Canadian Journal of Fisheries and Aquatic Sciences
Journal canadien des sciences halieutiques et aquatiques

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

Journal:	*Canadian Journal of Fisheries and Aquatic Sciences*
Manuscript ID	cjfas-2016-0306.R1
Manuscript Type:	Article
Date Submitted by the Author:	10-Nov-2016
Complete List of Authors:	Evans, Nathan T.; Florida International University, Southeast Environmental Research Center Li, Yiyuan; University of Notre Dame, Biological Sciences Renshaw, Mark A.; Oceanic Institute at Hawai'i Pacific University Olds, Brett P.; Oceanic Institute at Hawai'i Pacific University Deiner, Kristy; Cornell University, Ecology and Evolutionary Biology Turner, Cameron R.; EcoSystem Genetics LLC Jerde, Christopher L.; University of California, Santa Barbara, Marine Science Institute Lodge, David M.; Cornell University, Ecology and Evolutionary Biology and Atkinson Center for a Sustainable Future Lamberti, Gary A.; University of Notre Dame Pfrender, Michael E.; University of Notre Dame, Biological Sciences
Keyword:	Environmental DNA, eDNA, eDNA Metabarcoding, Species Richness, Fisheries Assessment

SCHOLARONE™
Manuscripts

1 **Title:** Fish community assessment with eDNA metabarcoding: effects of sampling design and
2 bioinformatic filtering

3 **Authors:** Nathan T. Evans[1,2*], Yiyuan Li[1], Mark A. Renshaw[1,3], Brett P. Olds[1,3], Kristy
4 Deiner[1,4], Cameron R. Turner[1,5], Christopher L. Jerde[1,6], David M. Lodge[1,4], Gary A.
5 Lamberti[1], and Michael E. Pfrender[1]

6 **Affiliations:** [1]*University of Notre Dame, Department of Biological Sciences and Environmental*
7 *Change Initiative, Notre Dame, Indiana, USA*

8 [2]*Florida International University, Southeast Environmental Research Center,*
9 *Miami, Florida, USA*

10 [3]*Oceanic Institute at Hawai'i Pacific University, Shrimp Department, Waimanalo,*
11 *Hawaii, USA*

12 [4]*Cornell University, Department of Ecology and Evolutionary Biology and*
13 *Atkinson Center for a Sustainable Future, Ithaca, New York, USA*

14 [5]*EcoSystem Genetics LLC, South Bend, Indiana, USA*

15 [6]*University of California, Santa Barbara, Marine Science Institute, Santa Barbara,*
16 *California, USA*

17

18 *Correspondence: Nathan T. Evans; Telephone: 305-919-4110; Fax: 305-919-403; Email:

19 nevans@fiu.edu

Draft

20 **Abstract**

21 Species richness is a metric of biodiversity that represents the number of species present in a
22 community. Traditional fisheries assessments that rely on capture of organisms often
23 underestimate true species richness. eDNA metabarcoding is an alternative tool, which infers
24 species richness by collecting and sequencing DNA present in the ecosystem. Our objective was
25 to determine how spatial distribution of samples and "bioinformatic stringency" affected eDNA-
26 metabarcoding estimates of species richness compared to capture-based estimates in a 2.2-ha
27 reservoir. When bioinformatic criteria required species to only be detected in a single sample,
28 eDNA metabarcoding detected all species captured with traditional methods plus an additional
29 11 non-captured species. However, when we required species to be detected with multiple
30 markers and in multiple samples, eDNA metabarcoding detected only seven of the captured
31 species. Our analysis of the spatial patterns of species detection indicated that eDNA was
32 distributed relatively homogenously throughout the reservoir, except near the inflowing stream.
33 We suggest that interpretation of eDNA metabarcoding data must consider the potential effects
34 of water body type, spatial resolution, and bioinformatic stringency.

Draft

35 **Introduction**

36 Species richness is a biodiversity metric used in community ecology to describe the number
37 of species in a given area at a given time, and has strong underpinnings in ecological theory
38 (William 1964; MacArthur and Wilson 1967; Connell 1978; Hubbell 2001; Holyoak et al. 2005).
39 Further, the effectiveness of human management efforts are commonly assessed using species
40 richness metrics (Bailey et al. 2004a; Hubert and Quist 2010). Traditionally, assessment of fish
41 species richness has relied on capture-based sampling of organisms via netting, trapping, or
42 electrofishing (Murphy and Willis 1996; Bonar et al. 2009). However, due to difficulties related

43 to underwater sampling and the mobility of fishes, traditional capture-based sampling often
44 limits the accuracy of species richness estimates (Bayley and Peterson 2001; Gu and Swihart
45 2004; Mackenzie and Royle 2005).

46 A progressive increase in sampling effort should theoretically eventually detect all of the
47 species present in the community (McDonald 2004). However, increased effort combined with
48 multiple sampling approaches may be needed to accurately measure species richness if all
49 species in the community are not biologically or behaviorally susceptible to a single sampling
50 modality (Peterson and Paukert 2009). For example, both active and passive sample methods are
51 often required to estimate freshwater fish species richness, as passive gears such as fyke nets and
52 gill nets tend to select for mobile species (Hubert 1996) while sedentary species are more
53 susceptible to active gear types such as electrofishing and trawl nets (Hayes et al. 1996).
54 Therefore, as a result of practical limitations in cost and effort, traditional sampling methods in
55 many contexts can be suboptimal in generating estimates of species richness. A potential
56 alternative for estimating species richness is the use of environmental DNA (eDNA)
57 metabarcoding (Lodge et al. 2012).

58 Environmental DNA metabarcoding infers taxa richness through the identification of taxa-
59 specific DNA fragments collected in relatively small environmental samples (e.g., 250 mL of
60 water). This bioassessment technique is highly sensitive (Ficetola et al. 2008; Bohmann et al.
61 2014; Rees et al. 2014) and capable of detecting multiple species (Thomsen and Willerslev
62 2015). Although a relatively recent technological development, eDNA metabarcoding expands
63 eDNA analysis beyond species-specific detection and allows for *en mass* detection of
64 assemblage-level species richness.

Draft

65 Previous research has shown that eDNA can be effective in determining the identity of fish
66 species in freshwater ecosystems (Dejean et al. 2011; Goldberg et al. 2011; Jerde et al. 2011;
67 Takahara et al. 2012; Thomsen et al. 2012; Wilcox et al. 2013). Evans et al. (2016) illustrated
68 that eDNA metabarcoding could effectively measure the complete fish and amphibian species
69 richness in experimental mesocosms with varying densities and relative abundances. Olds et al.
70 (2016) used eDNA metabarcoding to measure the complete fish species richness of a natural
71 stream ecosystem, and were able to identify DNA from an additional four species not captured
72 via electrofishing but likely present in the ecosystem. Similarly, Valentini et al. (2016) detected
73 at least as many fish as traditional sampling methods in 89% of 23 aquatic sampling sites which
74 included ponds, rivers, mountain lakes, streams, and ditches. Likewise, using eDNA
75 metabarcoding, Hänfling et al. (2016) detected 14 of 16 historically known fish species in a
76 1480-ha natural lake. Lastly, using eDNA metabarcoding, Shaw et al. (2016) detected all fyke-
77 net captured fish species in each of two Australian river systems. To date, however, with the
78 exception of post-hoc evaluation (Ficetola et al. 2015) and the influence of water column depth
79 and shoreline proximity (Hänfling et al. 2016), studies have provided little guidance on eDNA
80 metabarcoding sampling design or bioinformatic criteria necessary to infer detection.
81 The ability to use eDNA metabarcoding as an ecological research and conservation tool
82 requires a clear understanding of the data filtering steps that occur throughout the analysis
83 process. Data filtering takes places at multiple steps in the eDNA metabarcoding process.
84 Initially, the raw sequence data is processed to remove low quality and non-target reads (Schloss
85 et al. 2011; Nguyen et al. 2015; Thomsen and Willerslev 2015). There is little consensus, across
86 studies, about what criteria constitutes a species detection. This lack of consensus is a result of
87 context dependency (influenced by total species diversity, sequencing depth, marker specificity,

Draft

88 etc.) and the trade-off that exists between stringency and uncertainty during the interpretation of
89 eDNA metabarcoding results (Fig. 1). This tradeoff in stringency versus uncertainty results from
90 choices about how many and what markers are used to infer detection (Fig. 1). Furthermore, the
91 tradeoff occurs when requirements are set on the frequency with which DNA from an organism
92 must be observed in a sample or across samples before it is considered detected.
93 A full continuum of the stringency-uncertainty tradeoff is illustrated in recent eDNA
94 metabarcoding studies on freshwater fish communities with each study 'defining' what
95 constitutes a species detection in a unique way (Supplemental Table S1). These studies exhibit
96 diversity in both their filtering steps and in the types and number of markers used. The studies
97 also demonstrate varied ways in which the detection of species can be inferred from post-filtered
98 results. The lack of consensus among these eDNA metabarcoding studies provides little
99 guidance about how to optimize filtering stringency to best define species detections during
100 eDNA metabarcoding.
101 The overall objective of this study was to test the effectiveness of eDNA metabarcoding to
102 estimate the fish species richness of a small freshwater reservoir by comparing species richness
103 estimates derived from capture-based sampling and eDNA metabarcoding. Specifically, we
104 investigated three research questions: (1) What species does eDNA metabarcoding detect
105 relative to traditional capture-based sampling? (2) What is the effect of sample size and the
106 spatial distribution of samples on our ability to estimate species richness using eDNA
107 metabarcoding? (3) How does the stringency of bioinformatic criteria applied to species
108 detections, in terms of samples and genetic markers, influence our ability to measure species
109 richness via eDNA metabarcoding?

Draft

110 **Materials and Methods**

111 *Study site*

112 Lawler Pond is a 2.2-ha surface-area reservoir contained within the Fort Custer Training
113 Center (FCTC) of the Michigan Army National Guard located near Battle Creek, Michigan,
114 USA. Lawler Pond is a shallow impoundment (maximum depth ≈ 3 m) created by a dirt levee
115 and containing a warm-water fish assemblage. Fish habitat within Lawler Pond is relatively
116 homogeneous with a sand-bottom and abundant submerged aquatic vegetation (predominantly
117 *Chara* spp.) throughout the reservoir. A small 1st-order stream flows into and out of Lawler
118 Pond, which drains a watershed area of approximately 1.4 km^2. An approximately 2-m wide and
119 3-m deep channel is located along the northern edge of the reservoir beginning where the stream
120 enters the reservoir from the east and outflows in the northwest corner (Fig. 2). Prior to our
121 sampling, 26 fish species were known to inhabit aquatic ecosystems at FCTC (Michele Richard,
122 FCTC Environmental Biologist, personal communication); however, the fish assemblage of
123 Lawler Pond had not previously been surveyed.

124 *Capture-based sampling*

125 We directly assessed fish species richness in Lawler Pond using a combination of 17
126 unbaited metal minnow traps and three unbaited modified-fyke nets, a 2-m diameter cast net, and
127 handheld dip nets. Modified-fyke nets were constructed from two rectangular 91 X 183-cm steel
128 frames, four 76-cm diameter steel hoops, and 13-mm knotless nylon bar mesh. From June 2-6,
129 2014, all minnow traps and modified-fyke nets were deployed at approximately noon (1200 H),
130 emptied at approximately 1030 H the following morning, then redeployed for a total of four net-
131 nights per net (n=12 total net-nights) and trap (n=68 total trap-nights). Twenty cast net throws
132 were conducted from a boat on the morning June 6 after the completion of fyke netting.

Draft

133 Handheld dip nets were used to target schools of small (<2 cm TL) fishes whenever they were
134 observed. It is important to note that we were not permitted to electrofish in Lawler Pond due to
135 military regulations and safety concerns (i.e., unexploded munitions). All captured fish were
136 identified to species based on morphological features (and knowledge of local fish fauna),
137 measured for total length and mass, and then returned to the center of the reservoir.

138 *eDNA sampling*

139 On June 1, 2014, one day prior to the start of our capture-based sampling, we collected one
140 250-mL water sample (Evans et al. 2016; Olds et al. 2016) from each of 30 locations distributed
141 throughout Lawler Pond (Fig. 2). In addition, we collected one 250-mL water sample from the
142 stream inflow into Lawler Pond (Fig. 2). Each water sample was collected from the surface of
143 the reservoir by a researcher in a kayak. Prior to sampling, the kayak was decontaminated via a
144 10-minute exposure to 10% bleach solution and then rinsed with reverse osmosis water as
145 recommended by Prince and Andrus (1992) to remove any viable DNA on the surface of the
146 kayak. To minimize the potential for vectoring eDNA among sampling locations within Lawler
147 Pond, samples were collected, immediately upon arriving at each sampling location, from the
148 bow of the kayak at arms-length (~0.5 m). Additionally, to avoid disturbing future sampling
149 locations, samples were collected starting near the Lawler Pond outflow then proceeded along a
150 single zig-zag pattern ending in the southeast corner of reservoir.The location of each sample
151 was recorded with a handheld GPS (Garmin Corp, Lenexa, Kansas, USA). Each water sample
152 (250-mL bottle) was wiped with a 10% bleach solution and immediately placed in a cooler
153 containing ice for transport back to the laboratory.

Draft

154 *Sample processing and extraction*

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

163 *PCR–based Illumina library preparation and sequencing*

164 We amplified three mitochondrial gene fragments: the Cytochrome B gene (Cyt B; primer
165 set: L14735/H15149c), 12S rRNA (primer set: Am12S), and 16S rRNA (primer set: Ac16S) as
166 described in Evans et al. (2016). Amplified gene fragments were prepped for Illumina
167 sequencing following a two-step PCR-based approach as outlined in the Illumina 16S
168 Metagenomic Sequencing Library Preparation guidelines (Illumina, Inc., San Diego, CA, USA).
169 PCR products were electrophoresed through a 2% agarose gel, stained with ethidium bromide
170 then visualized on a UV light platform. Each amplified product was manually excised from the
171 gels using single-use razor blades, cleaned with the QIAquick Gel Extraction Kit (Qiagen,
172 Venlo, Netherlands), and eluted from spin columns with 30μL of Buffer EB. We excised a band
173 from the agarose gel at the expected amplicon size for the extraction and PCR negative controls
174 and, regardless of visual confirmation of amplification, carried each through the remaining
175 library prep for subsequent Illumina sequencing per the recommendation of Nguyen et al.
176 (2015). DNA concentration of each elution was quantified via Qubit dsDNA HS Assay.

Draft

177 Libraries were pooled in equal molar concentrations along with 25% PhiX (v3, Illumina, San
178 Diego, California, USA), then paired-end sequenced on an Illumina MiSeq in a single MiSeq
179 flow cell by the University of Notre Dame's Genomics and Bioinformatics Core Facility
180 (http://genomics.nd.edu/) using a MiSeq Reagent Kit v3 (600-cycle; Illumina, San Diego,
181 California, USA). To ensure sufficient read depth, libraries were sequenced via two MiSeq runs
182 with 17 libraries per run.

183 *Positive and Negative Controls*

184 Three types of controls where used to monitor potential contamination during the filtration
185 and laboratory analysis of samples. First, a single mock community sample was constructed
186 (Schloss et al. 2011) and run in parallel from the DNA extraction step. The mock community
187 sample was composed of equal amounts of tissue derived DNA (measured with Qubit dsDNA
188 HS Assay, Life Technologies, Carlsbad, California, USA) from six Indo-Pacific marine fishes:
189 Ocellaris Clownfish *Amphiprion ocellaris*, Jewelled Blenny *Salarias fasciatus*, Bicolor Blenny
190 *Ecsenius bicolor*, Twospined Angelfish *Centropyge bispinosa*, Dispar Anthias *Pseudanthias*
191 *dispar*, and Black Leopard Wrasse *Macropharyngodon negrosensis*. Second, a single extraction
192 blank was constructed by using only extraction reagents without a filter and subsequently
193 processed alongside the 31 eDNA samples for all laboratory steps. Lastly, a PCR no-template
194 control (NTC) was used for each of the three gene regions amplified and pooled as described
195 above during library preparation. The NTC consisted of sterile water that was added as template
196 during the first round of PCR amplification. A band was then excised from the agarose gel at the
197 anticipated amplicon size, cleaned, and used as template for the second round of PCR
198 amplification, which included the addition of a unique barcode.

Draft

199 *Bioinformatics analysis*

200 Raw sequence reads were filtered based on their quality (Q20), merged (Q0.5), and clustered
201 (97%) to species information following the procedure and parameters detailed in Olds et al.
202 (2016). In brief, to detect non-target (non-vertebrate) operational taxonomic units (OTUs),
203 usually of bacterial origin, we filtered with HMMER (Wheeler and Eddy 2013) using the same
204 parameter values as Olds et al. (2016). Centroid sequences from each OTU were assigned to
205 species with two different approaches. First, we used SAP v1.9.3 (Munch et al. 2008) to assign
206 species using the NCBI NR database (95% match to reference). Second, we used USEARCH
207 v8.0.1623 (Edgar 2010) to confirm species assignments using an in-house reference database
208 (Supplemental Table S2) of regional species (97% match to reference). Our in-house database,
209 included sequences for additional species, previously identified as present on Fort Custer, not
210 available on Genbank. Sequences for the in-house database were obtained via in-house Sanger
211 sequencing of tissue samples and have since been uploaded to Genbank (accession numbers
212 provided in Supplemental Table S2) We manually checked all OTUs that had a closely related
213 OTU (90-96.9% similarity) against NCBI Genbank.

214 Following species assignment, we assessed potential cross-sample contamination, on a per
215 marker basis, by screening for the presence of any of species detected in the 31 Lawler Pond
216 samples in the mock community, extraction blank, and NTC sample libraries. If sequence reads
217 from any species were detected in the three control libraries, we applied a threshold correction
218 (Hänfling et al. 2016; Valentini et al. 2016). For the correction, the cumulative relative
219 frequency of contaminant reads for the detected species in the control libraries functioned as a
220 minimum detection threshold. For the Lawler Pond samples, any species with a frequency of
221 occurrence (relative proportion of reads) less than that of the detection threshold were discarded

Draft

222 (Supplemental Table S3). This correction is similar to the procedure performed by Hänfling et
223 al. (2016), but is based on the false positive reads found in the negative control samples rather
224 than false positive reads found from their mock community species being detected in their field
225 samples.

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

234 *Species accumulation and richness estimation*

235 We estimated species richness based on the Chao II bias-corrected estimator (Chao 2005;
236 Colwell 2013). We calculated all species richness estimates and 95% confidence intervals using
237 EstimateS v9 (Colwell 2013). The number of samples necessary to measure both the total
238 observed (S_{obs}; detected) and the estimated (Chao II) species richness were calculated via
239 rarefaction analysis with 1000 sample order randomizations for each of the three bioinformatic
240 criteria scenarios. Sample-based species accumulation curves and 95% confidence intervals
241 were analytically derived using the S_{est} 'Mao Tau' estimator in EstimateS v9 (Colwell 2013).
242 The motivation for including both directly observed species richness (S_{obs}) and an estimator, such
243 as the Chao II bias-corrected, is to evaluate the effects of variable community composition,
244 sampling size, spatial sampling effort, and bioinformatics criteria have on the measured

Draft

245 uncertainty in our estimation of species richness, including those not directly observed in the
246 sampling effort.

247 *Sample similarity and spatial analysis*

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

Draft

256 The effect of sample spatial distribution on our ability to estimate species richness was
257 evaluated by subsampling 15 of the 30 available (stream sample omitted) Lawler Pond eDNA
258 samples using four spatial sampling designs: (1) subsampling the samples from the periphery of
259 the reservoir, (2) subsampling the samples from the interior of the reservoir, (3) subsampling the
260 upper (N) half of the reservoir relative to the inflow, (4) subsampling the lower (S) half of the
261 reservoir relative to the inflow (Fig. 2). The stream sample was excluded from the subsampling
262 as it was located outside of the analysis's scope of inference (Lawler Pond). Chao II species
263 richness estimates were calculated via rarefaction analysis of 1000 sample-order randomizations
264 for each sampling design. The resulting species richness estimates and rarefaction curves were
265 then compared across the four sampling designs and using the three bioinformatic stringency
266 criteria used to determine species richness in a sample.

267 **Results**

268 *Traditional capture-based sampling*

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

Draft

278 *High-throughput sequencing (HTS) statistics and effect on species detection*

279 We generated 30.3 million paired-end reads from two Illumina MiSeq runs. After primer
280 demultiplexing, 19.8 million paired-end reads were retained (Supplemental Table S4). The
281 demultiplexing rate was 71.4% for the Lawler Pond samples and 27.5% for the control samples
282 due to a large proportion of non-specific amplicons in the PCR negative controls and extraction
283 blanks. In total, 41.3% of the raw reads passed the stringent filtering criteria. USEARCH
284 analysis for OTUs on the combined pools of amplicon specific sequences and subsequent
285 HMMER modeling (to remove non-vertebrate OTUs) for each of the three markers resulted in
286 detection of 32 OTUs from the 16S fragment, 42 OTUs from the 12S fragment, and 29 OTUs
287 from the Cyt B fragment (Supplemental Table S4). Several OTUs occurred in low abundance
288 (≤1% of the total number of reads) and matched a reference sequence with only 90-96%
289 similarity. When manually checked, none of the low-abundance, low-similarity OTUs matched

290 a more similar reference on NCBI Genbank. Therefore, these low-similarity OTUs were
291 excluded from further analysis. Species assignment (see below) further reduced the number of
292 OTUs included in the bioinformatic stringency analysis. Prior to subtracting potential cross-
293 library contamination and removing species with only 1 read per sample, a total of 28 fish
294 species, two turtle species, and humans (all non-fish species were excluded from further
295 analysis) were detected in at least one of the 31 Lawler Pond samples with at least one marker
296 (Table 2).

297 *Comparison of genetic marker species assignments*

298 Based on both the initial species assignment to NCBI NR using SAP, and the secondary
299 species assignment to our in-house reference database using USEARCH, we matched 22 OTUs
300 with species-level assignments to the 16S marker (including four mock community species), 19
301 OTUs with species-level assignments to the 12S marker (including 6 mock community and
302 human), and 24 OTUs with species-level assignments to the Cyt B marker (including five mock
303 community, human, and two turtle species) (Table 2). For the 16S and 12S markers, one OTU
304 was assigned to eastern mudminnow (*Umbra pygmaea*), a species that is not believed to occur in
305 Michigan (Bailey et al. 2004b). However, the genetic distance of central mudminnow (*Umbra*
306 *limi*) and eastern mudminnow is less than 3%. Therefore, we were unable to distinguish between
307 the two species using the three markers employed in this study. We consider all *Umbra* spp.
308 detections to be central mudminnow, which is known to occur at Fort Custer. Another species,
309 chain pickerel (*Esox niger*), was detected in multiple samples by both the 16S and 12S markers;
310 however, for these two markers, no reference exists for American pickerel (*Esox americanus*),
311 which was captured via traditional sampling at the time of our sampling. In fact, in 15 of 16
312 samples where American pickerel were detected via the Cyt B marker, chain pickerel was

Draft

313 detected in the same sample with the 16S or 12S markers. Because chain pickerel is not known
314 to occur in inland Michigan (Bailey et al. 2004b), it is likely that these detections were a
315 misidentification of American pickerel due to a lack of NCBI reference data. We did not
316 consider chain pickerel detections to be accurate identifications and considered all chain pickerel
317 identifications to be American pickerel detections.

318 The number of species detected varied among the three genetic markers. No single marker
319 discovered all 21 of the eDNA-detected species under our low stringency scenario. The highest
320 number of species detected by a single marker was 16 species detected by 16S marker. Similarly
321 effective was the Cyt B marker that detected 15 species. The 12S marker was the least effective,
322 detecting just 10 species (Table 2). Of the 16 species detected by the 16S marker, five species
323 were unique to the gene region and not identified by either of the other two markers. Of the 15
324 species detected by the Cyt B marker, five species were unique to that gene region and not
325 identified by either of the other two markers. In total, nine species were identified by all three
326 markers, three species were identified by just two markers, and nine species were identified by a
327 single marker (Table 2). Overall, all 21 species could be detected with just the 16S and the Cyt
328 B markers. All species detected with the 12S marker were identified by at least one of the other
329 two markers.

330 *Effects of bioinformatic stringency on species detections and richness estimation*

331 In our low stringency scenario, eDNA metabarcoding detected 21 species of fishes including
332 the 10 species observed using traditional sampling (Table 1). Environmental DNA
333 metabarcoding at this stringency level detected an additional 11 fish species. The moderate
334 bioinformatic stringency scenario resulted in detection of 15 fish species, including the 10

Draft

335 species directly observed. Our high bioinformatic stringency scenario resulted in detection of
336 eight fish species, including only seven of the 10 species directly observed.
337 In the low and moderate stringency scenarios, sample-specific richness ranged from 3 to 11
338 species (Supplemental Figs. S1, S2). Under the high stringency scenario sample-specific
339 richness ranged from two to seven species (Supplemental Fig. S3). In the low stringency
340 scenario, six 'singleton' species were detected in only one sample (Supplemental Table S5):
341 brook trout (*Salvelinus fontinalis*), brown trout (*Salmo trutta*), channel catfish (*Ictalurus*
342 *punctatus*), johnny darter (*Etheostoma nigrum*), mottled sculpin (*Cottus bairdii*), and white
343 sucker (*Catostomus commersonii*). All singleton species were excluded by the high and
344 moderate stringency scenarios as each of the six singleton species were detected by a single
345 marker (three species by 16S and three species by Cyt B). Moreover, two 'doubleton' species;
346 green sunfish (*Lepomis cyanellus*) and creek chub (*Semotilus atromaculatus*) were detected in
347 only two samples (Supplemental Table S5). Despite only being detected in two samples, both
348 green sunfish and creek chub were detected in the moderate stringency scenario (Supplemental
349 Table S6). However, neither green sunfish, creek chub, lake chubsucker (*Erimyzon sucetta*),
350 bluntnose minnow (*Pimephales notatus*), yellow bullhead (*Ameiurus natalis*), blackchin shiner
351 (*Notropis heterodon*), or least darter (*Etheostoma microperca*) were detected in the high
352 stringency scenario (Supplemental Table S7).
353 For the low stringency scenario, the mean Chao II species richness estimate using all 31
354 Lawler Pond samples (including the one upstream sample) was 25.8 species present with a 95%
355 confidence interval of 21.8 to 49.1 species compared to 10 species captured via traditional
356 sampling (Fig. 4a). For the moderate stringency scenario, the mean Chao II species richness
357 estimate for the metabarcoding approach was 15 species present with a 95% confidence interval

Draft

358 of 15.0 to 16.2 species (Fig. 4b). For the high stringency scenario, the mean Chao II species
359 richness estimate for the metabarcoding approach was 8 species present with a 95% confidence
360 interval of 8.0 to 8.3 species (Fig. 4c).

361 *Effects of sample size on estimated species richness*

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

375 For the moderate-stringency scenario, the species accumulation curve illustrated that
376 observed species richness accumulated rapidly (>2% relative increase in the estimate) up through
377 the inclusion of eight samples (Fig. 4b). The rarefaction curve stabilized after the inclusion of
378 nine samples and reached an asymptote of 15.0 species with the inclusion of 29 samples.
379 Correspondingly, the 95% confidence intervals narrowed following inclusion of just three
380 samples with the upper and lower confidence bounds converging after the inclusion of 30

Draft

381 samples. The mean Chao II estimate increased rapidly through the inclusion of eight samples.
382 Increasing the number of samples in the analysis to include between nine and 26 samples yielded
383 a mean Chao II estimate that increased slowly from 14.0 to 15.0 species. Addition of the final
384 five samples resulted in the mean Chao II estimate remaining steady at 15.0 species.
385 Corresponding to these changes in the mean Chao II estimate, the range of the 95% confidence
386 intervals began to narrow with the inclusion of six samples.

387 For the high-stringency scenario, the species accumulation curve illustrated that observed
388 species richness accumulated steadily up through the inclusion of nine samples (Fig. 4c).
389 Accumulated species richness increased slightly from 7.9 to an asymptote of 8.0 with the
390 inclusion of 17 to 22 samples. Correspondingly, the 95% confidence intervals began to narrow
391 following inclusion of just two samples with the upper and lower confidence bounds converging
392 after the inclusion of 22 samples. The mean Chao II estimate increased through the inclusion of
393 19 samples. Increasing the number of samples in the analysis beyond 19 samples resulted in the
394 same asymptotic species richness estimate of 8.0 species. Corresponding to these changes in the
395 mean Chao II estimate, the range of the 95% confidence intervals began to narrow with the
396 inclusion of only seven samples.

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

399 Under the low-stringency scenario, Sørensen coefficients for the 435 pairwise comparisons
400 between each of the 30 Lawler Pond eDNA samples ranged from 27% to 91% with an overall
401 mean similarity of 61%. Under the moderate-stringency scenario, Sørensen coefficients for the
402 435 pairwise comparisons between each of the 30 Lawler Pond eDNA samples (excluding the
403 upstream sample) ranged from 33% to 94% with an overall mean similarity of 64%. Under the

Draft

404 high-stringency scenario, Sørensen coefficients for the 435 pairwise comparisons between each
405 of the 30 Lawler Pond eDNA samples ranged from 0% to 100% with an overall mean similarity
406 of 69%. Euclidean distance between each of the eDNA water samples ranged from 4 m to 192
407 m. We found no relationship between sample dissimilarity (D_s) and distance between the
408 samples under the low-stringency scenario (Mantel's r = -0.06, *P* = 0.79; Fig. 5), moderate-
409 stringency scenario (Mantel's r = -0.01, *P* = 0.5), or high-stringency scenario (Mantel's r = -0.64,
410 *P* = 0.98).

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

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

424 The range in the mean estimates was smaller for the moderate-stringency scenario, where the
425 mean species richness estimates for each of the sampling designs fell between 13.0 and 15.0
426 compared to the randomly-selected subsample mean richness estimate of 15.9 species (Fig. 6b).

Draft

427 Only the mean species richness estimates from the periphery and lower reservoir sampling
428 designs fell within the 95% confidence interval for the random-subsample estimate. The mean
429 estimate for the upper and interior reservoir sampling designs were below the lower 95%
430 confidence bound of the random-subsample estimate.

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

Draft

439 Discussion

440 *The effectiveness of eDNA metabarcoding relative to capture-based sampling*

441 The eDNA-metabarcoding approach employed in this study was able to detect all of the
442 species captured via traditional sampling. In addition, under the low-stringency scenario eDNA
443 metabarcoding detected 11 fish species that were not detected by traditional sampling. The
444 detection of coldwater species and species with lotic life histories (Table 2) may indicate that we
445 detected species that inhabit areas upstream of Lawler Pond and that eDNA from upstream
446 species is transported downstream where it can be detected in the reservoir. Previous studies
447 have illustrated that eDNA can be transported relatively long distances downstream (Deiner and
448 Altermatt 2014; Jane et al. 2015). For example, Jane et al. (2015) detected the eDNA of brook
449 trout at 239 m (the farthest distance they measured) downstream of experimentally-caged brook

450 trout. We did not sample the inflowing stream using traditional sampling and are, therefore,
451 unable to confirm the upstream presence of the additional species. However, our results indicate
452 that five of the six singleton species (all of which exhibit some degree of lotic life histories) were
453 only detected in samples collected from within the channelized portion (the primary flow
454 pathway) of Lawler Pond and thus may be the result of downstream transport of viable eDNA
455 into the reservoir. Increasing the bioinformatic stringency resulted in the lotic species not being
456 detected. In hindsight, having additional upstream eDNA samples to more fully characterize
457 species identity of the inflowing eDNA would have been ideal. This highlights an eDNA
458 transport phenomenon that needs to be accounted for adequately in eDNA sampling schemes.

459 *Effect of bioinformatic stringency on species detection*

460 As expected, increasing the stringency of our eDNA bioinformatic criteria resulted in a
461 decrease in the number of species detected. Our use of three markers to determine taxa presence
462 improved our assessment and the reliability of our conclusions about species richness. Similarly,
463 confidence in our species richness estimates increased with increasing bioinformatic stringency
464 (Fig. 4). However, under the high-stringency scenario, our failure to detect three species that
465 were captured by traditional sampling suggests that it is possible to underestimate species (via
466 species elimination) when bioinformatic criteria are too stringent. The magnitude of this effect
467 likely depends on the detection probabilities of the individual markers, the number of markers
468 used, and the quality of the reference database used for species identifications. For example,
469 when only a small number of markers are used, the relative effects of any differences in PCR
470 dynamics and primer binding affinity on species detection are likely to be greater. This would be
471 especially true if one of the markers has particularly good or poor species detection efficiency.
472 Although our three markers (targeting the 16S, 12S, and Cyt B gene regions) performed

Draft

473 similarly, with each detecting 10 to 15 fish species, eight species were detected by only a single
474 marker including the six singleton species that were each only detected in a single sample.
475 These eight species were responsible for the decrease in the number of detected species when
476 bioinformatic stringency was increased.

477 *Effect of sample distribution and sample size on species richness estimation*

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

487 The homogeneous distribution of eDNA in Lawler Pond may be the result of water column
488 mixing in this shallow reservoir. Previous research has illustrated that surface water in small
489 shallow lakes can mix rapidly due to wind-induced circulation (George and Edwards 1976;
490 Hilton 1985; Spigel and Imberger 1987). Another potential explanation for the homogeneous
491 distribution of eDNA in Lawler Pond is that fishes are dispersed throughout the reservoir
492 consistent with the relatively homogeneous habitat. While a potential, our sampling design, that
493 involved collecting samples away from the kayak immediately upon arriving at a sampling
494 location, minimizes the likelihood that the observed homogeneous distribution of eDNA in

Draft

495 Lawler Pond is an artifact of vectoring of eDNA between sampling locations during sample
496 collection.
497 Despite our overall finding that eDNA is relatively homogenously distributed within Lawler
498 Pond, the spatial heterogeneity that was observed appears to be related to the distribution of
499 where the 'singleton' and 'doubleton' species were detected among the 30 Lawler Pond samples
500 and the one upstream sample. The concentration of the singleton and doubleton species
501 detections in the reservoir channel explains the observed performance differences among the four
502 sampling zones (i.e., periphery, interior, upper, and lower reservoir). The unbalanced
503 distribution of the singletons and doubletons in the periphery (the location of the reservoir
504 channel) relative to the interior of the reservoir resulted in the underestimation of species
505 richness by the interior reservoir samples. This result is similar to the findings of Hänfling et al.
506 (2016) who detected the greater fish species richness in samples collected closest to the shoreline
507 of a 1480-ha natural lake than in samples collected nearer the center of the lake.

Draft

508 *Sample size effect*
509 Our evaluation of the effect of sample size on our ability to estimate asymptotic species
510 richness in Lawler Pond, under the lowest bioinformatic stringency, suggests that at least 26
511 water samples must be sequenced with eDNA metabarcoding before species richness can be
512 estimated with accuracy and precision, as indicated by the flattening of the curve and decreasing
513 confidence intervals. The number of water samples decreases under the moderate-stringency (19
514 samples) and high-stringency scenarios (14 samples). These estimates of necessary samples
515 apply to Lawler Pond only and may differ from the number of samples needed to estimate
516 species richness in larger and more heterogeneous ecosystems. As noted above, Lawler Pond is
517 a small relatively homogeneous body of water making it likely that eDNA would be evenly

518 distributed. In larger bodies of water with distinct spatial structuring, eDNA may be
519 heterogeneously distributed (Hänfling et al. 2016) and an increased numbers of independent
520 samples may be required to capture the maximum eDNA signal. This outcome is consistent with
521 previous research illustrating that diversity and similarity indices tend to underestimate
522 community similarity when calculated with sample sizes that fail to subsample a relatively large
523 proportion of the community (Lande 1996; Cao et al. 1997). The actual sample size needed to
524 accurately and precisely estimate asymptotic species richness also varies according the diversity
525 of assemblage (Chao et al. 2009). It is likely that had we collected additional samples beyond
526 31, we would have observed greater precision in our species richness estimate. The decrease in
527 the 95% confidence intervals with inclusion of additional samples (e.g., samples 26 to 31 under
528 the low stringency scenario) suggests that additional samples would likely continue to increase
529 the precision of the estimate.

Draft

530 Our study illustrates that eDNA metabarcoding can be an effective means of determining
531 species richness in areas that may be difficult to sample via traditional fish-capture methods.
532 These challenging areas can include military installations, remote wilderness areas, and sensitive
533 sites where traditional sampling approaches such as electrofishing may not be feasible or
534 permitted. Our results demonstrate that eDNA metabarcoding can, relative to capture-based
535 sampling, accurately measure and estimate species richness in a small reservoir. Further, eDNA
536 was relatively homogeneously distributed at the spatial scale of Lawler Pond (i.e., 2.2 ha),
537 suggesting that the number of accumulated samples may be more important than the spatial
538 distribution of samples when attempting to quantify species richness via eDNA metabarcoding in
539 small systems. Moreover, the detection of stream-dwelling species in the impoundment suggests
540 that eDNA can also detect species from water transported into the reservoir via streamflow.

541 Further research on the dynamics of eDNA transport is needed to better understand how
542 downstream transport may affect species richness estimation in impoundments and other
543 downstream habitats.

544 Our results illustrate that the stringency of bioinformatic criteria can have substantial effects
545 on the conclusions about the inferred species richness of the study system. Future research
546 should focus on determining how to optimize the number of markers for estimating species
547 richness via eDNA metabarcoding in diverse ecosystem of varying complexity and size. An
548 improved knowledge of the necessary sample replication would enable the design of more
549 effective and efficient sampling protocols for fish management and conservation. Lastly, while
550 our results illustrate that eDNA metabarcoding can be used to provide robust estimates of species
551 richness, eDNA cannot provide the same types of population structure data that is readily
552 obtained with capture-based methods where fish can be handled and measured. Therefore,
553 eDNA metabarcoding should be viewed as an additional tool in the fisheries professional's
554 sampling toolbox that can provide improved sensitivity for determining species richness rather
555 than a replacement for demographic sampling via capture-based sampling. However, rapidly
556 advancing genetic and genomic technology provides the promise for even greater utility and
557 interpretive power of eDNA data in the future.

Draft

558 **Acknowledgments**

559 Funding for this study was provided by United States Department of Defense Strategic
560 Environmental Research and Development Program Grant W912HQ-12-C-0073 (RC-2240).
561 The authors thank Crysta Gantz and Matthew Dougherty for their assistance with field sampling
562 and eDNA sample processing. The authors also thank Melissa Stephens and the Notre Dame
563 Genomics & Bioinformatics Core Facility for consultation and sequencing. This research was

completed in accordance with University of Notre Dame Institutional Animal Care and Use Protocol #16-015. This article is a publication of the University of Notre Dame Environmental Change Initiative.

Literature Cited

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

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, Michigan, USA.

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.

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

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, USA. pp. 7909-7916.

Draft

586 Colwell, R.K. 2013. EstimateS: Statistical estimation of species richness and shared species from
587 samples. http://purl.oclc.org/estimates.
588 Connell, J.H. 1978. Diversity in Tropical Rain Forests and Coral Reefs. Science **199**: 1302-1310.
589 doi:10.1126/science.199.4335.1302.
590 Deiner, K., and Altermatt, F. 2014. Transport distance of invertebrate environmental DNA in a
591 natural river. Plos One **9**: e88786. doi:10.1371/journal.pone.0088786.
592 Dejean, T., Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P., and Miaud, C.
593 2011. Persistence of environmental DNA in freshwater ecosystems. Plos One **6**: e23398.
594 doi:10.1371/journal.pone.0023398.
595 Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics
596 **26**: 2460-2461. doi:10.1093/bioinformatics/btq461.
597 Evans, N.T., Olds, B.P., Renshaw, M.A., Turner, C.R., Li, Y., Jerde, C.L., Mahon, A.R.,
598 Pfrender, M.E., Lamberti, G.A., and Lodge, D.M. 2016. Quantification of mesocosm fish and
599 amphibian species diversity via environmental DNA metabarcoding. Mol. Ecol. Resour **16**:
600 29-41. doi:10.1111/1755-0998.12433.
601 Ficetola, G.F., Miaud, C., Pompanon, F., and Taberlet, P. 2008. Species detection using
602 enivrionmental DNA from water samples. Biol. Lett. **4**: 423-425.
603 doi:10.1098/rsbl.2008.0118.
604 Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L.,
605 Lopes, C.M., Boyer, F., Pompanon, F., Raye, G., and Taberlet, P. 2015. Replication levels,
606 false presences and the estimation of the presence/absence from eDNA metabarcoding data.
607 Mol. Ecol. Resour **15**: 543-556. doi:10.1111/1755-0998.12338.

Draft

608 George, D.G., and Edwards, R.W. 1976. The Effect of Wind on the Distribution of Chlorophyll
609 A and Crustacean Plankton in a Shallow Eutrophic Reservoir. J. Appl. Ecol. **13**: 667-690.
610 doi:10.2307/2402246.
611 Goldberg, C.S., Pilliod, D.S., Arkle, R.S., and Waits, L.P. 2011. Molecular detection of
612 vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho
613 giant salamanders. Plos One **6**: e22746. doi:10.1371/journal.pone.0022746.
614 Gu, W., and Swihart, R.K. 2004. Absent or undetected? Effects of non-detection of species
615 occurrence on wildlife–habitat models. Biol. Conserv. **116**: 195-203. doi:10.1016/S0006-
616 3207(03)00190-3.
617 Hänfling, B., Lawson Handley, L., Read, D.S., Hahn, C., Li, J., Nichols, P., Blackman, R.C.,
618 Oliver, A., and Winfield, I.J. 2016. Environmental DNA metabarcoding of lake fish
619 communities reflects long-term data from established survey methods. Mol. Ecol.
620 doi:10.1111/mec.13660.
621 Hayes, D.B., Ferreri, C.P., and Taylor, W.W. 1996. Active fish capture methods. *In* Fisheries
622 techniques, 2nd edition. *Edited by* B.R. Murphy and D.W. Willis. American Fisheries
623 Society, Bethesda, Maryland, USA. pp. 193-220.
624 Hilton, J. 1985. A conceptual framework for predicting the occurrence of sediment focusing and
625 sediment redistribution in small lakes. Limnology and Oceanography **30**: 1131-1143.
626 doi:10.4319/lo.1985.30.6.1131.
627 Holyoak, M., Leibold, M.A., Mouquet, N., Holt, R.D., and Hoopes, M.F. 2005.
628 Metacommunities: a framework for large-scale community ecology. *In* Metacommunities:
629 spatial dynamics and ecological communities. *Edited by* M. Holyoak, M.A. Leibold and
630 R.D. Holt. University of Chicago Press, Chicago, IL, USA. pp. 1-32.

Draft

631 Hubbell, S.P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton
632 University Press, Princeton, NJ, USA.
633 Hubert, W.A. 1996. Passive capture techniques. *In* Fisheries techniques, 2nd edition. *Edited by*
634 B.R. Murphy and D.W. Willis. American Fisheries Society, Bethesda, Maryland, USA. pp.
635 157-192.
636 Hubert, W.A., and Quist, M.C. 2010. Inland Fisheries Management in North America, 3rd
637 edition. American Fisheries Society, Bethesda, Maryland, USA.
638 Jane, S.F., Wilcox, T.M., McKelvey, K.S., Young, M.K., Schwartz, M.K., Lowe, W.H., Letcher,
639 B.H., and Whiteley, A.R. 2015. Distance, flow, and PCR inhibition: eDNA dynamics in two
640 headwater streams. Mol. Ecol. Resour **15**: 216-227. doi:10.1111/1755-0998.12285.
641 Jerde, C.L., Mahon, A.R., Chadderton, W.L., and Lodge, D.M. 2011. "Sight-unseen" detection
642 of rare aquatic species using environmental DNA. Conserv. Lett. **4**: 150-157.
643 doi:10.1111/j.1755-263X.2010.00158.x.
644 Lande, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple
645 communities. Oikos **76**: 5-13. doi:10.2307/3545743.
646 Lodge, D.M., Turner, C.R., Jerde, C.L., Barnes, M.A., Chadderton, L., Egan, S.P., Feder, J.L.,
647 Mahon, A.R., and Pfrender, M.E. 2012. Conservation in a cup of water: estimating
648 biodiversity and population abundance from environmental DNA. Mol. Ecol. **21**: 2555-2558.
649 doi:10.1111/j.1365-294X.2012.05600.x.
650 MacArthur, R.H., and Wilson, E.O. 1967. The theory of island biogeography. Princeton
651 University Press, Princeton, NJ, USA.

Draft

652 Mackenzie, D.I., and Royle, J.A. 2005. Designing occupancy studies: general advice and
653 allocating survey effort. J. Appl. Ecol. **42**: 1105-1114. doi:10.1111/j.1365-
654 2664.2005.01098.x.
655 McDonald, L.L. 2004. Sampling rare populations. *In* Sampling Rare or Elusive Species. *Edited*
656 *by* W.L. Thompson. Island Press, Washington, D.C., USA. pp. 11-42.
657 Munch, K., Boomsma, W., Huelsenbeck, J.P., Willerslev, E., and Nielsen, R. 2008. Statistical
658 assignment of DNA sequences using Bayesian phylogenetics. Systematic Biol. **57**: 750-757.
659 doi:10.1080/10635150802422316.
660 Murphy, B.R., and Willis, D.W. 1996. Fisheries Techniques, 2nd edition. American Fisheries
661 Society, Bethesda, MD, USA.
662 Nguyen, N.H., Smith, D., Peay, K., and Kennedy, P. 2015. Parsing ecological signal from noise
663 in next generation amplicon sequencing. New Phytologist **205**: 1389-1393.
664 doi:10.1111/nph.12923.
665 Olds, B.P., Jerde, C.L., Renshaw, M.A., Li, Y., Evans, N.T., Turner, C.R., Deiner, K., Mahon,
666 A.R., Brueseke, M.A., Shirey, P.D., Pfrender, M.E., Lodge, D.M., and Lamberti, G.A. 2016.
667 Estimating species richness using environmental DNA. Ecol. Evol. **6**: 4214-4226.
668 doi:10.1002/ece3.2186.
669 Peterson, J.T., and Paukert, C.P. 2009. Converting nonstandard fish sampling data to
670 standardized data. *In* Standard methods for sampling North American freshwater fishes.
671 *Edited by* S.A. Bonar, W.A. Hubert and D.W. Willis. American Fisheries Society, Bethesda,
672 Maryland, USA. pp. 195-215.
673 Prince, A.M., and Andrus, L. 1992. PCR: how to kill unwanted DNA. Biotechniques **12**: 358-
674 360.

Draft

675 Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R.M., Gough, K.C., and Crispo, E.
676 2014. The detection of aquatic animal species using environmental DNA - a review of eDNA
677 as a survey tool in ecology. J. Appl. Ecol. **51**: 1450-1459. doi:10.1111/1365-2664.12306.
678 Renshaw, M.A., Olds, B.P., Jerde, C.L., McVeigh, M.M., and Lodge, D.M. 2015. The room
679 temperature preservation of filtered environmental DNA samples and assimilation into a
680 phenol–chloroform–isoamyl alcohol DNA extraction. Mol. Ecol. Resour **15**: 168-176.
681 Schloss, P.D., Gevers, D., and Westcott, S.L. 2011. Reducing the effects of PCR amplification
682 and sequencing artifacts on 16S rRNA-based studies. Plos One **6**: e27310.
683 Shaw, J.L.A., Clarke, L.J., Wedderburn, S.D., Barnes, T.C., Weyrich, L.S., and Cooper, A. 2016.
684 Comparison of environmental DNA metabarcoding and conventional fish survey methods in
685 a river system. Biol. Conserv. **197**: 131-138. doi:10.1016/j.biocon.2016.03.010.
686 Spigel, R.H., and Imberger, J. 1987. Mixing processes relevant to phytoplankton dynamics in
687 lakes. New Zealand Journal of Marine and Freshwater Research **21**: 361-377.
688 doi:10.1080/00288330.1987.9516233.
689 Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., and Kawabata, Z.i. 2012. Estimation of fish
690 biomass using environmental DNA. Plos One **7**: e35868. doi:10.1371/journal.pone.0035868.
691 Thomsen, P.F., Kielgast, J.O.S., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P.,
692 Orlando, L., and Willerslev, E. 2012. Monitoring endangered freshwater biodiversity using
693 environmental DNA. Mol. Ecol. **21**: 2565-2573. doi:10.1111/j.1365-294X.2011.05418.x.
694 Thomsen, P.F., and Willerslev, E. 2015. Environmental DNA – An emerging tool in
695 conservation for monitoring past and present biodiversity. Biol. Conserv. **183**: 4-18.
696 doi:10.1016/j.biocon.2014.11.019.

Draft

697 Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E.,
698 Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp,
699 G.H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.M., Peroux, T., Crivelli, A.J., Olivier, A.,
700 Acqueberge, M., Le Brun, M., Moller, P.R., Willerslev, E., and Dejean, T. 2016. Next-
701 generation monitoring of aquatic biodiversity using environmental DNA metabarcoding.
702 Mol. Ecol. **25**: 929-942. doi:10.1111/mec.13428.
703 Wheeler, T.J., and Eddy, S.R. 2013. nhmmer: DNA homology search with profile HMMs.
704 Bioinformatics **29**: 2487-2489.
705 Wilcox, T.M., McKelvey, K.S., Young, M.K., Jane, S.F., Lowe, W.H., Whiteley, A.R., and
706 Schwartz, M.K. 2013. Robust detection of rare species using environmental DNA: the
707 importance of primer specificity. Plos One **8**: e59520. doi:10.1371/journal.pone.0059520.
708 William, C.B. 1964. Patterns in the balance of nature: and related problems in quantitative
709 ecology. Academic Press, London, UK.
710
711

Draft

Table 1. 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). Black blocks indicate species detected via traditional sampling and/or eDNA metabarcoding. Gray blocks indicate eDNA metabarcoding false negative detections (i.e., species captured via traditional sampling but not detected with eDNA). White blocks indicate species not detected with either traditional sampling or eDNA metabarcoding.

Species	Capture-based	Low	Moderate	High
American Pickerel (*Esox americanus*)	X	X	X	X
Blackchin Shiner (*Notropis heterodon*)	X	X	X	
Bluegill Sunfish (*Lepomis macrochirus*)	X	X	X	X
Bluntnose Minnow (*Pimephales notatus*)		X	X	
Brook Trout (*Salvelinus fontinalis*)		X		
Brown Trout (*Salmo trutta*)		X		
Central Mudminnow (*Umbra limi*)	X	X	X	X
Channel Catfish (*Ictalurus punctatus*)		X		
Common Carp (*Cyprinus carpio*)	X	X	X	X
Creek Chub (*Semotilus atromaculatus*)		X	X	
Green Sunfish (*Lepomis cyanellus*)	X	X	X	
Iowa Darter (*Etheostoma exile*)		X	X	X
Johnny Darter (*Etheostoma nigrum*)		X		
Lake Chubsucker (*Erimyzon sucetta*)		X	X	
Largemouth Bass (*Micropterus salmoides*)	X	X	X	X
Least Darter (*Etheostoma microperca*)		X	X	
Mottled Sculpin (*Cottus bairdii*)		X		
Pumpkinseed Sunfish (*Lepomis gibbosus*)	X	X	X	X
Warmouth Sunfish (*Lepomis gulosus*)	X	X	X	X
White Sucker (*Catostomus commersonii*)		X		
Yellow Bullhead (*Ameiurus natalis*)	X	X	X	
Cumulative Species Richness	*10*	*21*	*15*	*8*

719 Table 2. Species identified with OTU species assignment for each marker under the low bioinformatic stringency scenario. Primary
720 habitats for each species were identified based on information available at www.natureserve.org.

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

721 **Figure Captions**

722 Fig. 1. Conceptual diagram illustrating the relationship between bioinformatic
723 stringency and strength of certainty about the presence of eDNA metabarcoding-
724 detected species.

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

Draft

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

737 Fig. 4. Mean species accumulation curve (eDNA detected; grey circles) and mean Chao II
738 species richness estimator curve (Chao estimated; black diamonds) derived from rarefaction
739 analysis of the 31 Lawler Pond eDNA samples libraries under the (a) low stringency scenario,
740 (b) moderate stringency scenario, and (c) high stringency scenario. Error bars represent 95%
741 confidence intervals.

742 Fig. 5. Euclidean distance (m) between eDNA water samples versus Sørensen dissimilarity (D_s).
743 Each point represents one of the 435 pairwise comparisons between all 30 Lawler Pond samples
744 (upstream sample excluded) under the low-stringency scenario. The dashed line in each plot,
745 illustrates the generally expected negative relationship (slope < 0) if sample dissimilarity were
746 predicted by distance, however, no significant relationship was found between Euclidean
747 distance and D_s (Mantel's r = -0.06, $P = 0.79$).

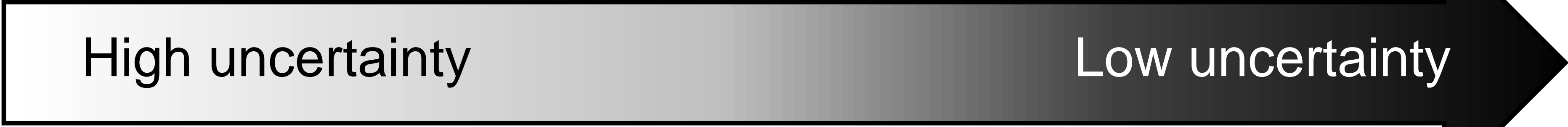
748 Fig. 6. Mean Chao II species richness estimator curves derived from rarefaction
749 analysis of the eDNA samples selected via each of the four 15-sample spatial designs
750 (upper, lower, periphery, interior) and from a randomly-selected subset of all 30
751 available samples (random) under the (a) low-stringency scenario, (b) moderate-
752 stringency scenario, and (c) high-stringency scenario. Error bars represent 95%
753 confidence intervals of the randomly-selected samples.

Draft

754 Fig. S1. Species detection by sample for all three markers combined using the low
755 stringency bioinformatic criteria.

756 Fig. S2. Species detection by sample for all three markers combined using the
757 moderate stringency bioinformatic criteria.

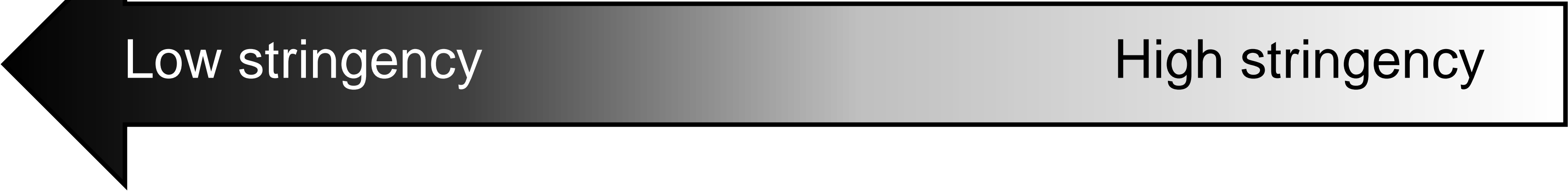
758 Fig. S3. Species detection by sample for all three markers combined using the high
759 stringency bioinformatic criteria.

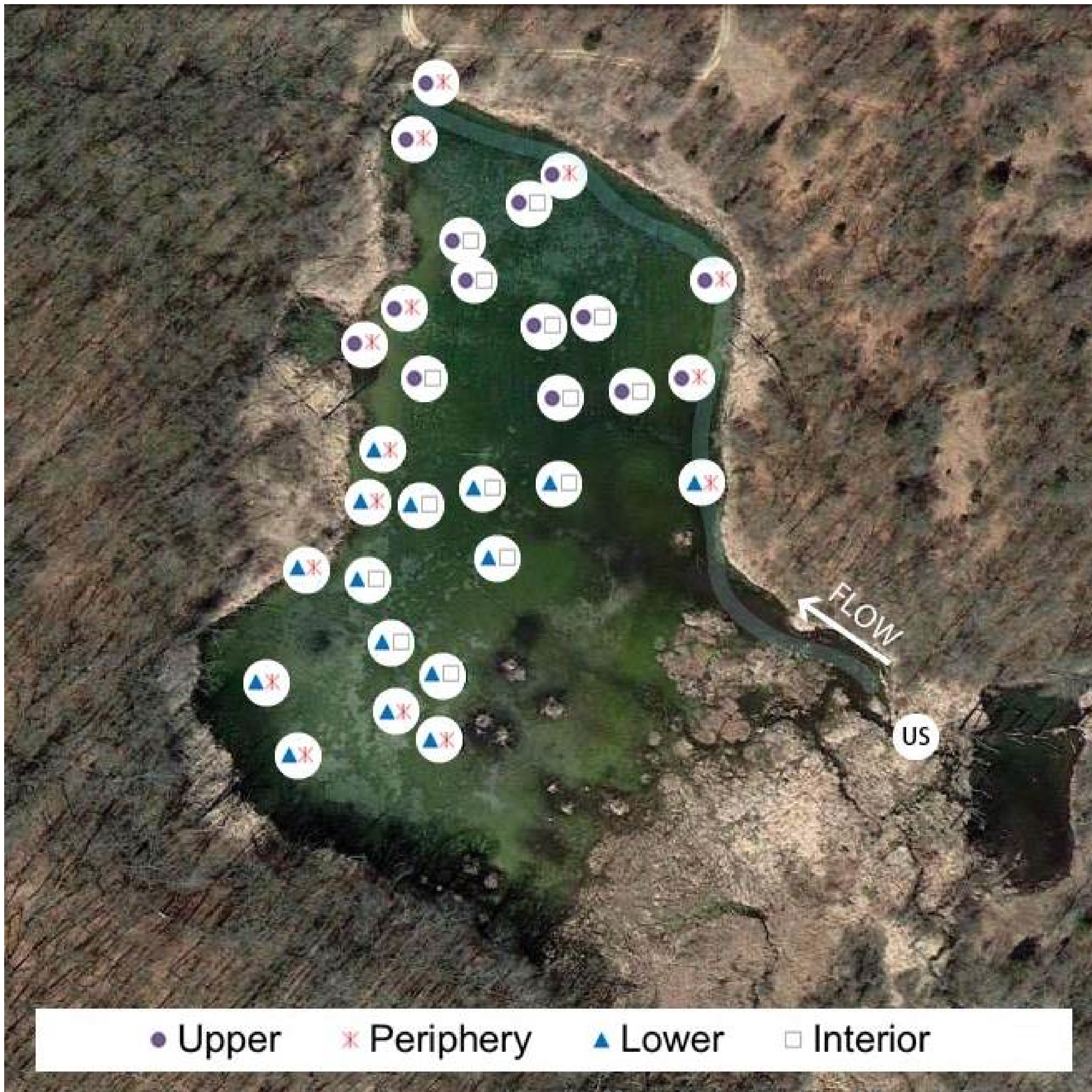


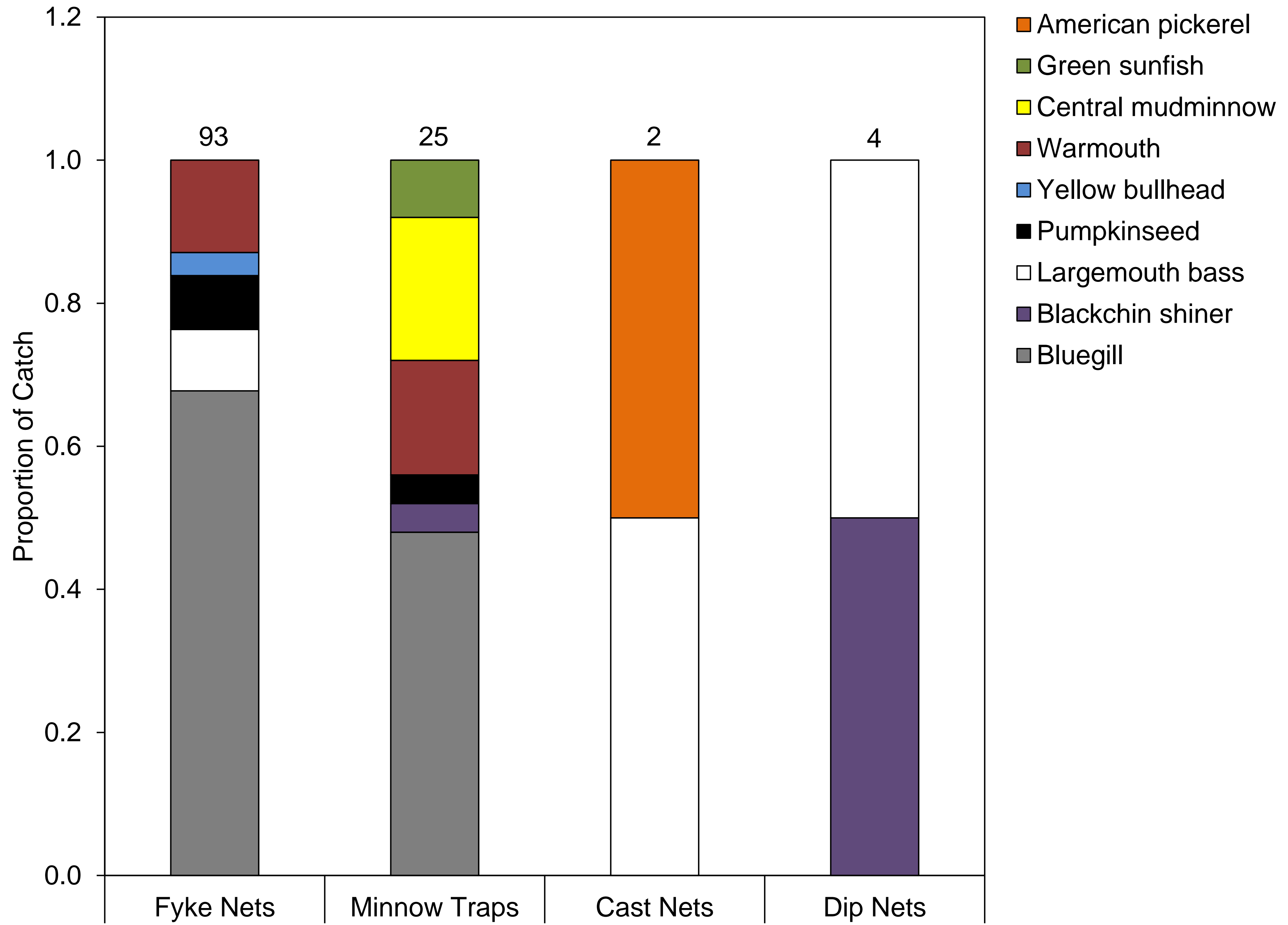
- Detected by a single marker
- Detected in a single sample

- Detected by multiple markers OR in multiple samples

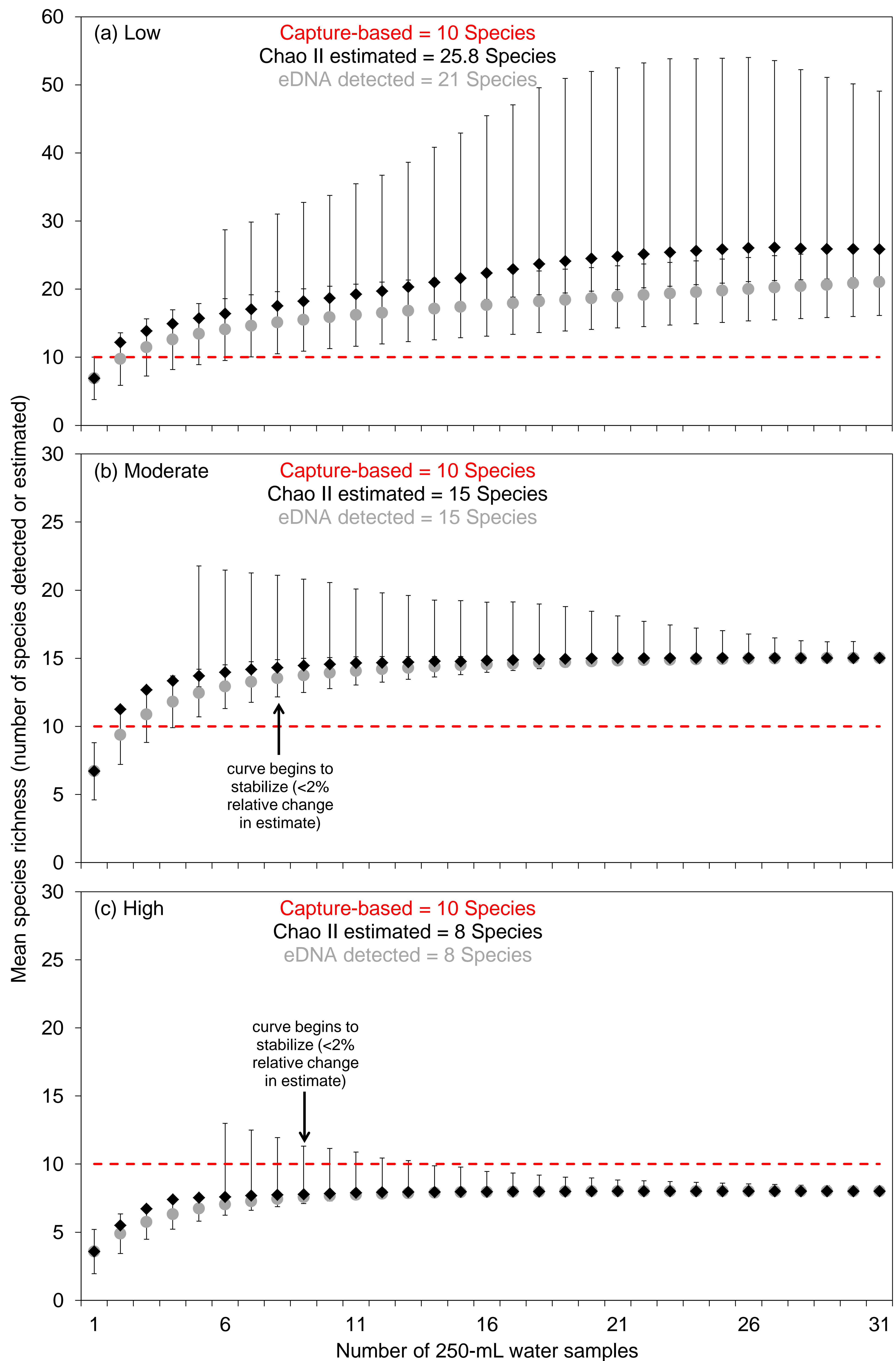
- Detected by multiple markers AND in multiple samples
- Greatest certainty when species physically captured

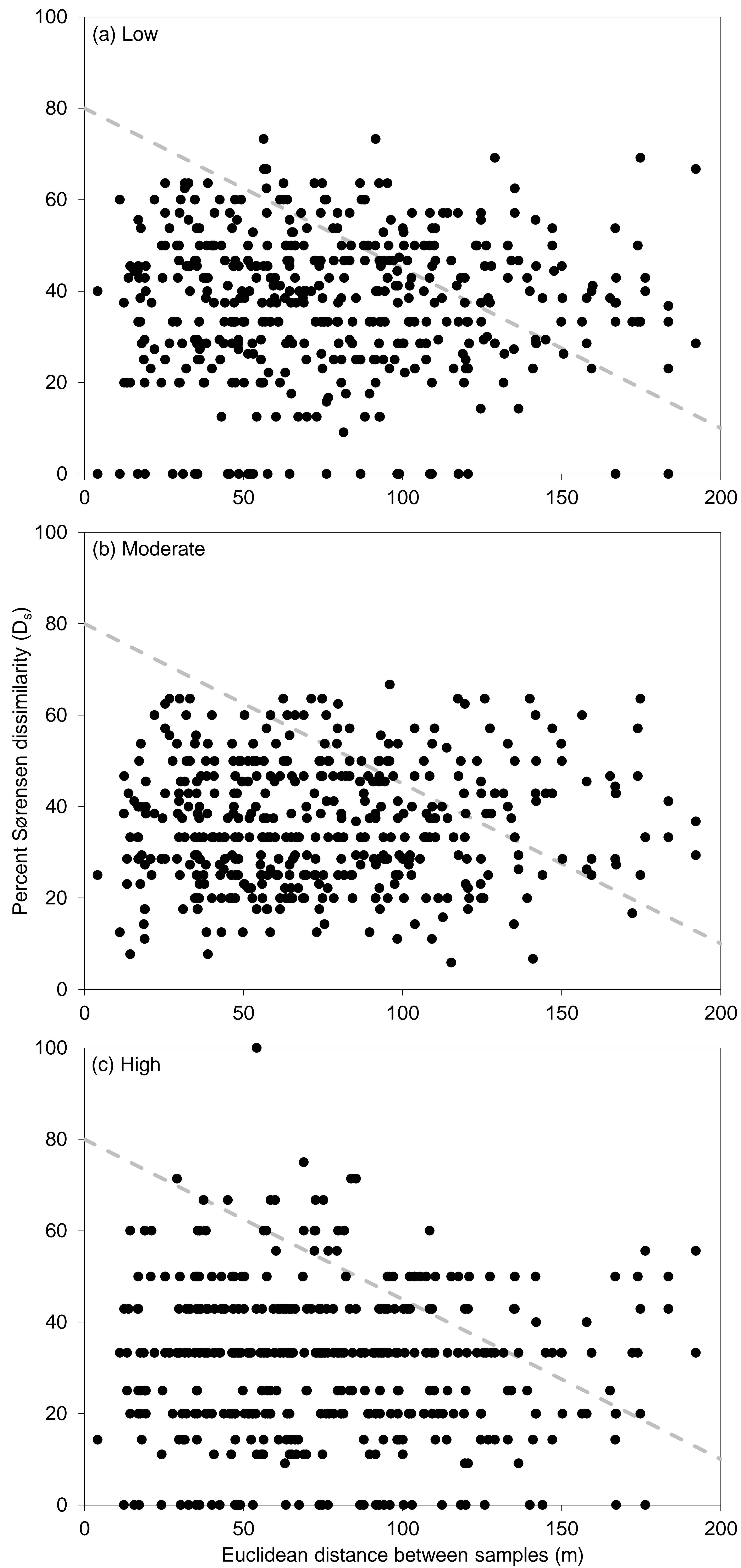


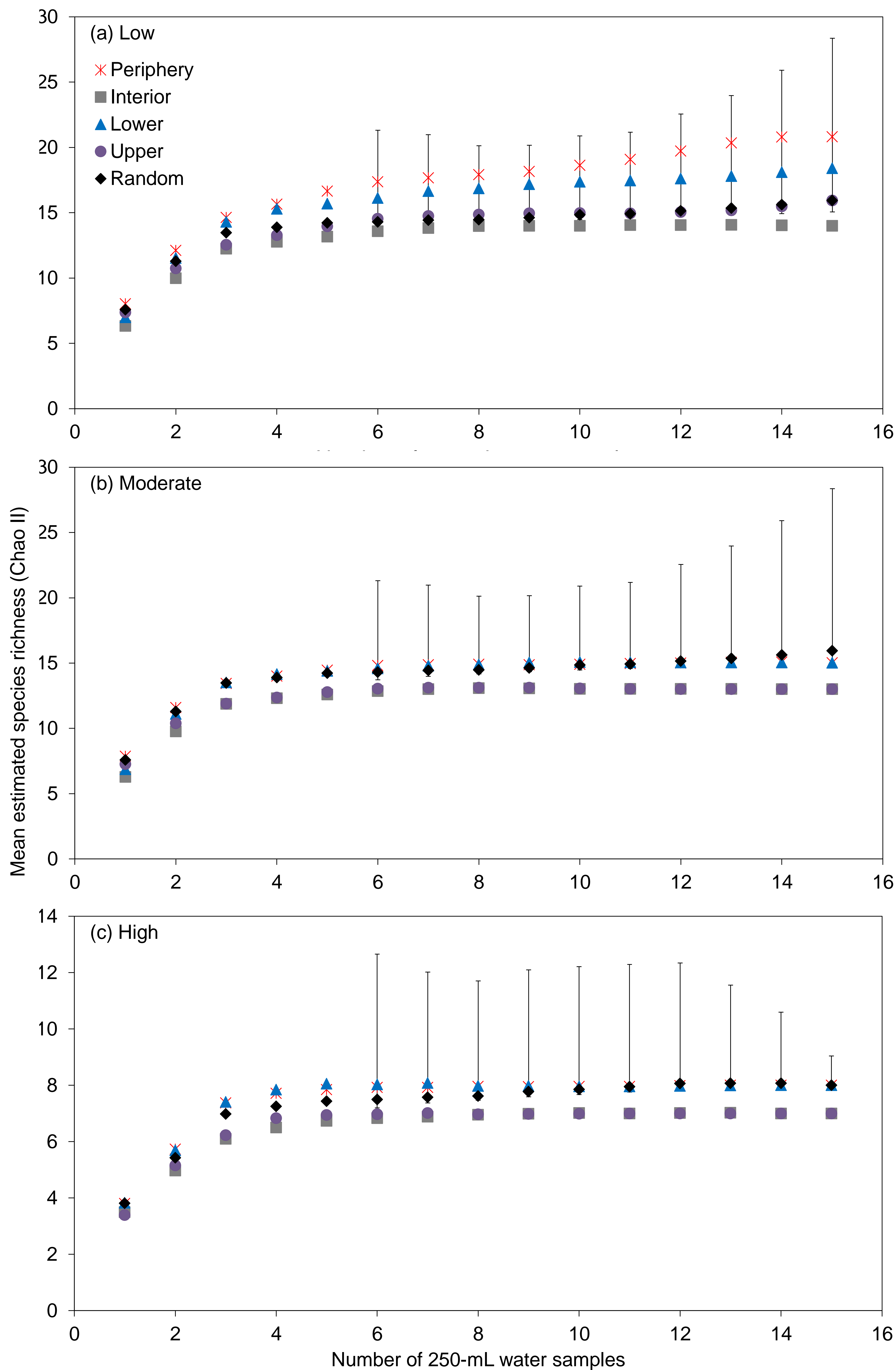

FLOW
US
Upper
Periphery
Lower
Interior

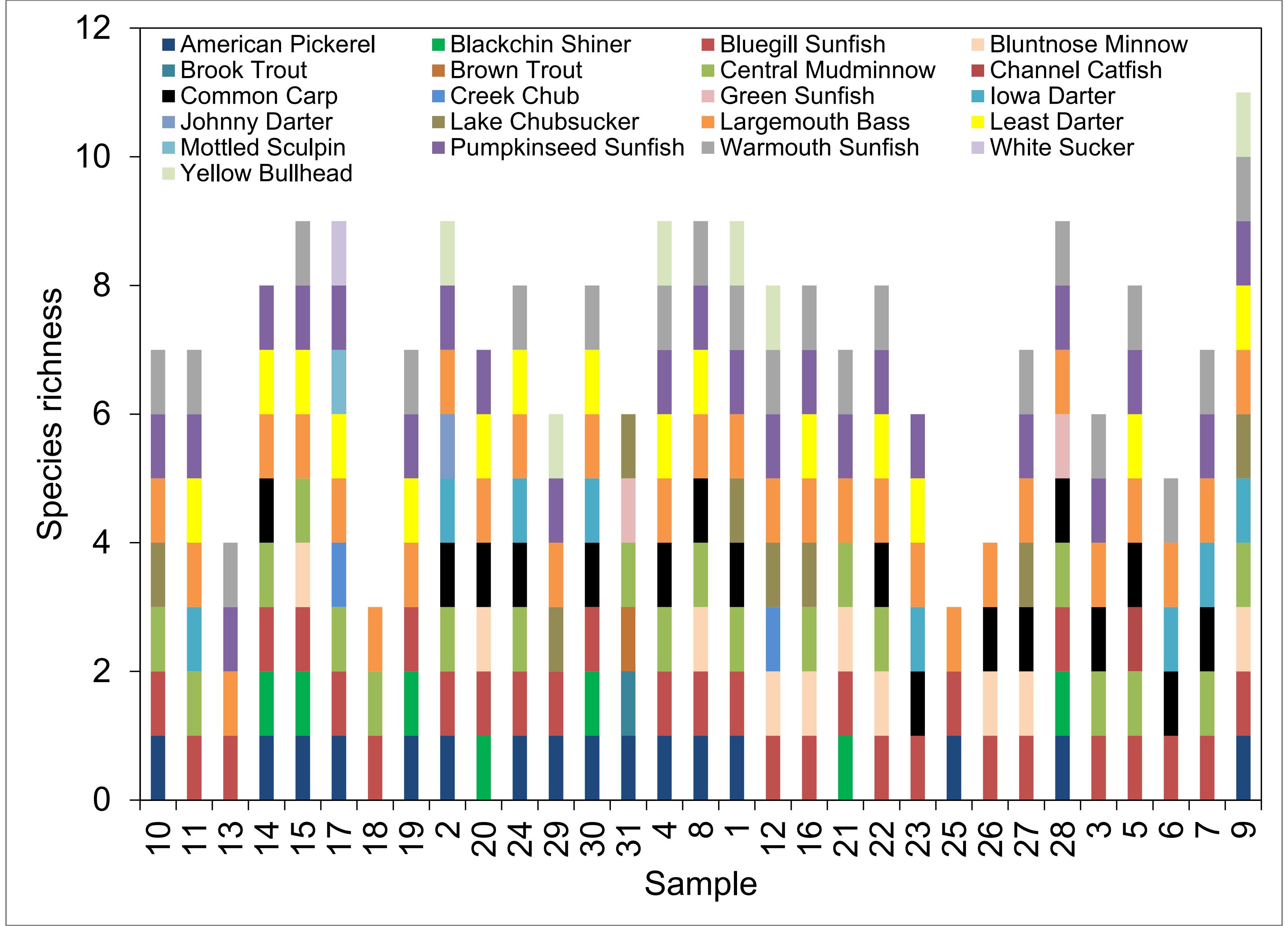


1.2
1.0
0.8
0.6
0.4
0.2
0.0
Proportion of Catch
93
25
2
4
Fyke Nets
Minnow Traps
Cast Nets
Dip Nets
American pickerel
Green sunfish
Central mudminnow
Warmouth
Yellow bullhead
Pumpkinseed
Largemouth bass
Blackchin shiner
Bluegill

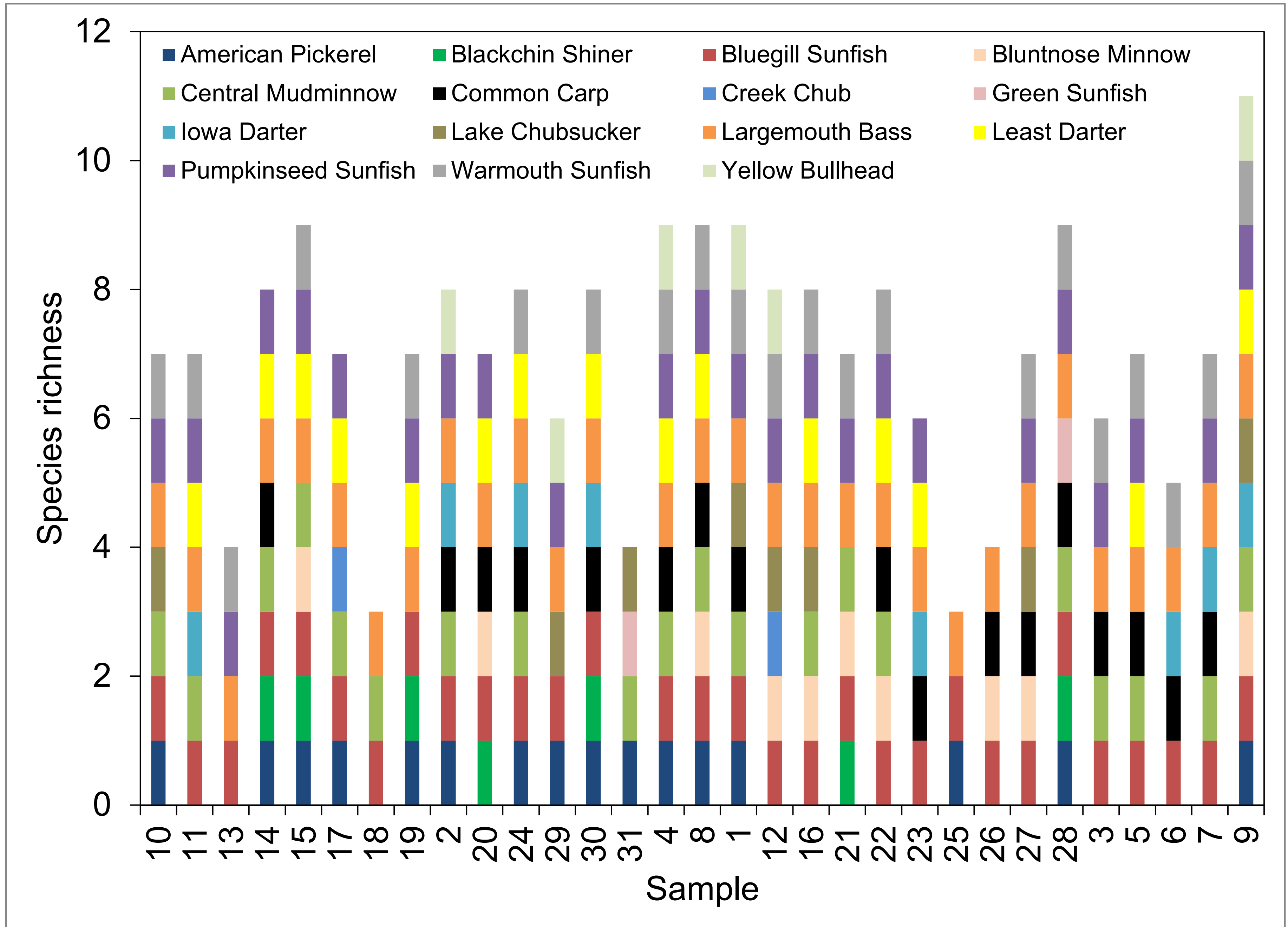

(a) Low
Capture-based = 10 Species
Chao II estimated = 25.8 Species
eDNA detected = 21 Species
(b) Moderate
Capture-based = 10 Species
Chao II estimated = 15 Species
eDNA detected = 15 Species
curve begins to stabilize (<2% relative change in estimate)
(c) High
Capture-based = 10 Species
Chao II estimated = 8 Species
eDNA detected = 8 Species
curve begins to stabilize (<2% relative change in estimate)
Mean species richness (number of species detected or estimated)
Number of 250-mL water samples

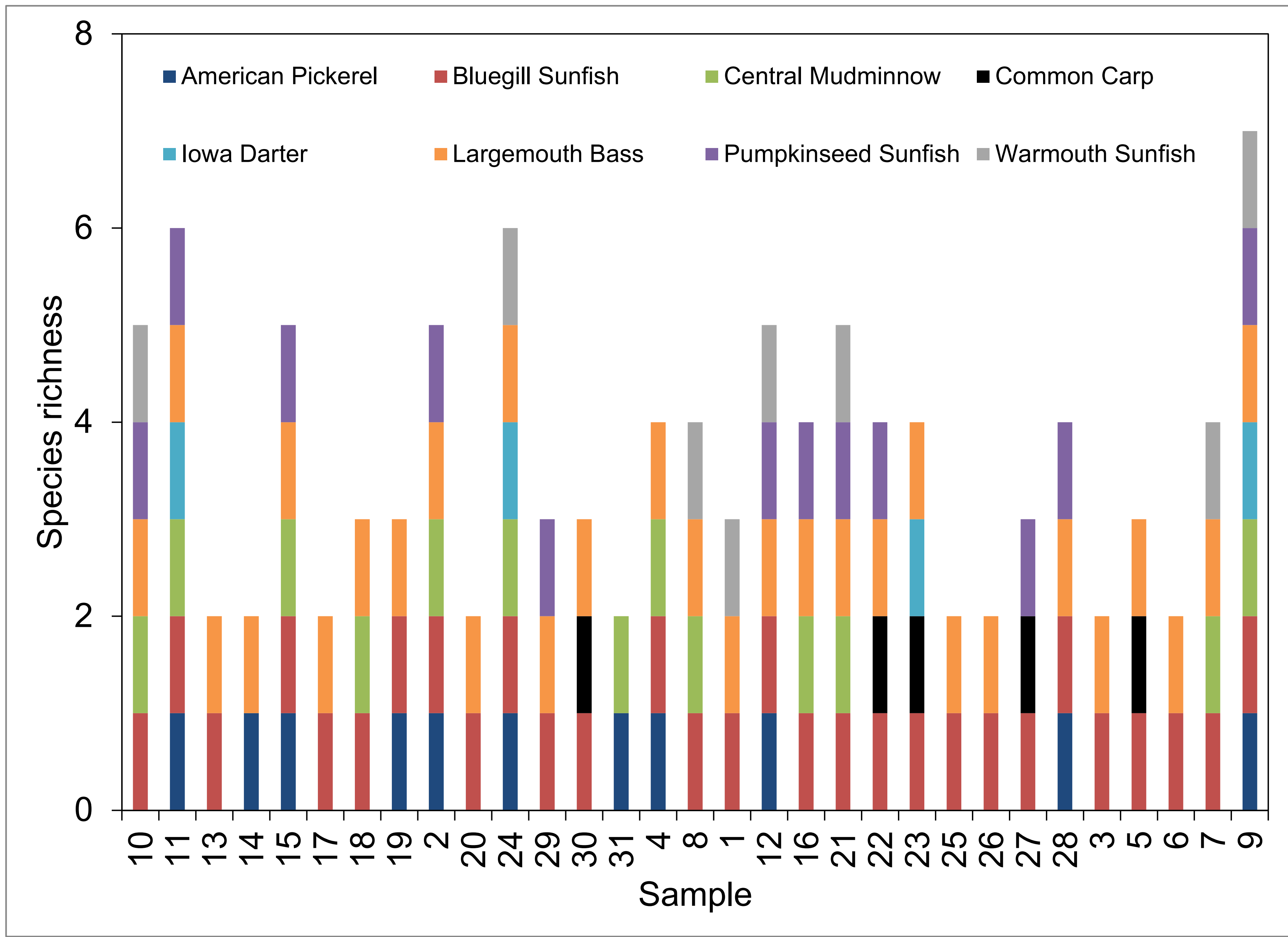

(a) Low
(b) Moderate
(c) High
Percent Sørensen dissimilarity (D_s)
Euclidean distance between samples (m)
0
20
40
60
80
100
0
50
100
150
200


(a) Low
Periphery
Interior
Lower
Upper
Random
(b) Moderate
(c) High
Mean estimated species richness (Chao II)
Number of 250-mL water samples



American Pickerel
Blackchin Shiner
Bluegill Sunfish
Bluntnose Minnow
Brook Trout
Brown Trout
Central Mudminnow
Channel Catfish
Common Carp
Creek Chub
Green Sunfish
Iowa Darter
Johnny Darter
Lake Chubsucker
Largemouth Bass
Least Darter
Mottled Sculpin
Pumpkinseed Sunfish
Warmouth Sunfish
White Sucker
Yellow Bullhead
Species richness
0
2
4
6
8
10
12
10
11
13
14
15
17
18
19
2
20
24
29
30
31
4
8
1
12
16
21
22
23
25
26
27
28
3
5
6
7
9
Sample


American Pickerel
Blackchin Shiner
Bluegill Sunfish
Bluntnose Minnow
Central Mudminnow
Common Carp
Creek Chub
Green Sunfish
Iowa Darter
Lake Chubsucker
Largemouth Bass
Least Darter
Pumpkinseed Sunfish
Warmouth Sunfish
Yellow Bullhead
Species richness
0
2
4
6
8
10
12
Sample
10 11 13 14 15 17 18 19 2 20 24 29 30 31 4 8 1 12 16 21 22 23 25 26 27 28 3 5 6 7 9


American Pickerel
Bluegill Sunfish
Central Mudminnow
Common Carp
Iowa Darter
Largemouth Bass
Pumpkinseed Sunfish
Warmouth Sunfish
Species richness
0
2
4
6
8
Sample
10
11
13
14
15
17
18
19
2
20
24
29
30
31
4
8
1
12
16
21
22
23
25
26
27
28
3
5
6
7
9

Supplemental Table S1. Bioinformatic filtering steps used in eDNA metabarcode studies on fish communities

			Bioinformatic filtering steps								
Study	Habitat type	Sample type	Read quality control	Primer trimming	Merged reads	Abundance removal	Clustering of OTUs	Removal of non-target sequences	Taxonomic assignment method	Post-filter for contamination	Genetic markers used	# sequences to establish presence in a sample post filtering
Valentini et al. 2015	River, stream, pond	varied by habitat see referance	Average quality: none stated Minimum read length: 20	2 mis-matches	yes, no min overlap length given	less than 10 reads & reads assessed as PCR error based on OBICLEAN	all samples run independent, 98% similarity to reference sequence	assignment less than 98% similarity with self-made references and public database (EMBL)	ECOTAG using self-made reference database from EMBL	sequences with frequency of occurance at 0.3% per taxon (calculated as proportion of all sequence reads in the sample)	12S	none reported less than 12 sequences
Shaw et al. 2015	River, wetland	30 cm below surface	Average quality: 20 phred Minimum read length: 80 bp	100% match	no parameters given	less than 20 reads	all samples run independent, 97% similarity	assignment bit score below 120 or less than 98% similar with top hit in NCBI nr/nt database	blast using NCBI's nucleotide database	no parameters given	12S, 16S	none reported less than 20 sequences for either gene
Hänfling et al. 2016	Lake	surface water, & at depths of 2, 10 & 20 m	Average quality: 30 phred Minimum read length: 90 bp for 12S & 100 bp for CytB	yes, no parameters given	no parameters given	less than 3 reads & chimeric removal using usearch	unstated pooled or separate, 100% similarity	assignemnt bit score below 80 or less than 100% (12S) /95% (CytB) similarity with self-currated database	lowest common ancester approach using self-made database from NCBI	sequences with frequency of occurance at 0.1% (12s) and 0.2% (CytB) per taxon and based on mock community false positives (calculated as proportion of all sequence reads in the sample)	12S, CytB	greater than zero sequences in either gene
Olds et al. 2016	River	surface water	Average quality: 20 phred Minimum read length: 50 bp	100% match	overlap: 16 bp	singletons	all samples pooled, 97% similarity	HMMER filter using self-made reference database for each gene	SAP using NCBI's nucleotide database	based on probability (p = 0.95) of presence using statistical model built from appearance of reads in negative controls	16S, 12S, CytB	greater than two seqeunces in two markers

Draft

Supplemental Table S2. List of species included in the in-house reference sequence database. Reference sequences taken from previously existing GenBank records are highlighted in blue; reference sequences generated in-house are highlighted in green.

Species	16S	12S	Cyt B
Acris_crepitans	AY843559	AY843559	
Acris_crepitans_blanchardi			EF988145
Amboplites_rupestris	KM282459	KM282394	KM523260
Ambystoma_laterale	NC006330	NC006330	NC006330
Ambystoma_maculatum			KM523263
Ambystoma_tigrinum	NC006887	NC006887	NC006887
Ameiurus_natalis	AY458872		AY184265
Amphiprion_ocellaris	NC009065	NC009065	NC009065
Bufo_americanus	AY680206	AY680206	AF171190
Bufo_fowleri	AY680224	AY680224	
Catostomus_commersonii	KM282461	KM282400	KM523268
Centropyge_bispinosa	NC028287	NC028287	NC028287
Cottus_bairdii	KM282462	KM282401	KM523269
Cyprinus_carpio	KM282467	KM282406	KM523272
Ecsenius_bicolor	NC028295	NC028295	NC028295
Erimyzon_sucetta	KM282468	KM282408	KM523274
Esox_americanus_vermiculatus			AY497430
Etheostoma_caeruleum	KM282469	KM282409	KM523275
Etheostoma_exile	KM282471	KM282411	KM523277
Etheostoma_nigrum	KM282474	KM282412	KM523280
Etheostoma_radiosum	NC005254	NC005254	NC005254
Hemidactylium_scutatum	DQ283120	DQ283120	NC006342
Hyla_chrysoscelis	EF566949	EF566949	
Hyla_versicolor	AY843682	AY843682	AY843928
Lepomis_cyanellus	KM282484	KM282423	KP013087
Lepomis_gibbosus	KM282485	KM282424	KM523290
Lepomis_gulosus	AY742526		
Lepomis_macrochirus	KM282486	KM282426	KM523292
Lepomis_megalotis	AY742533		AY828977
Lepomis_microlophus	AY742535		JF742834
Macropharyngodon_negrosensis	NC028289	NC028289	NC028289
Micropterus_dolomieu	NC011361	KM282429	KM523294
Micropterus_salmoides	KM282489	KM282430	KM523295
Necturus_maculosus			DQ283412
Necturus_maculosus_maculosus	KM282431	KM523296	
Notemigonus_crysoleucus	KM282490	KM282432	KM523297
Notophthalmus_viridescens	EU880323	EU880323	EU880323
Notropis_anogenus			KF744334
Notropis_heterodon	KM282491	KM282434	KM523298
Notropis_stramineus	KM282492	NC008110	KM523299
Oncorhynchus_mykiss	KM282499	KM282441	KM523306
Perca_flavescens	KM282501	KM282443	KM523308
Phoxinus_eos	NC015364	NC015364	NC015364
Pimephales_notatus	AY216556	AY216556	U66606
Pimephales_promelas	KM282503	KM282445	KM523310
Plethodon_cinereus_cinereus	NC006343	NC006343	NC006343
Pomoxis_nigromaculatus	AY742557	KM282446	KM523311
Pseudacris_crucifer			AY210883
Pseudacris_crucifer_crucifer	AY843735	AY843735	
Pseudacris_triseriata	AY843738	AY843738	KJ536224
Pseudanthias_dispar	NC028286	NC028286	NC028286
Rana_catesbeiana	KM282504	NC022696	KM523312
Rana_clamitans	KM282506	DQ283185	KM523314
Rana_palustris	AY779228		
Rana_pipiens	DQ283123	DQ283123	
Rana_sylvatica	DQ283387	DQ283387	AY083271
Rhinichthys_atratulus	AF038495		
Rhinichthys_obtusus	KM282509	KM282447	JX442984
Salarias_fasciatus	AP004451	AP004451	AP004451
Salmo_trutta	KM282510	KM282448	KM523316
Semotilus_atromaculatus	KM282512	AF023199	KM523318
Umbra_limi	KM282516	KM282453	KM523322
Umbra_pygmaea	NC022456	NC022456	NC022456

Supplemental Table S3. Relative frequency distributions of species reads in each of the 31 Lawler Pond samples for each of the three markers. Species with relative frequencies less than that of the cumulative relative frequency of contaminant reads in the negative control libraries were discarded (red). Only species with relative frequencies greater than the cumulative relative frequency of negative control reads were retained (green).

16S

Species	MC	EB	NTC	Sum of MC, EB, NTC	Cumulative Relative Frequency of Negative Control Reads (Threshold)	FC27_10	FC27_11	FC27_13	FC27_14	FC27_15	FC27_17	FC27_18	FC27_19	FC27_2	FC27_20	FC27_24	FC27_29	FC27_30	FC27_31	FC27_4	FC27_8	FC28_1	FC28_12	FC28_16	FC28_21	FC28_22	FC28_23	FC28_25	FC28_26	FC28_27	FC28_28	FC28_3	FC28_5	FC28_6	FC28_7	FC28_9
Mottled Sculpin (*Cottus bairdii*)	0	0	2	2	0.037%	0.000%	0.000%	0.000%	0.000%	0.000%	16.667%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Green Sunfish (*Lepomis cyanellus*)	0	0	4	4	0.075%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.427%	0.000%	0.000%	0.000%	0.000%	0.000%
Lake Chubsucker (*Erimyzon sucetta*)	0	0	0	0	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.015%
Rainbow Trout (*Oncorhynchus mykiss*)	0	0	2	2	0.037%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Brown Trout (*Salmo trutta*)	0	0	124	124	2.321%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	85.043%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Bluegill Sunfish (*Lepomis macrochirus*)	21	5	3384	3410	63.822%	31.414%	26.870%	61.685%	4.410%	78.269%	44.444%	69.399%	67.374%	58.409%	1.162%	90.335%	0.075%	80.776%	0.460%	51.099%	90.764%	0.044%	52.925%	82.448%	13.926%	75.619%	95.471%	90.122%	77.394%	95.656%	95.077%	49.903%	45.932%	47.241%	46.890%	54.099%
Largemouth Bass (*Micropterus salmoides*)	0	3	237	240	4.492%	65.699%	49.751%	35.266%	90.611%	7.714%	0.000%	26.188%	22.155%	13.648%	34.512%	1.134%	0.007%	3.229%	0.046%	41.314%	5.569%	99.949%	6.991%	2.092%	11.340%	9.921%	1.399%	7.392%	19.755%	0.433%	1.893%	46.978%	52.878%	50.541%	0.357%	3.849%
Pumpkinseed Sunfish (*Lepomis gibbosus*)	10	0	936	946	17.705%	0.002%	0.000%	2.681%	0.023%	0.962%	0.000%	0.036%	2.035%	17.342%	23.465%	0.643%	99.912%	0.400%	0.196%	0.128%	0.000%	0.000%	32.855%	8.328%	0.028%	7.531%	0.002%	2.467%	0.000%	0.099%	1.644%	1.796%	0.439%	0.921%	0.021%	2.406%
American Pickerel (*Esox americanus*)	0	0	30	30	0.561%	0.013%	15.875%	0.000%	3.978%	4.807%	0.000%	0.012%	8.381%	2.860%	0.063%	1.066%	0.001%	0.000%	13.279%	1.793%	0.000%	0.004%	1.977%	0.153%	9.824%	0.160%	0.935%	0.000%	0.000%	0.000%	0.000%	0.231%	0.081%	0.000%	0.003%	17.625%
Central Mudminnow (*Umbra limi*)	0	0	59	59	1.104%	1.169%	2.583%	0.089%	0.943%	1.564%	0.000%	4.310%	0.013%	3.718%	0.051%	5.499%	0.002%	0.238%	0.893%	5.487%	1.497%	0.000%	0.280%	2.075%	51.333%	0.070%	0.000%	0.005%	0.000%	0.000%	0.203%	1.093%	0.055%	0.309%	14.998%	1.410%
Bluntnose Minnow (*Pimephales notatus*)	0	0	43	43	0.805%	0.004%	0.000%	0.000%	0.006%	6.684%	0.000%	0.036%	0.042%	0.000%	40.747%	0.000%	0.001%	0.433%	0.009%	0.003%	1.016%	0.003%	2.759%	4.887%	9.911%	2.098%	0.001%	0.000%	2.851%	0.892%	0.223%	0.000%	0.282%	0.582%	0.003%	3.888%
Common Carp (*Cyprinus carpio*)	2	0	454	456	8.535%	1.688%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	2.874%	0.000%	0.000%	0.002%	14.750%	0.075%	0.000%	0.745%	0.000%	0.000%	0.000%	0.000%	4.380%	0.002%	0.000%	0.000%	2.920%	0.000%	0.000%	0.000%	0.000%	37.728%	0.001%
Yellow Bullhead (*Ameiurus natalis*)	0	0	2	2	0.037%	0.012%	0.005%	0.000%	0.019%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	1.247%	0.002%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	15.624%
Iowa Darter (*Etheostoma exile*)	0	0	13	13	0.243%	0.000%	4.913%	0.000%	0.008%	0.000%	0.000%	0.000%	0.000%	1.149%	0.000%	1.147%	0.001%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	2.191%	0.014%	0.000%	0.000%	0.000%	0.000%	0.229%	0.396%	0.000%	1.082%
Warmouth Sunfish (*Lepomis gulosus*)	0	0	6	6	0.112%	0.000%	0.000%	0.279%	0.000%	0.000%	0.000%	0.020%	0.000%	0.000%	0.000%	0.176%	0.000%	0.174%	0.000%	0.175%	0.409%	0.000%	0.950%	0.015%	3.638%	0.220%	0.000%	0.000%	0.000%	0.000%	0.534%	0.000%	0.104%	0.010%	0.000%	0.000%
Creek Chub (*Semotilus atromaculatus*)	0	0	0	0	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	22.222%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.007%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
White Sucker (*Catostomus commersonii*)	0	0	6	6	0.112%	0.000%	0.003%	0.000%	0.000%	0.000%	16.667%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
				Total Reads	**5343**	**108424**	**78696**	**17904**	**125503**	**374311**	**18**	**84485**	**30761**	**173792**	**25472**	**110181**	**243948**	**167393**	**58687**	**57152**	**438252**	**74357**	**41784**	**230794**	**63334**	**153547**	**198410**	**36677**	**36437**	**23318**	**54311**	**54911**	**43264**	**40399**	**57729**	**210083**

12S

Common name	MC	EB	NTC	Sum of MC, EB, NTC	Cumulative Relative Frequency of Negative Control Reads (Threshold)	FC27_10	FC27_11	FC27_13	FC27_14	FC27_15	FC27_17	FC27_18	FC27_19	FC27_2	FC27_20	FC27_24	FC27_29	FC27_30	FC27_31	FC27_4	FC27_8	FC28_1	FC28_12	FC28_16	FC28_21	FC28_22	FC28_23	FC28_25	FC28_26	FC28_27	FC28_28	FC28_3	FC28_5	FC28_6	FC28_7	FC28_9
Human (*Homo sapien*)	**0**	**0**	**0**	**0**	0.000%	3.254%	0.000%	0.000%	0.000%	0.000%	7.607%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.005%	0.000%	0.000%	0.011%	0.000%	0.039%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.023%	0.000%
Mottled Sculpin (*Cottus bairdii*)	**2**	**0**	**3**	**5**	1.259%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Blackchin Shiner (*Notropis heterodon*)	**0**	**0**	**0**	**0**	0.000%	0.000%	0.000%	0.000%	0.002%	0.004%	0.000%	0.000%	0.007%	0.000%	12.653%	0.000%	0.000%	0.004%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.048%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Bluegill Sunfish (*Lepomis macrochirus*)	**0**	**2**	**285**	**287**	72.292%	50.865%	19.316%	62.034%	10.107%	70.313%	0.064%	58.320%	45.528%	65.042%	24.904%	72.838%	63.957%	88.386%	0.013%	52.075%	87.313%	0.562%	63.844%	80.832%	44.549%	86.467%	85.273%	86.083%	84.158%	94.795%	85.651%	28.591%	42.960%	41.227%	42.101%	61.169%
Largemouth Bass (*Micropterus salmoides*)	**0**	**0**	**65**	**65**	16.373%	37.861%	39.893%	36.903%	75.505%	20.641%	92.305%	36.501%	30.713%	23.462%	14.287%	18.699%	35.142%	9.187%	0.009%	41.603%	10.251%	72.686%	9.101%	10.850%	25.259%	3.810%	5.083%	13.909%	15.829%	2.637%	5.034%	66.064%	54.244%	50.110%	22.009%	7.659%
American Pickerel (*Esox americanus*)	**0**	**0**	**14**	**14**	3.526%	1.376%	15.367%	7.063%	14.380%	1.780%	0.016%	4.430%	15.420%	2.309%	48.100%	0.000%	0.010%	2.007%	43.126%	3.887%	0.011%	0.057%	3.105%	0.000%	3.203%	0.663%	0.000%	0.000%	0.000%	0.062%	4.118%	0.000%	0.002%	0.000%	0.108%	13.837%
Pumpkinseed Sunfish (*Lepomis gibbosus*)	**0**	**0**	**6**	**6**	1.511%	2.759%	4.828%	0.000%	0.004%	3.793%	0.004%	0.000%	4.290%	9.182%	0.005%	0.000%	0.000%	0.000%	0.000%	0.005%	0.826%	0.051%	18.364%	6.061%	3.589%	7.067%	3.874%	0.000%	0.000%	1.805%	5.137%	3.434%	0.786%	1.422%	0.016%	7.446%
Central Mudminnow (*Umbra limi*)	**0**	**0**	**5**	**5**	1.259%	0.540%	11.804%	0.000%	0.002%	3.470%	0.013%	0.742%	0.310%	0.006%	0.008%	0.120%	0.005%	0.227%	14.455%	2.421%	0.623%	0.008%	0.000%	7.208%	23.295%	0.366%	0.000%	0.000%	0.000%	0.000%	0.000%	0.308%	0.087%	0.000%	20.466%	2.520%
Warmouth Sunfish (*Lepomis gulosus*)	**0**	**0**	**6**	**6**	1.511%	3.032%	0.617%	0.000%	0.000%	0.000%	0.000%	0.000%	3.534%	0.000%	0.000%	3.550%	0.000%	0.457%	0.000%	0.002%	0.014%	26.640%	2.896%	0.702%	0.000%	0.777%	0.000%	0.000%	0.000%	0.555%	0.000%	1.003%	1.655%	5.848%	15.174%	6.282%
Green Sunfish (*Lepomis cyanellus*)	**0**	**0**	**3**	**3**	0.756%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.003%	0.000%	0.005%	0.000%	42.029%	0.000%	0.000%	0.000%	0.000%	0.000%	0.003%	0.000%	0.000%	0.000%	0.000%	0.138%	0.000%	0.450%	0.244%	0.000%	0.000%	0.000%
Iowa Darter (*Etheostoma exile*)	**0**	**0**	**6**	**6**	1.511%	0.007%	8.775%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.007%	4.643%	0.000%	1.526%	0.000%	0.000%	0.005%	0.008%	0.000%	0.000%	0.010%	0.644%	5.617%	0.000%	0.000%	0.004%	0.000%	0.000%	0.000%	1.392%	0.043%	0.010%
Lake Chubsucker (*Erimyzon sucetta*)	**0**	**0**	**0**	**0**	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.006%	0.678%	0.347%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	1.076%
Common Carp (*Cyprinus carpio*)	**0**	**0**	**0**	**0**	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.144%	0.000%	0.000%	0.039%	0.000%	0.000%	0.000%	0.000%	0.010%	0.145%	0.000%	0.000%	0.004%	0.000%	0.000%	0.023%	0.000%	0.000%	0.000%
				Total Reads	**397**	**29653**	**42130**	**44406**	**160506**	**55051**	**112575**	**56588**	**30020**	**36126**	**59909**	**41522**	**41056**	**69546**	**45402**	**40805**	**61955**	**35608**	**53070**	**99050**	**67002**	**30577**	**37870**	**93276**	**71466**	**80168**	**92756**	**20674**	**117329**	**77158**	**30446**	**19245**

Cyt B

Common name	MC	EB	NTC	Sum of MC, EB, NTC	Cumulative Relative Frequency of Negative Control Reads (Threshold)	FC27_10	FC27_11	FC27_13	FC27_14	FC27_15	FC27_17	FC27_18	FC27_19	FC27_2	FC27_20	FC27_24	FC27_29	FC27_30	FC27_31	FC27_4	FC27_8	FC28_1	FC28_12	FC28_16	FC28_21	FC28_22	FC28_23	FC28_25	FC28_26	FC28_27	FC28_28	FC28_3	FC28_5	FC28_6	FC28_7	FC28_9
Human (*Homo sapien*)	0	9457	0	9457	99.474%	0.004%	0.002%	0.000%	0.002%	0.032%	0.027%	10.445%	0.000%	0.044%	0.003%	0.002%	0.004%	0.000%	0.000%	0.000%	0.007%	0.074%	63.886%	0.000%	2.418%	3.857%	0.149%	4.112%	0.000%	0.003%	0.007%	57.826%	0.003%	0.002%	0.003%	0.014%
Spiny Softshell Turtle (*Apalone spinifera*)	0	0	0	0	0.000%	10.294%	0.000%	0.000%	0.000%	0.000%	0.005%	0.000%	0.000%	10.439%	0.002%	0.000%	96.351%	0.000%	0.000%	0.000%	0.004%	0.015%	0.002%	0.000%	4.294%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	2.051%	0.000%	0.000%	0.003%	0.010%
Common Snapping Turtle (*Chelydra serpentina*)	0	0	0	0	0.000%	0.000%	0.000%	0.000%	0.002%	0.000%	0.069%	0.004%	13.165%	0.000%	0.001%	0.000%	0.000%	0.008%	14.490%	0.000%	1.151%	0.000%	0.000%	0.005%	0.000%	0.000%	0.000%	0.000%	0.000%	3.201%	0.004%	0.000%	0.000%	0.000%	0.197%	0.000%
Channel Catfish (*Ictalurus punctatus*)	0	0	0	0	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.407%	0.000%	0.000%	0.000%
Johnny Darter (*Etheostoma nigrum*)	0	0	0	0	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.002%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Blackchin Shiner (*Notropis heterodon*)	0	0	0	0	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.007%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.116%	0.000%	0.000%	0.000%	0.000%	0.000%
Bluegill Sunfish (*Lepomis macrochirus*)	0	15	13	28	0.295%	25.952%	33.824%	72.529%	0.051%	72.911%	0.463%	36.239%	28.016%	14.376%	76.668%	76.632%	1.617%	72.171%	0.003%	31.205%	71.995%	69.769%	36.027%	77.094%	12.733%	91.686%	97.712%	95.805%	99.999%	94.570%	94.793%	11.596%	94.186%	56.293%	85.842%	42.358%
Central Mudminnow (*Umbra limi*)	0	4	5	9	0.095%	40.114%	59.711%	0.000%	79.910%	20.425%	0.113%	53.193%	0.000%	65.258%	0.012%	13.183%	0.004%	0.004%	14.488%	19.980%	3.159%	7.341%	0.007%	15.612%	78.105%	1.576%	0.002%	0.000%	0.000%	0.000%	2.423%	6.149%	1.146%	0.000%	0.023%	44.118%
Least Darter (*Etheostoma microperca*)	0	0	2	2	0.021%	0.003%	4.982%	0.000%	19.689%	4.174%	98.257%	0.004%	18.406%	0.005%	20.830%	9.916%	0.000%	0.482%	0.000%	2.076%	1.238%	0.015%	0.000%	2.481%	0.009%	0.639%	1.638%	0.000%	0.000%	0.000%	0.000%	0.000%	1.059%	0.000%	0.003%	7.423%
Largemouth Bass (*Micropterus salmoides*)	0	0	0	0	0.000%	2.436%	0.040%	25.950%	0.126%	0.083%	0.983%	0.115%	29.178%	4.478%	2.480%	0.198%	0.012%	0.605%	0.000%	3.663%	21.795%	0.059%	0.069%	0.131%	0.200%	1.610%	0.433%	0.068%	0.000%	0.000%	0.028%	22.342%	0.801%	0.004%	4.913%	0.695%
Lake Chubsucker (*Erimyzon sucetta*)	0	2	2	4	0.042%	20.280%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.046%	0.012%	34.759%	0.000%	0.003%	0.015%	0.000%	0.000%	0.005%	0.000%	0.000%	0.000%	0.000%	0.411%	0.007%	0.000%	0.000%	0.002%	0.005%	0.026%
Common Carp (*Cyprinus carpio*)	0	0	0	0	0.000%	0.000%	0.000%	0.000%	0.002%	0.000%	0.000%	0.000%	0.000%	0.025%	0.002%	0.001%	0.000%	26.679%	0.000%	0.017%	0.000%	0.096%	0.000%	0.000%	0.000%	0.195%	0.002%	0.000%	0.001%	1.454%	0.005%	0.005%	1.122%	43.700%	0.000%	0.000%
American Pickerel (*Esox americanus*)	0	0	0	0	0.000%	0.019%	0.000%	0.000%	0.216%	0.003%	0.072%	0.000%	11.235%	3.198%	0.000%	0.053%	1.941%	0.003%	10.431%	14.064%	0.006%	0.015%	0.000%	0.000%	0.000%	0.000%	0.000%	0.015%	0.000%	0.000%	0.004%	0.000%	0.000%	0.000%	0.000%	0.006%
Yellow Bullhead (*Ameiurus natalis*)	0	0	0	0	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.007%	0.000%	0.000%	0.008%	0.000%	0.000%	27.641%	0.000%	0.030%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Pumpkinseed Sunfish (*Lepomis gibbosus*)	0	0	0	0	0.000%	0.244%	0.419%	1.521%	0.003%	2.232%	0.006%	0.000%	0.000%	2.167%	0.000%	0.000%	0.013%	0.000%	0.000%	1.354%	0.283%	22.401%	0.005%	4.434%	1.927%	0.437%	0.000%	0.000%	0.000%	0.348%	2.600%	0.000%	1.276%	0.000%	1.095%	3.239%
Warmouth Sunfish (*Lepomis gulosus*)	0	0	0	0	0.000%	0.653%	0.617%	0.000%	0.000%	0.139%	0.000%	0.000%	0.000%	0.000%	0.000%	0.016%	0.000%	0.000%	0.000%	0.000%	0.358%	0.171%	0.000%	0.243%	0.309%	0.000%	0.000%	0.000%	0.000%	0.012%	0.000%	0.030%	0.000%	0.000%	7.899%	2.065%
Iowa Darter (*Etheostoma exile*)	0	0	0	0	0.000%	0.000%	0.404%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.063%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.016%	0.047%
Mottled Sculpin (*Cottus bairdii*)	0	5	0	5	0.053%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.001%	0.000%	0.000%	0.001%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.003%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Brook Trout (*Salvelinus fontinalis*)	0	2	0	2	0.021%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.007%	25.806%	0.000%	0.000%	0.000%	0.002%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.003%	0.000%	0.000%	0.000%	0.000%	0.000%
				Total Reads	**9507**	**178519**	**181892**	**219867**	**152183**	**71940**	**66504**	**199839**	**181792**	**98190**	**178636**	**342184**	**52076**	**134497**	**192439**	**97349**	**122538**	**13486**	**117335**	**41230**	**43015**	**38009**	**120060**	**197772**	**165028**	**125219**	**75740**	**62885**	**146761**	**130431**	**114954**	**72532**

Draft

Supplemental Table S4. Number of reads from each sample (i.e. library) run on the Illumina MiSeq platform for each step of the bioinformatic pipeline.

	FC27_10	FC27_11	FC27_13	FC27_14	FC27_15	FC27_17	FC27_18	FC27_19	FC27_2	FC27_20	FC27_24	FC27_29
Raw data	784,956	900,597	753,690	1,046,078	1,129,135	727,933	819,859	746,354	664,130	1,095,999	1,199,104	804,502
Demultiplexing rate	83.85%	76.39%	83.91%	77.72%	85.09%	48.85%	79.61%	75.65%	102.25%	43.68%	76.52%	63.37%
Merging rate	66.60%	56.82%	61.91%	64.27%	69.18%	37.11%	62.46%	56.73%	84.75%	34.55%	61.38%	52.12%
Quality Filtering rate	48.01%	37.99%	42.20%	49.54%	54.53%	26.25%	44.83%	37.70%	65.23%	24.66%	44.22%	42.24%
Ac16S												
Matching F primer	215,579	150,035	37,727	256,692	662,166	56	137,672	67,325	388,728	43,045	172,985	392,763
Matching R primer	214,515	147,782	38,003	261,186	654,917	52	136,632	67,339	387,325	40,856	172,288	381,403
Paired reads	207,645	142,377	36,419	247,947	636,596	50	132,327	64,821	375,438	38,488	166,883	352,068
Merged reads	187,576	128,028	33,140	223,296	578,318	43	121,571	58,860	342,297	33,220	152,738	290,669
After quality filtering	155,651	103,439	27,505	185,266	472,026	32	101,430	47,362	285,689	27,188	125,923	244,052
Am12S												
Matching F primer	39,868	56,984	63,254	206,834	76,676	149,441	72,734	52,013	49,949	71,691	55,067	50,892
Matching R primer	39,185	54,897	61,513	202,239	74,644	152,725	75,398	50,347	48,759	70,592	53,012	49,310
Paired reads	38,101	53,786	60,383	198,391	73,201	141,389	69,684	49,184	47,902	68,840	52,157	48,412
Merged reads	37,427	52,498	59,159	194,304	71,845	138,062	68,430	48,060	47,032	66,983	51,103	47,550
After quality filtering	34,462	47,937	54,274	178,312	65,916	124,350	63,070	43,656	43,417	60,967	46,696	43,600
L14735/H15149c												
Matching F primer	440,755	556,935	572,524	392,877	267,705	234,563	479,756	485,664	273,016	397,305	743,829	116,809
Matching R primer	445,937	534,303	557,553	432,645	261,579	311,274	469,718	503,232	280,491	388,196	726,974	249,787
Paired reads	412,428	491,837	535,594	366,684	250,978	214,159	450,698	450,592	255,723	371,361	698,485	109,306
Merged reads	297,779	331,198	374,306	254,725	131,016	132,055	322,083	316,516	173,514	278,456	532,126	81,119
After quality filtering	186,776	190,797	236,303	154,685	77,820	66,701	203,003	190,364	104,076	182,165	357,629	52,140

	FC27_30	FC27_31	FC27_4	FC27_8	FC28_1	FC28_12	FC28_16	FC28_21	FC28_22	FC28_23	FC28_25	FC28_26	FC28_27	FC28_28
Raw data	758,534	801,170	743,727	1,202,782	873,977	768,830	1,013,274	720,539	665,194	791,803	808,073	684,380	630,655	699,014
Demultiplexing rate	84.55%	78.69%	78.00%	80.78%	57.32%	62.65%	70.52%	60.61%	67.07%	69.37%	67.75%	66.16%	71.57%	70.46%
Merging rate	73.75%	59.69%	53.93%	70.86%	37.86%	46.67%	57.89%	42.26%	57.47%	61.01%	58.86%	57.53%	56.22%	50.20%
Quality Filtering rate	58.67%	40.25%	32.61%	56.24%	27.45%	30.82%	46.13%	29.27%	45.89%	47.11%	44.79%	43.03%	39.60%	36.19%
Ac16S														
Matching F primer	279,326	94,738	132,018	669,389	294,594	91,961	462,694	133,637	362,785	327,496	84,684	77,214	42,044	95,961
Matching R primer	278,453	89,047	130,020	648,082	254,421	80,185	404,796	115,988	315,219	280,834	73,445	65,791	36,784	83,822
Paired reads	269,369	86,325	125,814	628,291	245,756	77,580	390,483	109,359	303,339	272,475	71,156	63,572	35,405	81,085
Merged reads	248,941	76,308	109,025	576,861	222,296	71,339	360,944	99,805	278,687	254,669	65,911	58,959	32,683	75,415
After quality filtering	210,392	61,649	78,095	476,956	183,102	59,194	298,723	81,640	230,928	211,222	54,207	48,492	26,448	62,422
Am12S														
Matching F primer	98,025	59,380	65,836	80,295	54,307	78,953	161,161	96,142	43,104	53,251	144,267	107,412	119,344	136,231
Matching R primer	95,026	57,995	63,188	77,112	46,326	66,134	134,439	83,509	36,451	44,994	121,220	89,545	99,677	114,266
Paired reads	93,289	56,800	61,966	75,427	45,352	65,016	132,182	79,205	35,766	44,064	118,955	88,030	97,858	112,153
Merged reads	91,730	55,243	60,327	74,015	42,967	64,087	130,260	77,741	35,236	43,406	117,377	86,963	96,465	110,472
After quality filtering	85,033	50,738	53,260	67,574	39,548	59,841	120,851	72,070	32,672	40,366	109,212	80,961	89,188	102,837
L14735/H15149c														
Matching F primer	297,284	514,874	422,762	288,157	262,044	436,134	242,396	310,105	135,877	295,465	447,816	382,929	402,625	373,921
Matching R primer	289,602	508,388	429,371	320,331	260,019	392,632	225,524	287,973	132,598	256,126	371,371	313,398	330,021	310,768
Paired reads	278,697	487,329	392,351	267,841	209,889	339,107	191,916	248,190	107,023	232,726	357,379	301,187	318,114	299,314
Merged reads	218,727	346,675	231,732	201,398	65,605	223,360	95,394	126,970	68,393	185,017	292,310	247,771	225,395	165,050
After quality filtering	149,615	210,110	111,190	131,871	17,282	117,935	47,800	57,158	41,685	121,425	198,510	165,052	134,130	87,680

	FC28_3	FC28_5	FC28_6	FC28_7	FC28_9		FC28_32 Extraction Blank	FC27_33 Mock Community	FC28_34 PCR Negative Control	Total Number of Reads After Quality Filtering	Unique Reads	Unique Reads (abundance>1)	# OTUs before HMMer	# OTUs after HMMer
Raw data	605,841	1,029,107	827,812	827,521	974,189	26,098,759	343,922	617,978	3,197,888					
Demultiplexing rate	67.11%	66.42%	62.39%	62.86%	70.58%		11.57%	81.46%	18.79%					
Merging rate	46.26%	51.12%	50.17%	45.78%	54.55%		6.99%	62.90%	2.79%					
Quality Filtering rate	27.56%	37.83%	36.21%	31.00%	39.99%		3.27%	44.94%	1.63%					
Ac16S										268,945	114,943	14,985	44	39
Matching F primer	127,034	100,455	90,950	125,602	442,880		67	48,681	397,447					
Matching R primer	110,591	87,748	79,425	108,782	383,555		63	42,257	371,435					
Paired reads	106,617	84,716	76,565	105,251	372,544		51	40,635	291,858					
Merged reads	96,216	77,910	68,617	96,935	343,749		39	37,403	88,272					
After quality filtering	76,512	64,991	55,900	79,850	286,556	4,422,842	24	31,319	51,117					
Am12S										504,913	143,813	27,466	30	25
Matching F primer	34,980	210,251	135,086	59,932	28,387		4	104,572	36,437					
Matching R primer	29,334	177,360	113,450	50,881	24,005		3	97,438	37,028					
Paired reads	28,767	174,277	111,597	49,776	23,439		3	95,624	29,626					
Merged reads	28,212	171,433	109,533	48,747	23,076		3	93,057	958					
After quality filtering	25,755	158,286	100,747	45,071	21,633	2,162,300	3	85,022	921					
L14735/H15149c										1,114,112	81,442	21,860	44	43
Matching F primer	347,347	550,192	426,639	457,312	364,471		50,916	389,829	334,948					
Matching R primer	285,421	466,563	343,447	428,590	303,734		131,897	386,369	295,364					
Paired reads	271,222	424,515	328,299	365,128	291,609		39,739	367,169	279,421					
Merged reads	155,859	276,691	237,126	233,135	164,577		23,988	258,250	103					
After quality filtering	64,686	166,001	143,075	131,644	81,356	4,180,664	11,210	161,388	50					
						0.412502602				185,400	26,112	5,839	18	18

Draft

Supplemental Table S5. Species detection by sample for all three markers combined using the low stringency bioinformatic criteria. The 31 Lawler Pond samples libraries were randomly divided among two Illumina MiSeq runs: FC27 and FC28.

	Sample Library																													
Detected Species	**FC27_10**	**FC27_11**	**FC27_13**	**FC27_14**	**FC27_15**	**FC27_17**	**FC27_18**	**FC27_19**	**FC27_2**	**FC27_20**	**FC27_24**	**FC27_29**	**FC27_30**	**FC27_31**	**FC27_4**	**FC27_8**	**FC28_1**	**FC28_12**	**FC28_16**	**FC28_21**	**FC28_22**	**FC28_23**	**FC28_25**	**FC28_26**	**FC28_27**	**FC28_28**	**FC28_3**	**FC28_5**	**FC28_6**	**FC28_7**	**FC28_9**
American Pickerel (*Esox americanus*)	X			X	X	X		X	X		X	X	X	X	X	X	X						X			X					X
Blackchin Shiner (*Notropis heterodon*)				X	X			X		X			X							X						X					
Bluegill Sunfish (*Lepomis macrochirus*)	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Bluntnose Minnow (*Pimephales notatus*)					X					X						X		X	X	X	X			X	X						X
Brook Trout (*Salvelinus fontinalis*)														X																	
Brown Trout (*Salmo trutta*)														X																	
Central Mudminnow (*Umbra limi*)	X	X		X	X	X	X		X		X			X	X	X	X		X	X	X					X	X	X		X	X
Channel Catfish (*Ictalurus punctatus*)																												X			
Common Carp (*Cyprinus carpio*)				X					X	X	X		X		X	X	X				X	X		X	X	X	X	X	X	X	
Creek Chub (*Semotilus atromaculatus*)						X												X													
Green Sunfish (*Lepomis cyanellus*)														X												X					
Iowa Darter (*Etheostoma exile*)		X							X		X		X									X							X	X	X
Johnny Darter (*Etheostoma nigrum*)									X																						
Lake Chubsucker (*Erimyzon sucetta*)	X											X		X			X	X	X						X						X
Largemouth Bass (*Micropterus salmoides*)	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Least Darter (*Etheostoma microperca*)		X		X	X	X		X		X	X		X		X	X			X		X	X						X			X
Mottled Sculpin (*Cottus bairdii*)						X																									
Pumpkinseed Sunfish (*Lepomis gibbosus*)	X	X	X	X	X	X		X	X	X		X			X	X	X	X	X	X	X	X			X	X	X	X		X	X
Warmouth Sunfish (*Lepomis gulosus*)	X	X	X		X			X			X		X		X	X	X	X	X	X	X				X	X	X	X	X	X	X
White Sucker (*Catostomus commersonii*)						X																									
Yellow Bullhead (*Ameiurus natalis*)									X			X			X		X	X													X
Total Species Richness	*7*	*7*	*4*	*8*	*9*	*9*	*3*	*7*	*9*	*7*	*8*	*6*	*8*	*6*	*9*	*9*	*9*	*8*	*8*	*7*	*8*	*6*	*3*	*4*	*7*	*9*	*6*	*8*	*5*	*7*	*11*

Draft

Supplemental Table S6. Species detection by sample for all three markers combined using the moderate stringency bioinformatic criteria. The 31 Lawler Pond samples libraries were randomly divided among two Illumina MiSeq runs: FC27 and FC28.

	Sample Library																														
Detected Species	**FC27_10**	**FC27_11**	**FC27_13**	**FC27_14**	**FC27_15**	**FC27_17**	**FC27_18**	**FC27_19**	**FC27_2**	**FC27_20**	**FC27_24**	**FC27_29**	**FC27_30**	**FC27_31**	**FC27_4**	**FC27_8**	**FC28_1**	**FC28_12**	**FC28_16**	**FC28_21**	**FC28_22**	**FC28_23**	**FC28_25**	**FC28_26**	**FC28_27**	**FC28_28**	**FC28_3**	**FC28_5**	**FC28_6**	**FC28_7**	**FC28_9**
American Pickerel (*Esox americanus*)	X			X	X	X		X	X		X	X	X	X	X	X	X						X			X					X
Blackchin Shiner (*Notropis heterodon*)				X	X			X		X			X							X						X					
Bluegill Sunfish (*Lepomis macrochirus*)	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Bluntnose Minnow (*Pimephales notatus*)					X					X						X		X	X	X	X			X	X						X
Central Mudminnow (*Umbra limi*)	X	X		X	X	X	X		X		X			X	X	X	X		X	X	X					X	X	X		X	X
Common Carp (*Cyprinus carpio*)				X					X	X	X		X		X	X	X				X	X		X	X	X	X	X	X	X	
Creek Chub (*Semotilus atromaculatus*)						X												X													
Green Sunfish (*Lepomis cyanellus*)														X												X					
Iowa Darter (*Etheostoma exile*)		X							X		X		X									X							X	X	X
Lake Chubsucker (*Erimyzon sucetta*)	X											X		X			X	X	X						X						X
Largemouth Bass (*Micropterus salmoides*)	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Least Darter (*Etheostoma microperca*)		X		X	X	X		X		X	X		X		X	X			X		X	X						X			X
Pumpkinseed Sunfish (*Lepomis gibbosus*)	X	X	X	X	X	X		X	X	X		X			X	X	X	X	X	X	X	X			X	X	X	X		X	X
Warmouth Sunfish (*Lepomis gulosus*)	X	X	X		X			X			X		X		X	X	X	X	X	X	X				X	X	X	X	X	X	X
Yellow Bullhead (*Ameiurus natalis*)									X			X			X		X	X													X
Total Species Richness	*7*	*7*	*4*	*8*	*9*	*7*	*3*	*7*	*8*	*7*	*8*	*6*	*8*	*4*	*9*	*9*	*9*	*8*	*8*	*7*	*8*	*6*	*3*	*4*	*7*	*9*	*6*	*7*	*5*	*7*	*11*

Draft

Supplemental Table S7. Species detection by sample for all three markers combined using the high stringency bioinformatic criteria. The 31 Lawler Pond samples libraries were randomly divided among two Illumina MiSeq runs FC27 and FC28.

	Sample Library															
Detected Species	**FC27_10**	**FC27_11**	**FC27_13**	**FC27_14**	**FC27_15**	**FC27_17**	**FC27_18**	**FC27_19**	**FC27_2**	**FC27_20**	**FC27_24**	**FC27_29**	**FC27_30**	**FC27_31**	**FC27_4**	**FC27_8**
American Pickerel (*Esox americanus*)		X		X	X			X	X		X			X	X	
Bluegill Sunfish (*Lepomis macrochirus*)	X	X	X		X	X	X	X	X	X	X	X	X		X	X
Central Mudminnow (*Umbra limi*)	X	X			X		X		X		X			X	X	X
Common Carp (*Cyprinus carpio*)													X			
Iowa Darter (*Etheostoma exile*)		X									X					
Largemouth Bass (*Micropterus salmoides*)	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X
Pumpkinseed Sunfish (*Lepomis gibbosus*)	X	X			X				X			X				
Warmouth Sunfish (*Lepomis gulosus*)	X										X					X
Total Species Richness	*5*	*6*	*2*	*2*	*5*	*2*	*3*	*3*	*5*	*2*	*6*	*3*	*3*	*2*	*4*	*4*

	Sample Library														
Detected Species	**FC28_1**	**FC28_12**	**FC28_16**	**FC28_21**	**FC28_22**	**FC28_23**	**FC28_25**	**FC28_26**	**FC28_27**	**FC28_28**	**FC28_3**	**FC28_5**	**FC28_6**	**FC28_7**	**FC28_9**
American Pickerel (*Esox americanus*)		X								X					X
Bluegill Sunfish (*Lepomis macrochirus*)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Central Mudminnow (*Umbra limi*)			X	X										X	X
Common Carp (*Cyprinus carpio*)					X	X			X			X			
Iowa Darter (*Etheostoma exile*)						X									X
Largemouth Bass (*Micropterus salmoides*)	X	X	X	X	X	X	X	X		X	X	X	X	X	X
Pumpkinseed Sunfish (*Lepomis gibbosus*)		X	X	X	X				X	X					X
Warmouth Sunfish (*Lepomis gulosus*)	X	X		X										X	X
Total Species Richness	*3*	*5*	*4*	*5*	*4*	*4*	*2*	*2*	*3*	*4*	*2*	*3*	*2*	*4*	*7*

Draft