNA samples (Fig. 4a). The width of the 95% confidence interval was relatively consistent along the length of the rarefaction curve. The mean Chao II estimated richness increased steadily with the addition of samples up through the inclusion of 27 samples. Inclusion of the final four samples (samples 28–31) resulted in a 0.0%–0.6% relative decrease in the mean Chao II estimate. Corresponding to these changes in the mean Chao II estimate were changes in the 95% confidence interval. The 95% confidence interval generally increased in range with the addition of each sample through the inclusion of 26 samples. The range of the 95% confidence interval narrowed with the addition of each sample following the inclusion of 27 samples.

For the moderate-stringency scenario, the species accumulation curve illustrated that observed species richness accumulated rapidly (>2% relative increase in the estimate) up through the inclusion of eight samples (Fig. 4b). The rarefaction curve stabilized after the inclusion of nine samples and reached an asymptote of 15.0 species with the inclusion of 29 samples. Correspondingly, the 95% confidence intervals narrowed following inclusion of just three samples, with the upper and lower confidence bounds converging after the inclusion of 30 samples. The mean Chao II estimate increased rapidly through the inclusion of eight samples. Increasing the number of samples in the analysis to include be-

tween nine and 26 samples yielded a mean Chao II estimate that increased slowly from 14.0 to 15.0 species. Addition of the final five samples resulted in the mean Chao II estimate remaining steady at 15.0 species. Corresponding to these changes in the mean Chao II estimate, the range of the 95% confidence intervals began to narrow with the inclusion of six samples.

For the high-stringency scenario, the species accumulation curve illustrated that observed species richness accumulated steadily up through the inclusion of nine samples (Fig. 4c). Accumulated species richness increased slightly from 7.9 to an asymptote of 8.0 with the inclusion of 17–22 samples. Correspondingly, the 95% confidence intervals began to narrow following inclusion of just two samples, with the upper and lower confidence bounds converging after the inclusion of 22 samples. The mean Chao II estimate increased through the inclusion of 19 samples. Increasing the number of samples in the analysis beyond 19 samples resulted in the same asymptotic species richness estimate of 8.0 species. Corresponding to these changes in the mean Chao II estimate, the range of the 95% confidence intervals began to narrow with the inclusion of only seven samples.

Spatial similarity of eDNA-inferred species richness and the effect of sampling design on estimated species richness

Under the low-stringency scenario, Sørensen coefficients for the 435 pairwise comparisons between each of the 30 Lawler Pond eDNA samples ranged from 27% to 91%, with an overall mean similarity of 61%. Under the moderate-stringency scenario, Sørensen coefficients for the 435 pairwise comparisons between each of the 30 Lawler Pond eDNA samples (excluding the upstream sample) ranged from 33% to 94%, with an overall mean similarity of 64%. Under the high-stringency scenario, Sørensen coefficients for the 435 pairwise comparisons between each of the 30 Lawler Pond eDNA samples ranged from 0% to 100%, with an overall mean similarity of 69%. Euclidean distance between each of the eDNA water samples ranged from 4 to 192 m. We found no relationship between sample dissimilarity (D_s) and distance between the

Fig. 4. Mean species accumulation curve (eDNA detected; grey circles) and mean Chao II species richness estimator curve (Chao estimated; black diamonds) derived from rarefaction analysis of the 31 Lawler Pond eDNA samples libraries under the (*a*) low-stringency scenario, (*b*) moderate-stringency scenario, and (*c*) high-stringency scenario. Error bars represent 95% confidence intervals. [Colour online.]
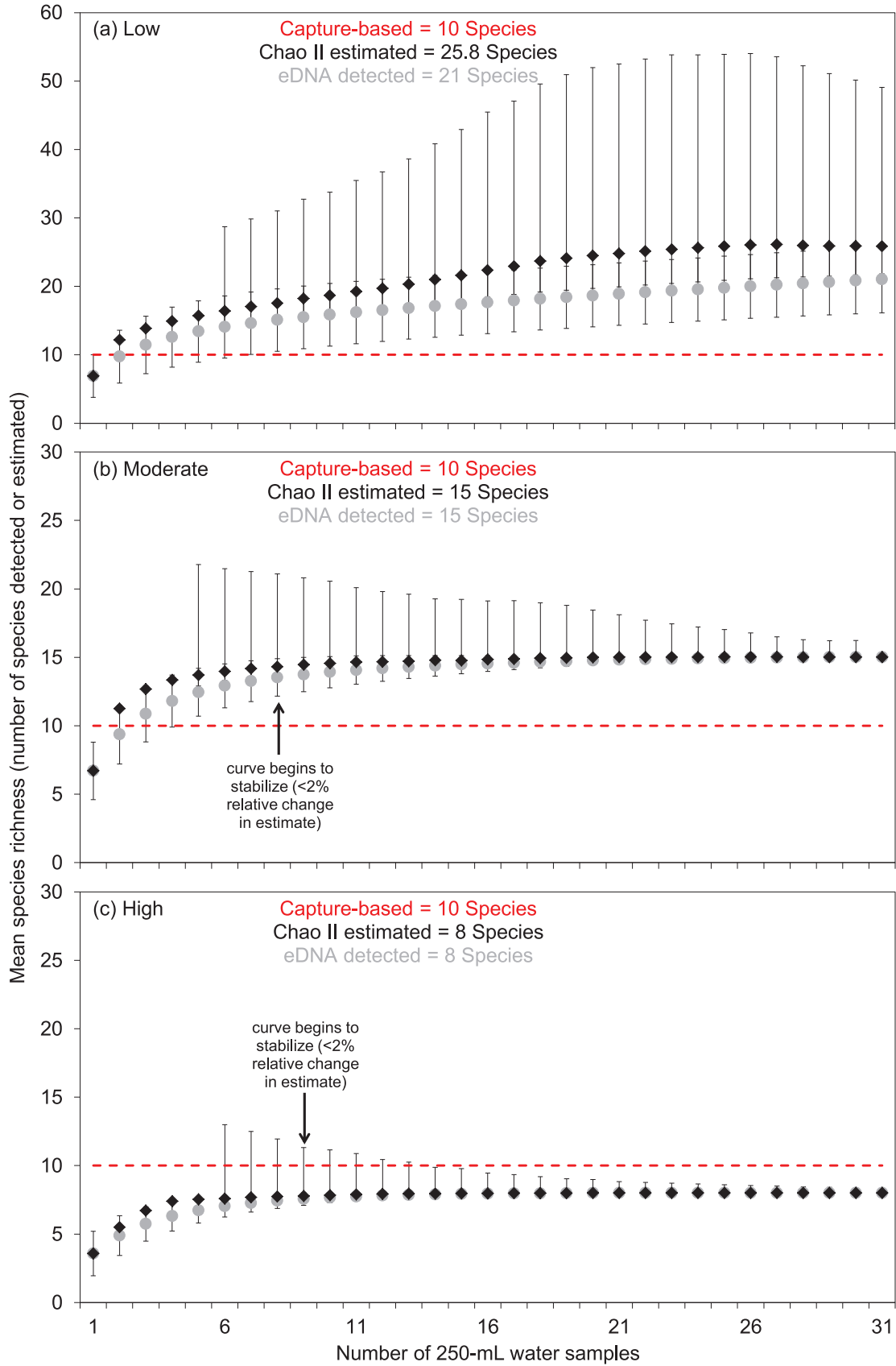
samples under the low-stringency scenario (Mantel's $r = -0.06$, $P = 0.79$; Fig. 5), moderate-stringency scenario (Mantel's $r = -0.01$, $P = 0.5$), or high-stringency scenario (Mantel's $r = -0.64$, $P = 0.98$).

Chao II species richness estimates varied among the three bioinformatic stringency scenarios and the four spatial sampling designs (Fig. 6). Three of the six singleton species (white sucker, channel catfish, and mottled sculpin) were detected in samples collected within the reservoir channel. Additionally, two species (brook trout and brown trout) were not included in the subsampling because they were only detected in the sample collected from the stream flowing into Lawler Pond.

For the low-stringency scenario, the mean species richness estimates for each of the sampling designs ranged from 14.0 to 20.8 compared with a mean estimate of 15.9 species derived from a randomly selected subsample of 15 samples from throughout Lawler Pond (Fig. 6a). The mean estimates of species richness for the upper, periphery, and lower reservoir sampling designs fell within the 95% confidence interval for the random-subsample estimate. The mean estimate for the interior reservoir sampling design was less than the lower 95% confidence bound of the random-subsample estimate.

The range in the mean estimates was smaller for the moderate-stringency scenario, where the mean species richness estimates for each of the sampling designs fell between 13.0 and 15.0 compared with the randomly selected subsample mean richness estimate of 15.9 species (Fig. 6b). Only the mean species richness estimates from the periphery and lower reservoir sampling designs fell within the 95% confidence interval for the random-subsample estimate. The mean estimates for the upper and interior reservoir sampling designs were below the lower 95% confidence bound of the random-subsample estimate.

For the high-stringency scenario, the mean species richness estimates for each of the sampling designs ranged from 6.0 to 7.0 relative to the randomly selected subsample mean species richness estimate of 8.0 (Fig. 6c). The mean species richness estimates from the periphery and lower reservoir sampling designs were both equal to the random-subsample estimate. The mean estimate for the upper and interior reservoir sampling designs were below the lower 95% confidence bound of the random-subsample estimate. Under all three bioinformatic stringency scenarios, the 95% confidence intervals for all the mean estimates overlapped among the spatial sampling designs.

Discussion

The effectiveness of eDNA metabarcoding relative to capture-based sampling

The eDNA-metabarcoding approach employed in this study was able to detect all the species captured via traditional sampling. In addition, under the low-stringency scenario, eDNA metabarcoding detected 11 fish species that were not detected by traditional sampling. The detection of cold-water species and species with lotic life histories (Table 1) may indicate that we detected species that inhabit areas upstream of Lawler Pond and that eDNA from upstream species is transported downstream where it can be detected in the reservoir. Previous studies have illustrated that eDNA can be transported relatively long distances downstream (Deiner and Altermatt 2014; Jane et al. 2015). For example, Jane et al. (2015) detected the eDNA of brook trout at 239 m (the farthest distance they measured) downstream of experimentally caged brook trout. We did not sample the inflowing stream using traditional sampling and are therefore unable to confirm the upstream presence of the additional species. However, our results indicate that five of the six singleton species (all of which exhibit some degree of lotic life histories) were only detected in samples collected from within the channelized portion (the primary flow pathway) of Lawler Pond and thus may be the result of downstream transport of viable eDNA into the reservoir. Increasing the

bioinformatic stringency resulted in the lotic species not being detected. In hindsight, having additional upstream eDNA samples to more fully characterize the species identity of the inflowing eDNA would have been ideal. This highlights an eDNA transport phenomenon that needs to be accounted for adequately in eDNA sampling schemes.

Effect of bioinformatic stringency on species detection

As expected, increasing the stringency of our eDNA bioinformatic criteria resulted in a decrease in the number of species detected. Our use of three markers to determine taxa presence improved our assessment and the reliability of our conclusions about species richness. Similarly, confidence in our species richness estimates increased with increasing bioinformatic stringency (Fig. 4). However, under the high-stringency scenario, our failure to detect three species that were captured by traditional sampling suggests that it is possible to underestimate species (via species elimination) when bioinformatic criteria are too stringent. The magnitude of this effect likely depends on the detection probabilities of the individual markers, the number of markers used, and the quality of the reference database used for species identifications. For example, when only a small number of markers are used, the relative effects of any differences in PCR dynamics and primer binding affinity on species detection are likely to be greater. This would be especially true if one of the markers has particularly good or poor species detection efficiency. Although our three markers (targeting the 16S, 12S, and CytB gene regions) performed similarly, with each detecting 10–15 fish species, eight species were detected by only a single marker, including the six singleton species that were each only detected in a single sample. These eight species were responsible for the decrease in the number of detected species when bioinformatic stringency was increased.

Effect of sample distribution and sample size on species richness estimation

Overall, we observed relatively low spatial heterogeneity in species richness among the 30 Lawler Pond eDNA samples. The low heterogeneity in species richness among the samples and the lack of a relationship between Euclidean distance and D_s suggest that eDNA is distributed relatively homogeneously in Lawler Pond. If eDNA were heterogeneously distributed throughout the pond, we would expect to find a positive relationship between sample dissimilarity and distance, with spatially near samples being more similar and distant samples being less similar. This observed low spatial heterogeneity in eDNA distribution within Lawler Pond suggests that the accumulation of water samples was more important than sample location when attempting to estimate species richness in Lawler Pond.

The homogeneous distribution of eDNA in Lawler Pond may be the result of water column mixing in this shallow reservoir. Previous research has illustrated that surface water in small shallow lakes can mix rapidly due to wind-induced circulation (George and Edwards 1976; Hilton 1985; Spigel and Imberger 1987). Another potential explanation for the homogeneous distribution of eDNA in Lawler Pond is that fishes are dispersed throughout the reservoir consistent with the relatively homogeneous habitat. Lastly, the homogeneous distribution of eDNA in Lawler Pond could be an artifact of the vectoring of eDNA between sampling locations during sample collection. However, our sampling design minimized the likelihood of such vectoring by collecting sampling away from the kayak immediately upon arriving at each sampling location.

Despite our overall finding that eDNA is relatively homogeneously distributed within Lawler Pond, the spatial heterogeneity that was observed appears to be related to the distribution of where the singleton and doubleton species were detected among

Fig. 5. Euclidean distance (m) between eDNA water samples versus Sørensen dissimilarity (D_s). Each point represents one of the 435 pairwise comparisons between all 30 Lawler Pond samples (upstream sample excluded) under the low-stringency scenario. The dashed line in each plot illustrates the generally expected negative relationship (slope < 0) if sample dissimilarity were predicted by distance; however, no significant relationship was found between Euclidean distance and D_s (Mantel's $r = -0.06$, $P = 0.79$).
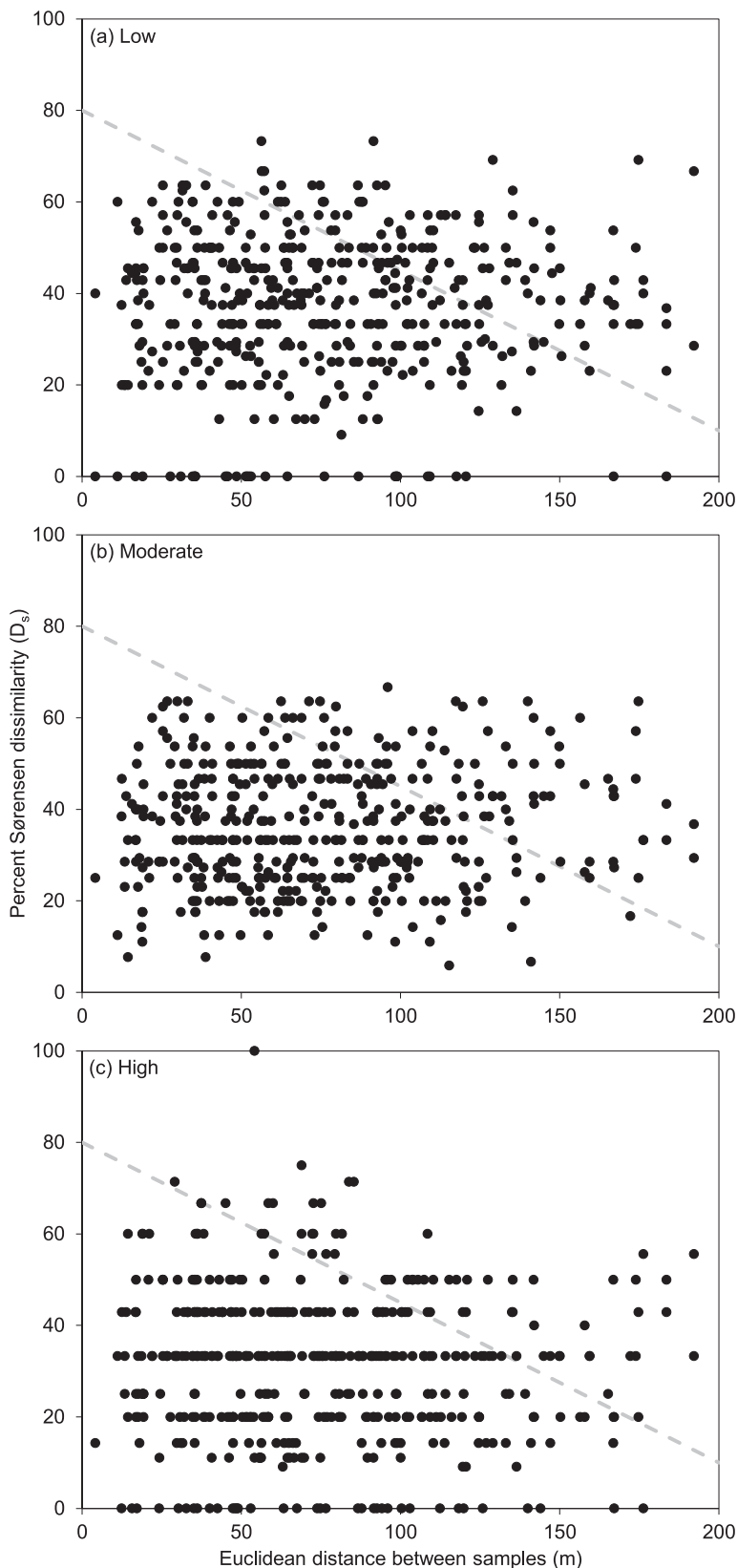
Fig. 6. Mean Chao II species richness estimator curves derived from rarefaction analysis of the eDNA samples selected via each of the four 15-sample spatial designs (upper, lower, periphery, interior) and from a randomly selected subset of all 30 available samples (random) under the (*a*) low-stringency scenario, (*b*) moderate-stringency scenario, and (*c*) high-stringency scenario. Error bars represent 95% confidence intervals of the randomly selected samples.
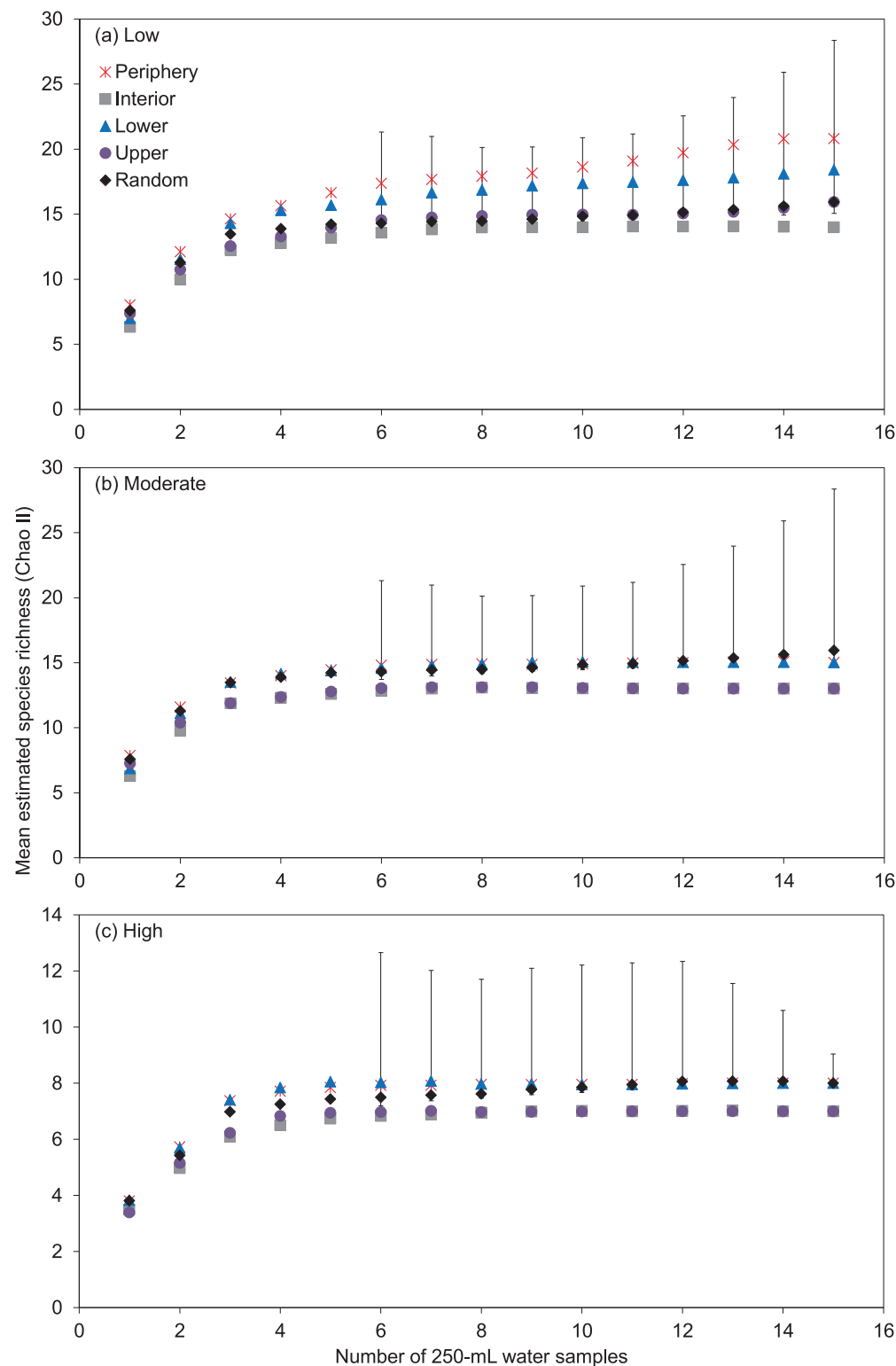
the 30 Lawler Pond samples and the one upstream sample. The concentration of the singleton and doubleton species detections in the reservoir channel explains the observed performance differences among the four sampling zones (i.e., periphery, interior,

upper, and lower reservoir). The unbalanced distribution of the singletons and doubletons in the periphery (the location of the reservoir channel) relative to the interior of the reservoir resulted in the underestimation of species richness by the interior reser-

voir samples. This result is similar to the findings of Hänfling et al. (2016), who detected the greater fish species richness in samples collected closest to the shoreline of a 1480 ha natural lake than in samples collected nearer the center of the lake.

Effect of sample size

Our evaluation of the effect of sample size on our ability to estimate asymptotic species richness in Lawler Pond, under the lowest bioinformatic stringency, suggests that at least 26 water samples must be sequenced with eDNA metabarcoding before species richness can be estimated with accuracy and precision, as indicated by the flattening of the curve and decreasing confidence intervals. The number of water samples decreases under the moderate-stringency (19 samples) and high-stringency (14 samples) scenarios. These estimates of necessary samples apply to Lawler Pond only and may differ from the number of samples needed to estimate species richness in larger and more heterogeneous ecosystems. As noted above, Lawler Pond is a small, relatively homogeneous body of water, making it likely that eDNA would be evenly distributed. In larger bodies of water with distinct spatial structuring, eDNA may be heterogeneously distributed (Hänfling et al. 2016) and an increased numbers of independent samples may be required to capture the maximum eDNA signal. This outcome is consistent with previous research illustrating that diversity and similarity indices tend to underestimate community similarity when calculated with sample sizes that fail to subsample a relatively large proportion of the community (Lande 1996; Cao et al. 1997). The actual sample size needed to accurately and precisely estimate asymptotic species richness also varies according to the diversity of the species assemblage (Chao et al. 2009). It is likely that had we collected additional samples beyond 31, we would have observed greater precision in our species richness estimate. The decrease in the 95% confidence intervals with inclusion of additional samples (e.g., samples 26–31 under the low-stringency scenario) suggests that additional samples would likely continue to increase the precision of the estimate.

Our study illustrates that eDNA metabarcoding can be an effective means of determining species richness in areas that may be difficult to sample via traditional fish-capture methods. These challenging areas can include military installations, remote wilderness areas, and sensitive sites where traditional sampling approaches such as electrofishing may not be feasible or permitted. Our results demonstrate that eDNA metabarcoding can, relative to capture-based sampling, accurately measure and estimate species richness in a small reservoir. Further, eDNA was relatively homogeneously distributed at the spatial scale of Lawler Pond (i.e., 2.2 ha), suggesting that the number of accumulated samples may be more important than the spatial distribution of samples when attempting to quantify species richness via eDNA metabarcoding in small systems. Moreover, the detection of stream-dwelling species in the impoundment suggests that eDNA can also detect species from water transported into the reservoir via streamflow. Further research on the dynamics of eDNA transport is needed to better understand how downstream transport may affect species richness estimation in impoundments and other downstream habitats.

Our results illustrate that the stringency of bioinformatic criteria can have substantial effects on the conclusions about the inferred species richness of the study system. Future research should focus on determining how to optimize the number of markers for estimating species richness via eDNA metabarcoding in diverse ecosystems of varying complexity and size. An improved knowledge of the necessary sample replication would enable the design of more effective and efficient sampling protocols for fish management and conservation. Lastly, while our results illustrate that eDNA metabarcoding can be used to provide robust estimates of species richness, eDNA cannot provide the same

types of population structure data that is readily obtained with capture-based methods where fish can be handled and measured. Therefore, eDNA metabarcoding should be viewed as an additional tool in the fisheries professional's sampling toolbox that can provide improved sensitivity for determining species richness rather than a replacement for demographic sampling via capture-based sampling. However, rapidly advancing genetic and genomic technology provides the promise for even greater utility and interpretive power of eDNA data in the future.

Acknowledgements

References

Bailey, R.C., Norris, R.H., and Reynoldson, T.B. 2004a. Bioassessment of freshwater ecosystems. Kluwer Academic Publisher, New York.

Bailey, R.M., Latta, W.C., and Smith, G.R. 2004b. An atlas of Michigan fishes with keys and illustrations for their identification. Museum of Zoology, Unversity of Michigan, Miscellaneous Publications No. 192, Ann Arbor, Mich.

Bayley, P.B., and Peterson, J.T. 2001. An approach to estimate probability of presence and richness of fish species. Trans. Am. Fish. Soc. **130**: 620–633. doi:10.1577/1548-8659(2001)130<0620:AATEPO>2.0.CO;2.

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Yu, D.W., and de Bruyn, M. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. Trends Ecol. Evol. **29**: 358–367. doi:10.1016/j.tree.2014.04.003. PMID:24821515.

Bonar, S.A., Hubert, W.A., and Willis, D.W. 2009. Standard methods for sampling North American freshwater fishes. American Fisheries Society, Bethesda, Md.

Cao, Y., Williams, W.P., and Bark, A.W. 1997. Effects of sample size (replicate number) on similarity measures in river benthic aufwuchs community analysis. Water Environ. Res. **69**: 107–114. doi:10.2175/106143097X125236.

Chao, A. 2005. Species richness estimation. *In* Encyclopedia of statistical sciences. *Edited by* N. Balakrishnan, C. Read, and B. Vidakovic. Wiley, New York. pp. 7909–7916.

Chao, A., Colwell, R.K., Lin, C.W., and Gotelli, N.J. 2009. Sufficient sampling for asymptotic minimum species richness estimators. Ecology, **90**: 1125–1133. doi:10.1890/07-2147.1. PMID:19449706.

Colwell, R.K. 2013. EstimateS: Statistical estimation of species richness and shared species from samples [online]. Available from http://purl.oclc.org/estimates.

Connell, J.H. 1978. Diversity in tropical rain forests and coral reefs. Science, **199**: 1302–1310. doi:10.1126/science.199.4335.1302. PMID:17840770.

Deiner, K., and Altermatt, F. 2014. Transport distance of invertebrate environmental DNA in a natural river. PLoS ONE, **9**: e88786. doi:10.1371/journal.pone.0088786. PMID:24523940.

Dejean, T., Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P., and Miaud, C. 2011. Persistence of environmental DNA in freshwater ecosystems. PLoS ONE, **6**: e23398. doi:10.1371/journal.pone.0023398. PMID:21858099.

Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics, **26**: 2460–2461. doi:10.1093/bioinformatics/btq461. PMID:20709691.

Evans, N.T., Olds, B.P., Renshaw, M.A., Turner, C.R., Li, Y., Jerde, C.L., Mahon, A.R., Pfrender, M.E., Lamberti, G.A., and Lodge, D.M. 2016. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Mol. Ecol. Resour. **16**: 29–41. doi:10.1111/1755-0998.12433. PMID:26032773.

Ficetola, G.F., Miaud, C., Pompanon, F., and Taberlet, P. 2008. Species detection using environmental DNA from water samples. Biol. Lett. **4**: 423–425. doi:10.1098/rsbl.2008.0118. PMID:18400683.

Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De, Barba, M., Gielly, L., Lopes, C.M., Boyer, F., Pompanon, F., Raye, G., and Taberlet, P. 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Mol. Ecol. Resour. **15**: 543–556. doi:10.1111/1755-0998.12338. PMID:25327646.

George, D.G., and Edwards, R.W. 1976. The effect of wind on the distribution of chlorophyll A and crustacean plankton in a shallow eutrophic reservoir. J. Appl. Ecol. **13**: 667–690. doi:10.2307/2402246.

Gu, W., and Swihart, R.K. 2004. Absent or undetected? Effects of non-detection

1374

Can. J. Fish. Aquat. Sci. Vol. 74, 2017

of species occurrence on wildlife–habitat models. Biol. Conserv. **116**: 195–203. doi:10.1016/S0006-3207(03)00190-3.

Hänfling, B., Lawson Handley, L., Read, D.S., Hahn, C., Li, J., Nichols, P., Blackman, R.C., Oliver, A., and Winfield, I.J. 2016. Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Mol. Ecol. **25**: 3101–3119. doi:10.1111/mec.13660. PMID:27095076.

Hayes, D.B., Ferreri, C.P., and Taylor, W.W. 1996. Active fish capture methods. *In* Fisheries techniques. 2nd ed. *Edited by* B.R. Murphy and D.W. Willis. American Fisheries Society, Bethesda, Md. pp. 193–220.

Hilton, J. 1985. A conceptual framework for predicting the occurrence of sediment focusing and sediment redistribution in small lakes. Limnol. Oceanogr. **30**: 1131–1143. doi:10.4319/lo.1985.30.6.1131.

Holyoak, M., Leibold, M.A., Mouquet, N., Holt, R.D., and Hoopes, M.F. 2005. Metacommunities: a framework for large-scale community ecology. *In* Metacommunities: spatial dynamics and ecological communities. *Edited by* M. Holyoak, M.A. Leibold, and R.D. Holt. University of Chicago Press, Chicago, Ill. pp. 1–32.

Hubbell, S.P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, N.J.

Hubert, W.A. 1996. Passive capture techniques. *In* Fisheries techniques. 2nd ed. *Edited by* B.R. Murphy and D.W. Willis. American Fisheries Society, Bethesda, Md. pp. 157–192.

Hubert, W.A., and Quist, M.C. 2010. Inland fisheries management in North America. 3rd ed. American Fisheries Society, Bethesda, Md.

Jane, S.F., Wilcox, T.M., McKelvey, K.S., Young, M.K., Schwartz, M.K., Lowe, W.H., Letcher, B.H., and Whiteley, A.R. 2015. Distance, flow, and PCR inhibition: eDNA dynamics in two headwater streams. Mol. Ecol. Resour. **15**: 216–227. doi:10.1111/1755-0998.12285. PMID:24890199.

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

Lande, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos, **76**: 5–13. doi:10.2307/3545743.

Lodge, D.M., Turner, C.R., Jerde, C.L., Barnes, M.A., Chadderton, L., Egan, S.P., Feder, J.L., Mahon, A.R., and Pfrender, M.E. 2012. Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. Mol. Ecol. **21**: 2555–2558. doi:10.1111/j.1365-294X.2012.05600.x. PMID:22624944.

MacArthur, R.H., and Wilson, E.O. 1967. The theory of island biogeography. Princeton University Press, Princeton, N.J.

Mackenzie, D.I., and Royle, J.A. 2005. Designing occupancy studies: general advice and allocating survey effort. J. Appl. Ecol. **42**: 1105–1114. doi:10.1111/j.1365-2664.2005.01098.x.

McDonald, L.L. 2004. Sampling rare populations. *In* Sampling rare or elusive species. *Edited by* W.L. Thompson. Island Press, Washington, D.C. pp. 11–42.

Munch, K., Boomsma, W., Huelsenbeck, J.P., Willerslev, E., and Nielsen, R. 2008. Statistical assignment of DNA sequences using Bayesian phylogenetics. Systematic Biol. **57**: 750–757. doi:10.1080/10635150802422316.

Murphy, B.R., and Willis, D.W. 1996. Fisheries techniques. 2nd ed. American Fisheries Society, Bethesda, Md.

Nguyen, N.H., Smith, D., Peay, K., and Kennedy, P. 2015. Parsing ecological signal

from noise in next generation amplicon sequencing. New Phytol. **205**: 1389–1393. doi:10.1111/nph.12923. PMID:24985885.

Olds, B.P., Jerde, C.L., Renshaw, M.A., Li, Y., Evans, N.T., Turner, C.R., Deiner, K., Mahon, A.R., Brueseke, M.A., Shirey, P.D., Pfrender, M.E., Lodge, D.M., and Lamberti, G.A. 2016. Estimating species richness using environmental DNA. Ecol. Evol. **6**: 4214–4226. doi:10.1002/ece3.2186. PMID:27516876.

Peterson, J.T., and Paukert, C.P. 2009. Converting nonstandard fish sampling data to standardized data. *In* Standard methods for sampling North American freshwater fishes. *Edited by* S.A. Bonar, W.A. Hubert, and D.W. Willis. American Fisheries Society, Bethesda, Md. pp. 195–215.

Prince, A.M., and Andrus, L. 1992. PCR: how to kill unwanted DNA. Biotechniques, **12**: 358–360. PMID:1571142.

Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R.M., Gough, K.C., and Crispo, E. 2014. The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. J. Appl. Ecol. **51**: 1450–1459. doi:10.1111/1365-2664.12306.

Renshaw, M.A., Olds, B.P., Jerde, C.L., McVeigh, M.M., and Lodge, D.M. 2015. The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA extraction. Mol. Ecol. Resour. **15**: 168–176. doi:10.1111/1755-0998.12281. PMID:24834966.

Schloss, P.D., Gevers, D., and Westcott, S.L. 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS ONE, **6**: e27310. doi:10.1371/journal.pone.0027310. PMID:22194782.

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

Spigel, R.H., and Imberger, J. 1987. Mixing processes relevant to phytoplankton dynamics in lakes. N.Z. J. Mar. Freshw. Res. **21**: 361–377. doi:10.1080/00288330.1987.9516233.

Thomsen, P.F., and Willerslev, E. 2015. Environmental DNA — an emerging tool in conservation for monitoring past and present biodiversity. Biol. Conserv. **183**: 4–18. doi:10.1016/j.biocon.2014.11.019.

Thomsen, P.F., Kielgast, J.O.S., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L., and Willerslev, E. 2012. Monitoring endangered freshwater biodiversity using environmental DNA. Mol. Ecol. **21**: 2565–2573. doi:10.1111/j.1365-294X.2011.05418.x. PMID:22151771.

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., Peroux, T., Crivelli, A.J., Olivier, A., Acqueberge, M., Le Brun, M., Moller, P.R., Willerslev, E., and Dejean, T. 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol. Ecol. **25**: 929–942. doi:10.1111/mec.13428. PMID:26479867.

Wheeler, T.J., and Eddy, S.R. 2013. nhmmer: DNA homology search with profile HMMs. Bioinformatics, **29**: 2487–2489. doi:10.1093/bioinformatics/btt403. PMID:23842809.

William, C.B. 1964. Patterns in the balance of nature: and related problems in quantitative ecology. Academic Press, London, UK.