eDNA metabarcoding in lakes to quantify influences of landscape features and human activity on aquatic invasive species prevalence and fish community diversity

Lilian Pukk1 | Jeannette Kanefsky1 | Amanda L. Heathman1 | Ellen M. Weise1 | Lucas R. Nathan2 | Seth J. Herbst2 | Nicholas M. Sard3 | Kim T. Scribner1,4 | John D. Robinson1

Abstract

Aim: Our goal was to use eDNA metabarcoding to characterize fish community diversity, detect aquatic invasive species (AIS) and assess how measures of community (or AIS) diversity are influenced by lake physical and environmental covariates, measures of hydrological connectivity and human accessibility.

Location: Michigan, USA.

Methods: eDNA samples collected from 22 lakes were sequenced using two mitochondrial gene regions (12S and 16S rRNA). Metabarcoding data were compared to traditional fisheries survey data for a subset of lakes, and data from all 22 lakes were combined with environmental information to identify significant associations with community diversity and AIS relative abundance.

Results: Occupancy modelling indicated that detection probabilities were generally higher with eDNA than traditional fisheries gear. Measures of connectivity with upstream aquatic habitats were positively associated with both AIS relative abundance and fish species diversity. We also demonstrate the use of spatial interpolation methods to map distributions of species diversity and AIS relative abundance within lakes.

Main conclusions: eDNA metabarcoding methods provided information on the composition and diversity of fish assemblages and the presence of AIS in freshwater lakes that varied greatly in drainage connectivity and anthropogenic development. Our case study identified associations between environmental covariates and fish diversity or AIS relative abundance across lakes. This information is of particular importance given increasing anthropogenic disturbance, invasive species spread and associated declines in aquatic biodiversity. Incorporating eDNA metabarcoding as a supplement to traditional fisheries surveys will permit managers to identify greater numbers of taxa, including early detection of AIS, with less field effort and fish mortality. Further, eDNA methods may more accurately identify physical and biological
Accurate and efficient monitoring of species community composition and quantification of causal environmental factors associated with species presence, abundance and distribution in aquatic environments necessitates effective sampling methodology and rigorous and innovative analytical designs. There is increasing demand for large-scale studies to quantify aquatic biological diversity at the community or ecosystem scale (Green, 1979; Kennedy et al., 2020; Magurran, 2013). In freshwater ecosystems, environmental DNA (eDNA) metabarcoding is an increasingly popular method for surveying biodiversity without the need to directly observe individual organisms (Hänfling et al., 2016; Olds et al., 2016). The utility of eDNA assays has rapidly been recognized by managers and policymakers (McElroy et al., 2020; Rees et al., 2014). Monitoring approaches that allow for accurate representation of rare and invasive species are important in conservation of biological diversity (Rice et al., 2012; Yates et al., 2019). Unfortunately, traditional survey methods (e.g. seines, electrofishing) are often laborious (in the case of multiple sites and gears), invasive and rely on taxonomic expertise (Creer et al., 2016). The detection of low abundance species can be difficult because of gear biases (Ruetz et al., 2007), as well as the large effort required to detect small numbers of individuals in large aquatic systems (Jerde et al., 2011). Using molecular methods for species identification, however, provides alternatives to traditional fisheries gears for surveying fish biodiversity that yield consistent estimates of species richness for aquatic communities (Hänfling et al., 2016; McElroy et al., 2020; Sard et al., 2019).

The probability of detection can vary greatly from species to species depending on the biology and behaviour of the target organism, and the environmental characteristics and complexity of the lakes they inhabit (Bracken et al., 2019; Li et al., 2019; Rees et al., 2014). In turn, diversity of aquatic habitats may also depend on environmental characteristics, including aquatic connectivity (Fagan, 2002; Lamy et al., 2013; Perkin & Gido, 2012), land-use/land-cover of surrounding landscapes (Jocelyn et al., 2005; Schlosser, 1991) and levels of human disturbance (Esselman et al., 2013). Additionally, the likelihood of false negatives is often high for rare and/or difficult to detect species (Kumar et al., 2020; Mathieu et al., 2020; Pinfield et al., 2019). To address this problem, occupancy modelling has been used to provide improved estimates of species occurrences by integrating data from multiple samples while evaluating the probability of detecting a given species (Kumar et al., 2020; MacKenzie et al., 2002; McClennaghan et al., 2020; Schmidt et al., 2013), in combination with limnological characteristics of aquatic habitats (Harper et al., 2020).

Introduction, establishment and increases in abundance and distribution of AIS are of concern within the Great Lakes region because of the increasing risk of secondary spread into inland water bodies, and the ecological impacts that these species may have on native communities (Coble et al., 1990; Escobar et al., 2018; Sard et al., 2019). Threats in Michigan are substantial because of the large amount of Great Lakes shoreline (~3,000 miles) and abundance of lakes (~11,000) in close proximity to source Great Lakes AIS populations. Early detection of AIS when populations are in low abundance and restricted in distribution is widely recognized as a critical component of AIS eradication and control (Vander Zanden & Olden, 2008). Environmental DNA methodology, whether on a single species or multiple species level of interrogation, provides sensitive and accurate early detection for AIS surveillance programs and has the added benefit of being able to simultaneously characterize community diversity when using metabarcoding. Of equal or greater importance, and as a call for future work in the field, hypothesis-oriented research is needed to develop predictive relationships between eDNA-based species presence, abundance and diversity measures and the biological and physical features of aquatic ecosystems and adjacent landscapes.

In addition to their utility as methods for early detection of AIS (McElroy et al., 2020; Sepulveda et al., 2020), eDNA metabarcoding surveys also provide an appealing approach for addressing questions related to the assembly of aquatic communities (Sard et al., 2019). These questions are of increasing importance in aquatic systems as species invasions, climate change and habitat fragmentation pose substantial threats to biodiversity in the 21st century. Theory suggests environmental features (e.g. the size and isolation of habitat patches; MacArthur & Wilson, 1967) play an important role in determining the composition and structure of ecological communities. Thus, lake size and patterns of connectivity may drive differences in species diversity and AIS presence and prevalence across habitats. Similarly, biotic interactions (including those with AIS) and anthropogenic influences may also influence the number of species in a habitat or the establishment of invasive species (Gallien & Carboni, 2017). At the same time, aquatic connectivity could lead to downstream dispersal of DNA molecules from established populations upstream. Previous studies have documented declining eDNA copy number and rates of detection with distance (Deiner & Altermatt, 2014; Jane et al., 2015; Tillotson et al., 2018). However, results from Deiner and Altermatt (2014) suggest that positive detections may extend as much as 50 km downstream in flowing systems, raising concerns about false positive detections in highly connected networks of aquatic habitats. eDNA metabarcoding surveys, particularly when they are combined with traditional fisheries survey gear and data on
the abiotic characteristics of lakes and surrounding landscapes, can thus help to clarify the relative influences and importance of characteristics such as area and isolation on aquatic community composition and the presence or prevalence of AIS.

When conducting landscape-scale analyses that quantify associations between measures of ecological diversity and environmental covariates or measures of anthropogenic disturbance, grain size, extent size and the spatial distance among sampling locations are important aspects of landscape data sampling (Anderson et al., 2010). For example, previous studies have demonstrated strong associations between the composition of fish communities and environmental features when sampling was conducted over expansive areas (Sharma et al., 2011). In this study, we sample aquatic communities at a state-wide spatial scale. Information was gathered for 22 lakes in Michigan to compare eDNA metabarcoding estimates of fish community composition and detection of invasive species across a range of environments differing in lake area, connectivity to other aquatic systems, and degree of anthropogenic disturbance. The goals of this project were to identify associations between physical lake characteristics and fish community diversity or relative abundance of AIS, compare detection probabilities of traditional fish sampling gears and eDNA, and use eDNA to identify “hot spots” of AIS to inform proactive management actions in Michigan’s inland lake habitats.

2 | METHODS

2.1 | Lake selection

Twenty-two lakes in Michigan were sampled over three consecutive years (2016–2018): eight lakes in 2016 (Sard et al., 2019), two lakes in 2017 and twelve lakes in 2018 (Figure 1; Table 1). Field work was conducted in collaboration with the Michigan Department of Natural Resources (MDNR), Michigan Department of Environment, Great Lakes, and Energy, and the Department of Fisheries and Wildlife at Michigan State University.

Lakes were chosen based on physical and abiotic attributes including lake size (ha), depth (m), and the lake classification scheme of Wehrly et al. (2012), which incorporates information on lake depth, growing degree days and mean water temperature during the ice-free period, to include a representative collection of lakes with diverse fish communities and limnological characteristics (Table 1). Additional landscape features were also considered during lake selection, including hydrologic connectivity and land use patterns (forest, agriculture, urban).

Physical and environmental information was collected for all 22 sampled lakes from the Lake multi-scaled geospatial and temporal database (LAGOS-NE, https://lagoslakes.org/; Soranno et al., 2017). Selected lakes include those frequented by anglers (Houghton Lake and Higgins Lake) as well as representatives from counties with the highest numbers of invasive species detections in the state (Cass Lake, Oakland Co. and Thompson Lake, Livingston Co.; Midwest Invasive Species Information Network, https://www.misin.msu.edu).

2.2 | Sampling and eDNA extraction

The study also included MDNR Fisheries Status and Trends (ST) survey information for 13 lakes that were sampled in 2016 and 2018 (Table 1). MDNR’s ST survey deploys multiple gear types and standardized techniques to characterize fish community composition and detect invasive species in Michigan’s lakes (Schneider, 2000). ST survey data were used to compare eDNA metabarcoding and traditional sampling gear (Table 2). Traditional gear types included boat electrofishing (boomshocker), seines, large mesh fyke nets, small mesh fyke nets, trap nets and experimental gill nets (for details on net dimensions and mesh sizes, see Schneider, 2000). Given the variation in sizes and maximum depths of lakes sampled in this study (Table 1), not all traditional gear types were used in all lakes and there was substantial variation in the total number of gear deployments across lakes (Table 2). A total of 499 traditional gear samples were collected across the 13 lakes and the number of sampling locations varied between 18 and 88 locations per lake. Fish collected by the MDNR field crews were identified to species. Fish counts per species were tabulated by gear type, sampling location and lake.

eDNA and traditional gear sampling were conducted in the same locations within a lake when possible, with additional randomly selected eDNA samples in lakes with fewer traditional gear sampling locations. In lakes where traditional gear surveys were not conducted, eDNA sample coordinates were randomized. Sample numbers ranged from 30 to 59 across the surveyed lakes, with more sampling effort allocated to larger lakes, although samples were not proportionally distributed based on lake size (Table 1). The collection
| Lake Name         | County         | Sampling date | Total eDNA samples | Surface / benthic eDNA samples | Max. depth (m) | Lake area (ha) | Lake Class* | IWS stream density (m/ha)† | No. upstream lakes (>4 ha)‡ | Area (ha) upstream lakes >4 ha§ | % Surrounding land area in dev. and ag.¶ | No. boat ramps§ | Distance weighted population h |
|------------------|----------------|---------------|--------------------|--------------------------------|----------------|---------------|-------------|---------------------------|-----------------------------|-------------------------------|---------------------------------|----------------|-----------------------------|
| Austin Lake      | Kalamazoo      | 7/30/2018     | 40                 | 29/11                         | 3.4            | 445.98        | 2           | 1.3                       | 5                           | 456.3                        | 62.6                             | 2                | 0.306                       |
| Brevoort Lake    | Mackinac       | 8/10/2018     | 39                 | 28/11                         | 9.1            | 1,745.73      | 5           | 4.7                       | 1                           | 66.1                         | 7.3                               | 4                | 0.092                       |
| Cass Lake        | Oakland        | 7/17/2018     | 40                 | 28/12                         | 38.1           | 530.72        | 2           | 10.8                      | 48                          | 2,525.8                      | 73.0                             | 8                | 1                           |
| Dumont Lake      | Allegan        | 6/2/2016      | 48                 | 33/15                         | UNK            | 97.59         | 1           | 4.1                       | 2                           | 25.4                         | 41.6                             | 2                | 0.378                       |
| Five Channels Dam Pond | Iosco | 6/28/2016    | 48                 | 35/13                         | 9.1            | 90.04         | 1           | 2.1                       | 68                          | 4,502                        | 9.1                              | 1                | 0.155                       |
| Fourth Lake      | Hillsdale      | 5/19/2016     | 38                 | 26/12                         | 15.5           | 19.17         | 1           | 12.6                      | 1                           | 28.6                         | 27.4                             | 0                | 0.679                       |
| Haithco Lake     | Saginaw        | 5/27/2016     | 34                 | 24/10                         | UNK            | 14.89         | 1           | 0.0                       | 0                           | 0                             | 92.0                             | 0                | 0.285                       |
| Higgins Lake     | Roscommon/Crawford | 8/30/2018   | 52                 | 37/15                         | 41.2           | 4,126.62      | 3           | 0.9                       | 0                           | 0                             | 33.2                             | 14               | 0.159                       |
| Holloway Reservoir | Genesee       | 6/9/2016      | 45                 | 31/14                         | 7.6            | 486.13        | 2           | 10.1                      | 58                          | 1,128.6                      | 37.6                             | 2                | 0.363                       |
| Houghton Lake    | Roscommon      | 8/14/2018     | 48                 | 32/16                         | 6.4            | 8,126.64      | 5           | 2.7                       | 11                          | 4,930.4                      | 49.9                             | 19               | 0.17                        |
| Kimball Lake     | Newaygo        | 5/17/2018     | 30                 | 22/8                          | 16.2           | 58.64         | 1           | 12.1                      | 4                           | 145.7                        | 20.3                             | 1                | 0.226                       |
| Lake George      | Clare          | 8/22/2017     | 36                 | 23/13                         | 7.6            | 52.4          | 1           | 0.0                       | 0                           | 0                             | 24.0                             | 1                | 0.202                       |
| Long Lake        | Hillsdale      | 5/24/2018     | 37                 | 28/9                          | 12.2           | 86.22         | 1           | 7.4                       | 3                           | 30.7                         | 50.1                             | 1                | 0.234                       |
| Manistique Lake  | Mackinac       | 8/9/2018      | 55                 | 39/16                         | 6.1            | 4,186.12      | 6           | 3.8                       | 10                          | 2,619.1                      | 19.0                             | 6                | 0.081                       |
| Mullett Lake     | Cheboygan      | 8/15/2017     | 55                 | 44/11                         | 36.6           | 6,762.28      | 3           | 6.7                       | 31                          | 11,005.8                     | 24.7                             | 9                | 0.116                       |
| Ocqueoc Lake     | Presque Isle   | 5/23/2016     | 48                 | 34/14                         | 34             | 50.7          | 1           | 6.5                       | 13                          | 409.2                        | 11.8                             | 2                | 0.108                       |
| Pentwater Lake   | Oceana         | 6/17/2016     | 55                 | 38/17                         | 12.2           | 198.08        | 2           | 9.8                       | 250                         | 8,194,571.2                  | 29.9                             | 6                | 0.157                       |
| Pickerel Lake    | Newaygo        | 5/17/2018     | 31                 | 23/8                          | 22.3           | 124.13        | 1           | 3.6                       | 5                           | 204.3                        | 16.9                             | 2                | 0.227                       |
| Thompson Lake    | Livingston     | 8/2/2018      | 25                 | 18/7                          | 16.8           | 106.85        | 1           | 8.5                       | 2                           | 147                          | 92.8                             | 1                | 0.956                       |
| Torch Lake       | Antrim         | 8/7/2018      | 59                 | 42/17                         | 86.9           | 7,578.99      | 2           | 6.6                       | 19                          | 2,065.5                      | 26.9                             | 10               | 0.132                       |
| Walloon Lake     | Charlevoix     | 6/23/2016     | 57                 | 41/16                         | 30.5           | 1,853.42      | 2           | 6.0                       | 0                           | 0                             | 22.1                             | 10               | 0.124                       |
| Wycamp Lake      | Emmet          | 8/6/2018      | 33                 | 23/10                         | 2.1            | 246.75        | 6           | 2.3                       | 0                           | 0                             | 1.5                              | 1                | 0.105                       |

Note: Lake sizes (ha) and landscape-scale environmental covariates are derived from the LAGOS-NE database (Soranno et al., 2017). Lake locations are shown in Figure 1.

*The maximum depth of the lake in metres (UNK—unknown, max depth for Dumont, Haithco, and Ocqueoc Lakes were not available in LAGOS-NE. Ocqueoc max lake depth was retrieved from the MDNR Aquatic Habitat Viewer).

†IWS Stream density (m/ha)—The density of streams within the individual watershed (IWS—inter watershed zone) of each lake in metres per hectare.

‡No. upstream lakes (>4 ha)—The total number of all lakes upstream that are larger than 4 ha in size.

§Area (ha) upstream lakes >4 ha—The total area of all lakes upstream that are larger than 4 ha in size, in hectares.

¶Percentage based on land within a 500 m buffer surrounding each lake. % Surrounding land area in dev. and ag.—The sum of the percentage of development and the percentage of agriculture on a 500 m buffer zone spatial scale for 2011. Total development: 1—the sum proportion of undeveloped land use categories on a 500 m buffer zone spatial scale for 2011.

hNo. boat ramps—The total number of boat accesses that were seen on satellite images of the lakes via Google Earth that meet criteria for public accessibility.

Using estimates of county population from 2017 (from MI Dept. Health & Human Services) and Great Circle distance between county centroids and lake centroids. For each lake, the sum (across counties) of the population size of the county divided by the distance between the county and lake centroids (Euclidean distance, in km). Values in the column are converted to a 0–1 scale by dividing each by the maximum of these summed, distance-weighted, population sizes per lake.
| Lake name          | Traditional Gear | 12S | 16S | 12S | 16S |
|-------------------|------------------|-----|-----|-----|-----|
| Austin Lake       | NA               | 19  | 23  | 1.230 | 1.496 | 15  | NA |
| Brevoort Lake     | 60               | 24  | 25  | 1.064 | 0.974 | 16  | 11 |
| Cass Lake         | 34               | 26  | 27  | 1.093 | 1.500 | 17  | 12 |
| Dumont Lake       | 23               | 21  | 28  | 1.391 | 1.417 | 17  | 7  |
| Five Channels Dam Pond | 29            | 30  | 39  | 1.767 | 1.901 | 20  | 11 |
| Fourth Lake       | 21               | 23  | 20  | 0.775 | 0.755 | 15  | 10 |
| Lake George       | NA               | 11  | 16  | 1.217 | 1.310 | 10  | NA |
| Haithco Lake      | 18               | 18  | 20  | 1.276 | 1.119 | 12  | 5  |
| Higgins Lake      | NA               | 20  | 21  | 0.708 | 1.054 | 12  | NA |
| Holloway Reservoir| 48               | 24  | 30  | 0.983 | 0.882 | 17  | 9  |
| Houghton Lake     | NA               | 24  | 27  | 1.097 | 1.374 | 16  | NA |
| Kimball Lake      | NA               | 22  | 25  | 1.310 | 1.510 | 16  | NA |
| Long Lake         | 24               | 33  | 36  | 1.419 | 1.771 | 23  | 12 |
| Manistique Lake   | NA               | 21  | 23  | 0.582 | 1.004 | 12  | NA |
| Mullet Lake       | NA               | 20  | 23  | 0.293 | 0.674 | 14  | NA |
| Ocqueoc Lake      | 24               | 29  | 37  | 1.722 | 1.927 | 19  | 12 |
| Pentwater Lake    | 27               | 38  | 42  | 1.773 | 1.911 | 29  | 19 |
| Pickerel Lake     | NA               | 19  | 23  | 1.249 | 1.486 | 14  | NA |
| Thompson Lake     | 27               | 16  | 22  | 1.075 | 1.155 | 13  | 7  |
| Torch Lake        | 88               | 23  | 27  | 0.646 | 0.884 | 14  | 7  |
| Walloon Lake      | 76               | 27  | 28  | 0.968 | 1.237 | 16  | 7  |
| Wycamp Lake       | NA               | 17  | 21  | 1.454 | 1.642 | 12  | NA |

Note: NA indicates that traditional gear was not used on the lake. Lake locations are shown in Figure 1.

aMean estimated over all eDNA samples per lake.
bNumber of fish species detected using both 12S and 16S metabarcoding.
cNumber of fish species detected by all three approaches (12S, 16S, and traditional gear).
of 2016 water samples (n = 446, from eight lakes), including benthic and surface eDNA sampling, was conducted according to Sard et al. (2019). In 2017, 95 water samples were collected and filtered from two lakes, and in 2018, 656 water samples were collected from 12 lakes. On average ~25.1% of samples per lake were of benthic zone water (range 20.0% to 29.6% per lake) and were collected using a Van Dorn sampler near the lake bottom. For each lake, six to nine negative control samples, consisting of one litre of distilled water ("no-DNA"), were filtered to provide a measure of quality control and quantify any contamination that occurred during water sampling. A Van Dorn negative (distilled water "no-DNA") was taken prior to use in each lake to quantify levels of contamination associated with the sampling device. Filtering (field) negatives were interspersed randomly with water samples to quantify contamination associated with the filtering apparatus. One litre of distilled water was also filtered as a negative control immediately after the last sample was processed to quantify levels of contamination arising during the sample processing. For each of the negatives, distilled water was poured into sterile 1-L wide-mouth Nalgene bottles and filtered for eDNA analysis. Prior to each sampling event, all bottles were soaked in a 20% (V/V) bleach solution for 10 min to eliminate residual DNA contamination (Prince & Andrus, 1992), rinsed with distilled water, and left to dry for 24–48 hr at room temperature. Water samples were filtered on the boat immediately after collection using a Smith-Root ANDe™ Backpack eDNA Sampler (Thomas et al., 2018) with single-use filter housings. Samples in 2016 were filtered through 0.45-μm nitrocellulose filters, while those from 2017 and 2018 were filtered through 1.2-μm polyethersulfone (PES) filters. Filters were then removed from filter housings using sterile, single-use forceps and placed in pre-labelled tubes containing 95% ethanol. Gloves were changed between each sample. The boat and the trailer were sterilized using 2% Virkon Aquatic solution after each sampling event. Prior to DNA extraction, filters from all water samples were removed from ethanol and air dried in a sterile designated eDNA hood. DNA extractions followed a protocol developed by Laramie et al. (2015), which employs a combination of a QIAshredder homogenization kit (QIAGEN), DNeasy Blood and Tissue extraction kit (QIAGEN), and a OneStep PCR inhibitor removal kit (Zymo Research). eDNA extractions were carried out in a separate room away from the laboratory where genomic DNA was handled. All necessary materials and bench spaces were cleaned prior to use with either 25% (V/V) bleach, UV light, and/or DNA Away (Thermo Fisher Scientific, Inc.). All pipetting was done with pipettors used exclusively for eDNA samples and sterile filter tips. Also, one extraction negative control was included in each extraction event for each lake (some lakes had up to three extraction negatives) to test for contamination during the DNA extraction procedure.

### 2.3 eDNA library preparation and sequencing

eDNA samples were amplified using primer sets targeting two mitochondrial rRNA gene regions, 12S (Riaz et al., 2011) and 16S (Deagle et al., 2009), to characterize fish community composition. PCR amplifications followed a modified protocol developed by Sard et al. (2019). Briefly, amplification reactions were carried out in total volume of 25 μl that included 1x AmpliTaq Gold PCR Buffer II (no Mg²⁺), 1 μg/μl of BSA, 1.25 U AmpliTaq Gold DNA polymerase and 0.32 μM of forward and reverse primer. The 12S and 16S assays also included 0.24 and 0.32 mM dNTPs, and 2.0 and 2.5 mM MgCl₂, respectively. Each forward and reverse primer included CS1 and CS2 oligo tails (Fluidigm), respectively, which enabled each sample to be dual-indexed for sequencing in a subsequent PCR. Each 96-well plate included one “no template” PCR negative as well as one positive control sample of pooled DNA from ten fish species to confirm successful PCR amplification. During preparation for sequencing, blanks (randomly chosen wells that were left empty) were included on the sequencing plate to account for possible contamination or incorrect sample indexing during sequencing.

All PCR products were visualized by gel electrophoresis to confirm successful PCR amplification. DNA concentrations were quantified using Quant-IT PicoGreen assay kits (Thermo Fisher Scientific) on a QuantStudio 6 (Applied Biosystems) instrument and subsequently diluted to 7 ng/μl. Sequencing was conducted on an Illumina MiSeq platform at Michigan State University’s Research Technology Support Facility. Sequencing conditions were the same as in Sard et al. (2019) with minor modification. The 2018 12S and 16S amplicons were pooled for sequencing on two Illumina 2 × 150 bp paired-end flow cells (each individual sample was sequenced two times). Base calling was done by Illumina Real Time Analysis (RTA) v1.18.54 and output of RTA was de-multiplexed and converted to FastQ format with Illumina Bcl2fastq v2.19.1.

### 2.4 Database development

The 12S and 16S databases included native Michigan fishes and aquatic invasive species. Altogether, the 12S database was comprised of 139 fish species (including 34 AIs) and the 16S database was comprised of 140 fish species (including 36 AIs). Finally, 98 and 102 non-fish vertebrate (amphibian, reptile, bird, mammal) sequences were included in the 12S and 16S databases, respectively, to identify and remove any non-fish sequences amplified with the metabarcoding primers. Each species in the database was represented with 1–5 sequences. Sequence databases developed separately for 12S and 16S loci were analysed to estimate intra- and inter-specific sequence divergence using Mega v6 (Tamura et al., 2013). Taxonomic databases for both markers are available on Dryad (https://doi.org/10.5061/dryad.1zcrjdfs9).

### 2.5 Analysis of eDNA community matrix

Sequencing data from all 22 lakes were combined for subsequent analysis in Mothur v1.39.5 (Schloss et al., 2009) as described by Sard et al. (2019), with minor modifications. All sequencing data were analysed using computational resources provided by Michigan State
University's High-Performance Computing Center at the Institute for Cyber-Enabled Research. In short, paired-end sequences were merged for each sample based on quality scores, removing poorly merged sequences from the subsequent analysis. Sequences were aligned to the taxonomic databases, separately for 12S (size range 123–147 bp) and 16S (size range 73–129 bp). Chimeric sequences were removed using VSEARCH v.1.8.0 (Rognes et al., 2016). Remaining reads were clustered into operational taxonomic units (OTUs) with a sequence similarity cut-off of 99% using default settings. OTUs within each sample were enumerated and assembled into community matrices. In situations where the same classifications were represented by multiple OTUs, the initial community matrix was condensed to only unique classifications. To account for the lack of interspecific sequence variation within a genus, some OTU classifications were collapsed to their respective genera (e.g. *Ameiurus natalis*, *Ameiurus melas* and *Ameiurus nebulosus* classifications were changed to *Ameiurus spp*). Nucleotide BLAST was used to query all OTUs that were not classified to species, to account for other possible fish species missing from the taxonomic databases.

Prior to analyses, contamination was considered for each sample by calculating the median number of non-zero reads per classification from the different negative control samples included in the study. Sequences from all eDNA lake samples in the community matrix were evaluated and if a species had a read count less than the median number of reads per classification (excluding classifications with 0 reads) in negative and blank samples associated with the sampled lake, its read count was changed to zero as a conservative way to avoid false positive detections.

All community matrix analyses were conducted in R v3.5.3 (R Core Team, 2020) using the “tidyverse” (Wickham, 2017) package. For each sample within a lake, the number of species (species richness) detected by 12S and 16S markers was calculated. Estimates of the Shannon index of diversity (Shannon, 1948) for each lake were also generated, under the assumption that metabarcoding read counts reflect relative abundance of fish species.

### 2.6 Generalized Linear Modelling

Generalized linear models (GLMs) were used to identify landscape and lake covariates that may explain variation in alpha diversity and invasive species prevalence across the 22 sampled lakes. Dependent variables evaluated included species richness, Shannon diversity, the proportion of sequencing reads attributed to AIS, and the number of AIS per lake. All analyses were conducted separately for the two eDNA metabarcoding markers (12S and 16S). We also used species richness from traditional gear surveys as a response variable, for the subset of lakes (n = 13) where MDNR ST survey data were available. Shannon diversity, number of AIS and AIS prevalence from traditional gear survey data were not used as response variables in our GLMs, given expected gear selectivity and variation in effort and gear deployments across lakes.

Climatological, lake morphometric and connectivity data from the LAGOS-NE database (which includes information for more than 50,000 lakes >4 ha in surface area across 17 states; Soranno et al., 2017) were included as independent variables in modelling analyses for measures of alpha diversity. Specific variables in our modelling analyses included: log-transformed lake area (in hectares), lake class (from Wehrly et al., 2012), stream density in the inter-lake watershed, log-transformed area of upstream lakes (greater than 4 hectares) and per cent developed or agricultural land in a 500 m buffer around each lake. Modelling analyses for AIS species richness and read proportions also included square-root transformed boat ramp density per lake and distance-weighted human population density (calculated as the sum of county census numbers from 2017, weighted by distance to each lake) as surrogates for potential anthropogenic influences. See Table 1 for complete descriptions of the environmental variables used in GLM analyses.

For each response variable, we evaluated single-variable models and two candidate models focused on “natural” (i.e. area, class, stream density, upstream area) and “anthropogenic” characteristics (distance-weighted population size, boat ramps per area and per cent development) of the sampled lakes. Poisson regressions were used for count-based response variables (i.e. number of AIS, total species richness), while normal regressions were used for continuous response variables (i.e. the proportion of reads attributed to AIS, Shannon diversity). Variables were log- or square root-transformed to meet model assumptions. GLMs were generated and plotted by using R functions from the “ggplot2” (Wickham, 2016), “tidyrr” (Wickham, 2020), “dplyr” (Wickham et al., 2020) and “varhandle” (Mahmoudian, 2020) packages. For each set of models, we conducted AICc and BIC model selection using functions from the “MuMIn” (Barton, 2020) and “stats” (R Core Team, 2020) packages.

### 2.7 Occupancy modelling

A Bayesian hierarchical occupancy model (Kery, 2010; MacKenzie et al., 2002) was used to estimate gear-specific detection probabilities for three individual fish species in lakes that were sampled using both eDNA and traditional gear (n = 13). We focus this analysis on three surrogate species to represent species with similar body morphology and patterns of habitat usage (Gallien & Carboni, 2017; Procheş et al., 2008) that are comparable to common AIS in Michigan. Using the surrogate species approach also allows us to facilitate broader application of our findings and provide recommendations for future AIS monitoring efforts. The species chosen for this analysis include two common benthic generalist AIS across the sampled lakes (*Neogobius melanostomus*, round goby and *Cyprinus carpio*, common carp), and *Amia calva* (bowfin), which was chosen as a surrogate for the invasive *Channa argus* (snakehead). The goal of the analysis was to compare gear types to determine which gears provide the greatest detection probability for AIS with different life histories and patterns of habitat usage. It was assumed that the four eDNA methods (both markers included surface and benthic samples) represented independent gear types. Therefore, a total of ten gear types were compared: benthic eDNA 12S, surface eDNA 12S, benthic eDNA 16S, surface eDNA
16S, boomshocker, experimental gill net, large mesh fyke net, seine, small mesh fyke net and trap net.

In the model, occupancy was defined at the lake level and it was assumed that occupancy did not change between sampling events. Therefore, the detection probabilities generated from the analysis can be interpreted as the probability of a species being detected with a single sample of a given gear type, given that the species was present in the lake. The biological model of occupancy was defined as $Z_s - \text{Bernoulli}(\psi)$, where $\psi$ is the probability of occupancy for each species ($s$) and $Z_s$ is the true occupancy for each species at each lake ($i$). The observation model was defined as $Y_i - \text{Binomial}(p, Z_s)$, where $Y_i$ was the detection at a given sample ($i$) and $p$ was the detection probability for each species ($s$) and gear ($g$). The model was analysed using Markov Chain Monte Carlo (MCMC) methods in R with the “jagsUI” package (Keller, 2018). Three independent Markov chains were run with 4,000 iterations following a 1,000 iteration burn-in. Model convergence was assessed by visually inspecting chains and based on $Rhat$ values near one.

2.8 | Heat maps

We used spatial interpolation methods to visualize patterns in species diversity and relative abundance of $N$. melanostomus (a common AIS in the Great Lakes drainage basin; Jude et al., 1992) within lakes (similar to maps in Jo et al., 2017). Given the limited number of water samples collected from each lake, surfaces of total fish community diversity (species richness) and relative abundance (to account for unequal sequencing effort across samples) of $N$. melanostomus were interpolated using inverse distance weighting (IDW; Bartier & Keller, 1996). Optimal weighting coefficients were identified by leave-one-out cross-validation for each lake (similar to the approach applied for interpolation of rainfall data in Chen & Liu, 2012). Briefly, species richness (or $N$. melanostomus relative abundance) values associated with each sampled location were sequentially removed from the dataset and predicted by IDW interpolation from the remaining data. Then the root mean squared error (RMSE) of the predictions associated with each weighting coefficient was calculated as,

$$RMSE = \sqrt{\frac{\sum (\hat{SR}_i - SR_i)^2}{n}}$$

where $SR_i$ is the observed species richness (or relative abundance) associated with location $i$, $\hat{SR}_i$ is the predicted species richness when that sample was removed, and $n$ is the total number of water samples collected from a lake. Evaluated IDW coefficients ranged from 2 to 21 (larger values more severely limit the influence of distant samples) and we chose the coefficient that minimized RMSE for each lake. Interpolated surfaces were generated and plotted in R using functions from the “gstat” (Benedikt et al., 2016; Pebesma, 2004), “sp” (Bivand et al., 2013; Pebesma & Bivand, 2005), “rgdal” (Bivand, Keitt, et al., 2020), “prevR” (Larmarange et al., 2011), “raster” (Hijmans, 2020), “prettymapr” (Dunnington, 2017), “rgeos” (Bivand, Rundel, et al., 2020) and “colorRamps” (Keitt, 2012) packages.

3 | RESULTS

The total number of sequence reads (excluding no-DNA controls) generated for the 12S and 16S markers were 9,165,523 and 9,437,972, respectively, while the average over all lakes was 416,614 ($\pm$47,830.9 SE; 12S) and 428,998 ($\pm$36,186.3 SE; 16S). The mean number of reads per eDNA sample per lake ranged from 2,646 to 18,522 ($\pm$1,109.1 SE) for 12S, and from 1,600 to 19,523 ($\pm$978.1 SE) for 16S. The high number of total reads observed across all samples for each lake (whole lake range 137,597–791,077 and 60,815–722,331 for 12S and 16S, respectively) enabled us to evaluate relative sequence abundance of each fish species per sample. For 12S and 16S eDNA metabarcoding markers, a small proportion of the reads collected from each lake were attributable to contamination. For both markers, the mean number of reads per no-DNA (negative) sample was less than 0.2% of the mean number of reads in filtered environmental samples (12S: 0.085%; 16S: 0.19% of the total sequence reads from a sample).

3.1 | Alpha diversity

The number of fish species detected per lake using eDNA markers varied by over 3-fold (11 to 38 species) for 12S, while for 16S the number of species detected varied from 16 to 42 species (Table 2). The highest number of species was detected in Pentwater Lake, which has a direct connection to Lake Michigan, and Five Channels Dam Pond. The number of species detected per lake was highly positively correlated between 12S and 16S markers (Pearson correlation 0.904, $p < .001$). eDNA metabarcoding data were also used to estimate Shannon diversity for each of the 22 lakes (Table 2). Shannon diversity was highly positively correlated between the 12S and 16S markers (Pearson correlation coefficient = 0.918, $p < .001$). There was no significant correlation between Shannon diversity (estimated from either eDNA marker) and the number of sampling locations per lake or number of sequence reads generated per lake ($p > .05$), nor was there a correlation between species richness and the proportion of sequencing reads attributed to AIS for either of the two eDNA markers ($p > .05$). Shannon diversity measures varied slightly between the lakes for 12S and 16S (mean = 1.128 and 1.317; min = 0.359 and 0.674; max = 1.773 and 1.927; SE = 0.082 and 0.080, respectively). However, both measures of alpha diversity (species richness and Shannon diversity) were highest in Pentwater Lake (12S = 38, 1.773; 16S = 42, 1.911), Ocqueoc Lake (12S = 29, 1.722; 16S = 37, 1.927), Five Channels Dam Pond (12S = 30, 1.767; 16S = 39, 1.901), and Long Lake (12S = 33, 1.471; 16S = 36, 1.771; Table 2).

3.2 | Aquatic invasive species detection

Aquatic invasive fish species were detected in 21 lakes using at least one sampling method (Table S1). One to four invasive species
were detected per lake (except Manistique Lake where no AIS were detected). Preliminary examination of detection patterns suggested that eDNA sampling was a more sensitive means of AIS detection than traditional gear approaches, as both eDNA markers detected AIS in a larger proportion of lakes (10 out of 13 lakes; Table 2).

The proportions of native and aquatic invasive fish species for all 22 lakes are shown in Figure 2. Across all lakes surveyed, nine AIS were detected including *C. carpio*, *Hypophthalmichthys* sp. (bighead or silver carp), *Mylopharyngodon* piceus (black carp), *Ctenopharyngodon* sp. (grass carp), *Carassius auratus* (goldfish), *Mylopharyngodon* piceus (black carp), *Osmerus mordax* (rainbow smelt), *Alosa* sp. (alewife), and *Petromyzon marinus* (sea lamprey). The detections of *Hypophthalmichthys* sp. led to subsequent eDNA surveys that were ongoing at the time of publication. Several AIS were especially widespread, being detected in multiple lakes (*C. carpio*, *n* = 20; *N. melanostomus*, *n* = 9; and *Alosa* sp., *n* = 4; Table S2) and particularly abundant in some lakes (e.g. Holloway Reservoir and Mullet Lake), comprising a substantial proportion of all sequences (62.9% - 65.1% for 12S and 16S, and 38.2% - 62.9% for 16S and 12S, respectively; Table S1).

The number of fish AIS detected based on 12S and 16S metabarcoding markers was concordant (Pearson correlation \( R = 0.647, p = .001 \)) differing by at most one species across the sampled lakes (Table 2). In cases where one marker detected an additional AIS, the detection was based on either one or two positive samples and few total reads (<0.1% of sequencing reads in all instances; Table S1), suggesting that differences are due to detections of low-abundance species (based on eDNA detections and proportional sequence counts). In some cases, the lack of detection may be explained by differences in the power of the markers to resolve sequences to different species or their failure to amplify certain taxa (e.g. *P. marinus* in Ocqueoc Lake; Table S2), which serves as a further justification for recommendations to use multiple markers in eDNA surveys (McClanaghan et al., 2020; McElroy et al., 2020; Zhang et al., 2018). The fish taxonomic databases for 12S and 16S also differed in species representation, depending on the availability of the genetic information in GenBank. Fish AIS sequence reads as a proportion of total sequences (an eDNA surrogate measure of species biomass; Kelly et al., 2019) were positively correlated between 12S and 16S metabarcoding markers (Pearson correlation 0.928, *p* < .001) and varied among lakes from 0.000 (both 12S and 16S; Manistique Lake) to 0.688 (12S, Mullett Lake) and 0.719 (16S, Holloway Reservoir). Thus, at the high extreme, in Holloway Reservoir and Mullet Lake,

**FIGURE 2** Proportion of aquatic invasive fish species reads per lake (a—12S, b—16S) for 22 Michigan lakes. More detailed information on the number of aquatic invasive species (AIS) and native species (Non_AIS) is given in Table 2.
over half of the lake fish community sequences were from AIS (Figure 2, Table S1).

3.3 | Associations between environmental variables and fish community diversity or AIS presence

Upstream lake area, a measure of hydrological connectivity, was the top supported model for 12S and 16S species richness (Akaike weights = 0.79 and 0.97, respectively), and exhibited a significant positive relationship with observed species richness from both markers (p < .05; Figure 3a,d, Table S3). Lake area was the top supported model for both 12S and 16S Shannon diversity (Akaike weights = 0.91 and 0.85, respectively) and exhibited a significant negative relationship (p < .05; Figure 3b,e). Upstream lake area was the top supported model for 12S and 16S proportion of AIS reads (Akaike weights = 0.88 and 0.63, respectively) and exhibited a significant positive relationship (p < .05; Figure 3c,f, Table S3). No significant covariates were identified for the number of AIS, as the intercept only model was best supported for both 12S and 16S (Table S3).

We analysed the same set of models with the number of species detected in ST surveys using traditional gears. As with the eDNA model results, area upstream was still the top supported model. However, it was not as strongly supported compared to the eDNA model results, likely as a result of the lower sample size of lakes where ST survey data were available (13 lakes compared to 22 lakes for eDNA). Area upstream had Akaike weights of 0.79 and 0.97 for 12S and 16S (Table S3), indicating that it was clearly the top supported model. For the traditional gear data, the model including area upstream had an Akaike weight of 0.36, and two other models (lake area and per cent developed land cover) had delta AICc <2. These results further support our interpretation that area upstream is positively associated with increased species richness in the sampled lakes.

3.4 | Occupancy analysis

Occupancy analyses were conducted for a subset of species detected by traditional gear and eDNA metabarcoding as a means of comparing detection probabilities associated with these complementary survey methods. The species selected for these analyses had sufficient detections across the sampled lakes for reliable estimation of detection probability. Species selected included *C. carpio* and *N. melanostomus*, two invasive species found in the majority (n = 9) of the 13 lakes where eDNA and traditional gear were jointly deployed, and *A. calva* (as a surrogate for invasive *C. argus*). Combined, the three species represent different ecologies and traits (life history, body size, etc.) that may influence detection probability or gear susceptibility bias.

For all three species, eDNA gear types (benthic 12S and 16S, surface 12S and 16S) consistently had among the highest detection probabilities (0.46–0.48, 0.40–0.49 and 0.21–0.23 for *N. melanostomus*, *C. carpio* and *A. calva*, respectively; Figure 4) compared to...
the traditional gear types deployed for ST surveys (where detection probabilities ranged from 0.01–0.53, 0.01–0.15 and 0.01–0.31 for the three species, respectively; Figure 4). However, eDNA was only significantly better than all six traditional gear types for *C. carpio* (Figure 4, centre). eDNA detection probability for *N. melanostomus* was not significantly higher than that of small mesh fyke nets and only the surface eDNA sample detection probabilities were higher than those for seines. In contrast, eDNA detection probability for *A. calva* was not significantly different from trap net, boomshocker, or large mesh fyke net sampling gears. Jerde (2021) provided a formulation for estimating the probability that a species would be detected by eDNA prior to traditional gear sampling based on estimates of detection probability like those provided above. Based on this formulation, the probability of eDNA detection before traditional fisheries survey gear ranged from 0.47 to 0.97 for *N. melanostomus*, from 0.74 to 0.98 for *C. carpio*, and from 0.43 to 0.95 for *A. calva*, illustrating the robust performance of eDNA metabarcoding for early detection of invasive species.

### 3.5 Spatial patterns of species richness and AIS abundance within lakes

Data from Pentwater Lake are presented here as an example of how information from eDNA metabarcoding surveys can be used to characterize spatial patterns in overall fish community diversity and AIS relative sequence abundance within a lake (Figure 5; see Figure S1 for interpolated surfaces from other lakes). Interpolations for Pentwater Lake suggest higher species richness near the Pentwater River inlet (Figure 5; see Table S4 for IDW power information), while reads attributed to *N. melanostomus* appear to be concentrated on the southern and western portions of the lake. Spatially explicit “hot spots” like these could be used in the future to prioritize AIS response efforts by identifying potential locations that would maximize the effectiveness of removal efforts.

### 4 DISCUSSION

The objectives and results from this case study extend historical single species and multi-species descriptive characterizations of species presence/absence, estimates of relative abundance and community diversity, and comparative analyses of the efficiency of different gear types (Hänfling et al., 2016; Sard et al., 2019). Specifically, our research goals were to spatially characterize fish biodiversity and AIS dispersion within lakes and to quantify detection probabilities and evaluate hypotheses regarding associations between AIS presence or species diversity and measures of hydrological connectivity, surrounding landscape features, and measures of human access to lakes. Regression analyses and modelling as conducted here have been embraced as effective ways to identify causal explanations for invasive species distribution and abundance (Gallien & Carboni, 2017). Novel applications and results described for our case study here may guide future lakescape-scale monitoring and modelling that will advance agency abilities to identify factors that can negatively affect aquatic biodiversity and impact AIS distribution and abundance.

Analyses of associations between environmental variables and measures of fish community diversity and AIS presence demonstrate the importance of watershed connectivity. Best-supported models highlight the influence of upstream lake area (a measure of aquatic connectivity) on species richness of the fish community and the proportion of reads that are attributable to AIS (Figure 3, Table S3). In this study, the highest species richness was detected in relatively small Pentwater Lake, which has one of the largest upstream lake networks (250 lakes larger than 4 ha) among lakes sampled in this study. One potential explanation for the relationship between upstream area and fish species richness is that larger, and presumably more diverse, upstream habitats provide more sources for downstream dispersal and establishment in highly connected lakes (i.e. increasing colonization rates; MacArthur & Wilson, 1967). Alternatively, the pattern could be the result of downstream transport of eDNA (Shogren et al., 2017) leading to increased detections.
in lakes with large amounts of upstream aquatic habitat. However, when species richness from ST survey data (using traditional sampling gears) was used as the response in our GLM analysis, the upstream area model was still favored (Table S3), supporting our interpretation that increased connectivity, rather than eDNA transport, is a driver of fish diversity in the sampled lakes. Anecdotally, Pentwater Lake (where species richness was highest in both eDNA metabarcoding datasets and ST survey data) has a direct connection to the Great Lakes, which could also increase the diversity of the fish community by allowing for connectivity with, and immigration from, the diverse fish assemblage present in Lake Michigan. Our results also suggest an inverse relationship between lake area and measures of alpha diversity that incorporate relative abundance of community members (Shannon Diversity; Figure 3, Table S3). The relationship between upstream area and proportion of AIS reads is consistent with findings that species diversity, as quantified using either traditional fisheries survey gear or eDNA, scales positively with increasing upstream area. Larger upstream areas could also result in increased human access and use, additional suitable habitat for AIS, and propagate pressure into the focal lake. However, we were not able to explicitly quantify human use in lakes upstream of those sampled in this study. Future work could look to quantify human use and impacts at multiple spatial scales to explore this relationship. Results such as these illustrate the potential for eDNA metabarcoding data to address basic questions in community ecology, including those related to community assembly and invasibility.

In addition to their potential applications in community ecology, eDNA approaches have been widely used as early detection methods for invasive species. Our occupancy modelling results highlight the utility of different eDNA markers (12S and 16S) and sampling approaches (surface and benthic, however not significantly different in this analysis) for detecting individual species in environmental samples. It has been shown that marker choice greatly affects species identification and detection accuracy, therefore using multiple markers is highly recommended in metabarcoding studies (Alberdi et al., 2018; McClenaghan et al., 2020; Zhang et al., 2018). In addition, to improve species detection probabilities during large-scale metabarcoding studies, multiplexing different markers is a cost-effective and efficient way to survey natural ecosystems (McClenaghan et al., 2020). Estimates of detection probability from eDNA metabarcoding were comparable to the most efficient traditional gear types for all three surrogate species considered (Figure 4), suggesting that eDNA metabarcoding approaches are amenable for community-wide AIS detection. We note that multiple conventional gears may need to be deployed for adequate surveillance of AIS with different ecologies, which stands in contrast to the more consistent detection across species associated with eDNA metabarcoding. In traditional surveys, multiple gear types improve overall detection probability, and the same is likely true eDNA metabarcoding surveys employing multiple markers. Notably, differences in detection probability across the surrogate species considered here likely reflect differences in abundance and/or spatial distribution within the sampled lakes, a technical artefact of the assumptions of our occupancy analysis. However, comparisons of detection probability across gear types illustrate known gear-related biases (Ruetz et al., 2007; Zale et al., 2012), where certain species are more susceptible to capture with certain traditional gears, stressing the importance of matching gears with species’ traits. In some cases, such as with *N. melanostomus*, a traditional gear like small mesh fyke nets may be reliable for species detection. Given the high relative ranking of eDNA across the species, the results of these analyses suggest that eDNA detection probability is less influenced by species-specific variation. Therefore, eDNA metabarcoding provides a more robust method for simultaneously detecting members of a diverse community.

FIGURE 5 Heatmaps showing the spatial distribution of (a) overall fish species richness and (b) relative sequence abundance of an aquatic invasive species (*N. melanostomus*). Surfaces are based on the 12S eDNA metabarcoding marker.
compared to traditional gears (McElroy et al., 2020; Sard et al., 2019). Importantly, estimated detection probabilities for the four eDNA sample types (surface and benthic, 125 and 165) were similar for each species (indicating little selectivity of eDNA metabarcoding approaches for species detection; Sard et al., 2019).

Previous studies have suggested that differences in physical environments may affect the distribution of species and that visualization of fish communities within each lake would help to determine representative study sites (Sakata et al., 2020). Here, novel applications of spatial interpolation of eDNA signal (species richness and relative abundance of *N. melanostomus*) are demonstrated, which could be used in the future to prioritize AIS response efforts by identifying potential locations that would maximize the effectiveness of removal efforts. Similar spatial interpolation approaches were applied to examine relationships between eDNA signal and abundance of Japanese jack mackerel (*Trachurus japonicus*) in Maizuru Bay, Japan (Jo et al., 2017). These authors demonstrated a positive correlation between longer (719 bp) eDNA fragments and echo sounder surveys of *T. japonicus* distribution, suggesting that eDNA data can provide information on patterns of fish abundance and distribution (Jo et al., 2017). Other studies have demonstrated the usefulness of snapshot sampling for species monitoring (Bracken et al., 2019). As in our study, Bracken et al. (2019) used eDNA to reveal “hot spots” of vital habitat locations (including spawning sites) of their target species *P. marinus*, to help with possible conservation efforts. Implementing similar spatial interpolation approaches, potentially coupled with extensive sampling efforts, as necessary for more sophisticated interpolation methods, to identify “hot spots” of diversity or AIS prevalence would improve future large-scale community diversity studies. While these interpolated surfaces represent hypotheses relating to spatial patterns in relative abundance and species richness, managers may be able to make use of these patterns to guide control and eradication efforts targeting AIS. Specifically, future work should focus on determining the mechanisms that describe the distribution of eDNA in lakes, similar to what has been evaluated for lotic and marine systems (Andruszkiewics et al., 2019; Jane et al., 2015; Shogren et al., 2017; Wilcox et al., 2016). For example, environmental covariates, such as in-lake physical and biological characteristics should be collected to better quantify the relationship between eDNA detections and realized species occupancy and relative abundance (Jerde et al., 2016; Pilliod et al., 2014).

## 5 | CONCLUSIONS

The development and implementation of eDNA metabarcoding methods allow the rapid and cost-efficient collection of information on species diversity and composition of fish assemblages in aquatic habitats, which is of particular importance given the current increase in anthropogenic disturbance and associated declines in aquatic biodiversity in those ecosystems. Data in this study demonstrate that eDNA metabarcoding provides a sensitive means of collecting fish community information with several advantages relative to traditional survey gears. Our study also documents significant positive relationships between the area of upstream lakes and the diversity of the fish community inhabiting a given lake, stressing the role of connectivity in determining fish community diversity. Our case study on 22 Michigan lakes provides several recommendations for future projects using eDNA metabarcoding datasets to address questions related to community assembly and invasibility. Data collected for this project identified several aquatic invasive species in Michigan’s lakes that were not detected with traditional gear (Table S1), highlighting another advantage associated with eDNA metabarcoding. While traditional survey data are indispensable in conservation and management (for life history information, population genetic studies, and confirmation of species presence, among others) eDNA metabarcoding provides a valuable and complementary survey technique that is well suited for addressing questions in the field of community ecology.

## ACKNOWLEDGEMENTS

We thank the Michigan Department of Natural Resources and Department of Environment, Great Lakes, and Energy personnel for collecting specimens in the field. Seth Smith provided help with bioinformatic analysis. Two anonymous reviewers provided helpful suggestions for the manuscript. Computational resources for this project were provided by Michigan State University and the Institute for Cyber-Enabled Research. Funding was provided by the Michigan Department of Natural Resources and the Great Lakes Restoration Initiative, administered through the United States Fish and Wildlife Service.

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

## PEER REVIEW

The peer review history for this article is available at https://publon.com/publon/10.1111/ddi.13370.

## DATA AVAILABILITY STATEMENT

Datasets and R code for statistical analyses and raw eDNA metabarcoding sequencing data are available on Dryad (https://doi.org/10.5061/dryad.1zcrjdfs9).

## ORCID

John D. Robinson https://orcid.org/0000-0002-9782-1598

## REFERENCES

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, 9, 134–137. https://doi.org/10.1111/2041-210X.12849

Anderson, C., Epperson, B. K., Fortin, M.-J., Holderegger, R., James, P., Rosenberg, M., Scribner, K. T., & Spear, S. (2010). Consideration of scale in landscape genetics can improve predictive relationships between measures of spatial genetic structure and landscape pattern. *Molecular Ecology*, 19, 3565–3575.

Andruszkiewics, E. A., Koseff, J. R., Fringer, O. B., Ouellette, N. T., Lowe, A. B., Edwards, C. A., & Boehm, A. B. (2019). Modeling environmental
DNA transport in the coastal ocean. Using Lagrangian particle tracking. Mar. Pollut. Bull., 78, 1268–1277.

Bartier, P. M., & Keller, C. P. (1996). Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW). Computers & Geosciences, 22(7), 795–799. https://doi.org/10.1016/0098-3004(96)00021-0

Barton, K. (2020). MuMIn: Multi-model inference. R package version 1.43.17.

Benedikt, G., Pebesma, E., & Heuvelink, G. (2016). Spatio-Temporal Interpolation using gstat. The R Journal, 8(1), 204–218. https://doi.org/10.32614/RJ-2016-014

Bivand, R. S., Keitt, T., & Rowlingson, B. (2020). rgdal: Binding to the ‘Geospatial’ Data Abstraction Library. R package version 1.5-18.

Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). Applied spatial data analysis with R (2nd ed.). Springer.

Bivand, R. S., Rundel, C., Pebesma, E., Stuetz, R., Hufthammer, K. O., Giraudoux, P., & Santilli, S. (2020). rgdal: Interface to Geometry Engine - Open Source (‘GEOS’). R package version 0.5-5.

Bracken, F. S. A., Rooney, S. M., Kelly-Quinn, M., King, J. J., & Carlsson, J. (2019). Identifying spawning sites and other critical habitat in lotic systems using eDNA “snapshots”: A case study using the sea lamprey Petromyzon marinus L. Ecology and Evolution, 9(1), 553–567.

Chen, F.-W., & Liu, C.-W. (2012). Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. Paddy and Water Environment, 10, 209–222. https://doi.org/10.1007/s10333-012-0319-1

Coble, D. W., Buresewitz, R. E., Fratt, T. W., & Scheirer, J. W. (1990). Lake trout, sea lampreys, and overfishing in the upper Great Lakes Region: An overview and reanalysis. Transactions of the American Fisheries Society, 119, 985–995. https://doi.org/10.1577/1548-8659(1990)119<0985:TLSTLO>2.0.CO;2

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

Deagle, B. E., Kirkwood, R., & Jarman, S. N. (2009). Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. Molecular Ecology, 18(9), 2022–2038. https://doi.org/10.1111/j.1365-294X.2009.04158.x

Deiner, K., & Alternatt, F. (2014). Transport distance of invertebrate environmental DNA in a natural river. PLoS One, 9(2), e88786. https://doi.org/10.1371/journal.pone.0088786

Dunnington, D. (2017). prettypmap: Scale bar, North arrow, and pretty margins in R. R package version 0.2.2.

Escobar, L. E., Mallez, S., McCartney, M., Lee, C., Zielinski, D. P., Ghosal, R., Phelps, N. B. D. (2018). Aquatic invasive species in the Great Lakes Region: An overview. Reviews in Fisheries Science & Aquaculture, 26(1), 121–138. https://doi.org/10.1080/233082849.2017.1363715

Esselmann, P. C., Infante, D. M., Wang, L., Cooper, A. R., Wieferich, D., Tsang, Y.-P., Thornbrugh, D. J., & Taylor, W. W. (2013). Regional fish community indicators of landscape disturbance to catchments of the conterminous United States. Ecological Indicators, 26, 163–173. https://doi.org/10.1016/j.ecolind.2012.10.028

Fagan, W. F. (2002). Connectivity, fragmentation, and extinction risk in dendritic metapopulations. Ecology, 83(12), 3243–3249.

Gallien, L., & Carboni, M. (2017). The community ecology of invasive species: Where are we and what’s next. Ecology, 94, 335–352. https://doi.org/10.1111/ecocep.02446

Green, R. H. (1979). Sampling design and statistical methods for environmental biologists. Wiley.

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

Harper, L. R., Lawson Handley, L., Hahn, C., Boonham, N., Rees, H. C., Lewis, E., ... Häfling, B. (2020). Generating and testing ecological hypotheses at the pondscape with environmental DNA metabarcoding: A case study on a threatened amphibian. Environmental DNA, 2(2), 184–199. https://doi.org/10.1002/edn3.57

Hijmans, R. J. (2020). raster: Geographic data analysis and modeling.

Jane, S. F., Wilcox, T. M., McKelvey, K. S., Young, M. K., Schwartz, M. K., Lowe, W. H., ... Whiteley, A. R. (2015). Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. Molecular Ecology Resources, 15, 216–227. https://doi.org/10.1111/1755-0998.12285

Jerde, C. L. (2021). Can we manage fisheries with the inherent uncertainty from eDNA? Journal of Fish Biology, 98, 341–353. https://doi.org/10.1111/jfb.14218

Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). “Sight-unseen” detection of rare aquatic species using environmental DNA. Conservation Letters, 4(2), 150–157. https://doi.org/10.1111/j.1755-263X.2010.00158.x

Jo, T., Murakami, H., Masuda, R., Sakata, M. K., Yamamoto, S., & Minamoto, T. (2017). Rapid degradation of longer DNA fragments enables the improved estimation of distribution and biomass using environmental DNA. Molecular Ecology Resources, 17(6), e25–e33.

Jocelyn, P., Knight, T. W., & Ferguson, M. M. (2005). Nonequilibrium conditions following landscape rearrangement: The relative contribution of past and current hydrological landscapes on the genetic structure of a stream-dwelling fish. Molecular Ecology, 14(5), 1321–1331. https://doi.org/10.1111/j.1365-294X.2005.02500.x

Jude, D. J., Reider, R. H., & Smith, G. R. (1992). Establishment of Gobiidae in the Great Lakes basin. Canadian Journal of Fisheries and Aquatic Sciences, 49(2), 416–421. https://doi.org/10.1139/f92-047

Keitt, T. (2012). colorRamps: Builds color tables. R package version 2.3.

Keller, K. (2018). jagsUI: A Wrapper Around ’jags’ to Streamline ’JAGS’ Analyses. R package version 1.5.0. https://CRAN.R-project.org/package=jagsUI

Kelly, R. P., Shelton, A. O., & Gallego, R. (2019). Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. Scientific Reports, 9(1), 1–14. https://doi.org/10.1038/s41598-019-48546-x

Kennedy, S. R., Prost, S., Overcast, I., Rominger, A. J., Gillespie, R. G., & Krehenwinkel, H. (2020). High-throughput sequencing for community analysis: The promise of DNA barcoding to uncover diversity, relatedness, abundances and interactions in spider communities. Development Genes and Evolution, 230, 185–201. https://doi.org/10.1007/s00427-020-00652-x

Kerry, M. (2010). Introduction to WinBUGS for Ecologists: A Bayesian approach to regression, ANOVA, mixed models, and related analyses. Academic Press.

Kumar, G., Eble, J. E., & Gaither, M. R. (2020). A practical guide to sample preservation and pre-PCR processing of aquatic environmental DNA. Molecular Ecology Resources, 20(1), 29–39. https://doi.org/10.1111/1755-0998.13107

Lamy, T., Jarne, P., Laroche, F., Pointier, J. P., Huth, G., Segard, A., & David, P. (2013). Variation in habitat connectivity generates positive correlations between species and genetic diversity in a metacommunity. Molecular Ecology, 22(17), 4445–4456. https://doi.org/10.1111/mec.12399

Laramie, M. B., Pilliod, D. S., & Goldberg, C. S. (2015). Characterizing the distribution of an endangered salmonid using environmental DNA analysis. Biological Conservation, 183, 29–37. https://doi.org/10.1016/j.biocon.2014.11.025
Larmaneg, J., Roselyne, V., Seydou, Y., Philippe, M., & Nicolas, M. (2011). Methods for mapping regional trends of HIV prevalence from Demographic and Health Surveys (DHS). *Cybergeo: European Journal of Geography*, 558.

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

MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biology*. Princeton University Press.

Mackenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8), 2248–2255.

Magurran, A. E. (2013). *Ecological diversity and its measurement*. Princeton University Press.

Mahmoudian, M. (2020). *VarHandle: Functions for robust variable handling*. R package version 2.0.5.

Mathieu, C., Hermans, S. M., Lear, G., Buckley, T. R., Lee, K. C., & Buckley, H. L. (2020). A systematic review of sources of variability and uncertainty in eDNA data for environmental monitoring. *Frontiers in Ecology and Evolution*, 8, 135. https://doi.org/10.3389/feevo.2020.00135

McClanaghan, B., Compson, Z. G., & Hajibabaei, M. (2020). Validating metabarcoding-based biodiversity assessments with multispecies occupancy models: A case study using coastal marine eDNA. *PloS One*, 15(3), e0224119. https://doi.org/10.1371/journal.pone.0224119

McElroy, M. E., Dressler, T. L., Titcomb, G. C., Wilson, E. A., Deiner, K., Dudley, T. L., ... Jerde, C. L. (2020). Calibrating Environmental DNA Metabarcoding to Conventional Surveys for Measuring Fish Species Richness. *Frontiers in Ecology and Evolution*, 8(276). https://doi.org/10.3389/feevo.2020.00276

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

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences*, 30(7), 683–691. https://doi.org/10.1016/j.cageo.2004.03.012

Pebesma, E. J., & Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5, 9–13.

Perkin, J. S., & Gido, K. B. (2012). Fragmentation alters stream fish community structure in dendritic ecological networks. *Ecological Applications*, 22(8), 2176–2187. https://doi.org/10.1890/12-0318.1

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., & Waits, L. P. (2014). Factors influencing detection of eDNA from a stream-dwelling amphibian. *Molecular Ecology Resources*, 14, 109–116. https://doi.org/10.1111/1755-0998.12159

Pinfield, R., Dillane, E., Runge, A. K. W., Evans, A., Mirimin, L., Niemann, J., ... Foote, A. D. (2019). False-negative detections from environmental DNA collected in the presence of large numbers of killer whales (*Orcinus Orca*). *Environmental DNA*, 1, 316–328.

Prince, A. M., & Andrus, L. (1992). PCR: How to kill unwanted DNA. *BioTechniques*, 12(3), 358–359.

Procháš, Š., Wilson, J. R. U., Richardson, D. M., & Rejmánek, M. (2008). Searching for phylogenetic pattern in biological invasions. *Global Ecology and Biogeography*, 17, 5–10.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., & Gough, K. C. (2014). REVIEW The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51(5), 1450–1459. https://doi.org/10.1111/1365-2664.12306

Riaz, T., Shehzad, W., Viari, A., Pomponio, F., Taberlet, P., & Coissac, E. (2011). ecoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, 39(21), e145. https://doi.org/10.1093/nar/gk732

Rice, J., Moksness, E., Attwood, C., Brown, S. K., Dahle, G., Gjerde, K. M., & Westlund, L. (2012). The role of MPAs in reconciling fisheries management with conservation of biological diversity. *Ocean & Coastal Management*, 69, 217–230.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahe, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2584. https://doi.org/10.7717/peerj.2584

Ruetz, C. R., Uzarski, D. G., Krueger, D. M., & Rutherford, E. S. (2007). Sampling a Littoral Fish Assemblage: Comparison of small-mesh fyke netting and boat electrofishing. *North American Journal of Fisheries Management*, 27(3), 825–831. https://doi.org/10.1002/nafm.147.1

Sakata, M. K., Watanabe, T., Maki, N., Ikeda, K., Kosuge, T., Okada, H., ... Minamoto, T. (2020). Determining an effective sampling method for eDNA metabarcoding: A case study for fish biodiversity monitoring in a small, natural river. *Limnology*, 22, 221–235. https://doi.org/10.1007/s10201-020-00645-9

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

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., ... Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23), 7537–7541. https://doi.org/10.1128/AEM.01541-09

Schlosser, J. I. (1991). Stream fish ecology: A landscape perspective. *BioScience*, 41(10), 704–712. https://doi.org/10.2307/1311765

Schmidt, B. R., Kery, M., Ursenbacher, S., Hyman, O. J., & Collins, J. P. (2013). Site occupancy models in the analysis of environmental DNA presence/absence surveys: A case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution*, 4(7), 646–653. https://doi.org/10.1111/2041-210X.12052

Schneider, J. C. (2000). *Manual of Fisheries Survey Methods II: With periodic updates*. Fisheries Division, Michigan Department of Natural Resources.

Sepulveda, A. J., Nelson, N. M., Jerde, C. L., & Luikart, G. (2020). Are environmental DNA methods ready for aquatic invasive species management? *Trends in Ecology & Evolution*, 35(8), 668–678. https://doi.org/10.1016/j.tree.2020.03.011

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423. https://doi.org/10.1002/10022772

Sharma, S., Legendre, P., DeCaceres, M., & Boisclair, D. (2011). The role of environmental and spatial processes in structuring native and non-native fish communities across thousands of lakes. *Ecography*, 34, 762–771.

Shogren, A. J., Tank, J. L., Andruszkiewicz, E., Olds, B., Mahon, A. R., Jerde, C. L., & Bolster, D. (2017). Controls on eDNA movement in streams: Transport, retention, and resuspension. *Scientific Reports*, 7, 5065. https://doi.org/10.1038/s41598-017-05223-1

Soranno, P. A., Bacon, L. C., Beauchene, M., Bednar, K. E., Bissell, E. G., Boudreau, C. K., … Yuan, S. (2017). LAGOS-NE: A multi-scaled geo-spatial and temporal database of lake ecological context and water quality for thousands of US lakes. *GigaScience*, 6(12), gix101. https://doi.org/10.1093/gigascience/gix101

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). *MEGA6*: Molecular evolutionary genetics analysis Version 6.0.
BIOSKETCH

Our inter-disciplinary eDNA research team is composed of academic and agency scientists with shared interests in using advanced assessment technology and analytical methods to quantify lake fish community diversity and species relative abundance, including aquatic invasive species (AIS). A major focus of our team has been to incorporate landscape-scale environmental variables characterizing surrounding lakes, measures of human disturbance, spatial mapping, statistical methods, and occupancy modelling to interpret patterns in fish community diversity and to inform AIS monitoring and control.

Author contributions: The experimental design was developed and funding secured by JDR, KTS, LN, and SJH. ALH, EW, and JDR collected landscape environmental covariate data. LP, NMS, SJH, ALH, JDR, and EW participated in collection of field water samples. LP, JK, and ALH conducted laboratory analyses. LP, LN, NMS, KTS, and NMS conducted statistical analyses. All authors were involved with manuscript preparation and review.

SUPPORTING INFORMATION

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

How to cite this article: Pukk, L., Kanefsky, J., Heathman, A. L., Weise, E. M., Nathan, L. R., Herbst, S. J., Sard, N. M., Scribner, K. T., & Robinson, J. D. (2021). eDNA metabarcoding in lakes to quantify influences of landscape features and human activity on aquatic invasive species prevalence and fish community diversity. Diversity and Distributions, 27, 2016–2031. https://doi.org/10.1111/ddi.13370