Received: 22 October 2021 | Revised: 12 May 2022 | Accepted: 22 June 2022

DOI: 10.1002/eap.2730

ARTICLE

ECOLOGICAL APPLICATIONS
ECOLOGICAL SOCIETY OF AMERICA

Integrating invasive species risk assessment into environmental DNA metabarcoding reference libraries

Andrew R. Mahon[1] | Erin K. Grey[2] | Christopher L. Jerde[3]

[1]Department of Biology, Central Michigan University, Mount Pleasant, Michigan, USA

[2]School of Biology and Ecology and Maine Center for Genetics in the Environment, University of Maine, Orono, Maine, USA

[3]Marine Science Institute, University of California Santa Barbara, Santa Barbara, California, USA

Correspondence
Christopher L. Jerde
Email: cjerde@ucsb.edu

Funding information
BOEM, Grant/Award Number: MC15AC00006; NASA, Grant/Award Number: NNX14AR62A; NOAA; National Science Foundation, Grant/Award Number: OIA-1849227

Handling Editor: Thomas W. Therriault

Abstract

Environmental DNA (eDNA) metabarcoding has shown promise as a tool for estimating biodiversity and early detection of invasive species. In aquatic systems, advantages of this method include the ability to concurrently monitor biodiversity and detect incipient invasions simply through the collection and analysis of water samples. However, depending on the molecular markers chosen for a given study, reference libraries containing target sequences from present species may limit the usefulness of eDNA metabarcoding. To explore the extent of this issue and how it may be resolved to aid biodiversity and invasive species early detection goals, we focus on fishes in the well-studied Laurentian Great Lakes region. First, we provide a synthesis of species currently known from the region and of non-indigenous species identified as threats by international, national, regional, and introduction pathway-specific fish risk assessments. With these species lists, we then evaluate 23 primer pairs commonly used in fish eDNA metabarcoding with available databases of sequence coverage and species specificity. Finally, we identify established and potentially invasive non-indigenous fish that should be prioritized for genetic sequencing to ensure robust eDNA metabarcoding for the region. Our results should increase confidence in using eDNA metabarcoding for fisheries conservation and management in the Great Lakes region and help prioritize reference sequencing efforts. The ultimate utility of eDNA metabarcoding approaches will come when conservation management of existing fish communities is integrated with early detection efforts for invasive species surveillance to assess total fish biodiversity.

KEYWORDS

early detection, eDNA, Laurentian Great Lakes, non-indigenous species, sequence coverage, species specificity, threatened and endangered species

INTRODUCTION

From the earliest studies of American bullfrogs in French ponds (Ficetola et al., 2008) and Asian carps swimming towards the Laurentian Great Lakes (Jerde et al., 2011), environmental DNA (eDNA) applications have focused on the detection of aquatic invasive species. The motivation for these studies emerged from a need for early detection of invasive species before they have the potential to establish, spread, and cause irreversible ecological and economic impacts (Lodge et al., 2006). The increased detection sensitivity from single species eDNA methods versus conventional detection gears, most clearly demonstrated in fish (Jerde, 2021; Wilcox et al., 2016), has improved early detection capabilities and has been touted as a reliable and advantageous approach for early detection across many taxa (Sepulveda, Nelson, et al., 2020).

Shortly after the initial single species eDNA applications were developed and applied in the field, high throughput sequencing and genomics platforms offered a second approach allowing for multispecies communities to be examined (Thomsen et al., 2012; Valentini et al., 2016). This eDNA metabarcoding approach has resulted in reliable estimates of a system's species richness (McElroy et al., 2020; Olds et al., 2016) and beta-diversity patterns (Grey et al., 2018; Li et al., 2018; Mächler et al., 2020) and is primarily used for conservation management of entire communities (Deiner et al., 2017; Harper et al., 2019). As with single species eDNA approaches, invasive fish species were at the forefront of these multispecies surveillance applications, wherein the search for bighead and silver carp in the Ohio River basin, USA, the invasive Northern Snakehead was unexpectedly detected genetically, and its DNA was subsequently found again in a targeted single species approach (Simmons et al., 2016).

However, single and multispecies eDNA approaches are not without issues, particularly for invasive species management applications with costs for both false positive and false negative detections (Darling & Mahon, 2011; Sepulveda, Nelson, et al., 2020). A false positive detection, or a genetic detection of a live species when it is not actually present, can trigger unnecessary management actions for invasive species. Numerous reasons, such as contamination (Sepulveda, Hutchins, Forstchen, et al., 2020) and DNA transport in lentic systems (Shogren et al., 2017), can lead to sample-level false positives and site-level inferential errors (Darling et al., 2021). Alternatively, false negative detections, or failing to detect a species present in an area, can lead to the establishment and spread of an invasive species and associated ecological and economic harms. Although well-designed eDNA approaches typically have lower false-negative rates than other methods, they may still be insufficient for effective early detection (Erickson et al., 2019). Ultimately, genetic approaches (indirect detection) are compared to conventional survey approaches (direct detection such as nets and camera traps), which comes with inferential pitfalls if the conventional surveillance approach has a low probability of detection but is, nevertheless, used to assess the performance of single species (Jerde, 2021) or multispecies (McElroy et al., 2020) eDNA approaches. Given these issues, there has been some trepidation in building the interfaces for decision support tools to trigger active management for eradication or control and containment of species detected with eDNA methods (Sepulveda, Nelson, et al., 2020).

One additional issue for eDNA metabarcoding studies, particularly in systems with high species richness coupled with endemism and new species introductions, is the completeness of reference databases used to match recovered eDNA sequence data back to a vouchered genetic sequence (Marques et al., 2021; Stoeckle et al., 2020). Lack of reference sequences, coupled with the fact that primers do not necessarily amplify or distinguish between all species in a sample, increases false negative rates in eDNA metabarcoding applications. For example, Jerde et al. (2021) showed that of the 1345 fish species in the Mekong River Basin, only 782 (58.1%) had a reference sequence in one of the 23 fish eDNA metabarcoding primer pair loci assessed. When only one primer pair (i.e., a single targeted amplicon), albeit the primer pair with the best coverage (the mitochondrial 16S; Shaw et al., 2016) is used, only 643 (47.8%) fish had a reference sequence, and many of those were not differentiable at the species level. The much needed GAPeDNA application (Marques et al., 2021) uses species lists generated from a database of known occurrences by watershed (Tedesco et al., 2017) to evaluate the performance of 23 commonly used fish eDNA metabarcoding primer pairs in terms of coverage, or the availability of a reference sequence and the ability of a primer set to amplify that sequence. Among many interesting patterns identified, Marques et al. (2021) found that coverage of tropical species was relatively low compared to non-tropical species and that coverage of established non-indigenous species was generally higher than indigenous species, but with large gaps depending on the locus.

Along with species known to be present in a region, whether indigenous or non-indigenous, eDNA metabarcoding surveys used for early detection of invasive species should also consider species not currently present but likely to be introduced. Yet the question of how to connect the reference database to species that are likely not currently present but have a high chance of arriving and causing damages remains. The answer to this lies with another tool used in invasive species management—risk assessment. For invasive species, risk assessment has been

used to identify species likely to arrive at a novel location, with the life history and habitat matching characteristics to survive, and some properties that make them likely to cause damages (Andersen et al. 2004; Kolar & Lodge, 2001; Lodge, 1993). Fishes arriving at the Great Lakes were one of the first examples of quantitative invasive species risk assessment (Kolar & Lodge, 2002), but now qualitative and quantitative risk assessments have been conducted in different regions for many invasive taxa. For fish, perhaps the most widely used is the Fish Invasiveness Screening Kit (FISK; Copp, 2013), a question-based spreadsheet tool adapted from the Australian Weed Risk Assessment that has been applied over 1900 times across 45 countries and six continents (Vilizzi et al., 2019). Other fish risk assessments tools have been developed to meet regulatory needs (e.g., the probabilistic US Fish and Wildlife's Freshwater Fish Injurious Species Risk Assessment model, Marcot et al., 2019), to provide more quantitative trait-based predictions (Chan et al., 2021; Howeth et al., 2016), to consider vector-specific factors (Chan et al., 2013), or to incorporate more detailed knowledge of species environmental tolerances (Gallardo et al., 2013) and habitat requirements (Poulos et al., 2012). Risk assessments can inform policy to reduce propagule pressure and the probability of accidental introduction and robust early detection surveillance programs that use either conventional direct capture methods or, as we show, passive surveillance or indirect detection methods such as eDNA metabarcoding (Simmons et al., 2016).

In this study, we expand and evaluate reference libraries used for fish eDNA metabarcoding to include non-indigenous species using the Laurentian Great Lakes (Figure 1), a region with an ecologically and economically valuable fishery and multiple invasive fish risk assessments, as our model system. Despite the importance of monitoring for early detection for AIS management, there are currently no system-wide efforts to survey all fishes of the Great Lakes (Trebitz et al., 2017). The advantage of a metabarcoding method that we explore here is that a given water sample can be used to monitor existing fish biodiversity and provide a survey for the early detection of incipient invaders into the region. In this process, we (a) produce a comprehensive species list of fishes currently present in the Laurentian Great Lakes, (b) provide a synthesis of fish risk assessments relevant to the Great Lakes, (c) evaluate the primer pairs commonly used in fish eDNA metabarcoding that currently have the best coverage and species specificity to increase confidence by managers in using eDNA metabarcoding for fisheries conservation and management, and (d) identify established and potentially invasive fish that should be prioritized for genetic sequencing to ensure more robust eDNA metabarcoding

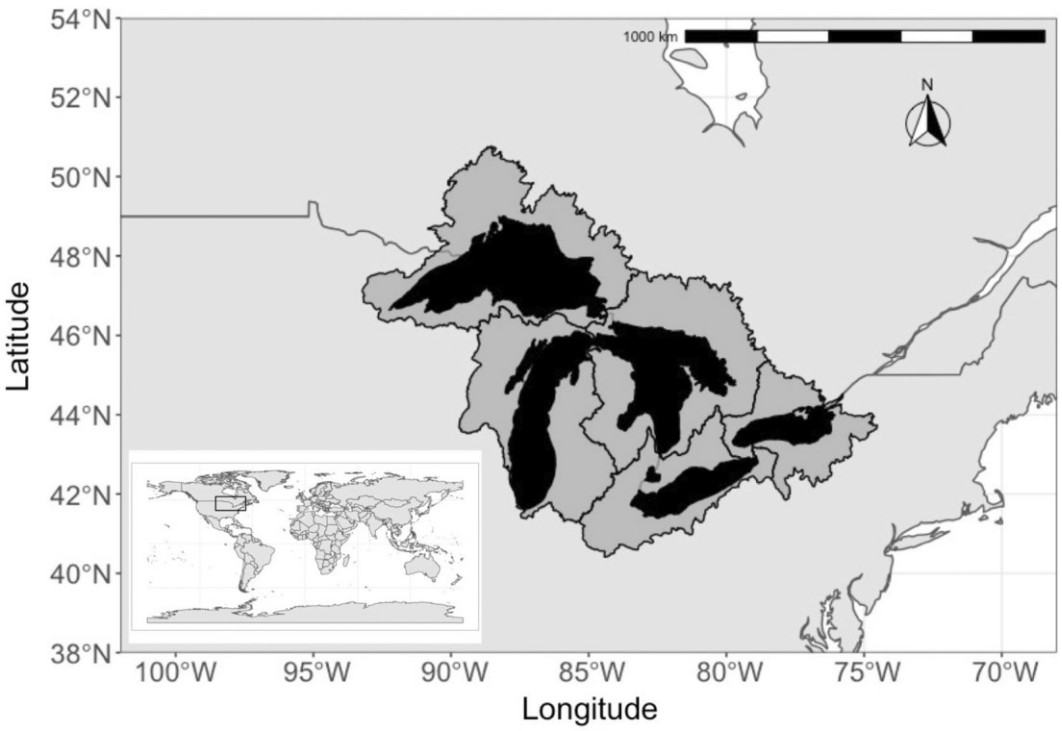

FIGURE 1 Map of the Laurentian Great Lakes (U.S. and Canada) basin delineation used for this study (black = lake, dark gray = watershed of the lake basin). Marine and estuarine species in the St. Lawrence Seaway are not included in this study. The black rectangle within the inset world map delimits the panel area.

going forward for the region. Our work outlines the current limitations and future potential of eDNA metabarcoding for joint biodiversity monitoring and early detection of harmful invasive species to support the conservation and management of freshwater ecosystems.

METHODS

Species lists for the Great Lakes

We merged two databases to generate an inclusive list of fish species currently present in the Great Lakes (GL) basin. The first species list is the default generated by the GAPeDNA program for freshwater fish (GAP; Marques et al., 2021). This list is sourced from a global database of freshwater fish occurrences by basin and was compiled by extensive searches of available peer-reviewed literature, reports, and theses (Tedesco et al., 2017). The second list was compiled by Roth et al. (2013) as a checklist of fish species found within the Great Lakes and their watersheds (ROTH). Both lists were compared to reconcile species name changes or synonyms using FishBase (Froese & Pauly, 2021).

Potential invaders

To complement our biodiversity database for fish already present in the GL basin, we identified fish risk assessments of potential invasive freshwater species conducted for the Laurentian Great Lakes, the United States, and globally. We considered both peer-reviewed literature and gray literature publications. For the Great Lakes, we identified three lists: Snyder et al. (2014), Davidson, Tucker, Chadderton, and Weibert (2021), and an application of the Howeth et al. (2016) quantitative risk assessment to the Great Lakes available at http://takeaim.org/wp-content/uploads/2016/11/FishRA_assesses_species_PC_May_19.pdf. The species lists for these risk assessments are denoted as Snyder, Davidson, and NDSTAIR, respectively. While a decision tree approach to Great Lakes fish risk assessment was developed by Kolar and Lodge (2002) for the Ponto Caspian region of origin, we used the update of this approach provided by Snyder et al. (2014) for this study. The Davidson list is comprehensive to identify invasive species to the entirety of the Great Lakes, but Snyder considers only potential invaders from the Ponto Caspian region. The NDSTAIR risk assessment emphasizes the trade pathway and considers only that vector for fish introduction.

At the national level, we identified two risk assessments: the US Fish and Wildlife High Risk Species (https://www.fws.gov/fisheries/ANS/erss_high_risk.html) used for rapid assessment of potentially invasive fish that is coupled to a Bayesian network decision tool (Marcot et al., 2019), and the species listed under the US Lacey Act found at https://www.fws.gov/injuriouswildlife/list-of-injurious-wildlife.html. The species lists for these risk assessments are denoted as USFWS and Lacey, respectively. The former is meant to inform resources managers, stakeholders, and the public about species with the potential to become invasive and motivate further research. It is based on two key factors, the similarity of climate between the native and established range and the United States and the history of invasiveness. The Lacey Act fish list is reactive to invasive species identified as causing damages (Fowler et al., 2007), but in our framework, we consider the listed fish that have not become established in the Great Lakes.

Two risk assessments identified potentially invasive freshwater fish at the global scale—the "100 of the World's Worst Invasive Alien Species Lists" developed by the IUCN's Global Invasive Species Database in 2014 and available at http://www.iucngisd.org/gisd/100_worst.php, and a list of potential invaders for current and future climates as determined by the FISK risk assessment tool (Copp, 2013; Vilizzi et al., 2019). The species lists for these risk assessments are denoted as WW and Vilizzi, respectively. In contrast to the USFWS risk assessment, these fish may not have an appropriate climate match for the Great Lakes, but given their history of global invasiveness and as Vilizzi et al. (2019) pointed out that with our changing climate, it is potentially advantageous to consider further some of these "unlikely to establish under current conditions but known to be damaging" fish for surveillance.

For purposes of reference library evaluation, we consider three composite lists: (1) Fish currently present in the Great Lakes (Roth and GAP), (2) Fish currently in the Great Lakes and fish likely to establish in the Great Lakes (Roth, GAP, Snyder, Davidson, and NDSTAIR), and (3) Fish currently in the Great Lakes, fish likely to establish in the Great Lakes, and fish broadly identified as invasive in the US or Globally (Roth, GAP, Snyder, Davidson, NDSTAIR, USFWS, Lacey, WW, and Vilizzi). To visualize the overlap of these databases, we constructed Euler diagrams in the R package EulerR (Larsson, 2021).

Reference library evaluation

Primer coverage

We screened for coverage of 23 common fish metabarcoding primer pairs using publicly available reference sequence data for each Great Lakes and potential invader species list composite. These primer pairs are within one of four

routinely targeted mitochondrial gene regions (12S, 16S, COI, cytB) or one nuclear gene region (18S) (Figure 2). Coverage was initially assessed using the GAPeDNA web interface (Marques et al., 2021). This program searches the European Nucleotide Archive for reference sequence data and uses the *ecoPCR* function (Ficetola et al., 2010) to align primers to each sequence, allowing up to three mismatches with each primer. We chose to follow the GAPeDNA default cutoff of three mismatches because a previous study found that ≥4 mismatches in a single primer was required to block a polymerase chain reaction (PCR) reaction completely (Lefever et al., 2013). However, we note that no single in silico primer evaluation method or cutoff accurately predicts all outcomes of in vitro PCR reactions, which themselves can vary by reagents and thermocycling conditions (So et al., 2020). Rank order bar charts of the number of species with reference sequences for each primer pair are used to visualize and assess coverage. The database with all fish species considered in this study, membership to established Great Lakes list or risk assessment list(s), and presence or absence of marker coverage for the 23 primer pairs is provided in Appendix S1.

One eDNA metabarcoding approach to improve the coverage of species having sequences information is to use multiple amplicons/primer pairs (Evans et al., 2016, 2017; McElroy et al., 2020; Pitz et al., 2017). We evaluated the use of multiple primer pairs for established Great Lakes species and fish identified using Great Lakes-specific risk assessments. By conducting stepwise forward selection, we first selected the primer pair that maximized species coverage and then subsequently selected the next primer pair that maximized the remaining species that did not have sequence coverage. We stepped through this process until all fish species with at least one sequence in a primer pair were included. We then plotted the species accumulation as a

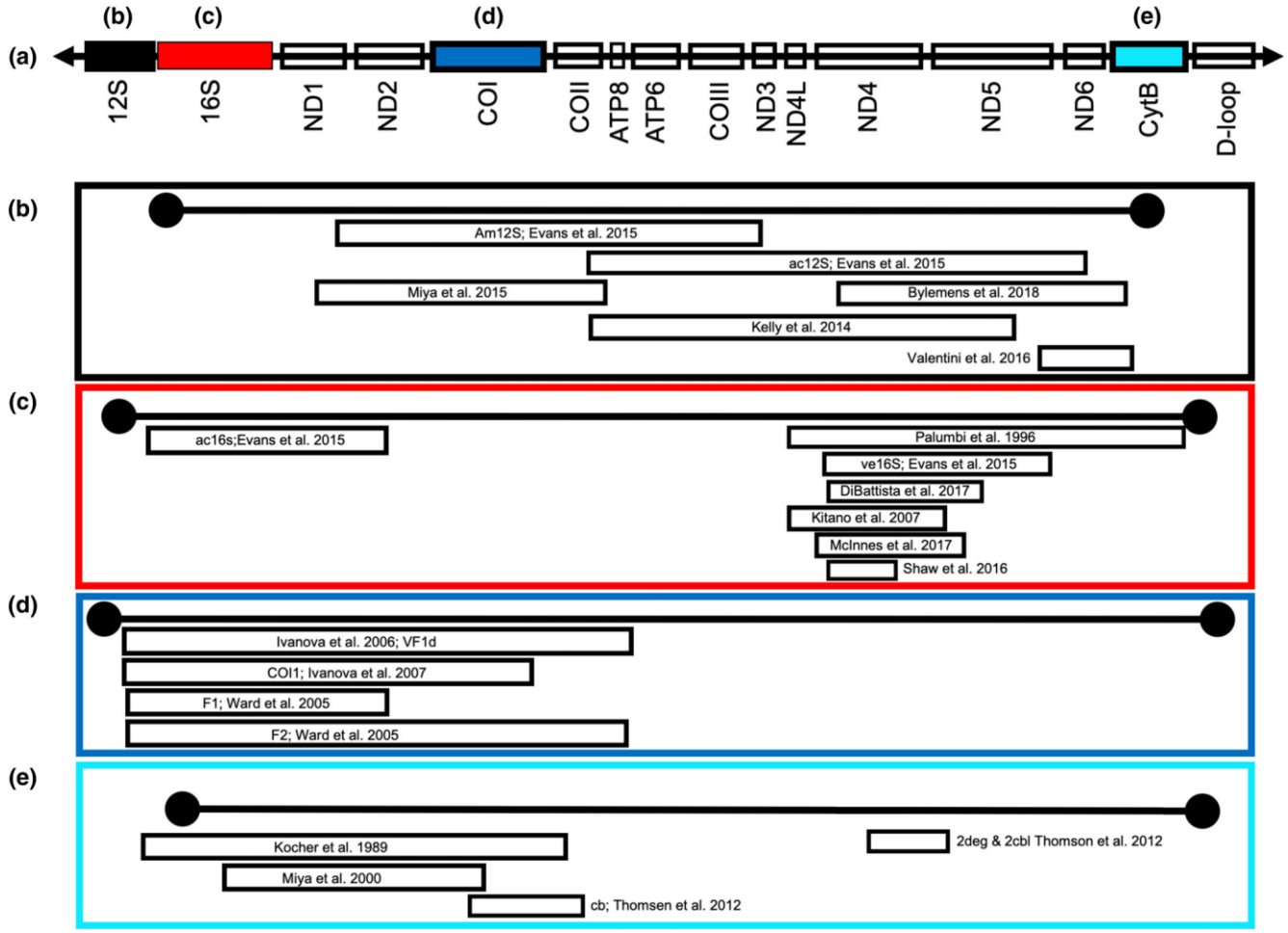

FIG URE 2 Distribution of primer pairs (as provided by Marques et al., 2021) and their relative size and location across a representative mitochondrial genome (based on the *Channa argus* mitochondrial genome) (a). Additionally, gene order is representative of that found in northern snakehead (*Channa argus*). Colors (filled boxes in panel [a] and colored outlines of panels [b–e]) denote the 12S region (black; panel), the 16S region (red; panel), the COI region (dark blue; panel), and the cytB region (light blue; panel). The representative location (i.e., location of where the amplicon is located on the gene) of each primer pair used in metabarcoding studies is shown in the panels. The size of each amplicon is also scaled relative to the gene and to the other included amplicons.

function of primer pair. This evaluation of coverage does not account for species specificity.

Species specificity

To investigate species-specificity in terms of distinguishing between species using a particular metabarcoding marker/amplicon, we quantified between-species sequence divergence for the three most commonly used target genes (16S, cytB and 12S) for the established Great Lakes species (Roth and GAP). Previous studies in other systems with higher levels of diversity found these mitochondrial regions had more coverage for fish species (Jerde et al., 2021; Marques et al., 2021), which was confirmed in our preliminary analysis of primer coverage for the Great Lakes. Comparing percent sequence divergence (as uncorrected *p* distance) is a conservative measure of specificity (i.e., the ability to distinguish between species; Meyer & Paulay, 2005; e.g., Mahon et al., 2008, for implementation). We note that species-level differences should be approximately 5% or greater for comparing gene fragments between different, distinct species. Sequence data for all Great Lakes fishes were downloaded for the three target genes from GenBank (http://ncbi.nlm.nih.gov). Complete gene fragments were used from whole mitochondrial genomes when available to provide a conservative estimate across the entire genetic element. The commonly used eDNA primer pairs used in fisheries metabarcoding studies (Marques et al., 2021) for these genes are not the same length as the entire gene itself; however using the whole (or maximum available portion) allowed us to provide a conservative estimate of sequence divergence to determine percent sequence divergence between species.

Alignments for each of the three individual datasets were completed using MAFFT v7.48 (Katoh & Standley, 2013). Aligned datasets were imported into MegaX (Kumar et al., 2018), and percent sequence divergences were calculated between each species (as uncorrected *p* distances). The percent divergences were summarized using descriptive statistics.

RESULTS

Species lists for the Great Lakes

The GAP ($n = 201$) and ROTH ($n = 176$) Great Lake fish species lists were very consistent with each other (Figure 3a). After binomial nomenclature reconciliation, two species were added to the GAP database: *Carpiodes carpio* and *Cyprinella whipplei*. Both species were justified in Roth et al., 2013 as being captured in the Great Lakes region. Four species were removed from the GAP database: *Microgadus tomcod*, *Morone saxatillis*, *Moxostoma hubbsi*, and *Myoxocephalus quadricornis*, as these species were endemic to rivers of the St. Lawrence Seaway. The GAP and ROTH shared 174 fish species; GAP had 27 species not in ROTH; ROTH had two species not in GAP. The total species richness (the union of GAP and ROTH) for the Great Lakes was 203 fishes.

The GAPeDNA interface also provides the International Union for Conservation of Nature (IUCN) red list conservation status for each species evaluated (Marques et al., 2021). In the Great Lakes, 15 species are identified as extinct, critically endangered, endangered, vulnerable, or near threatened. Of these 15 species of conservation concern, six species have no primer pair coverage (7.3% of known Great Lakes fishes). Of the 203 known fishes in the Great Lakes, 22 have no primer pair coverage (10.8%) and four of these 22 species without coverage are non-indigenous. Details and species lists are provided in Appendix S1.

Potential invaders

Within the Great Lakes risk assessments, the comprehensive Davidson assessment identified the most potential invaders (23 species), with the Snyder assessment identifying nine species from the Ponto–Caspian region and NDSTAIR identifying four species from the trade vector (Figure 3b). While Snyder and NDSTAIR largely overlapped with Davidson, each identified unique potential invaders (Snyder identified five such species and NDSTAIR one). Furthermore, no one species was identified as a risky invader in all three Great Lakes risk assessments. Taken together, these results highlight the benefit of taking a multi-faceted approach to risk assessment that considers both generally predictive factors like climate, but also the intricacies associated with sources regions and vectors.

At the scale of the USA, the USFWS risk assessment identified the most potential invaders (63 species), with the Lacey Act, previously criticized for being too slow and conservative for invasive species prevention (Fowler et al., 2007), comprising a small 16 species subset of the USFWS list. Potential invaders from global freshwater-fish risk assessments were also a small subset of the USFWS list, with the World Wildlife fund's Top 100 list (WW) identifying three species and the Villizi risk assessment identifying four species.

Seventy five species were identified as potentially invasive in at least one of the seven risk assessments considered here. The minority were found in multiple assessments: six species were found four assessments

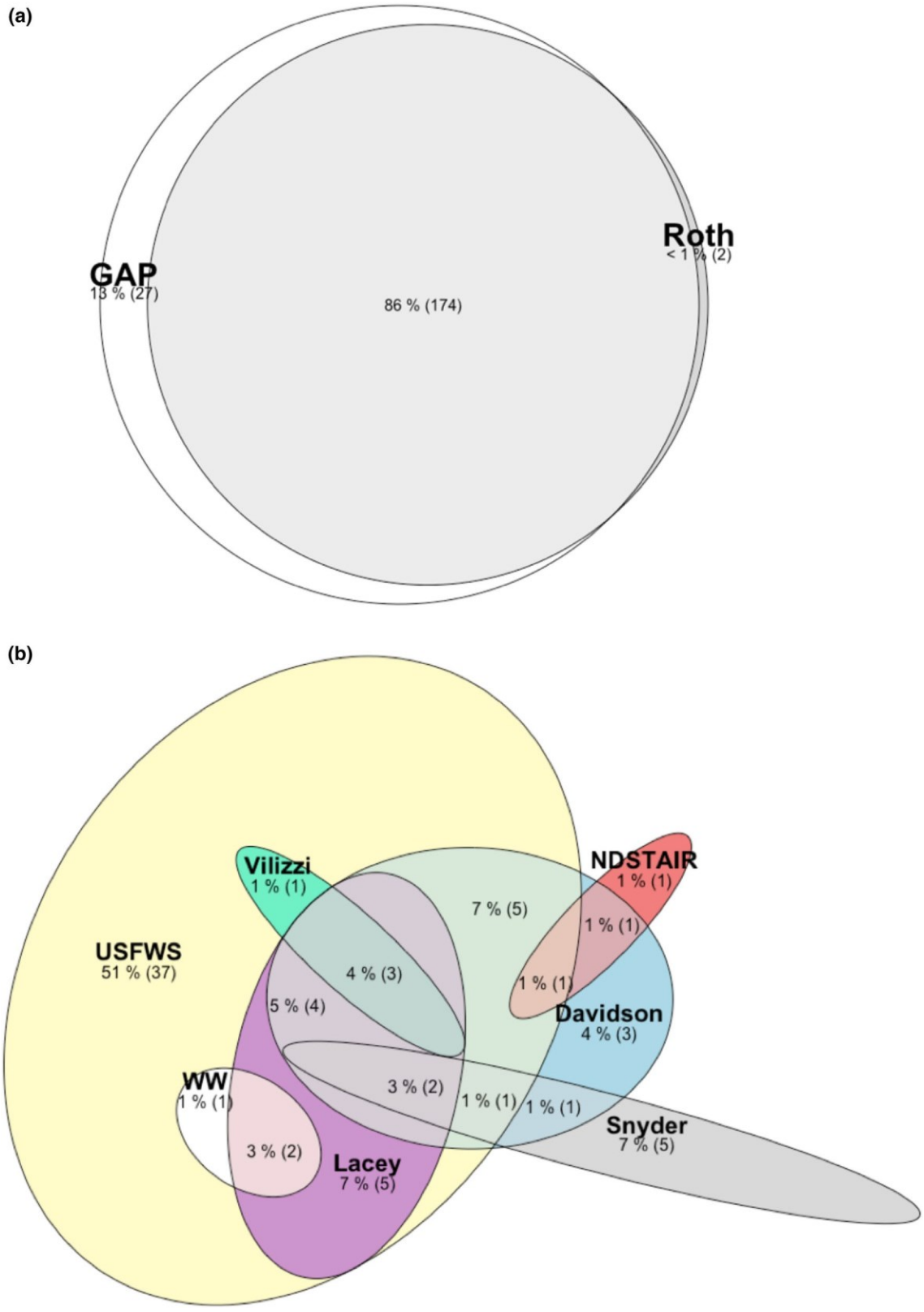

FIGURE 3 Euler diagrams showing overlap in established Great Lakes species lists (a) and invasive species (b) lists.

(*Carassius gibelio*, *Percottus glenii*, and *Pseudorasbora parva*, *Perca fluviatilis*, *Phoxinus phoxinus*, and *Silurus glanis)*, six species in three assessments, and 16 species in two assessments. Most species (47) were found in only one risk assessment, with 10 only found in one of the Great Lakes assessments (five in Snyder, three in

Davidson, and one in NDSTAIR) and 37 only found in the USFWS assessment.

Reference library evaluation

Sequence availability and primer coverage

If we consider only the established fish species found in the Great Lakes, then using either a 16S or 12S region primer pair, with the exception of the 16S DiBattista primer pair, has between 64% (16S Palumbi; 130/203) and 71.9% (12S Bylemans; 16S McInnes; 16S Shaw; 146/203) coverage (Figure 4a). The COI and cytB primer pairs have noticeably lower coverage with 18S having few reference sequences at 10% coverage (20/203). This pattern of 16S and 12S regions having similar coverage performance irrespective of primer pair is consistent with fish reference library studies in more diverse systems like the Mekong River, Cambodia (Jerde et al., 2021; Marques et al., 2021). As we add potentially invasive species from Great Lakes specific risk assessment (Figure 4b) and from US and Global risk assessments, the coverage of best performing primer pair remains somewhat consistent, 72.4% (168/232) and 75.2% (209/278), respectively. From a single primer pair, eDNA metagenetic approach however, it means that 28% ([203–146]/203) of established Great Lakes fish, 27.6% ([232–168]/232) of Great Lakes fish and Great Lakes risk assessed fish, and 24.8% ([278–209]/278) of Great Lakes fish and regionally, nationally, and globally risky fish are undetectable. While we expect in vitro outcomes to vary somewhat from our in silico predictions, we anticipate that overall rankings of primers by coverage will remain the same based on previous research (Ficetola et al., 2010).

Calibration of species richness estimation using eDNA metabarcoding to conventional gears, such as traps, nets, and electrofishing, has shown there is an advantage in using multiple primer pairs (McElroy et al., 2020). With the Great Lakes established fish and likely invaders ($n = 203 + 29 = 232$), a multiple primer pair approach could be useful, particularly if it aids in improving species specificity (Figure 5), with the caveat that there will be increased costs and DNA per sample may become limiting as primer pairs are added. With one primer pair, 72.4% (168/232) of the total species list has coverage. This percentage increases to 81.5%, 85.3%, 88.4%, and 89.2% as we increase the number of primer pairs used in the assessment to two, three, four, and five, respectively. At five primer pairs used, there is full coverage of all fish species with at least one genetic reference sequence in our study. Irrespective, with a maximum of 207 fish species of 232, there remain at least 25 fish that a Great Lakes eDNA metagenetic survey will be blind to. Furthermore, while there does appear to be good coverage of genetic information for Great Lake and invasive fish generally, many of the sequences may be unable to reconcile species-level identification.

Species specificity

Our sequence comparisons and analyses of whether amplicons could distinguish species from each other (i.e., did interspecific variation outweigh intraspecific variation) found differences between routinely targeted metabarcoding markers. From this, not only are there differences in detectability between the three target amplicons, but our ability to differentiate between species is also questionable if we use the routine 5% variation between species caveat (Table 1). For cytB, the gene with the most available reference sequences, 19 species pairs vary 5% or less, whereas 16S and 12S have 497 and 500 pairs that exhibit this same level of sequence differentiation. Additionally, those species pairs for cytB with 5% or less variation are all congeneric. For 16S and 12S evaluations, there are species with less than 5% differences (a typical "species-level percent difference" boundary used in published studies) that belong to different families (e.g., bighead carp, longnose dace). Additionally, there are even some species in different orders (e.g., paddlefish and *Ambloplites rupestris* in Perciformes and Acipenseriformes, respectively) that vary <5% (uncorrected *p* distance). These types of variations, or lack of variations, can be confounding and while some metabarcoding markers may perform better, they are not necessarily as species specific as needed to make direct comparisons.

DISCUSSION

eDNA metabarcoding is a promising approach for biodiversity monitoring and early detection of invasive species, but rarely are both aims considered concurrently. We found that reference libraries for monitoring established fish communities can be informed by invasive species risk assessment to serve early detection surveillance efforts. In the Great Lakes, where invasive species have been particularly well-studied, we found three risk assessments focused on the region and several more relevant assessments at the national and global scale. The Great Lakes assessments identified an additional 29 species that are currently not present in the Great Lakes but likely to invade, and the national and global assessments identified another 46 species that have some risk to invade the region. We found that considering these species in eDNA

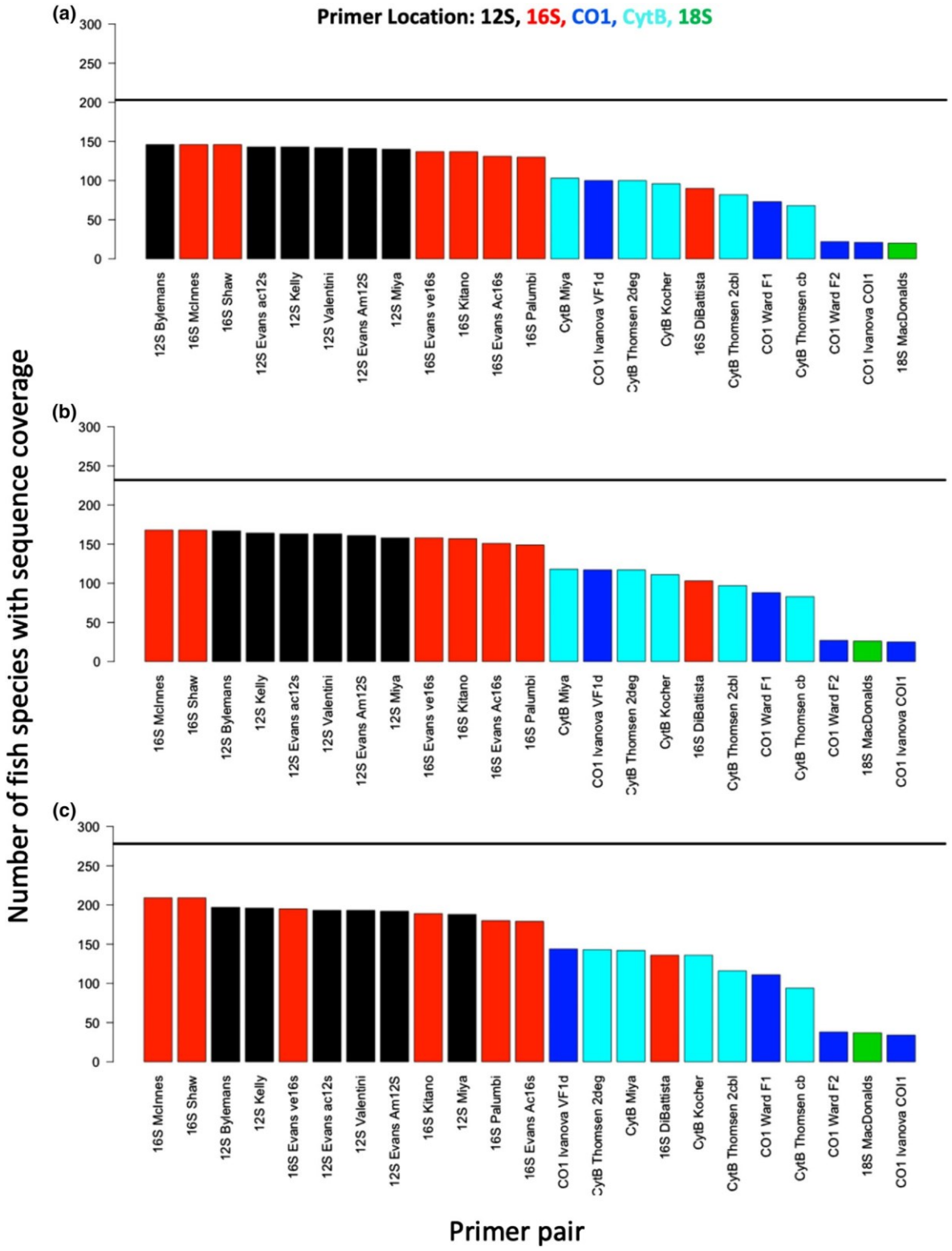

FIG URE 4 Coverage of rank ordered primer pairs for reference libraries containing only fish established in the Great Lakes (a), fish established in the Great Lakes with additional fish species identified from Great Lakes risk assessments (b), and fish established in the Great Lakes with additional fish species identified by all risk assessments considered (c). The solid line (species richness ceiling) in (a) is set at 203 species, the total number of established species in the Great Lakes. The solid line in (b) is at 232 species and in (c) is 278. The primer pairs are rank ordered and the coverage changes between (a), (b), and (c) with 12S and 16S consistently providing better coverage than COI, cytB, and 18S.

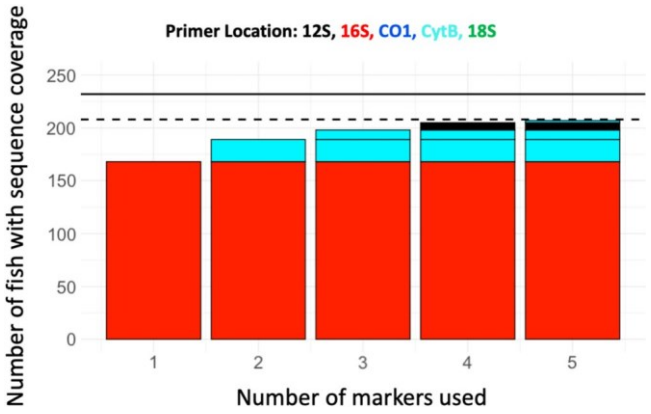

FIGURE 5 Species accumulation of eDNA metabarcoding with additional primer pairs. The solid line indicates the total number of fish species established in the Great Lakes and the fish species identified through risk assessment most that pose a threat to the Great Lakes (n = 232). Of those fish species, 207 have at least one sequence covered by a primer pair. Saturation occurs with the use of five markers. The stepwise addition of primer pair order that maximizes coverage is Shaw 16S (or McInnes 16S), followed by Miya CytB, Thomsen cb cytB, Bylemans 12S, and then Thomsen 2cbl cytB. COI and 18S were considered, so are included in the legend, but were not plotted. The dashed line is the number of species with one covered sequence, and the black line is the total number of species considered.

TABLE 1 A comparison of genetic divergence (calculated as uncorrected p distance) between fish species in the Laurentian Great Lakes for three routinely targeted metabarcoding genes.

Species pairs catagory	16S	12S	cytB
Total species pairs compared	11,030	10,585	14,196
Species pairs >5% divergence	10,529	10,085	14,177
Species pairs between 0% and 5%	497	500	19
Species pairs between 0% and 3%	135	147	12
Species pairs between 0% and 2%	64	41	10
Species pairs ≤1%	16	15	6

Note: The value of each cell is the number of Great Lakes species pairs distinguished by each of three target genes at various percent similarity thresholds.

metabarcoding surveys and the 203 species present in the Great Lakes would have large benefits for early detection of invasive species without significantly altering biodiversity-focused methods. For example, primer pair coverage across all species varied slightly when risky species were included in our Great Lakes reference library (Figure 4). This means that with a little extra effort, every water sample could aid multiple conservation priorities.

However, even in a well-studied region such as the Great Lakes, genetic coverage is limiting the eDNA metabarcoding approach. With current reference libraries and primer sets, a single-primer eDNA metabarcoding survey would not be able to detect 24.8%–28% of species present or potentially invasive to the Great Lakes, with a minimum of 10.8% undetectable if five primers pairs are used. Even in the best coverage scenarios, many species would likely not be able to be distinguished from each other, given that the primers with the best coverage (e.g., 12S, 16S) have relatively low specificity while primers with high specificity (e.g., cytB) have relatively low coverage (Figure 4; Table 1). Additional sequencing to increase the completeness of the freshwater fish reference sequence libraries will improve eDNA metabarcoding coverage to a point. For example, 30 out of 278 species present or potentially invasive to the Great Lakes had no coverage across all 23 primer pairs evaluated. Of these 30 no-coverage species, only 14 lacked references sequences at three or more loci (Table 2, seven species had no references available, seven had references only at cytB and COI loci). These species should be prioritized for sequencing to increase coverage of metabarcoding surveys in the Great Lakes. Moreover, sequencing of the Parana river stingray, *Potamotrygon schuhmacheri*, should be a priority for those in the United States as this species was identified as a risk to invade North America by the national USFWS risk assessment.

It is important to note that, while additional sequencing will likely improve coverage, it will not enable current eDNA metabarcoding approaches to achieve complete coverage. For example, more than half of species with zero coverage have reference sequences at four or more loci, with complete mitochondrial genomes available for 11 species (Table 2). Metabarcoding will likely remain blind to at least 11 species of established or potentially invasive Great Lakes species (4.0% of total) regardless of reference sequencing effort directed at current markers. To achieve complete coverage of Great Lakes fishes with eDNA, the metabarcoding methodology must be improved by developing metabarcoding primers targeting new loci or incorporating complementary approaches such as species-specific PCRs (e.g., Roy et al., 2018). Additionally, one basic research question needing more consideration is the issue of PCR primer bias (Kelly et al., 2019). Some of the primer pairs available for fish eDNA metabarcoding may disproportionally detect common fish over rare fish species, which ultimately can allow for incipient invaders to remain undetected. This will lower the utility of eDNA metabarcoding for biodiversity surveys to serve as an early detection tool for invasive species. However, the severity and sensitivity of this bias is largely unknown (but see Simmons et al., 2016).

The Laurentian Great Lakes are arguably the most studied freshwater fisheries globally. Yet, finding a

TABLE 2 Great Lakes fishes and potential invaders with no coverage across common primer pairs.

Species	Great Lakes status	IUCN status	Reference sequences available
Ammocrypta clara	Present	VU	cytB, COI
Catostomus utawana	Present	DD	
Coregonus alpenae	Present	EX	
Coregonus johannae	Present	EX	
Coregonus kiyi	Present	VU	cytB, COI
Coregonus nigripinnis	Present	EX	cytB, COI
Coregonus reighardi	Present	CR	
Esox americanus	Present	LC	Mitogenome
Etheostoma exile	Present	LC	cytB, COI, 12S, 16S
Fundulus diaphanus	Present	LC	Mitogenome
Fundulus heteroclitus	Present	LC	Mitogenome, whole genome
Ichthyomyzon castaneus	Present	LC	cytB, COI
Lepomis gulosus	Present	LC	Mitogenome
Lethenteron appendix	Present	LC	Mitogenome
Moxostoma macrolepidotum	Present	LC	Mitogenome
Myoxocephalus thompsonii	Present	LC	cytB, COI
Notropis buccatus	Present	LC	cytB, COI
Notropis dorsalis	Present	LC	Mitogenome
Opsopoeodus emiliae	Present	LC	Mitogenome
Osmerus mordax	Present	LC	Mitogenome
Phoxinus neogaeus	Present	LC	cytB, COI, 12S, 16S
Salvelinus alpinus	Present	LC	Mitogenome, 18S
Channa argus	Potential invader	NE	Mitogenome
Clupeonella caspia	Potential invader	LC	
Coptodon rendalli	Potential invader	LC	cytB, COI, 12S, 16S
Hyrcanogobius bergi	Potential invader	LC	
Potamotrygon falkneri	Potential invader	DD	cytB, COI
Potamotrygon schuhmacheri	Potential invader	NE	
Tilapia mariae	Potential invader	LC	cytB, COI, 12S, 16S, 18S
Tilapia zillii	Potential invader	NE	cytB, COI, 12S, 16S, 18S

Note: Reference sequence availability based on NCBI Genbank accessed September–October 2021.
Abbreviations: CR, Critically Endangered; DD, Data Deficient; EX, Extinct; IUCN, International Union for Conservation of Nature; LC, Least Concern; NCBI, National Center for Biotechnology Information; NE, Not Evaluated; VU, Vulnerable.

composite fish species list was not obvious beyond the default Tedesco et al. (2017) list provided by GAPeDNA and Roth et al. (2013). Given the dynamics of fisheries populations from anthropogenic spread and extinctions, maintaining an established species list to develop an eDNA metabarcoding surveillance program is critical for ensuring the coverage of a reference library is sufficient to provide meaningful inference (Marques et al., 2021). This knowledge gap of what fish species are present will be more pronounced in understudied systems with many endemics and highly diverse systems (Jerde et al., 2021). The 22 fish species in the Great Lakes without coverage are comprised mainly of small fish (<13 cm; e.g., *Fundulus diaphanus*) that are rare (e.g., *Coregonus reighardi*) with some that are presumably extinct (e.g., *Coregonus alpenae*; Appendix S1). It is worth noting that GAPeDNA does not assess IUCN listed extinct species (Marques et al., 2021). Still, for purposes of eDNA metabarcoding, it may be a conservation priority to have references sequences available on the chance that remnant populations exist and can be protected. In our study, three species are now considered extinct and had no reference sequences (*C. alpenae*, *C. johannae*, and *C. nigripinnis*).

Risk assessments were initially motivated to prevent the introduction of new invasive species by identifying

those species likely to cause damage and actively stopping introductions. However, policy to do this is often lacking (Davidson, Tucker, Chadderton, Jensen, et al., 2021; Peters & Lodge, 2009) and the same risk assessments can be valuable for prioritizing surveillance efforts. Here we demonstrate the utility of eDNA metabarcoding for potentially invasive fishes to the Great Lakes, but additional non-fish taxa should be considered as well. Dreissenid mussels, for example, have caused substantial harm to the Great Lakes and are also readily detectable and distinguishable with eDNA approaches (Sepulveda, Hutchins, Jackson, et al., 2020). Within the Laurentian Great Lakes, applicable risk assessments exist for aquatic plant and invertebrate taxa (e.g., Gantz et al., 2015; Keller et al., 2007; Zeng et al., 2015) as does a surveillance species list (Davidson, Tucker, Chadderton, & Weibert, 2021), a reference aquatic metazoan species inventory (Trebitz et al., 2019), and an analysis of COI reference sequence availability for current and potentially invasive Great Lakes metazoans (Trebitz et al., 2015). As we have done with fish, eDNA metabarcoding approaches for biodiversity and early detection could be evaluated by estimating the coverage and specificity of available primers. Although we suspect that regional species lists, risk assessments, and reference libraries will be less complete for non-fish taxa in the Great Lakes and other regions, such an effort could help coordinate sequencing and method development efforts to maximize the potential of eDNA for biodiversity monitoring and early detection across a broader range of taxa.

Irrespective of basic or applied science motivation for a study, methodological limitations, such as reference library incompleteness, has the potential to mislead our fundamental understanding of the processes shaping biodiversity and the practical management of invasive species. Incidental and false positive detections (Darling et al., 2021) for applied problems can lead to costly management actions that undermine confidence in the data even as best practices and protocols are evolving to add reproducibility and credibility to eDNA applications (Sepulveda, Hutchins, Jackson, et al., 2020). However, it is worth echoing two of the "interim solutions" provided in Darling et al. (2020) that can allow for use of eDNA metabarcoding detection in biodiversity surveys for species of concern management (i.e., invasive species): First, species lists, both established and potentially invasive, should come with discussions of the limitations in detectability and completeness. Second, having an international reference library of high scrutinized refence sequences for invasive species is critical. This would require a significant increase in sample collection, data generation, and additional quality control of new entries to publicly available databases.

We are in the infancy of the genomic revolution (Shokralla et al., 2012). Environmental DNA metabarcoding for fisheries management can advance quickly by starting with comprehensive lists of established species, ensuring those species have broad primer pair coverage or, ideally, whole mitochondria genome sequencing, and by justifying genetic reference libraries inclusive of likely invasive species. Here we have shown that invasive species risk assessments can be used to prioritize invasive species lists to add to those reference libraries across national, region, and local scales. Moving forward, we speculate that many of the global invasive species will have their entire genomes sequenced due the global impact and damages of the world's worst invaders (Lowe et al., 2000). Regionally and locally however, the burden of reference library coverage will likely fall to agencies tasked with managing invasive species working with academia and industry, unless agencies have the resources to support genomic research. By including potential invaders into the reference library, we add value to ongoing fisheries monitoring of endemic species by concurrently monitoring for invasive species within the same sample, and vice versa. Additionally, we now realize the recommendations of 15 years ago for invasive species management by providing an integrated avenue to manage invasive species through early detection and risk assessment (Lodge et al., 2006; Sepulveda, Nelson, et al., 2020).

ACKNOWLEDGMENTS

This work was partially funded by NASA (NNX14AR62A), BOEM (MC15AC00006), and NOAA's support of the Biodiversity Observation Network. Erin Grey was supported by National Science Foundation award #OIA-1849227 to Maine EPSCoR at the University of Maine. Thank you to Julie Hinderer, Reid Swanson, Andrew Muir, John Dettmers, Jeff Tyson, Rob McLaughlin, Randy Eshenroder, and Nicholas Mandrak for correspondence regarding sources for Great Lakes fish species lists.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The synthesized database (Jerde et al., 2022) is available in Dryad at https://doi.org/10.25349/D9KW49. Established species lists for the Great Lakes are found in Marques et al. (2021), Tedesco et al. (2017), and Roth et al. (2013). Potential invasive fish lists are found in Snyder et al. (2014), Davidson, Tucker, Chadderton, and Weibert (2021), Howeth et al. (2016), Marcot et al. (2019), Fowler et al. (2007), Copp (2013), and Vilizzi et al. (2019). Sequence data for all established fishes were downloaded

for the three target genes (16S, cytB and 12S) from GenBank (http://ncbi.nlm.nih.gov). Complete gene fragments were used from whole mitochondrial genomes when available to provide a conservative estimate across the entire genetic element.

ORCID

Andrew R. Mahon https://orcid.org/0000-0002-5074-8725
Erin K. Grey https://orcid.org/0000-0001-5883-0013
Christopher L. Jerde https://orcid.org/0000-0002-8074-3466

REFERENCES

Andersen, M. C., H. Adams, B. Hope, and M. Powell. 2004. "Risk Assessment for Invasive Species." *Risk Analysis* 24(4): 787–93. https://doi.org/10.1111/j.0272-4332.2004.00478.x.

Chan, F. T., S. A. Bailey, C. J. Wiley, and H. J. MacIsaac. 2013. "Relative Risk Assessment for Ballast-Mediated Invasions at Canadian Arctic Ports." *Biological Invasions* 15(2): 295–308. https://doi.org/10.1007/s10530-012-0284-z.

Chan, J., Y. Zeng, and D. C. Yeo. 2021. "Invasive Species Trait-Based Risk Assessment for Non-native Freshwater Fishes in a Tropical City Basin in Southeast Asia." *PLoS One* 16(3): e0248480. https://doi.org/10.1371/journal.pone.0248480.

Copp, G. H. 2013. "The Fish Invasiveness Screening Kit (FISK) for Non-native Freshwater Fishes: A Summary of Current Applications." *Risk Analysis* 33(8): 1394–6. https://doi.org/10.1111/risa.12095.

Darling, J. A., C. L. Jerde, and A. J. Sepulveda. 2021. "What Do you Mean by False Positive?" *Environmental DNA* 3(5): 879–83. https://doi.org/10.1002/edn3.194.

Darling, J. A., and A. R. Mahon. 2011. "From Molecules to Management: Adopting DNA-Based Methods for Monitoring Biological Invasions in Aquatic Environments." *Environmental Research* 111(7): 978–88. https://doi.org/10.1016/j.envres.2011.02.001.

Darling, J. A., X. Pochon, C. L. Abbott, G. J. Inglis, and A. Zaiko. 2020. "The Risks of Using Molecular Biodiversity Data for Incidental Detection of Species of Concern." *Diversity and Distributions* 26(9): 1116–21. https://doi.org/10.1111/ddi.13108.

Davidson, A. D., A. J. Tucker, W. L. Chadderton, E. Jensen, C. Weibert, and R. Death. 2021. "Assessing Progress in Regulation of Aquatic Nonindigenous Species across the Multijurisdictional Waters of the Laurentian Great Lakes, with Emphasis on the Live Trade Pathways." *Management of Biological Invasions* 12(3): 546–77.

Davidson, A. D., A. J. Tucker, W. L. Chadderton, and C. Weibert. 2021. "Development of a Surveillance Species List to Inform Aquatic Invasive Species Management in the Laurentian Great Lakes." *Management of Biological Invasions* 12(2): 272–93.

Deiner, K., H. M. Bik, E. Mächler, M. Seymour, A. Lacoursière-Roussel, F. Altermatt, S. Creer, et al. 2017. "Environmental DNA Metabarcoding: Transforming how we Survey Animal and Plant Communities." *Molecular Ecology* 26(21): 5872–95. https://doi.org/10.1111/mec.14350.

Erickson, R. A., C. M. Merkes, and E. L. Mize. 2019. "Sampling Designs for Landscape-Level eDNA Monitoring Programs." *Integrated Environmental Assessment and Management* 15(5): 760–71. https://doi.org/10.1002/ieam.4155.

Evans, N. T., Y. Li, M. A. Renshaw, B. P. Olds, K. Deiner, C. R. Turner, C. L. Jerde, D. M. Lodge, G. A. Lamberti, and M. E. Pfrender. 2017. "Fish Community Assessment with eDNA Metabarcoding: Effects of Sampling Design and Bioinformatic Filtering." *Canadian Journal of Fisheries and Aquatic Sciences* 74(9): 1362–74. https://doi.org/10.1139/cjfas-2016-0306.

Evans, N. T., B. P. Olds, M. A. Renshaw, C. R. Turner, Y. Li, C. L. Jerde, A. R. Mahon, M. E. Pfrender, G. A. Lamberti, and D. M. Lodge. 2016. "Quantification of Mesocosm Fish and Amphibian Species Diversity Via Environmental DNA Metabarcoding." *Molecular Ecology Resources* 16(1): 29–41. https://doi.org/10.1111/1755-0998.12433.

Ficetola, G. F., E. Coissac, S. Zundel, T. Riaz, W. Shehzad, J. Bessière, P. Taberlet, and F. Pompanon. 2010. "An *In Silico* Approach for the Evaluation of DNA Barcodes." *BMC Genomics* 11(1): 1–10.

Ficetola, G. F., C. Miaud, F. Pompanon, and P. Taberlet. 2008. "Species Detection Using Environmental DNA from Water Samples." *Biology Letters* 4(4): 423–5. https://doi.org/10.1098/rsbl.2008.0118.

Fowler, A. J., D. M. Lodge, and J. F. Hsia. 2007. "Failure of the Lacey Act to Protect US Ecosystems against Animal Invasions." *Frontiers in Ecology and the Environment* 5(7): 353–9. https://doi.org/10.1890/1540-9295(2007)5[353:FOTLAT]2.0.CO;2.

Froese, R., and D. Pauly. 2021. "FishBase. World Wide Web Electronic Publication." www.fishbase.org, version (06/2021).

Gallardo, B., P. S. Zu Ermgassen, and D. C. Aldridge. 2013. "Invasion Ratcheting in the Zebra Mussel (*Dreissena polymorpha*) and the Ability of Native and Invaded Ranges to Predict its Global Distribution." *Journal of Biogeography* 40(12): 2274–84. https://doi.org/10.1111/jbi.12170.

Gantz, C. A., D. R. Gordon, C. L. Jerde, R. P. Keller, W. L. Chadderton, P. D. Champion, and D. M. Lodge. 2015. "Managing the Introduction and Spread of Non-native Aquatic Plants in the Laurentian Great Lakes: A Regional Risk Assessment Approach." *Management of Biological Invasions* 6(1): 45–55. https://doi.org/10.3391/mbi.2015.6.1.04.

Grey, E. K., L. Bernatchez, P. Cassey, K. Deiner, M. Deveney, K. L. Howland, A. Lacoursière-Roussel, et al. 2018. "Effects of Sampling Effort on Biodiversity Patterns Estimated from Environmental DNA Metabarcoding Surveys." *Scientific Reports* 8(1): 1–10.

Harper, L. R., L. Lawson Handley, A. I. Carpenter, M. Ghazali, C. di Muri, C. J. Macgregor, T. W. Logan, et al. 2019. "Environmental DNA (eDNA) Metabarcoding of Pond Water as a Tool to Survey Conservation and Management Priority Mammals." *Biological Conservation* 238: 108225. https://doi.org/10.1016/j.biocon.2019.108225.

Howeth, J. G., C. A. Gantz, P. L. Angermeier, E. A. Frimpong, M. H. Hoff, R. P. Keller, N. E. Mandrak, et al. 2016. "Predicting Invasiveness of Species in Trade: Climate Match, Trophic Guild and Fecundity Influence Establishment and Impact of Non-native Freshwater Fishes." *Diversity and Distributions* 22(2): 148–60. https://doi.org/10.1111/ddi.12391.

Jerde, C. L. 2021. "Can we Manage Fisheries with the Inherent Uncertainty from eDNA?" *Journal of Fish Biology* 98(2): 341–53. https://doi.org/10.1111/jfb.14218.

Jerde, C. L., A. R. Mahon, T. Campbell, M. E. McElroy, K. Pin, J. N. Childress, M. N. Armstrong, et al. 2021. "Are Genetic Reference Libraries Sufficient for Environmental DNA Metabarcoding of Mekong River Basin Fish?" *Water* 13(13): 1767. https://doi.org/10.3390/w13131767.

Jerde, C. L., A. R. Mahon, W. L. Chadderton, and D. M. Lodge. 2011. "Sight-Unseen Detection of Rare Aquatic Species Using Environmental DNA: eDNA Surveillance of Rare Aquatic Species." *Conservation Letters* 4: 150–7. https://doi.org/10.1111/j.1755-263X.2010.00158.x.

Jerde, C. L., A. R. Mahon, and E. K. Grey. 2022. "Established and Potentially Invasive Fish of the Laurentian Great Lakes." Dryad. Dataset. https://doi.org/10.25349/D9KW49.

Katoh, K., and D. M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30(4): 772–80. https://doi.org/10.1093/molbev/mst010.

Keller, R. P., J. M. Drake, and D. M. Lodge. 2007. "Fecundity as a Basis for Risk Assessment of Nonindigenous Freshwater Molluscs." *Conservation Biology* 21(1): 191–200. https://doi.org/10.1111/j.1523-1739.2006.00563.x.

Kelly, R. P., A. O. Shelton, and R. Gallego. 2019. "Understanding PCR Processes to Draw Meaningful Conclusions from Environmental DNA Studies." *Scientific Reports* 9(1): 1–14.

Kolar, C. S., and D. M. Lodge. 2001. "Progress in Invasion Biology: Predicting Invaders." *Trends in Ecology & Evolution* 16(4): 199–204. https://doi.org/10.1016/S0169-5347(01)02101-2.

Kolar, C. S., and D. M. Lodge. 2002. "Ecological Predictions and Risk Assessment for Alien Fishes in North America." *Science* 298(5596): 1233–6. https://doi.org/10.1126/science.1075753.

Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms." *Molecular Biology and Evolution* 35(6): 1547–9. https://doi.org/10.1093/molbev/msy096.

Larsson, J. 2021. "eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses." R package version 6.1.1. https://CRAN.R-project.org/package=eulerr.

Lefever, S., F. Pattyn, J. Hellemans, and J. Vandesompele. 2013. "Single-Nucleotide Polymorphisms and Other Mismatches Reduce Performance of Quantitative PCR Assays." *Clinical Chemistry* 59(10): 1470–80. https://doi.org/10.1373/clinchem.2013.203653.

Li, Y., N. T. Evans, M. A. Renshaw, C. L. Jerde, B. P. Olds, A. J. Shogren, K. Deiner, D. M. Lodge, G. A. Lamberti, and M. E. Pfrender. 2018. "Estimating Fish Alpha-and Beta-Diversity along a Small Stream with Environmental DNA Metabarcoding." *Metabarcoding and Metagenomics* 2: e24262. https://doi.org/10.3897/mbmg.2.24262.

Lodge, D. M. 1993. "Biological Invasions: Lessons for Ecology." *Trends in Ecology & Evolution* 8(4): 133–7. https://doi.org/10.1016/0169-5347(93)90025-K.

Lodge, D. M., S. Williams, H. J. MacIsaac, K. R. Hayes, B. Leung, S. Reichard, R. N. Mack, et al. 2006. "Biological Invasions: Recommendations for US Policy and Management." *Ecological Applications* 16(6): 2035–54. https://doi.org/10.1890/1051-0761(2006)016[2035:BIRFUP]2.0.CO;2.

Lowe, S., M. Browne, S. Boudjelas, and M. De Poorter. 2000. *100 of the World's Worst Invasive Alien Species: A Selection from the Global Invasive Species Database*, Vol 12. Auckland: Invasive Species Specialist Group.

Mächler, E., J. C. Walser, and F. Altermatt. 2020. "Decision-Making and Best Practices for Taxonomy-Free Environmental DNA Metabarcoding in Biomonitoring Using Hill Numbers." *Molecular Ecology* 30(13): 3326–39.

Mahon, A. R., C. P. Arango, and K. M. Halanych. 2008. "Genetic Diversity of *Nymphon* (Arthropoda: Pycnogonida: Nymphonidae) along the Antarctic Peninsula with a Focus on *Nymphon australe* Hodgson 1902." *Marine Biology* 155(3): 315–23.

Marcot, B. G., M. H. Hoff, C. D. Martin, S. D. Jewell, and C. E. Givens. 2019. "A Decision Support System for Identifying Potentially Invasive and Injurious Freshwater Fishes." *Management of Biological Invasions* 10(2): 200–26. https://doi.org/10.3391/mbi.2019.10.2.01.

Marques, V., T. Milhau, C. Albouy, T. Dejean, S. Manel, D. Mouillot, and J. B. Juhel. 2021. "GAPeDNA: Assessing and Mapping Global Species Gaps in Genetic Databases for eDNA Metabarcoding." *Diversity and Distributions* 27: 1880–92. https://doi.org/10.1111/ddi.13142.

McElroy, M. E., T. L. Dressler, G. C. Titcomb, E. A. Wilson, K. Deiner, T. L. Dudley, E. J. Eliason, et al. 2020. "Calibrating Environmental DNA Metabarcoding to Conventional Surveys for Measuring Fish Species Richness." *Frontiers in Ecology and Evolution* 8: 276.

Meyer, C. P., and G. Paulay. 2005. "DNA Barcoding: Error Rates Based on Comprehensive Sampling." *PLoS Biology* 3(12): e422. https://doi.org/10.1371/journal.pbio.0030422.

Olds, B. P., C. L. Jerde, M. A. Renshaw, Y. Li, N. T. Evans, C. R. Turner, K. Deiner, et al. 2016. "Estimating Species Richness Using Environmental DNA." *Ecology and Evolution* 6(12): 4214–26. https://doi.org/10.1002/ece3.2186.

Peters, J. A., and D. M. Lodge. 2009. "Invasive Species Policy at the Regional Level: A Multiple Weak Links Problem." *Fisheries* 34(8): 373–80. https://doi.org/10.1577/1548-8446-34.8.373.

Pitz, K., C. Closek, A. Djurhuus, R. Michisaki, K. Walz, A. Boehm, M. Breitbart, R. Kelly, and F. Chavez. 2017. "Rewards and Challenges of eDNA Sequencing with Multiple Genetic Markers for Marine Observation Programs." *TDWGProc* 1: e20548.

Poulos, H. M., B. Chernoff, P. L. Fuller, and D. Butman. 2012. "Ensemble Forecasting of Potential Habitat for Three Invasive Fishes." *Aquatic Invasions* 7(1): 59–72. https://doi.org/10.3391/ai.2012.7.1.007.

Roth, B. M., N. E. Mandrak, T. R. Hrabik, G. G. Sass, and J. Peters. 2013. "Fishes and Decapod Crustaceans of the Great Lakes Basin." In *Great Lakes Fisheries Policy and Management*, edited by W. W. Taylor and A. Lynch, 105–35. East Lansing, MI: Michigan State University.

Roy, M., V. Belliveau, N. E. Mandrak, and N. Gagné. 2018. "Development of Environmental DNA (eDNA) Methods for Detecting High-Risk Freshwater Fishes in Live Trade in Canada." *Biological Invasions* 20(2): 299–314. https://doi.org/10.1007/s10530-017-1532-z.

Sepulveda, A. J., P. R. Hutchins, M. Forstchen, M. N. Mckeefry, and A. M. Swigris. 2020. "The Elephant in the Lab (and Field):

Contamination in Aquatic Environmental DNA Studies." *Frontiers in Ecology and Evolution* 8: 440.

Sepulveda, A. J., P. R. Hutchins, C. Jackson, C. Ostberg, M. B. Laramie, J. Amberg, T. Counihan, A. Hoegh, and D. S. Pilliod. 2020. "A Round-Robin Evaluation of the Repeatability and Reproducibility of Environmental DNA Assays for Dreissenid Mussels." *Environmental DNA* 2(4): 446–59.

Sepulveda, A. J., N. M. Nelson, C. L. Jerde, and G. Luikart. 2020. "Are Environmental DNA Methods Ready for Aquatic Invasive Species Management?" *Trends in Ecology & Evolution* 35(8): 668–78.

Shaw, J. L., L. J. Clarke, S. D. Wedderburn, T. C. Barnes, L. S. Weyrich, and A. Cooper. 2016. "Comparison of Environmental DNA Metabarcoding and Conventional Fish Survey Methods in a River System." *Biological Conservation* 197: 131–8. https://doi.org/10.1016/j.biocon.2016.03.010.

Shogren, A. J., J. L. Tank, E. Andruszkiewicz, B. Olds, A. R. Mahon, C. L. Jerde, and D. Bolster. 2017. "Controls on eDNA Movement in Streams: Transport, Retention, and Resuspension." *Scientific Reports* 7(1): 1–11.

Shokralla, S., J. L. Spall, J. F. Gibson, and M. Hajibabaei. 2012. "Next-Generation Sequencing Technologies for Environmental DNA Research." *Molecular Ecology* 21(8): 1794–805. https://doi.org/10.1111/j.1365-294X.2012.05538.x.

Simmons, M., A. Tucker, W. L. Chadderton, C. L. Jerde, and A. R. Mahon. 2016. "Active and Passive Environmental DNA Surveillance of Aquatic Invasive Species." *Canadian Journal of Fisheries and Aquatic Sciences* 73(1): 76–83. https://doi.org/10.1139/cjfas-2015-0262.

Snyder, R. J., L. E. Burlakova, A. Y. Karatayev, and D. B. MacNeill. 2014. "Updated Invasion Risk Assessment for Ponto-Caspian Fishes to the Great Lakes." *Journal of Great Lakes Research* 40(2): 360–9. https://doi.org/10.1016/j.jglr.2014.03.009.

So, K. Y. K., J. J. Fong, I. P. Y. Lam, and D. Dudgeon. 2020. "Pitfalls during In Silico Prediction of Primer Specificity for eDNA Surveillance." *Ecosphere* 11(7): e03193. https://doi.org/10.1002/ecs2.3193.

Stoeckle, M. Y., M. Das Mishu, and Z. Charlop-Powers. 2020. "Improved Environmental DNA Reference Library Detects Overlooked Marine Fishes in New Jersey, United States." *Frontiers in Marine Science* 7: 226.

Tedesco, P. A., O. Beauchard, R. Bigorne, S. Blanchet, L. Buisson, L. Conti, J. F. Cornu, et al. 2017. "A Global Database on Freshwater Fish Species Occurrence in Drainage Basins." *Scientific Data* 4(1): 1–6.

Thomsen, P. F., J. Kielgast, L. L. Iversen, C. Wiuf, M. Rasmussen, M. T. P. Gilbert, L. Orlando, and E. Willerslev. 2012. "Monitoring Endangered Freshwater Biodiversity Using Environmental DNA." *Molecular Ecology* 21(11): 2565–73. https://doi.org/10.1111/j.1365-294X.2011.05418.x.

Trebitz, A., M. Sykes, and J. Barge. 2019. "A Reference Inventory for Aquatic Fauna of the Laurentian Great Lakes." *Journal of Great Lakes Research* 45(6): 1036–46. https://doi.org/10.1016/j.jglr.2019.10.004.

Trebitz, A. S., J. C. Hoffman, J. A. Darling, E. M. Pilgrim, J. R. Kelly, E. A. Brown, W. L. Chadderton, et al. 2017. "Early Detection Monitoring for Aquatic Non-indigenous Species: Optimizing Surveillance, Incorporating Advanced Technologies, and Identifying Research Needs." *Journal of Environmental Management* 202: 299–310. https://doi.org/10.1016/j.jenvman.2017.07.045.

Trebitz, A. S., J. C. Hoffman, G. W. Grant, T. M. Billehus, and E. M. Pilgrim. 2015. "Potential for DNA-Based Identification of Great Lakes Fauna: Match and Mismatch between Taxa Inventories and DNA Barcode Libraries." *Scientific Reports* 5(1): 1–11.

Valentini, A., P. Taberlet, C. Miaud, R. Civade, J. Herder, P. F. Thomsen, E. Bellemain, et al. 2016. "Next-Generation Monitoring of Aquatic Biodiversity Using Environmental DNA Metabarcoding." *Molecular Ecology* 25(4): 929–42. https://doi.org/10.1111/mec.13428.

Vilizzi, L., G. H. Copp, B. Adamovich, D. Almeida, J. Chan, P. I. Davison, S. Dembski, et al. 2019. "A Global Review and Meta-Analysis of Applications of the Freshwater Fish Invasiveness Screening Kit." *Reviews in Fish Biology and Fisheries* 29(3): 529–68. https://doi.org/10.1007/s11160-019-09562-2.

Wilcox, T. M., K. S. McKelvey, M. K. Young, A. J. Sepulveda, B. B. Shepard, S. F. Jane, A. R. Whiteley, W. H. Lowe, and M. K. Schwartz. 2016. "Understanding Environmental DNA Detection Probabilities: A Case Study Using a Stream-Dwelling Char *Salvelinus fontinalis*." *Biological Conservation* 194: 209–16. https://doi.org/10.1016/j.biocon.2015.12.023.

Zeng, Y., K. Y. Chong, E. K. Grey, D. M. Lodge, and D. C. J. Yeo. 2015. "Disregarding Human Pre-introduction Selection can Confound Invasive Crayfish Risk Assessments." *Biological Invasions* 17(8): 2373–85.

SUPPORTING INFORMATION

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

How to cite this article: Mahon, Andrew R., Erin K. Grey, and Christopher L. Jerde. 2023. "Integrating Invasive Species Risk Assessment into Environmental DNA Metabarcoding Reference Libraries." *Ecological Applications* 33(1): e2730. https://doi.org/10.1002/eap.2730