PNAS RESEARCH ARTICLE | ECOLOGY OPEN ACCESS

Environmental DNA reveals the genetic diversity and population structure of an invasive species in the Laurentian Great Lakes

Kara J. Andres[a,b,c,1], David M. Lodge[a,d], and Jose Andrés[a]

Edited by Mary Power, University of California, Berkeley, CA; received May 3, 2023; accepted July 31, 2023

Environmental DNA (eDNA) has been established as a noninvasive and efficient approach to sample genetic material from aquatic environments. Although most commonly used to determine species presence and measure biodiversity, eDNA approaches also hold great potential to obtain population-level genetic information from water samples. In this study, we sequenced a panel of multiallelic microsatellite markers from filtered water and fish tissue samples to uncover patterns of intraspecific diversity in the freshwater Round Goby (*Neogobius melanostomus*) across their invaded range in the Laurentian Great Lakes region. Although we found that the concentration of nuclear eDNA is lower than mitochondrial eDNA, we nonetheless detected over two-thirds of all nuclear alleles identified from genotyped tissues in our eDNA samples, with the greatest recovery of common alleles in the population. Estimates of allele frequencies and genetic variability within and between populations were detected from eDNA in patterns that were consistent with individual tissue-based estimates of genetic diversity and differentiation. The strongest genetic differentiation in both eDNA and tissues exists in an isolation by distance pattern. Our study demonstrates the potential for eDNA-based approaches to characterize key population parameters required to effectively monitor, manage, or sustain aquatic species.

nuclear eDNA | biodiversity | population genetics | metabarcoding

Significance

Through the analysis of DNA shed by organisms into their environment, environmental DNA (eDNA) sampling is a powerful and efficient approach for collecting information about species across a wide range of ecosystems. While typically used for species detections, eDNA may also provide information about population-level genetic variability. Here, we explore the potential for eDNA sampling to estimate genetic diversity and structure of an invasive fish species throughout the Great Lakes region, uncovering genetic patterns from analysis of eDNA that are consistent with conventional tissue-based estimates. This work highlights the potential for eDNA sampling to reveal detailed population characteristics that may inform monitoring and management plans for species that are difficult to sample by conventional means.

The analysis of environmental DNA (eDNA) has revolutionized our ability to noninvasively study species composition and measure biodiversity (1). By capturing DNA present in environmental samples, eDNA approaches can identify the organisms that contribute genetic information to a mixed eDNA sample. Although eDNA was initially used for the detection of single species (2), developments in eDNA metabarcoding now allow for hundreds of species to be detected simultaneously from a single sample, with wide-ranging uses including biodiversity assessments and biomonitoring (3, 4). Furthermore, because eDNA approaches often have higher sensitivity than conventional capture-based methods for detecting species (5), they represent a rapid and cost-effective approach for detecting rare, elusive, and invasive species and for characterizing entire biological communities (6).

While eDNA is commonly used for detecting taxonomic diversity at the species level, recent studies also show promise for detecting within-species genetic variation (reviewed in refs. 7–9). In most of these studies, population genetic information has been revealed through eDNA sequencing of short variable regions in the mitochondrial genome (10, 11), although much longer mitochondrial DNA fragments and entire mitogenomes may also be recovered in eDNA samples (12, 13). Such research challenges the paradigm that only short degraded fragments are present in eDNA samples and demonstrates the potential for targeting longer and more variable DNA markers in population-level assessments using eDNA. Despite our ability to obtain informative mitochondrial haplotypes from environmental samples, mitochondrial markers represent a single independent locus and, in many cases, may not contain sufficiently detailed genetic information to characterize population genetic patterns (14).

To obtain higher-resolution population genetic information, sequence variation at multiple independent nuclear genetic markers should be assessed. Previous research showed that highly polymorphic, nuclear microsatellite markers can be amplified from eDNA samples in experimental mesocosms (15). Here, we investigate the potential for eDNA approaches to uncover the patterns of natural variation in nuclear genomes from field collections of water samples. We conducted eDNA-based and tissue-based population genetic assessments of a Eurasian fish, the Round Goby (*Neogobius melanostomus*), in sites across their invaded range in the Laurentian Great Lakes region. Round Gobies were first reported in North America in the St. Clair River in 1990 (16) and have subsequently spread rapidly through a combination of natural dispersal and ballast water exchange from shipping within the Great Lakes (17, 18). Because the Round Goby invasion exhibits high

Author contributions: K.J.A., D.M.L., and J.A. designed research; K.J.A. performed research; K.J.A., D.M.L., and J.A. contributed new reagents/analytic tools; K.J.A. and J.A. analyzed data; and K.J.A., D.M.L., and J.A. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: akara@wustl.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2307345120/-/DCSupplemental.

Published September 5, 2023.

levels of diversity and genetic structure (19), it is ideal for developing eDNA-based population genetics tools.

Our field results answer some of the key questions for population genetic analyses of nuclear eDNA: 1) Can nuclear markers be reliably amplified from environmental samples? 2) Do allele frequencies from eDNA reflect allele frequencies from tissue-based genetic analysis? 3) Are patterns from eDNA samples consistent with tissue-derived patterns? While the primary goal of this study is to answer these pivotal questions, our study design includes sites in the Great Lakes as well as several new sites at the Round Goby invasion front. We therefore provide an updated test of the proposed patterns of genetic variation associated with the Round Goby invasion (19–21), including increased levels of genetic diversity found at the invasion origin (i.e., leading edge hypothesis) and isolation by distance.

Results

mtDNA and nuDNA Concentrations in eDNA. To determine whether water samples contain enough nuclear DNA (nuDNA) for downstream analyses, we first estimated the relative abundance of nuclear and mitochondrial DNA (mtDNA) in environmental samples at 14 sites across the Great Lakes region (Fig. 1*A*). Comparisons of mtDNA and nuDNA concentrations revealed that both qPCR assays amplified Round Goby DNA from environmental samples. No PCR amplification was measured in any of the negative controls or in eDNA replicates taken from Seneca Lake, a site where Round Gobies have not yet been physically observed (this site was excluded from further analysis). We detected Round Goby mtDNA (COI assay) and nuDNA (microsatellite assay) in at least two of the three eDNA sample replicates at all remaining sites, with mtDNA detected in all samples (100%) and nuDNA detected in 92% of the samples. Overall, the concentration of mtDNA was significantly higher than the concentration of nuDNA ($F = 486.73$, $P < 0.001$), with an average of 32,512 ± 41,497 copies/L of mtDNA and 288 ± 478 copies/L of nuDNA (Fig. 1*B*). The amount of mtDNA in each sample was positively associated with the amount of nuDNA in each sample ($R^2 = 0.36$, $P < 0.001$; Fig. 1*C*). In samples where both DNA markers were detected, the average ratio of mtDNA:nuDNA concentration was 196:1, but there was substantial variation in this ratio within and among sites (range 20:1 to 1,132:1; *SI Appendix*, Fig. S1).

Tissue-Based Genetic Diversity and Population Structure. We identified 389 total alleles across 27 loci in the analyzed tissues from 285 Round Goby individuals, with an average of 178.9 ± 21.8 alleles per population (mean 6.63 ± 0.81 alleles per locus; *SI Appendix*, Table S1). Allelic richness significantly differed among populations ($F = 2.33$, $P = 0.007$) but was not associated with time since invasion ($t = 0.433$, $P = 0.674$). However, the number of private alleles per population was associated with time since invasion, with the highest number of private alleles found at the oldest populations near the site of original introduction

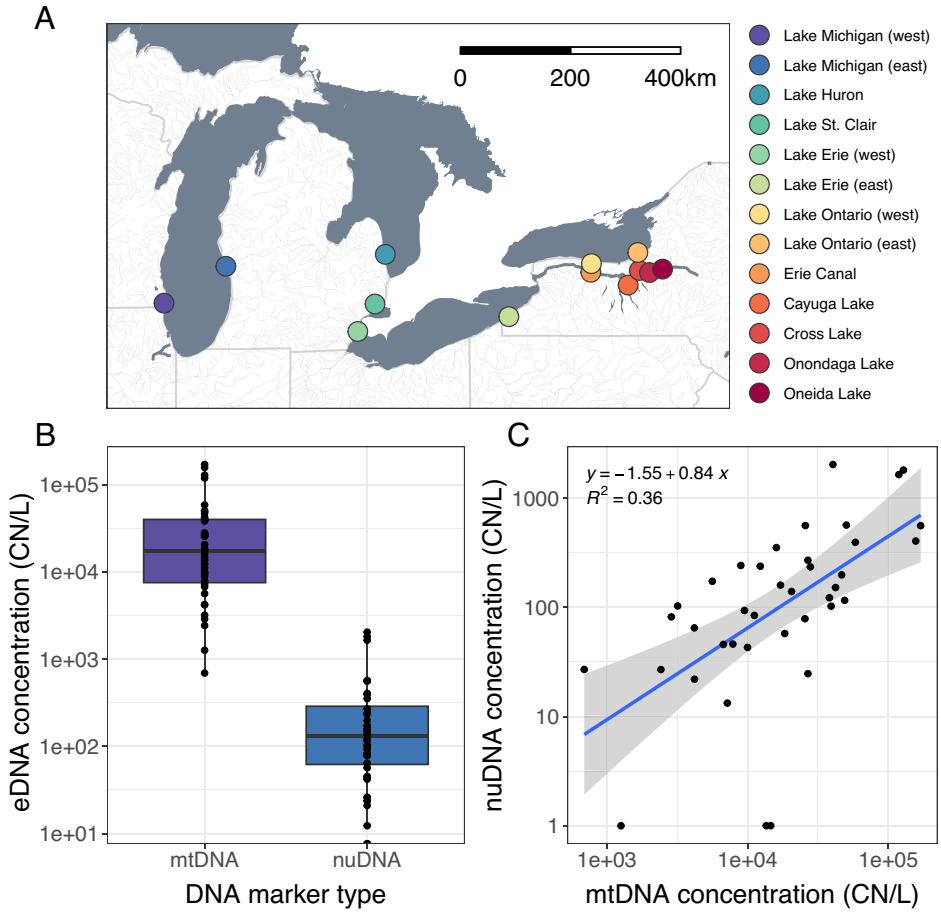

Fig. 1. Results of the qPCR assays measuring mtDNA and nuDNA concentrations in eDNA samples. (*A*) Locations sampled for Round Goby tissues and eDNA; (*B*) The concentration (copy number/L) of mtDNA (purple) in eDNA was much higher than the concentration of nuDNA (blue); (*C*) Log-transformed mtDNA and nuDNA concentrations are correlated with one another. The prediction (blue line) and 95% CI (shaded area) came from a linear model on the log-transformed DNA concentrations in each sample (*C*).

to North America (t = 2.234, P = 0.047). Pairwise F_{ST} values ranged from 0 to 0.12 and were significantly different from zero in most pairwise comparisons (*SI Appendix*, Fig. S2). We identified a significant relationship between pairwise genetic distance and geographic distance (r = 0.744, P = 0.001), indicating a pattern of isolation by distance among Round Goby populations.

Results from Bayesian assignment tests on Round Goby tissues identified the most likely number of genetic clusters at K = 2 (*SI Appendix*, Fig. S3). Populations at the periphery (i.e., westernmost and easternmost populations) exhibited the highest average assignment values, whereas sites near the center of the sampled region exhibited a mixture of assignments to the two genetic clusters. Visualization of genetic structure with Discriminant Analysis of Principal Components (DAPC) further supported the separation of populations along this longitudinal gradient, with distinct groupings of peripheral populations and overlapping central populations (*SI Appendix*, Fig. S3). Hierarchical analysis of molecular variance (AMOVA) based on putative genetic clusters and populations indicated that a significant proportion of the genetic variation could be attributed to differences among genetic clusters (3.97%, P = 0.002) and populations (5.37%, P < 0.001), although was much lower compared to the amount of variation attributed to within populations (90.66%, P < 0.001).

Detection and Quantification of Genetic Diversity from eDNA. There was substantial variation in the number of sequencing reads, observed alleles, and successfully amplified loci in eDNA samples within and among sites (*SI Appendix*, Fig. S4). Due to unpredictable microsatellite amplification in some of the eDNA samples, we only retained eDNA samples that exhibited an average allelic richness of ≥3 alleles in at least 80% of the loci (≥21 loci). This filtering led to the removal of all eDNA samples at six of the 13 sampled sites. In the remaining eDNA samples, we detected a total of 543 alleles, including 66.8% of all alleles known from genotyped tissues and 83.7% of all alleles that are present in genotyped tissues at greater than 10% frequency. Rarefaction curves of the accumulated number of alleles per sample suggest that the alleles detected with tissue sampling began to saturate in the sampled individuals across each locus and all loci combined (*SI Appendix*, Fig. S5). However, accumulation curves did not saturate at most loci for eDNA samples, and more sampling likely would have recovered more microsatellite alleles. Average allelic richness and expected heterozygosity were lower in eDNA samples than in genotyped tissues (*SI Appendix*, Fig. S6 and Table S1), although this difference was only significant for expected heterozygosity (t = 8.277, P < 0.001). On the other hand, the number of private alleles was significantly higher based on eDNA samples than tissues (t = −4.816, P = 0.003; *SI Appendix*, Fig. S6). None of these summary statistics were significantly correlated between eDNA samples and tissue samples from the same population (i.e., taken at the same site). The number of private alleles in eDNA samples did not exhibit significant association with time since invasion, as was the case based on genotyped tissues.

Across all sites, allele frequencies from genotyped tissues were positively correlated with eDNA read frequencies (r = 0.69, P < 0.001), and correlation coefficients per site ranged from 0.59 to 0.85 (*SI Appendix*, Fig. S7). principal components analysis (PCA) of population allele frequencies from genotyped tissues and eDNA samples show congruent patterns of genetic differentiation, where the first principal component axis differentiated populations along the longitudinal gradient (Fig. 2*A*). Neighbor-joining trees based on pairwise population AFD values showed similar patterns of population associations, with the greatest differentiation between the eastern and western sites (Fig. 2*B*). This pattern of isolation by distance was further supported in the eDNA data with a Mantel test, which found a significant association between pairwise AFD and geographic distance among all populations (r = 0.38, P = 0.037). Pairwise allele frequency distances (AFDs) calculated from eDNA read frequencies were positively correlated with AFD measured from tissue-based allele frequencies (r = 0.76, t = 5.147, P < 0.001), although AFD values were higher for eDNA than for tissues (Fig. 2*C*).

Discussion

eDNA is a sensitive and powerful tool for biodiversity assessments, but environmental samples contain much more genetic information than is typically analyzed. In this study, we used high-throughput sequencing of a microsatellite panel to show that eDNA approaches can be used to detect nuclear genetic variation within species in field settings, making it possible to estimate population allele frequencies and analyze genetic diversity and structure from environmental samples. The amplification of a panel of nuclear genetic markers in natural settings represents a substantial advancement toward population-level analyses from eDNA samples, which to date have been limited to analysis of mitochondrial markers. As expected, we found that nuDNA was less abundant in eDNA samples than mtDNA, but we were nonetheless able to amplify at least 80% of the loci in over half of the eDNA samples. The resulting estimates of population allele frequencies and genetic differentiation were correlated between eDNA and tissue-based approaches.

Round Gobies were introduced to North America over 30 y ago, and previous population genetic work identified high levels of diversity at the sites of initial invasion, limited gene flow among North American sites, and slightly higher diversity at the invasion core than at the periphery (19–21). We found similarly high levels of genetic diversity and structuring through the analysis of Round Goby tissues, with most populations genetically distinguishable from one another and the primary pattern of divergence following an isolation by distance pattern along a longitudinal gradient. Our tissue-based analysis did not find reduced genetic diversity at the expansion edge (i.e., the leading-edge model of invasion) (19), although we did report the greatest number of private alleles at the earliest sites of Round Goby invasion. We found similar patterns of genetic structuring and differentiation based on analysis of eDNA samples collected at the same sites as Round Goby tissues, with populations structured along the same longitudinal gradient and exhibiting differentiation from one another. We observed a pattern of isolation by distance in both tissue and eDNA samples, consistent with previous research documenting limited migration in this species (19, 21).

Although we sampled in nearshore habitats where Round Gobies are expected to be in highest abundance during the sampled time period (22), we observed relatively low concentrations of nuclear DNA in eDNA samples (mean mtDNA:nuDNA ratio = 196:1). This is in contrast to previous research investigating ratios of mtDNA:nuDNA, most of which targeted multicopy ribosomal nuDNA markers and found comparable concentrations of mtDNA and nuDNA in eDNA samples (23–26). However, much lower concentrations of single-copy nuDNA markers such as microsatellites are expected, as cells contain many more mitochondria than nuclei (27). Low nuDNA template copy numbers may lead to low or unreliable amplification of nuclear genetic markers during eDNA sequencing. Indeed, we did not observe plateauing of the accumulation curves of microsatellite alleles across eDNA samples (*SI Appendix*, Fig. S5), indicating that each eDNA sample provided additional genetic information, a trend that was not observed in the tissue samples. Similar

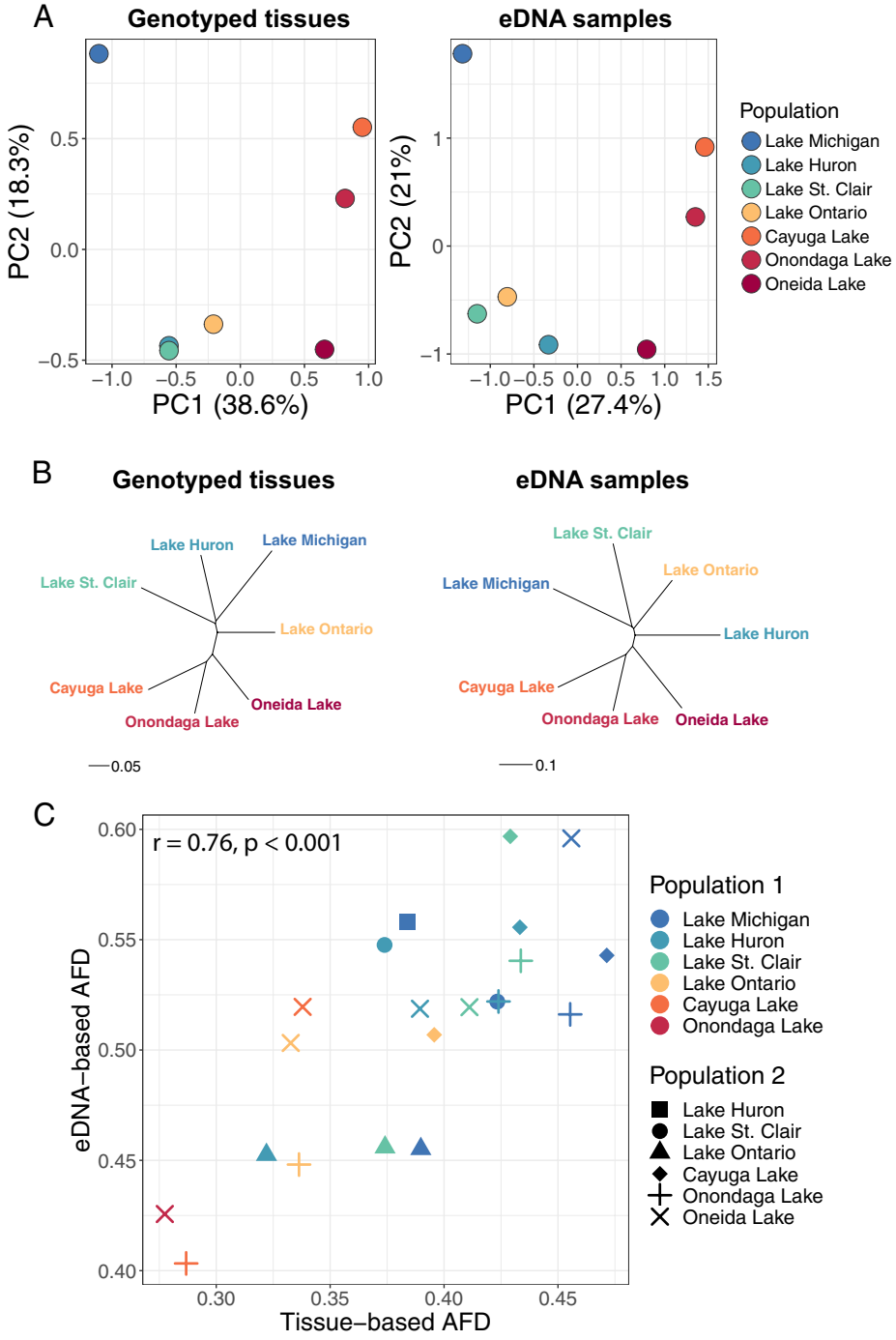

Fig. 2. Environmental DNA-based analysis of Round Goby genetic structure and differentiation reflects tissue-based genetic patterns. (*A*) PCA of tissue-based allele frequencies (up to 28 genotyped individuals per site) and eDNA-based allele frequencies ($n = 3$ water samples per site); (*B*) Unrooted neighbor-joining trees constructed from a matrix of genetic distance using allele frequencies derived from Round Goby tissues and eDNA samples; (*C*) Correlation between values of pairwise population genetic distance (allele frequency distance; AFD) calculated from eDNA read frequencies and from tissue allele frequencies. Each point represents a population pair, with color and shape indicating the measured populations.

accumulation patterns were observed in Sigsgaard et al (11), who recovered more haplotypes in eDNA and did not saturate the genetic diversity in the population with their eDNA sampling scheme. Our results also show that rare alleles in genotyped Round Goby tissues were less likely to be recovered in eDNA samples, likely due to stochastic sampling of low-concentration nuclear markers. Similar challenges of recovering rare genetic variants have been found in the analysis of mitochondrial haplotypes, where all common haplotypes but only one rare haplotype of the mollusk Blackfoot Pāua (*Haliotis iris*) were recovered from seawater samples (28). To reliably amplify rare genetic variants in nuclear markers from eDNA, the development and optimization of approaches to maximize target DNA concentrations may be needed, including targeted sampling where the species is known to occur, target capture of specific genes of interest (12), laboratory extraction and amplification protocols returning higher DNA yields, and/or sampling a greater number of larger-volume samples (29).

In addition to incomplete characterization of population genetic diversity, reduced detection of alleles in eDNA can impact

the resulting genetic analyses. In this study, eDNA samples did not produce estimates of allelic richness, expected heterozygosity, or number of private alleles per population that were similar to estimates obtained from tissue samples (*SI Appendix*, Fig. S6). Due to sporadic allele detections, the number of private alleles detected per site was much higher in eDNA samples, while high frequencies of one or few alleles per eDNA sample led to greatly reduced estimates of expected heterozygosity. Therefore, while eDNA-based population analyses based on allele frequencies were correlated with tissue-based estimates, analyses based on the presence of alleles were more prone to biases introduced from false negatives. The technical and analytical challenges faced by population genetic analysis of eDNA samples, as well as potential solutions to these challenges, should be considered when using eDNA to characterize genetic diversity (9).

Unlike tissue-based genotyping and population genetic methods, eDNA is a mixture of DNA from potentially many individuals, and eDNA datasets cannot be analyzed using the same theoretical frameworks and analyses developed for genotyped tissues. For instance, many diversity and differentiation parameters rely on knowledge of individual genotypes to assess pairwise measures of genetic identity among individuals. However, any allele detected in eDNA may have been derived from single or multiple individuals, with no ability to determine the source number of individuals.

Another challenge in the analysis of intraspecific genetic variation from eDNA is ensuring species specificity of the primers used, as coamplification of DNA from closely related species can have a strong influence on the conclusions of eDNA research. Although Round Gobies do not have any congeners in North America, the Tubenose Goby (*Proterorhinus semilunaris*) of the Gobiidae family is present in Lake St. Clair and western Lake Erie (30). We targeted a large panel of loci and rigorously filtered out any alleles from eDNA that did not align with high identity to alleles identified from Round Goby tissues, minimizing the possibility of Tubenose Goby alleles in our analysis. Researchers interested in using eDNA analysis to investigate population genetics should similarly take special care to exclude the possibility of coamplifying locally occurring close relatives of the species of interest. Genetic reference databases, particularly for nuclear markers, are unlikely to contain genetic information for all possible close relatives of the target species. For instance, although the Round Goby genome has been sequenced (31), the Tubenose Goby does not have a reference genome available for comparison. A conservative approach to addressing this issue would be to restrict all genetic variants detected in eDNA to alleles known from tissues (10), but this could lead to more incidences of false negatives in eDNA samples.

The Laurentian Great Lakes are among the most heavily invaded ecosystems in the world, with a documented 188 established nonindigenous species, of which over 30% have measurable socioeconomic or environmental impacts (32). The most effective strategy for minimizing the impact of invasive species is preventing the introduction and establishment of nonindigenous species altogether, an effort that requires the identification of introduction pathways and efficient monitoring methods (33). Monitoring the presence of nonindigenous species using eDNA approaches can result in higher detection probabilities, allowing for species detections at very low population densities when eradication efforts are less costly and most likely to succeed (34, 35). However, if intraspecific genetic variation can be detected in environmental samples, eDNA approaches may also be useful for identifying the geographic source of introduced individuals, elucidating introduction pathways and patterns of subsequent spread. The genetic variation recovered in eDNA samples can also indicate the presence of multiple individuals in the sampled environment, with DNA mixture models able to estimate the number of distinct individuals in a sample (9, 15). eDNA therefore provides new opportunities to evaluate population densities and introduction pathways for nonindigenous species, information that may be critical for developing effective management plans (36, 37).

In addition to invasive species, eDNA may be useful in the study of intraspecific diversity in species for which tissue collection may be difficult or risks harming the target organism. For instance, marine mammals are a common target in eDNA research (10, 11, 38, 39), at least partially due to the logistical challenges with sampling genetic material of large, potentially deep-diving mammals in remote locations at sea. Obtaining population genetic information from water samples therefore provides an efficient and noninvasive alternative to tissue sampling. Rare or endangered species may also be targeted in population genetic research using eDNA samples, as conventional methods of physically capturing and sampling such species may be ineffective or risk harm to vulnerable species. While this study focused on detecting nuclear genetic diversity in a single species, eDNA samples contain genetic material from many different species. The expansion of this approach to population genetic assessments of multiple species simultaneously is possible (40, 41), although this has not yet been demonstrated with nuclear genetic markers.

Biodiversity is declining at alarming rates, with populations of freshwater species declining at a disproportionately faster rate (42) and exhibiting particularly severe impacts from invasive species compared to terrestrial systems (43). Conserving freshwater biodiversity requires reliable monitoring of species and populations over space and time, but such data are challenging to collect using conventional sampling. eDNA provides an opportunity to conduct bioassessments and biomonitoring surveys with improved species detections over other sampling methods, often with lower effort and cost (44, 45). With the expansion of eDNA approaches into the realm of population genetics, detailed patterns of genetic diversity within and among species may be easily attainable with the collection of environmental samples alone. Although further efforts are needed to investigate and validate the potential for eDNA sequencing to assess within-species genetic diversity, such information will be vital for measuring, predicting, and preventing biodiversity declines across the globe.

Materials and Methods

Sample Collection and DNA Extraction. Water samples and Round Gobies were collected in July to September 2019 at each of 14 locations across the invasive range of Round Gobies (*SI Appendix*, Table S2). Sampled sites encompass locations across four of the five Laurentian Great Lakes (Michigan, Huron, Erie, and Ontario), a site near the first documented Round Goby introduction in Lake St. Clair, and sites along the more recent eastward invasion front in the Erie Canal and Finger Lakes region. We collected eDNA samples without accompanying Round Goby tissues at one site on Seneca Lake, NY, where Round Gobies are not yet known to occur.

At each site, triplicate 2-L surface water samples were collected from shore in sterile collection bags (Whirl-Pak) and filtered through Whatman cellulose nitrate membrane filters (47 mm diameter; 1 μm pore size) using a hand-operated vacuum pump and filter funnel (Pall Corporation). If at any time the filters became clogged or developed cracks due to pressure, filtration was halted, the filter was replaced, and the remaining water was filtered through a new membrane. Following sample collection, we filtered 2 L of distilled water at each site to serve as negative controls. All membrane filters were preserved in 700 μL of Longmire's buffer (100 mM Tris, 100 mM EDTA, 10 mM NaCl, 0.5% SDS) and kept at room temperature for up to 1 wk before storage at −20 °C until DNA extraction. Prior to sampling, all reusable eDNA sampling and filtration equipment was cleaned

with 50% household bleach, rinsed with DI water, and treated under UV light for 30 min. Immediately following eDNA sample collection, up to 28 Round Goby individuals were collected at each site (N = 285; *SI Appendix*, Table S2) via beach seining. Fish were humanely killed on site with buffered MS-222 according to Cornell Center for Animal Resources and Education (CARE) 306.01 (https://www.opefe.com/CARE306.pdf). Fin tissue was sampled from the caudal fin of each fish and stored in 95% EtOH until DNA extraction.

All pre-PCR laboratory work for eDNA samples was carried out in a dedicated PCR-free laboratory where all surfaces and reusable laboratory equipment (e.g., pipettes and tube racks) were bleached after each use and UV-irradiated daily. This laboratory is housed in a separate facility from the laboratory used to handle post-PCR eDNA samples. All eDNA laboratory work was carried out in a separate facility from the laboratory used to process Round Goby tissues.

For Round Goby tissues, we extracted genomic DNA using the HotSHOT protocol (46), which included incubating goby fin clips with 25 mM NaOH at 95 °C for 30 min followed by neutralization with 40 mM Tris-HCl. To extract eDNA, we used the DNeasy Blood and Tissue extraction kit (Qiagen Inc.) with a modified protocol that includes centrifugation of the preservation buffer, increased volumes of lysis and digestion reagents, and extended incubation periods to accommodate eDNA filters (47). A negative control was added at this stage by performing the same DNA extraction steps on 700 μL of molecular grade deionized water in place of Longmire's buffer. If more than one filter was used for a sample, each filter was extracted separately and the elution was pooled into a single sample. The final elution volume for all eDNA samples was 100 μL. Extracted DNA from eDNA samples and Round Goby tissues was stored at −20 °C until amplification.

qPCR assays. To test the relative quantities and detectability of mtDNA and nuDNA markers in eDNA samples, we measured the quantity of each type of DNA (copies/L of sampled water) in each sample using species-specific primer sets and fluorescent TaqMan probes in qPCR assays. The mtDNA marker, developed for previous research on eDNA-based detection of Round Gobies in their invasive range (48, 49), targets a 147 bp region of the Round Goby cytochrome oxidase I (COI) gene (*SI Appendix*, Table S3). Standard curves were generated from 10-fold serial dilutions (10 to 10^5 copies/μL) of a gBlock Gene Fragment (Integrated DNA Technologies). We designed the gBlock fragment based on the consensus sequence of aligned Round Goby COI sequences from GenBank (https://www.ncbi.nlm.nih.gov/genbank/), with the synthesized 191 bp DNA fragment overlapping the target sequences by at least 16 bp on either side. To identify possible sample cross-contamination by the gBlock fragment, position 61 of the synthesized DNA contains adenine in place of guanine, resulting in a stop codon that would not occur in DNA sequences derived from Round Gobies.

To quantify nuclear eDNA, we developed a marker targeting a short, conserved region (60 bp) of the *Nmel505* microsatellite locus (from ref. 15). After aligning all alleles of this locus from genotyped Round Goby tissues in Geneious Prime (2020.0.5), we designed qPCR primers and a TaqMan probe using Primer Express (v3.0.1). As with the mtDNA assay, we designed a gBlock fragment (125 bp) that encompassed the entire target sequence. A 10-fold serial dilution (10 to 10^5 copies/μL) of the synthesized gBlock was used to generate a standard curve, allowing for the quantification of nuDNA copies/L in eDNA samples.

We ran qPCR assays on the *QuantStudio 7 Pro at the* Cornell Institute of Biotechnology's Genomics Facility with three technical replicates of each eDNA sample, standard dilutions, and no-template negative controls in each run. Reactions were prepared in 20 μL volumes containing 10 μL TaqMan Environmental Master Mix 2.0 (ThermoFisher Scientific), 1.8 μL each 10-μM primer (forward and reverse), 0.05 μL of 100-μM Taqman probe, 4.35 μL of PCR-grade water, and 2 μL template DNA. Thermocycler conditions included an initial denaturation step at 95 °C for 10 min followed by 45 cycles of 95 °C for 15 s and either 60 °C (mtDNA) or 58 °C (nuDNA) for 1 min. We used the Design and Analysis software (v2.6.0) to determine amplification efficiency and R^2 for each assay (mean amplification efficiency across all runs was 94.6% ± 6.1 and mean R^2 was 0.995 ± 0.002). The average DNA concentration was calculated for each sample as the average of the three technical replicates and expressed as copy number/L of sampled water. Any replicate with no measurable amplification was treated as having zero DNA copies.

Statistical analysis for qPCR results and the remainder of this study was performed using R v4.1.2 (50) unless otherwise indicated (51). We tested the difference between mtDNA and nuDNA concentration in eDNA samples using a linear mixed model with the 'lmer' function in the lme4 package (52), with sample and site specified as nested random effects. The correlation between the amount of mtDNA and nuDNA was tested using Pearson's correlation test and a linear regression was fitted to the log-transformed data.

Microsatellite PCR, Library Preparation, and Sequencing. To investigate genetic diversity in eDNA samples, we amplified a panel of 28 microsatellite markers for Round Gobies developed in Andres et al (15). Each primer set targets a variable-length region of the nuclear genome, with microsatellite alleles ranging in size from 228 to 435 bp (*SI Appendix*, Table S4). Primer pairs were grouped into multiplexes, and eDNA samples were amplified with each multiplex assay in triplicate to account for PCR stochasticity. Each reaction contained 2 μL of each primer in the multiplex in equimolar concentrations (2 μM), 10 μL of Multiplex PCR Master Mix (Qiagen Inc.), 2 μL bovine serum albumin (BSA; 4 μg/μL), 3 μL PCR-grade water, and 3 μL of eDNA sample. A negative PCR control contained 3 μL molecular H_2O in place of DNA. The program for multiplex PCR included initial denaturation at 95 °C for 15 min followed by 40 cycles of 94 °C for 30 s, 58 °C for 3 min, and 72 °C for 60 s. PCR products were visualized with electrophoresis on a 1% agarose gel stained with ethidium bromide. Equal volumes (5 μL) of each technical replicate and multiplex were combined for each sample and uniquely indexed in a second-stage PCR using Illumina Nextera XT tags. This amplification step included 1.5 μL PCR product, 4 μL of One*Taq* 5× buffer (New England Biolabs, Inc.), 0.4 μL of 10 mM dNTPs, 0.1 μL of One*Taq* DNA polymerase, 12.4 μL of molecular H_2O, and 0.8 μL of each Index 1 (i7) and Index 2 (i5) adapters. The index PCR thermocycling program included six cycles of 95 °C for 30 s, 62 °C for 60 s, and 68 °C for 60 s, with a final extension step of 68 °C for 10 min. We pooled the final amplicons of each sample and negative control (field, extraction, and PCR) in equal volumes and purified the eDNA library with Agencourt AMPure XP beads (reaction ratio AMPure beads 0.7: PCR product 1; Beckman Coulter Genomics). We assessed library concentration with a Qubit 2.0 fluorometer (Thermo Fisher Scientific), diluted to 18 nM, and sequenced at Cornell University's Institute of Biotechnology Genomics Facility on the Illumina MiSeq platform with the MiSeq v2 500 bp kit (paired-end 2 × 250 bp). Due to low read counts in eDNA samples, we resequenced the eDNA library in two additional MiSeq runs (53).

Multiplex PCR amplification of Round Goby tissues followed the same protocol as eDNA samples but with reduced reaction volumes (1 μL of 2 μM primer pairs, 5 μL of Qiagen Multiplex PCR Master Mix, 3 μL PCR-grade water, and 1 μL of Round Goby DNA) and reduced PCR cycles (initial denaturation at 95 °C for 15 min and 35 cycles of 94 °C for 30 s, 58 °C for 90 s, and 72 °C for 60 s). PCR products from each multiplex were pooled (5 μL), and each sample was barcoded using the same index PCR protocol described above. The tissue library was pooled, purified, and diluted as above and sequenced with a single MiSeq v2 500 bp kit (paired-end 2 × 250 bp).

Bioinformatic Analysis. We processed a total of 81,822,102 demultiplexed sequence reads from eDNA samples and 30,255,948 reads from Round Goby tissue samples. Sequencing adapters were trimmed with Trimmomatic v0.39 (54), and sequence quality was assessed with FastQC version 0.11.8 (55). We analyzed all eDNA and tissue sequences using a custom script (https://bitbucket.org/cornell_bioinformatics/amplicon/src/master/amplicon_dada2.py) that involved a modification to the DADA2 pipeline (56) to allow for the processing of multiple loci. The script split sequences based on primers, removed primers from forward and reverse reads, and discarded sequences that were shorter than 100 bp or that had more than two expected errors (EE, calculated based on Phred scores). The DADA2 filtering algorithm was then used to identify and discard sequence errors from each sequence file. Denoised reads were merged with an overlap of at least 20 bp and one allowable mismatch in the overlap region. Although microsatellite primer specificity was inspected using NCBI Primer Blast during primer development (15), we further protected against coamplification of DNA from nontarget species by assessing the similarity of each eDNA sequence variant to Round Goby tissue sequences. To do so, we aligned all sequence variants at each locus to the most common allele from Round Goby tissues using the Smith–Waterman alignment algorithm on the first and last 25 bp of the query and reference sequences. The resulting dataset contained read counts of alleles at each locus for each tissue and eDNA sample. Although qPCR did not detect contamination in any of the negative controls, eDNA sequencing exhibited an average of 17.2 total reads across 12.6 alleles (mean read depth of 1.3 reads

per allele) in the negative controls. We therefore removed all alleles with fewer than 2 reads in a sample to account for the observed low levels of contamination.

Round Goby tissue samples were genotyped at each locus. Individuals were designated as heterozygous if the read ratio between the two most common alleles was at least 0.2:0.8; otherwise, the individual was considered homozygous. Individuals with fewer than 10 reads at a locus were specified as having missing data at that locus. We removed one locus with > 50% missing data across all individuals (*Nmel*60; *SI Appendix*, Table S4). The remaining 27 loci exhibited an average allelic richness of 14.4 alleles (range 3 to 29) with a total of 389 alleles across all loci. Reads for eDNA samples were combined across sequencing runs. eDNA read frequencies were then calculated per sample and per site by combining normalized reads from the three eDNA replicates.

Data Analysis. To investigate patterns of population genetic diversity and structure in the invasive range of Round Gobies, we first conducted population genetic analyses using high-quality data from individual tissues. Allele frequencies for all genotyped individuals and for each population were calculated in adegenet (57) and allelic richness, expected and observed heterozygosity, and the number of private alleles per population was calculated. To test whether time since invasion predicted genetic diversity, we determined the minimum time since invasion using the earliest published record of Round Goby presence within a 10-km radius of each sampled population (58) and assessed the fit with linear models for the effect of minimum time since invasion on allelic richness and number of private alleles. We calculated pairwise estimates of genetic diversity among sampled sites (F_{ST}) and assessed significance using 95% CI calculated using a bootstrap procedure with 1,000 bootstraps in hierfstat (59). An assessment of the correlation between pairwise genetic distances and Euclidean geographic distances (measured in geodist; ref. 60) was conducted using a Mantel test with 999 permutations in ade4 (61). Spatial patterns of genetic structure were visualized using DAPC in adegenet (62) and further inspected using Bayesian assignment tests in STRUCTURE v2.3 (63) using an admixture model with correlated allele frequencies among groups. For each value of K (i.e., potential number of genetic clusters) from 1 to 13, we ran fifteen independent runs of 10^6 iterations of a Markov chain Monte Carlo simulation following a burn-in period of 10^5 iterations. We determined the likely number of genetic clusters according to the Delta-K statistic (64). Hierarchical AMOVA was performed in poppr (65) to determine the proportion of genetic variation that could be attributed to the genetic clusters identified in STRUCTURE and sampling site.

We then assessed the ability of eDNA to detect these patterns of genetic diversity and structure by evaluating the similarity between eDNA-based and tissue-based estimates of allele frequencies and population genetic parameters. Because eDNA contains genetic information from potentially multiple individuals, we consider eDNA read frequencies to estimate population allele frequencies and therefore conduct genetic analyses that can be implemented at the population level. Although individual-level information is available for Round Goby tissues, we also conducted tissue-based analyses at the population level to allow for direct comparison between eDNA- and tissue-based population genetic analysis. To compare the efficiency of sampling individual tissues and collecting eDNA samples at a site, we constructed accumulation curves of allelic richness per eDNA sample and sampled individual using vegan 2.5 to 6 (66). Normalized reads of alleles were combined across triplicate eDNA samples per site and the read frequencies per site were compared to allele frequencies from genotyped tissues using a Pearson's correlation test. From these allele frequencies, we calculated population-level allelic richness, expected heterozygosity, and the number of private alleles. To visualize spatial patterns of genetic variation, we conducted a PCA of the centered population allele frequencies in eDNA and tissue samples.

Because eDNA samples do not contain individual genotypes, we calculated pairwise genetic distances using the absolute allele frequency difference (AFD; ref. 67) among populations. This metric is an intuitive alternative to F_{ST} that requires only population-level allele frequencies in its estimation and is therefore applicable to our eDNA dataset. Pairwise population estimates of AFD and F_{ST} are highly correlated in Round Goby tissues (Mantel test with 999 permutations; $r = 0.81$, $P < 0.001$). We assessed the similarity between tissue-based and eDNA-based pairwise measures of AFD using Pearson's correlation test and assessed the correlation between AFD and Euclidean geographic distances among eDNA samples using a Mantel test. Last, we used pairwise AFD estimates from allele frequencies estimated from both eDNA and Round Goby tissues to construct unrooted neighbor-joining trees using the method of Saitou and Nei (68) in ape (69).

Data, Materials, and Software Availability. Illumina MiSeq raw sequence data; code data have been deposited in NCBI Sequence Read Archive; GitHub (PRJNA997584 (53) and https://github.com/karaandres/eDNA_goby_popgen) (51).

ACKNOWLEDGMENTS. Funding for this project was provided by the University of Michigan Cooperative Institute for Great Lakes Research (CIGLR), through the National Oceanic and Atmospheric Administration Cooperative Agreement NA17OAR4320152 (CIGLR contribution), the NSF Coastal Science, Engineering, and Education for Sustainability program (grant number 1748389), the Department of Defense Strategic Environmental Research and Development Program (grant number RC19-1004), and the Cornell Atkinson Center for Sustainability. We thank Maria Haileselassie, Diana Tapia, George Rogalskyj, Miranda Andres, and Joe Boschert for assistance with collecting specimens in the field. All animals were collected according to local collecting permits.

Author affiliations: [a]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853; [b]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130; [c]Living Earth Collaborative, Washington University in St. Louis, St. Louis, MO 63130; and [d]Cornell Atkinson Center for Sustainability, Cornell University, Ithaca, NY 14853

1. K. Deiner *et al.*, Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Mol. Ecol.* **26**, 5872–5895 (2017).
2. G. F. Ficetola, C. Miaud, F. Pompanon, P. Taberlet, Species detection using environmental DNA from water samples. *Biol. Lett.* **4**, 423–425 (2008).
3. P. F. Thomsen, E. Willerslev, Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* **183**, 4–18 (2015).
4. A. Valentini *et al.*, Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol. Ecol.* **25**, 929–942 (2016).
5. J. Fediajevaite, V. Priestley, R. Arnold, V. Savolainen, Meta-analysis shows that environmental DNA outperforms traditional surveys, but warrants better reporting standards. *Ecol. Evol.* **11**, 4803–4815 (2021).
6. P. F. Thomsen *et al.*, Monitoring endangered freshwater biodiversity using environmental DNA. *Mol. Ecol.* **21**, 2565–2573 (2012).
7. C. I. Adams *et al.*, Beyond biodiversity: Can environmental DNA (eDNA) cut it as a population genetics tool? *Genes* **10**, 192 (2019).
8. E. E. Sigsgaard *et al.*, Population-level inferences from environmental DNA-Current status and future perspectives. *Evol. Appl.* **13**, 245–262 (2020).
9. K. J. Andres, D. M. Lodge, S. A. Sethi, J. Andrés, Detecting and analysing intraspecific genetic variation with eDNA: From population genetics to species abundance. *Mol. Ecol.* **32**, 4118–4132 (2023).
10. K. M. Parsons, M. Everett, M. Dahlheim, L. Park, Water, water everywhere: Environmental DNA can unlock population structure in elusive marine species. *R. Soc. Open Sci.* **5**, 180537 (2018).
11. E. E. Sigsgaard *et al.*, Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nat. Ecol. Evol.* **1**, 4 (2016).
12. M. R. Jensen *et al.*, Genome-scale target capture of mitochondrial and nuclear environmental DNA from water samples. *Mol. Ecol. Res.* **21**, 690–702 (2021).
13. K. Deiner *et al.*, Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods Ecol. Evol.* **8**, 1888–1898 (2017).
14. J. W. O. Ballard, M. C. Whitlock, The incomplete natural history of mitochondria. *Mol. Ecol.* **13**, 729–744 (2004).
15. K. J. Andres, S. A. Sethi, D. M. Lodge, J. Andrés, Nuclear eDNA estimates population allele frequencies and abundance in experimental mesocosms and field samples. *Mol. Ecol.* **30**, 685–697 (2021).
16. D. J. Jude, R. H. Reider, G. R. Smith, Establishment of gobiidae in the great lakes basin canadian. *J. Fish. Aquat. Sci.* **49**, 416–421 (1992).
17. J. E. Bronnenhuber, B. A. Dufour, D. M. Higgs, D. D. Heath, Dispersal strategies, secondary range expansion and invasion genetics of the nonindigenous round goby, Neogobius melanostomus, in great lakes tributaries. *Mol. Ecol.* **20**, 1845–1859 (2011).
18. A. Ricciardi, H. J. MacIsaac, Recent mass invasion of the north american great lakes by ponto-caspian species. *Trends Ecol. Evol.* **15**, 62–65 (2000).
19. J. E. Brown, C. A. Stepien, Invasion genetics of the Eurasian round goby in North America: Tracing sources and spread patterns. *Mol. Ecol.* **18**, 64–79 (2009).
20. E. A. LaRue, C. R. Ruetz, M. B. Stacey, R. A. Thum, Population genetic structure of the round goby in Lake Michigan: Implications for dispersal of invasive species. *Hydrobiologia* **663**, 71–82 (2011).
21. M. L. Johansson *et al.*, Human-mediated and natural dispersal of an invasive fish in the eastern Great Lakes. *Heredity* **120**, 533–546 (2018).
22. K. J. Andres *et al.*, Seasonal habitat use indicates that depth may mediate the potential for invasive round goby impacts in inland lakes. *Freshwater Biol.* **65**, 1337–1347 (2020).
23. A. Furtwängler *et al.*, Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis. *Sci. Rep.* **8**, 1–8 (2018).

24. T. Jo, M. Arimoto, H. Murakami, R. Masuda, T. Minamoto, Estimating shedding and decay rates of environmental nuclear DNA with relation to water temperature and biomass. *Environ. DNA* **2**, 140–151 (2020).
25. T. Jo, K. Takao, T. Minamoto, Linking the state of environmental DNA to its application for biomonitoring and stock assessment: Targeting mitochondrial/nuclear genes, and different DNA fragment lengths and particle sizes. *Environ. DNA* **4**, 271–283 (2022).
26. T. Minamoto *et al.*, Nuclear internal transcribed spacer-1 as a sensitive genetic marker for environmental DNA studies in common carp *Cyprinus carpio*. *Mol. Ecol. Res.* **17**, 324–333 (2017).
27. L. W. Cole, The evolution of per-cell organelle number. *Front. Cell Dev. Biol.* **4**, 85 (2016).
28. C. I. M. Adams *et al.*, Environmental DNA reflects common haplotypic variation. *Environ. DNA*, 10.1002/edn3.352 (2022).
29. C. Bessey *et al.*, Maximizing fish detection with eDNA metabarcoding. *Environ. DNA* **2**, 493–504 (2020).
30. P. M. Kocovsky *et al.*, Expansion of tubenose gobies *Proterorhinus semilunaris* into western Lake Erie and potential effects on native species. *Biol. Invasions* **13**, 2775–2784 (2011).
31. I. Adrian-Kalchhauser *et al.*, The round goby genome provides insights into mechanisms that may facilitate biological invasions. *BMC Biol.* **18**, 1–33 (2020).
32. R. A. Sturtevant *et al.*, Recent history of nonindigenous species in the Laurentian Great Lakes; An update to Mills et al., 1993 (25 years later). *J. Great Lakes Res.* **45**, 1011–1035 (2019).
33. D. M. Lodge *et al.*, Biological invasions: Recommendations for US policy and management. *Ecol. Appl.* **16**, 2035–2054 (2006).
34. M. M. Dougherty *et al.*, Environmental DNA (eDNA) detects the invasive rusty crayfish *Orconectes rusticus* at low abundances. *J. Appl. Ecol.* **53**, 722–732 (2016).
35. M. J. Vander Zanden, G. J. Hansen, S. N. Higgins, M. S. Kornis, A pound of prevention, plus a pound of cure: Early detection and eradication of invasive species in the Laurentian Great Lakes. *J. Great Lakes Res.* **36**, 199–205 (2010).
36. D. M. Lodge *et al.*, Risk analysis and bioeconomics of invasive species to inform policy and management. *Annu. Rev. Environ. Res.* **41**, 17.1–17.36.41 (2016).
37. M. J. Vander Zanden, J. D. Olden, A management framework for preventing the secondary spread of aquatic invasive species, Canadian. *J. Fish. Aquat. Sci.* **65**, 1512–1522 (2008).
38. L. Dugal *et al.*, Individual haplotyping of whale sharks from seawater environmental DNA. *Mol. Ecol. Res.* **22**, 56–65 (2022).
39. D. Székely *et al.*, Environmental DNA captures the genetic diversity of bowhead whales (*Balaena mysticetus*) in West Greenland. *Environ. DNA* **3**, 248–260 (2021).
40. M. Stat *et al.*, Ecosystem biomonitoring with eDNA: Metabarcoding across the tree of life in a tropical marine environment. *Sci. Rep.* **7**, 1–11 (2017).
41. K. Weitemier *et al.*, Estimating the genetic diversity of Pacific salmon and trout using multigene eDNA metabarcoding. *Mol. Ecol.* **30**, 4970–4990 (2021).
42. A. J. Reid *et al.*, Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biol. Rev.* **94**, 849–873 (2019).
43. T. P. Moorhouse, D. W. Macdonald, Are invasives worse in freshwater than terrestrial ecosystems? *Wiley Interdiscip. Rev. Water* **2**, 1–8 (2015).
44. M. Bálint *et al.*, Accuracy, limitations and cost efficiency of eDNA-based community survey in tropical frogs. *Mol. Ecol. Res.* **18**, 1415–1426 (2018).
45. V. Elbrecht, E. E. Vamos, K. Meissner, J. Aroviita, F. Leese, Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods Ecol. Evol.* **8**, 1265–1275 (2017).
46. G. E. Truett *et al.*, Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques* **29**, 52–54 (2000).
47. J. Spens *et al.*, Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: Advantage of enclosed filter. *Methods Ecol. Evol.* **8**, 635–645 (2017).
48. L. R. Nathan, C. L. Jerde, M. L. Budny, A. R. Mahon, The use of environmental DNA in invasive species surveillance of the Great Lakes commercial bait trade. *Conserv. Biol.* **29**, 430–439 (2015).
49. M. B. Nevers *et al.*, Environmental DNA (eDNA): A tool for quantifying the abundant but elusive round goby (*Neogobius melanostomus*). *PLoS One* **13**, e0191720 (2018).
50. R Core Team, R: A language and environment for statistical computing (2021). https://www.R-project.org/. Accessed 28 June 2021.
51. K. Andres *et al.*, eDNA_goby_popgen. GitHub. https://github.com/karaandres/eDNA_goby_popgen. Deposited 24 July 2023.
52. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Software* **67**, 1–48 (2015).
53. K. Andres *et al.*, "PRJNA997584 Environmental DNA reveals genetic diversity and population structure of an invasive species in the Laurentian Great Lakes." NCBI Sequence Read Archive (SRA). Available at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA997584. Deposited 22 July 2023.
54. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
55. S. Andrews, FastQC: A quality control tool for high throughput sequence data (2010). https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 28 June 2021.
56. B. J. Callahan *et al.*, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
57. T. Jombart, adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
58. J. A. Freedman *et al.*, Neogobius melanostomus (Pallas, 1814): U.S. geological survey, Nonindigenous aquatic species database (2022).
59. J. Goudet, Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005).
60. M. D. Padgham, Sumner, geodist: Fast, dependency-free geodesic distance calculations R package. Version 0.0.7 4 (2021). https://CRAN.R-project.org/package=geodist. Accessed 24 July 2023.
61. S. Dray, A.-B. Dufour, The ade4 package: Implementing the duality diagram for ecologists. *J. Stat. Software* **22**, 1–20 (2007).
62. T. Jombart, S. Devillard, F. Balloux, Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 1–15 (2010).
63. J. K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
64. G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
65. Z. N. Kamvar, J. F. Tabima, N. J. Grünwald, Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281 (2014).
66. J. Oksanen *et al.*, Package 'vegan'. Community ecology package (2013), Version 2, pp. 1–295. https://CRAN.R-project.org/package=vegan. Accessed 24 July 2023.
67. D. Berner, Allele frequency difference AFD–an intuitive alternative to FST for quantifying genetic population differentiation. *Genes* **10**, 308 (2019).
68. N. Saitou, M. Nei, The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
69. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).