Received: 22 April 2023 | Revised: 7 June 2023 | Accepted: 13 June 2023

DOI: 10.1002/edn3.444

Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

WILEY

ORIGINAL ARTICLE

How eDNA data filtration, sequence coverage, and primer selection influence assessment of fish communities in northern temperate lakes

Erik García-Machado[1] | Eric Normandeau[1] | Guillaume Côté[2] | Louis Bernatchez[1]

[1]Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Québec, Canada

[2]Ministère de l´Environnement, de la Lutte contre les changements climatiques, de la Faune et des Parcs, Québec, Québec, Canada

Correspondence
Erik García-Machado, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Pavillon Charles-Eugène-Marchand, Québec, QC G1V 0A6, Canada.
Email: erik.garcia-machado.1@ulaval.ca

Funding information
Ministère de l´Environnement, de la Lutte contre les changements climatiques, de la Faune et des Parcs; Société des Établissements de Plein Air du Québec (SÉPAQ); Strategic Partnership Grants for Projects from the Natural Sciences and Engineering Research Council of Canada (NSERC)

Abstract

For nearly 15 years now, environmental DNA has demonstrated its effectiveness in monitoring biodiversity. Methodological and technical improvements have significantly enhanced the field. However, the effect of factors such as sequence coverage, bioinformatic filtration, and primer choice have been less explored or need to be optimized according to specific survey objectives and study site characteristics. We evaluated these factors to help optimize monitoring fish biodiversity in North American temperate lakes. We sampled water for fish community eDNA analysis in 12 lakes from southwestern Québec, Canada. The lakes were selected to encompass a wide range of surface areas and species richness. We sampled water from a total of 520 sites (25–50 per lake) and analyzed three mitochondrial DNA regions (12S rRNA; 16S rRNA; and cytb) using NovaSeq sequencing. Our results, based on rarefied count matrices (from a sequencing depth of 100,000 to a minimum depth of 1000 reads per sample), showed that keeping only species in each sample if they represented at least one thousandth (species minimum read proportion threshold = 0.001) of the sample's reads was adequate to remove false positives and had a limited negative impact on true positives with low read counts. The sequencing depth was found to have a negligible impact on the accuracy of fish community assessment in a given lake. With the same sequencing depth and a complete local reference database for each primer set, a single primer set produced similar species richness medians than the combination of two or three primer sets. Overall, 12S and 16S detected more species and provided more consistent community profiles than cytb. Based on our observations, we suggest using the 12S MiFish-U primer set and applying a minimum proportion of 0.001 reads per species and site to monitor north-temperate lentic freshwater fish communities.

KEYWORDS
environmental DNA, false positive, fish, metabarcoding, optimization, sequencing depth

1 | INTRODUCTION

The cumulative effects of human activities such as overharvesting, habitat fragmentation, introduction of non-native species, and global climate change represent major threats to freshwater fish biodiversity across the world (Su et al., 2021). Biodiversity monitoring is a crucial task to assist in implementing best conservation practices (de Bello et al., 2010; Deiner et al., 2021). Notwithstanding the usefulness of traditional sampling methods, environmental DNA (eDNA) methods are more and more seen as robust and cost-effective surveillance tools (Blackman et al., 2020; Bush et al., 2017; Deiner et al., 2017; Ji et al., 2013). For fish in particular, eDNA metabarcoding allows assessing biodiversity with lower sampling efforts while offering a lower probability of missing species (Carvalho et al., 2022). However, standardizing protocols of diversity surveys from eDNA metabarcoding requires precise knowledge of the experimental parameters (e.g. sampling effort, water volume sampled, number of PCR replicates, etc.) affecting biodiversity estimates (Bruce et al., 2021; Dickie et al., 2018; Grey et al., 2018; Minamoto et al., 2021; Stauffer et al., 2021).

In this study, we consider three specific parameters; (i) the marker (primers) choice, (ii) the minimum proportion of reads in a sample to identify a species, and (iii) the sequencing depth to assess community composition from eDNA metabarcoding. Metabarcoding primers are ideally designed to target all the species of a given taxonomic group while not amplifying species outside of the group. Nowadays, consensus suggests using two or more markers as the best practice to counterbalance the potential biases of each single marker (Cole et al., 2022; Duke & Burton, 2020; Hajibabaei et al., 2019;). However, using multiple primer sets increase both sequencing and analytical costs. As a result, adequate primer selection should be considered an important step of any biomonitoring program (Zhang et al., 2020).

False-positive identifications are an important problem in metabarcoding surveys as they can bias species richness estimates (Jerde, 2021; Rodriguez-Martinez et al., 2023; Smith & Goldberg, 2020). Several methods have been proposed to deal with false positives and how to apply minimum read coverage and site occurrence thresholds to accept a species as present. For example, some authors have used the maximum (Doble et al., 2020; Gehri et al., 2021) or the average number of reads found in negative (Euclide et al., 2021; Sard et al., 2019; Zhang et al., 2019) or positive controls (Bista et al., 2017; De Barba et al., 2014; Hänfling et al., 2016) to correct for false or unwanted species, followed by the application of minimum reads thresholds per species and site (Berger et al., 2020; Civade et al., 2016; Doble et al., 2020; Evans et al., 2017; Gehri et al., 2021; Zhang et al., 2019), a minimum quantity of positive-presence sites for a species (Boivin-Delisle et al., 2021; Euclide et al., 2021; Sard et al., 2019), and/or a minimum number of positive PCR replicates containing a given species (Sard et al., 2019).

Previous studies evaluating sequencing depth revealed that a minimum sequencing depth is needed to accurately describe the diversity using eDNA analysis, particularly in highly diverse communities (Doble et al., 2020; Smith & Peay, 2014; but see Bylemans, Gleeson, Lintermans, et al., 2018). Recently, Shirazi et al. (2021) found that sequence sample read depths and filtering thresholds impact alpha and beta diversity, suggesting that both of these experimental factors should be considered to improve reliability of results.

In this study, we address three questions: (1) How does sequencing depth influence the proportion of sites a given species is detected in typical North American temperate lakes? (2) What minimal proportion of reads per species leads to an adequate balance between false positives removal and detection of rare but true positive species and (3) What primer set or primer set combinations better capture(s) fish diversity in these ecosystems? We assessed the performance of three primer sets targeting three mtDNA loci (12S, 16S, and cytb) to detect freshwater fishes from water samples collected from 12 lakes in Québec, Canada. Based on our results, we formulate recommendations for improving monitoring fish communities in north temperate lakes.

2 | MATERIALS AND METHODS

2.1 | Sampled lakes

We collected water samples in 12 lakes located in southern Québec (Figure 1) which encompasses limnological diversity in a relatively large temperate region (Table 1). We followed a grid-based sampling design, with each cell size set proportional to the size of the lake and took water samples at the middle point of each cell. We fixed the maximum number of samples per lake to 50. When the number of possible cells outnumbered the fixed number of samples in large lakes, we randomly selected which cell to sample by using the function runif() as a random number generator in R v4.2.0 (R Core Team, 2022).

A fish species list for each lake was provided by the Ministère de l'Environnement, de la Lutte contre les changements climatiques, de la Faune et des Parcs du Québec (MELCCFP) and the Société des Établissements de Plein Air du Québec (SÉPAQ) based on surveys conducted after each eDNA sampling events using traditional sampling gears. When available, we also used information from previous surveys (Appendix S1). No data of abundance was available.

2.2 | Water sampling and processing

A total of 520 1 L water samples were collected (Table 1). To avoid eDNA cross-contamination, we used a different boat on each lake. Before each sampling transect, we washed up the boat surface and work clothes by aspersion with 10% commercial bleach (Labo Pro6). Before taking each sample, we also decontaminated the metallic cords, bottle holders, coolers, and icepacks by soaking them in bleach solution followed by rinsing them with purified water. We used 2 L new disposable bottles to take the water samples, bringing the bottles from the sampling depth to the surface (Lacoursière-Roussel et al., 2016). In each lake, we obtained two to four field

Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

FIGURE 1 Map showing the geographical distribution of the 12 lakes investigated in this study. Lakes are presented at an approximate size scale as outsets in the figure.

negative controls consisting of 2 L bottles filled with ultrapure distilled water which were briefly opened during fieldwork and keep in the coolers with field samples. We changed nitrile disposable gloves between sampling stations to prevent cross-contamination. To preserve water samples, we kept them (from 1 to 5 h maximum) in coolers with ice packs until filtration was conducted. Before filtration, we sterilized all materials (filter funnel, connecting tubes, and plastic tweezers) by submerging them in 10% commercial bleach for 30 min followed by ultraviolet exposure for 30 min on both sides. We filtered 500 mL of each water sample and field blank using ultraviolet-sterilized 47 mm, 1.2 μm glass microfiber filter (GF/C Whatman), and a water pressure diaphragm pump. Field blanks were filtered at the middle and the end of each filtration batch. We immediately dried the filters in separate sterile plastic bags containing 30 g of silica gel with a humidity indicator (Merck) and kept them refrigerated (4–10°C). This method was revealed efficient in a pilot assay involving qPCR amplification of eDNA after preservation in different storage temperatures (−20°C; 4°C and room temperature) for a week (data not shown). We stored filters at −20°C upon arrival at the laboratory.

2.3 | eDNA extraction, amplification, and sequencing

We performed eDNA extractions in a safety cabinet (PCR workstation, WWR) placed in a room dedicated to eDNA work. Before each

TABLE 1 Names and characteristics of the 12 lakes sampled in 2019 and 2020 for freshwater fish eDNA in this study.

Lake	Acronym	Region	Coordinates	Surface area (ha)	Maximum deep (m)	Flagged species	Number of samples (*n*)
Sampling year 2019							
Chavoy	Chav	Laurentides	46.7270, −74.49805	36.6	6.2	*Salvelinus fontinalis*	31
Herman	Herm	Laurentides	46.4349, −74.3464	50.5	26	*Salvelinus fontinalis*	33
Sept-Frères	SFre	Outaouais	46.3460, −75.1586	335.7	40	*Salvelinus namaycush*	50
Marie-Le Franc	MLFr	Outaouais	46.14011, −74.9974	668.3	75	*Salvelinus namaycush*	50
Carrière	Carr	Abitibi	47.2549, −77.2080	1432.9	19	*Sander vitreus*	50
Aylmer	Aylm	Estrie	45.8145, −71.3700	3331.5	33.53	*Sander vitreus*	50
Sampling year 2020							
Tape	Tape	Laurentides	46.5478, −74.1498	30.1	27	*Salvelinus fontinalis*	25
Blair	Blai	Laurentides	47.0295, −74.6317	34.0	20	*Salvelinus fontinalis*	31
Papineau	Papi	Outaouais	45.8167, −74,7638	1290.0	72	*Salvelinus namaycush*	50
Brompton	Bromp	Estrie	45.4405, −72.1421	1191.0	42.4	*Salvelinus namaycush*	50
Atocas	Atoc	Laurentides	47.0537, −75.2660	161.3	26	*Sander vitreus*	50
Labyrinthe	Laby	Abitibi	48.2370, −79.4951	787.0	11	*Sander vitreus*	50

Note: Flagged species are the species of interest for recreational fisheries in each sampled lake according to SÉPAQ.

extraction batch, we carefully wiped the safety cabinet, pipettes, scissors, tweezers, and all other equipment using DNA AWAY™ Surface Decontaminant (Molecular BioProducts) followed by ultraviolet exposure for 30 min. We used filtered tips for all steps involving pipetting. We followed the extraction protocols of (Goldberg et al., 2011) and (Spens et al., 2017). Briefly, we cut the filters in half and placed both halves in a 5 mL previously identified tube and added 720 µL ATL buffer and 80 µL of proteinase K, vortexed, and incubated at 56°C overnight. We then used QIAshredder columns and followed the DNeasy Blood and Tissue Kit (Qiagen) protocol for the next extraction steps. In addition to the field negative controls, we included an extraction negative control (consisting of sterile water) for each extraction batch. Each eDNA extraction was diluted in 80 µL of sterile water and split into two tubes (40 µL each) for storage at −20°C.

Three primer sets producing similar amplicon sizes were selected to evaluate their performance at characterizing temperate freshwater fish communities in Québec: 12S Mifish-U (Miya et al., 2015); 16S rRNA (Deagle et al., 2009); cytb (developed for this study) (Appendix S2: Table S1). The MiFish-U universal fish primers (Miya et al., 2015) have proven useful in recent studies to characterize freshwater fish communities of lotic systems in Québec (Berger et al., 2020; Boivin-Delisle et al., 2021; García-Machado et al., 2021; Laporte et al., 2021). For each sample and field negative controls, we ran five PCR replicates that were pooled after amplification. We used a dual-indexing approach for each sample and an 8 bp barcode was added during PCR to the amplicon. The PCR reactions consisted of a total volume of 25 µL including 12.5 µL of Quiagen PCR Multiplex Master Mix, 2 µL of each primer (10 µM), 5.5 µL of diH20, and 3 µL of eDNA sample. For all three primer sets used, we set the amplification conditions as follows: 15 min at 95°C, 35 cycles of amplification (30 s at 94°C, 90 s at X°C, 60 s at 72°C), and a final step of 10 min at 72°C (see Appendix S2 for details in annealing temperature "X" for each gene segment). We added a non-template PCR control for each index combination to track laboratory contamination. To verify the PCR products and negative (field and non-template PCR) controls we ran 3 µL of each reaction on a 1.5% agarose gel stained with 0.5 µg/mL of ethidium bromide. We did not detect contamination in non-template PCR controls and discarded them from sequencing. At each round of amplified eDNA purification and pooling, we used Ultra AMPure PCR purification beads and measured DNA concentration using AccuClear® Ultra High Sensitivity dsDNA Quantitation Kit (Biotium). We pooled the 1683 amplified samples into two libraries (846 from 2019 and 837 from 2020 sampling, respectively) in equimolar concentrations to equalize sequencing depths across samples. We added the full volume of PCR amplification of field and laboratory-negative controls to the pools instead of the median volumes of the test library as suggested by Bruce et al. (2021). That decision might inflate contamination reads numbers but, in return, strengthened stringency to reduce false-positive detections due to laboratory or field contamination. We sent libraries (7.82 and 7.26 ng/µL PCR concentrations, as measured by Bioanalyzer, for libraries one and two, respectively) to the Centre d'expertise et de services Génome Québec (Montréal, Canada) for sequencing using Illumina NovaSeq 6000 in two NovaSeq6000 SP PE250 lanes.

2.4 | Reference databases

For the 12S metabarcoding marker, we updated the MiFish eDNA metabarcoding database (Miya et al., 2020) on June 14, 2021, to which we previously added in situ generated sequences of freshwater and marine fishes of Québec (García-Machado et al., 2021). For the 16S rRNA and cytb metabarcoding markers, we built a reference database based on the list of 118 freshwater fishes reported in Québec (Bernatchez & Giroux, 2012, MELCCFP) but excluded marine and euryhaline species. For 16S and cytb libraries, we first queried MitoFish (Iwasaki et al., 2013) focusing on the wanted species listed for the studied region. Then, we queried the NCBI nucleotide database by searching for species names and the terms 12S, 16S, cytb, and cytochrome. We used MEGA 10.2.5 (Kumar et al., 2018) to align the retrieved sequences and discard entries for species never reported in Québec. Geneious Prime 2020.0.3 was used to search for the homologous regions of the primers selected (i.e., 16S) or to design new primers (i.e., cytb). We then trimmed the sequences outside the 5′ and 3′ primer regions and used PrimerMiner (Elbrecht & Leese, 2015) for in silico evaluation and optimization of the amplification primers for 16S and cytb (Appendix S2). To improve species coverage in the reference databases, we sequenced 66 fish species (16S rRNA sequences, $n = 275$; cytb sequences, $n = 255$) that were not present or poorly represented in reference databases for those markers. Only lampreys (Petromyzontidae) are not covered by the cytb primer set, but these species have never been reported in the lakes studied. We obtained the tissue samples for DNA extraction from the freshwater fish collection in our laboratory (April et al., 2011) (Appendix S2 for protocols).

To evaluate the power of 16S and cytb markers to taxonomically distinguish the species and to pinpoint incorrect identifications, we constructed maximum likelihood phylogenetic trees using MEGA 10.2.5 (Kumar et al., 2018). To prune dubious or erroneous entries, we followed the criteria of Collins et al. (2019). We discarded all sequences from an individual of a given species when it showed the same sequences or clustered with a group of individuals from another species and for which there is consistent evidence that it belongs to a distinct species cluster. We also removed potentially erroneous entries when information from published studies suggested the species sequence identification from a single study conflicted with other conclusive studies. The 12S, 16S, and cytb databases are accessible in Barque v1.7.4 eDNA metabarcoding analysis pipeline (www.github.com/enormandeau/barque) developed in our group (see also Mathon et al., 2021).

2.5 | Bioinformatic procedures

Raw sequences were obtained as demultiplexed fastq files from the Centre d'expertise et de services Génome Québec and the Barque v1.7.4 pipeline. We used trimmomatic v0.36 (LEADING:20, TRAILING:20, SLIDINGWINDOW:20:20, MINLEN:100, CROP:200) (Bolger et al., 2014) for trimming forward and reverse sense sequences, keeping only amplicons of the expected size and containing the primers used. Then we used FLASH v1.2.11 (-z, -m 30, -M 280) (Magoč & Salzberg, 2011) to merge clean reads, following the identification of the expected primer combinations and the splitting of amplicons. We used VSEARCH (Rognes et al., 2016) to identify and exclude PCR chimeras and for taxonomic assignment at the species level with a 97% sequence similarity as a threshold (--qmask none, --dbmask none, --id 0.97, --maxaccepts 20, --maxrejects 20, --maxhits 20, --query_cov 0.6 --fasta_width 0, --minseqlength 20). To improve assignment and reduce spurious sequence assignation to species absent in the studied region, after a first Barque run, we pruned the 12S metabarcoding database (which includes a much larger number of fish species than 16S and cytb metabarcoding databases) from irrelevant species and re-ran the Barque pipeline, which reduced multiple hits and reassigned reads to the expected species. We used genus-level classification in the few cases where closely related species showed identical sequences (Appendix S1 and Table S3). We removed non-fish taxa (mammals, reptiles, birds, and amphibians). The reads detected in field and laboratory negative controls were used to attenuate the impact of false positives in our dataset. For each lake and marker, we corrected the number of reads assigned to each fish species by subtracting the maximum number of reads observed for that species in any of the lake's negative controls. After this correction, we found very low reads numbers for several species in the different lakes. The presence of some of these species was not supported by the available information of captures and species distribution obtained over the years by using a diversity of fishing gears in these same lakes. Since these species do inhabit other lakes included in our study, it is most likely that some of these reads are false positives possibly caused by artefactual tag jumping during the sequencing phase, although the one-step PCR method we used seems less likely affected by this process (Rodriguez-Martinez et al., 2023; Schnell et al., 2015). Therefore, we applied a minimum threshold as a proportion of reads assigned to a species from the total number of reads in the sample to accept species identifications as valid (see below).

2.6 | Statistical analyses

To illustrate the diversity in the studied lakes, we used a shading matrix diagram based on the 12S read counts per sample and species. Only the 12S primer set was used, as visual profiles are very similar for the 16S and cytb markers. We ordered the species in alphabetic order and the lakes by surface size, from the largest (left) to the smallest (right). The log-transformed ($\log10[x+1]$) read matrix was created using the decostand function implemented in the vegan package v 2.6-2 (Oksanen et al., 2022) in R v4.2.0 (R Core Team, 2022). To test for differences among fish species assemblages and the determining factor (i.e., lake and primer set), we performed a two-way ANOVA using the function *aov* in R v4.2.0 (R Core Team, 2022).

2.7 | Species minimum read proportions and sequencing depth

To explore how different species minimum read proportion thresholds and sequence depth per sample influence species detection and to help establish proper filters to define species community for subsequent analysis, we used the 12S reads matrix across lakes because the highest number of sampling sites with the highest number of reads were obtained with that marker (see Section 3). Based on previous surveys information, we performed the analysis on two groups of lakes according to their relative species richness: lakes with high species richness (Aylmer, Brompton, Blair, Carrière, Labyrinthe, Papineau, and Sept-Frères), and lakes with low species richness (Atocas, Chavoy, Herman, Marie-Le-Franc, and Tape). After negative control filtering, we retained samples with 50,000 or more reads. Then, we applied different species minimum read proportion stringencies (minimum read thresholds as a proportion of all reads in a given sample: 0.0001, 0.0002, 0.0005, 0.001) and used the *rrarefy* function from the vegan package v 2.6-2 (Oksanen et al., 2022) in R v4.2.0 (R Core Team, 2022) to generate rarefaction matrices going from a sequencing depth of 100,000 to a minimum of 1000 reads per sample, after which we plotted the evolution of the number of sites where each species was found as a function of the sequencing depth.

2.8 | Comparisons among primer sets

We compared community composition depicted using each primer set (12S, 16S, and cytb) and primer set combinations across and within lakes. To standardize the sequencing depth across primer sets and samples we rarefied (1000 repetitions each time) the data matrices of each primer set to 50,000; 20,000; and 10,000 reads per sample. As the number of species detected across lakes for each rarefaction level was similar for each primer set and because 16S primer set sequencing coverage was much lower compared to 12S and cytb (see Section 3), we kept the 20,000 reads per sample datasets for further analysis. We retained only sites that had a minimum of 20,000 reads and applied rarefaction to a level of 20,000 reads by sample. This was an acceptable compromise to the number of retained samples across lakes (291 samples versus 350 and 197 samples for 10,000 and 50,000, respectively). We use these data sets for the following analyses: (i) quantify the degree to which the tree primer sets produced similar read counts per MOTU, similar spatial patterns of reads distributions per species, and similar estimates of species richness per sample and (ii) to analyze the performance in species identification, diversity estimation, and community structure of the primer sets at a similar sequencing effort for the whole data set and in each lake.

We used Pearson's product–moment correlations to test if the three primer sets produced consistent values of species richness per sample, consistent estimates of sequence reads/species/sample, and consistent patterns of site occurrence per species across lakes. The correlations were based on log10[$x+1$] transformed reads matrices and the trend and 95% confidence level were inferred and plotted using the ggplot2 function geom_smooth (method = lm) to fit the linear model.

We also computed the species richness to test whether the primer set/primer set combinations produced similar species richness values across the 12 studied lakes. Since combinations of two or three primers then benefited from coverages of 40,000 and 60,000 reads per site, respectively, we also rarefied each dataset to 10,000 and 6667 reads so that two and three marker combinations also had a total read depth of 20,000 reads per site. We used pairwise Wilcoxon signed-rank tests with continuity correction to determine if the median number of species detected per lake differed between primer set and primer set combinations. To test for species richness differences among primer sets within lakes, we used the Kruskal–Wallis test followed by a Dunn post hoc test. We used (Benjamini & Hochberg, 1995) false discovery rates (FDR) method, with $\alpha = 0.05$ as cutoff, to correct significance values in all multiple comparisons.

2.9 | Species richness rarefaction curves

We plotted species accumulation curves to compare the performance of primer sets to detect the fish diversity in each lake. As the relationship between eDNA reads and the number of individuals is unknown, we translated the reads matrix into an occurrence matrix (i.e., "1" for presence and "0" for absence) as input data (datatype = "incidence-raw") and plotted interpolation and extrapolation sampling curves and with 95% confidence intervals (bootstrap100 replicates) (Chao & Jost, 2012) using iNEXT function from the R package iNEXT (Hsieh et al., 2016). Finally, to examine the completeness of the number of species inferred by each primer set, we computed Chao2 bias-corrected estimate (Chao, 2005; Colwell et al., 2012), which accounts for unobserved species in the sample, and its 95% confidence intervals after 1000 bootstrap replications using SPadeR (Chao et al., 2016).

2.10 | Similarity of community composition between primer sets

To evaluate differences in the freshwater fish assemblages characterized by each primer set, we first transformed our response matrices reads into matrices of relative abundance by applying the Hellinger transformation (Laporte et al., 2021; Legendre & Gallagher, 2001) using the decostand function. As a dissimilarity measure, we then calculated Euclidian distance matrices and applied the nonmetric multidimensional scaling (NMDS) method to plot the differences using metaMDS function. This was followed by redundancy analysis (RDA) using rda function to test if the communities detected within each lake by the three independent primer sets (i.e., used as an explanatory variable) were compositionally different. To test RDA significance (i.e., overall and of each canonical axis), we used the *F*

statistic as implemented in anova.cca function with 1000 permutations. All these analyses were conducted using vegan R package v 2.6-2 (Oksanen et al., 2022).

3 | RESULTS

3.1 | Data description

After removing sequences of non-fish species and fish species that cannot occur in the study area, the total numbers of reads obtained for each marker were quite similar for 12S (181,740,274 reads), and cytb (144,971,288) but lower for 16S (45,196,729). The lower proportion of fish sequences for 16S was caused by the amplification and sequencing of a high proportion (median = 44%, range 10%–90% per sample) of amplicons that belong to various invertebrate taxa (e.g., copepods). Thus, the 16S sequencing efficiency for similar DNA concentrations was much lower than for the other two primers sets, in addition to being more prone to amplify human DNA (18.3% of total reads) than 12S (0.15%), and cytb (0%). The proportion and number of reads in negative controls (field and laboratory) across lakes was low for all markers, accounting for 0.42% (776,778 reads), 0.09% (39,860 reads), and 0.08% (112,574 reads) for 12S, 16S, and cytb, respectively. The 12S libraries produced the highest read counts per sample, with an average of 351,385 reads (median = 326,827; range = 4767–2,374,084), followed by cytb (average = 276,976; median = 198,627; range = 1750–1,464,054), and 16S (average = 86,840; median = 32,829; range = 196–556,099). The 12S libraries produced the lowest percentage of samples (3.1% and 2.9%) with fewer than 50,000 and 20,000 reads, respectively, followed by cytb (11.7% and 5.7%, respectively), while 16S had the highest percentage of samples with low read counts (58.3% and 40%, respectively).

3.2 | Species minimum read proportion stringency and sequencing depth

The analysis of species minimum read proportion stringency and sequencing depth for the 12S region in terms of site occurrence is presented in Figure 2. When species minimum read proportion filtration was applied, for the sequencing efforts assayed (between 1000 to 100,000 reads per sample), the 500 samples retained and the variety of fish assemblages sampled in this study, none of the species detected at higher sequence depths (100,000) were lost at the lowest sequence depth (1000). However, with no or low species minimum read proportion stringency, we observed a strong decline in the number of sites in which each species was detected as sequencing depth was reduced (Figure 2). We detected the strongest reduction in the group of low species richness lakes (Atocas, Chavoy, Herman, Marie-Le-Franc, and Tape) where 47% of species disappeared from more than 50% of the sites where they were detected at a sequencing depth below 10,000 reads per site, compared to 19% of species in the high species richness lakes. In the low richness group, 47% included species with reads relative abundances ranging from medium-high (purple lines) to low (yellow lines), compared to medium (blue lines) to low in the high species richness lakes. This trend remained similar when applying different species minimum read proportion stringencies. However, the sequencing depth rapidly loses impact, and its effect becomes marginal after a species minimum read proportion threshold of 0.0002 or more. When we compared the list of species detected at each threshold with the species distribution detected with traditional gears in each lake, the best agreement between both methods was observed at a minimum read's threshold of 0.001 (Appendix S1). At this stringency level, a few true positives (i.e., species that were detected using sampling gears) were removed from 12S metabarcoding sequences. For example, the species *Notropis hudsonius* and *Luxilus cornutus* in Lake Tape; *Notemigonus crysoleucas* in Lake Chavoy; and *Lepomis macrochirus* and *Catostomus catostomus* in Lake Brompton. Nonetheless, at this stringency level we found a reasonable balance allowing the removal of most putative false positives with a minimum impact on true positives with low read counts. Therefore, we applied this species minimum read proportion threshold to the three markers for all subsequent analyses.

3.3 | Fish communities across lakes

The shading matrix diagram based on the full 12S reads matrix revealed pronounced differences in fish assemblages among lakes and the spatial heterogeneity in the distribution of sequence reads for the different species (Figure 3). In lakes where they were present, some species were detected in most samples (ex. *Catostomus commersonii* and *Perca flavescens*) while others showed a more localized distribution or were more sparsely distributed. For instance, Walleye (*Sander vitreus*) (lakes Atocas, Aylmer, Carrière, and Labyrinthe), Brook charr (*Salvelinus fontinalis*) (lakes Blair, Chavoy, Herman, and Tape), and Lake trout (*Salvelinus namaycush*) (lakes Brompton, Marie-Le-Franc, Papineau, Sept-Frères) were essentially detected in three groups of lakes where these species are the main targets of anglers (Figure 3). Other species (e.g., *Coregonus clupeaformis*, *Lepomis gibbosus*, and *Notemigonus crysoleucas*, and *Semotilus atromaculatus*) were more sparsely distributed both across and within lakes. A two-way ANOVA showed that lake ($F(2) = 14.186$), primer set ($F(11) = 240.463$), and their interaction ($F(22) = 2.128$) all significantly influenced ($p < 0.01$) the observed differences in species composition.

3.4 | Taxonomic coverage

Overall, the three primer sets were congruent in classifying most MOTUs at the species level. However, some intrageneric taxa were unresolved and, for a few species, assignment to the genus level was adopted across primer set comparisons (Table S3). *Cottus bairdii* and *C. cognatus* were distinguished by the cytb but not with the 12S and

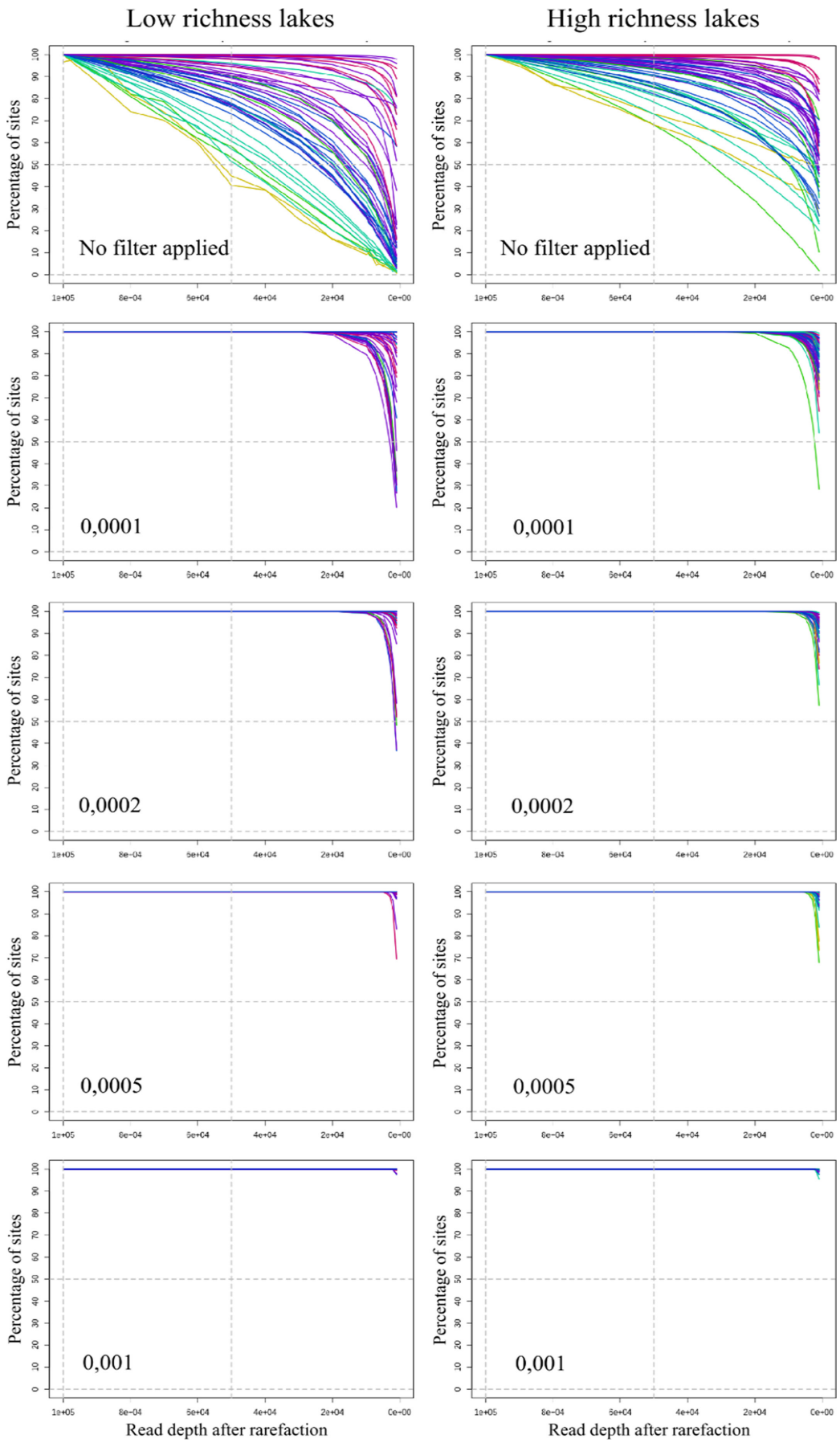

FIGURE 2 Percentage of sites (occurrence) where each species was detected as a function of the read depth and filter stringency. Color gradient graphically indicates the number of reads detected in the whole dataset for each species (reddish for higher and yellow for lowest). Each graphic indicates the filter stringency used (0.0001, 0.0002, 0.0005, 0.001) which is the minimum proportion of reads necessary for a species to be accepted as valid in a sample.

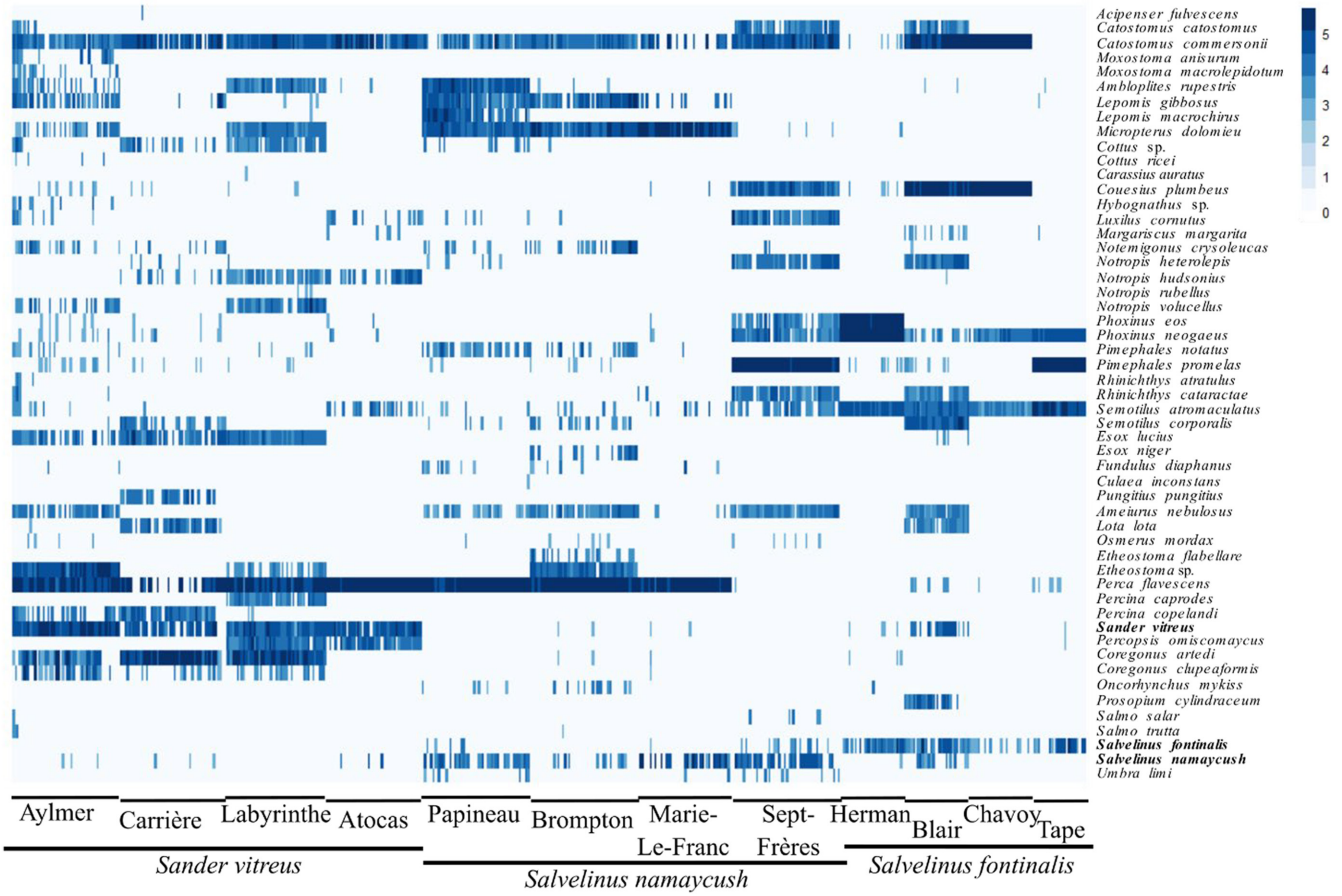

FIGURE 3 Shading matrix showing the species composition across lakes as inferred using the 50,000 reads rarefied eDNA metabarcoding 12S primer set. Blue color indicates the presence and white the absence. Color gradient indicates the number of reads (log[$x+1$] transformed) detected for each species at each site (darker blue for higher).

16S primer sets. Similarly, the 12S reads were assigned unambiguously to *Etheostoma olmstedi* while the 16S primer set associated *Etheostoma* sequences to *E. nigrum* and the cytb associated the sequences to either *E. nigrum* or an unresolved *E. nigrum/E. olmstedi* molecular taxonomic unit (MOTU). For several unresolved MOTUs, we were able to assign the reads to a single MOTU by analyzing the correlation between the reads of the unresolved MOTU and the reads assigned to the corresponding single MOTU. For instance, 12S assigned a large number of reads to an unresolved *Pimephelas notatus*+*Notropis volucellus* group. However, we could assign the reads to *N. volucellus* as a result of a strong correlation ($r=0.997$) between the unresolved MOTU reads' counts and those of *N. volucellus* but not with *P. notatus* read counts ($r=0.031$). Similarly, a number of reads in some lakes were associated to both *Coregonus artedi* and *C. clupeaformis* using 12S, and those number of reads strongly correlated with the number of reads for *C. clupeaformis* ($r=0.95$) and excluded *C. artedi* ($r=0.25$) as the correct species. Finally, cytb could not distinguish *Salvelinus alpinus* from *S. namaycush*, but *S. alpinus* is not present in any of the studied lakes (Rivière et al., 2018) and the reads were assigned to *S. namaycush*.

Surveys conducted using traditional capture methods found 47 species among the 12 study lakes (Appendix S1). Using our combined primer sets, we detected on average 16 ± 3 species per lake with the eDNA surveys, including 43/47 species detected using traditional capture methods and 10 additional species previously unrecorded but falling within their known range of distribution Bernatchez and Giroux (2012), for a total of 53 species detected by eDNA, of which 81% were shared with traditional sampling. In bigger lakes harboring higher species richness (e.g., Aylmer and Brompton), species known to be present based on historical records but not detected in the most recent sampling surveys using different fishing gears (2019 and 2020) were consistently detected by eDNA. For example, 18 of the 29 species historically recorded in Aylmer Lake were detected by eDNA although none were detected by sampling gears in 2019 or

2020. Moreover, 8 of the 19 historically reported species were detected by eDNA in Brompton Lake but in 2020 none were captured by the sampling gear. In lakes Labyrhinte and Marie-Le-Franc, eDNA detected *Percopsis omiscomaycus* in the former and *Salvelinus fontinalis* in the latter which were not detected using fishing methods in 2020 and 2019, respectively (but see https://www.sepaq.com/resources/docs/rf/pal/pal_stats_peche_2022.pdf for recent reports of *S. fontinalis* in Lake Marie-Le-Franc). In contrast, eDNA metabarcoding did not detect *Notropis atherinoides*, a species reported in lakes Carrière and Labyrinthe. Similarly, *N. heterodon* and *Micropterus salmoides* were not detected in Lakes Marie-Le-Franc and Papineau where they are known to occur. Finally, *Hiodon tergisus* was not detected in Lake Aylmer, but this species was only recorded in 1986. Note that all these species are represented in the three reference databases (12S, 16S, and cytb). Finally, cytb detect neither *Fundulus diaphanus* and *Margariscus margarita* in the lakes where the first species (lakes Brompton, Mari-Le-Franc, and Papineau) and the second species (Lake Tape) are known to occur.

3.5 | Comparisons among primer sets

We computed Pearson's product–moment correlations across lake samples and found that read counts per MOTU, species, and species richness per sample were highly correlated ($r > 0.8$) across the three markers (Figure S1A,B). This suggests that the number of reads per species obtained by each marker is proportional and that the distribution of the reads is also relatively similar among markers. The same trend was observed across within-lake comparisons of the three primer sets for sequence reads and site occurrence but not for species richness estimates, for which only 50% of correlations were significant (data not shown). Comparing primer sets across lakes, the 12S and 16S showed the highest correlation for the number of sequence reads per sample ($r = 0.9862$, $p < 0.001$) and the highest correlation for site occurrence ($r = 0.9637$, $p < 0.001$). The correlation graphics among sequence reads per species and among site occurrence revealed the taxonomic bias produced by rare species that are recovered stochastically by the three primer sets, as indicated by the high number of null values observed along both the abscises and ordinate axes of graphics. This impacted the species richness estimates per sample which showed correlation coefficients varying between 0.8003 and 0.8660 ($p < 0.001$) with the highest correlation value observed between 12S and 16S.

The comparison of the median number of species detected by the different primer sets/primer set combinations across lakes revealed that 62% (13/21) of the comparisons illustrated in Table 2 were significantly different (Wilcoxon signed-rank tests $p < 0.05$ with FDR correction). With only two exceptions (i.e., comparison 12S vs. 16S + cytb and comparison 16S vs. 12S + cytb), all primer pair combinations produced statistically significant higher richness in comparison with the single primer data across lakes, with the highest number produced by all three primer sets combined (median = 22, range 7–35) (Figure 4, Table 2). The 12S + 16S (median = 21, range 8–33) combination detected nearly as many species as the three primer sets combination ($p \sim 0.05$). Individually, the median number of species detected by 12S primer set (median = 18.5, range 5–32) overlapped 16S (median = 18.5, range 8–29) and did not differ statistically ($p > 0.05$) from the median of cytb (median = 17, range 5–31), and 16S median number of species was statistically higher ($p < 0.05$) than the median for cytb. Within lakes, 16S and 12S detected a higher proportion of species (87% and 82%) on average compared with cytb with 74% and 12S and 16S medians were higher (Dunn post hoc test $p < 0.05$ with FDR correction) than cytb ones in four lakes (Blair, Chavoy, Labyrinthe, and Papineau) whereas all three markers were statistically indistinguishable in six others (Aylmer, Atocas, Brompton, Herman, Marie-Le-Franc, and Tape). Overall, then, cytb primers tend to show the smallest medians for the number of detected species within a given lake, as observed for the comparisons across all lakes.

3.6 | Species richness rarefaction curves

The species richness rarefaction-based curves revealed variation in how each primer set captured the species diversity within lakes

TABLE 2 Results of the pairwise comparisons of the median number of species detected across lakes among each primer set/primer set combinations using the Wilcoxon signed-rank tests with continuity correction.

Data matrix normalized to equate sequence depth among primer sets							
	12S	**16S**	**Cytb**	**12S + 16S**	**12S + cytb**	**16S + cytb**	**12S + 16S + cytb**
No. species	5–32	8–29	5–26	8–33	5–34	8–30	7–35
Median	18.5	18.5	17	21	20.5	20	22
16S	0.3558						
cytb	0.0787	**0.0206**					
12S + 16S	**0.0411**	**0.0321**	**0.0110**				
12S + cytb	**0.0193**	0.2571	**0.0131**	0.6954			
16S + cytb	0.0638	**0.0197**	**0.0094**	0.3009	0.6623		
12S + 16S + cytb	**0.0159**	**0.0129**	**0.0226**	0.0485	**0.0202**	**0.0206**	

False discovery rates (FDR) correction was applied with $\alpha = 0.05$ as cutoff. The significant p-adjusted values are in bold.

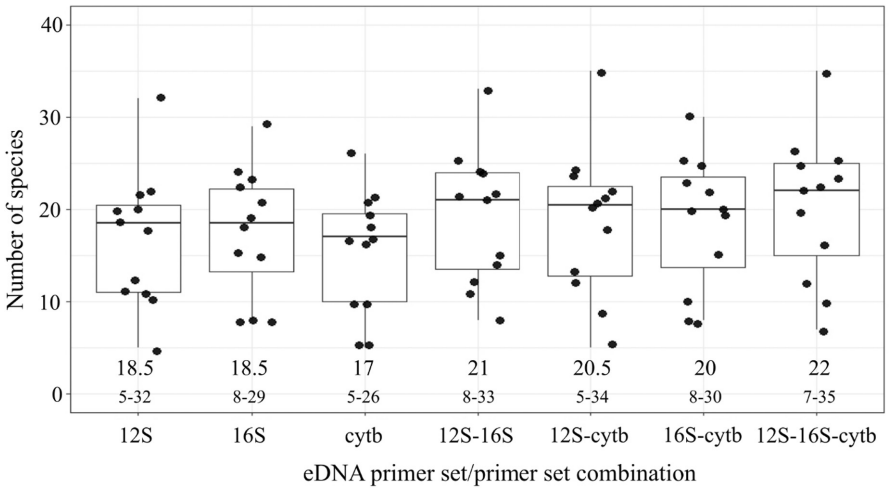

FIGURE 4 Boxplots illustrating the variation of the species richness estimates inferred from each primer set/primer set combinations across the 12 studied lakes. Species richness estimates were obtained after normalizing the sequencing depth per site for marker combinations (rarefied to 10,000 reads for two markers and 6667 reads for three markers). At the bottom are depicted the medians and ranges of species richness.

(Figure 5). Overall, 12S and 16S outperformed cytb in most lakes. For instance, in four lakes (Aylmer, Blair, Herman, and Tape), the curve for 12S consistently lied above the others. However, its 95% confidence intervals overlapped with the other two primer set curves, indicating no significant difference ($p < 0.05$) of expected diversity obtained with the 12S primer set and the other two primer sets. In other cases, the curve was lower (Lake Marie-Le-Franc). The 16S diversity curves were the highest in five lakes (Atocas, Carrière, Chavoy, Papineau, Sept-Frères) with no overlap of the 95% confidence intervals with the other two markers in Chavoy and Sept-Frères indicating higher and significant differences ($p < 0.05$) of expected diversity obtained with 16S primer set. Except for Lake Marie-Le-Franc, the cytb was lower in most of the comparisons or its 95% confidence intervals widely overlapped with 12S or 16S curves. The Chao2 bias-corrected estimates showed that the number of species inferred, based on rare species incidence, did differ from the number of species detected in 58% of the comparisons (7/12 lakes) with 12S, and 83% (10/12 lakes) with 16S and cytb (Table S3). According to Chao2, direct richness values (observed) underestimated species richness. Between one and 10 species remained undetected in seven lakes with 12S, between one and five in eight lakes with 16S, and between one and eight in 10 lakes with cytb.

3.7 | Similarity of community composition among primer sets

The analysis of dissimilarities within lakes revealed a strong consistency among the fish assemblages generated with 12S and 16S (Figure 6, Figure S2). The NMDS plots showed highly overlapping 95% confidence ellipses among these two primers sets in most lakes. In contrast, different freshwater fish assemblages were detected with cytb in several lakes (e.g., Blair, Chavoy, Herman, Labyrinthe, Papineau, Tape), as indicated by the distribution of sample dissimilarity values and the 95% confidence ellipses. Based on redundancy analysis (RDA), the variance explained by the "primer set" parameter ranged between 3.5% for Lake Marie-Le-Franc and 64.1% for Lake Chavoy (Figure S2). Except for the Aylmer and Marie-Le-Franc lakes, all lakes showed statistically significant RDA values, indicating that the primer set choice influenced fish community inference. In these cases, cytb was generally different from the other two. In those lakes with statistically significant RDA values, RDA axis 1 was also significant in all but one (i.e., Lake Atocas) and separated 12S and 16S primer sets from cytb with explained variance varying from 9.3% for Lake Brompton to 64.1% for Lake Chavoy. In Lake Carrière, the 12S primer set differed from the 16S and cytb primer sets, but explained variance was small (5.3%). Only one of the RDA showed a statistically significant axis 2 (Lake Labyrinthe), suggesting some dissimilarities between 12S and 16S. The within-lake differences among primer sets were largely driven by differences in the number of reads assigned to some species relative to the total number of reads for a given primer set. For example, the proportion of cytb reads (range 0.7%–20%) assigned to *Catostomus commersonii* in eight lakes (Atocas, Blair, Brompton, Carrière, Chavoy, Labyrinthe, Papineau, and Sept-Frères) was generally lower than the proportion of reads for 12S (1.8%–44.8%) and 16S (2.7%–42.3%). The same trend was observed for *Phoxinus neogaeus* (0.4%–14.5%) in four lakes (Herman, Chavoy, Sept-Frères, and Tape) compared to the other two primer sets (12S: 2.9%–44.8%, 16S: 3.0%–43.4%). Lower proportions of reads relative to the total number of reads for a given set were also observed for *Ambloplites rupestris*, *Lepomis gibbosus*, *Micropterus dolomieu*, *Notropis volucellus*, *Phoxinus eos*, *Pimephales promelas*, *Salvelinus fontinalis*, *Sander vitreus* in one or two lakes (Figure S2). Yet cytb primer set produced a higher proportion of reads in other lakes (from 13.4% to 76.9% vs. 12S: 6.2% to 50.6% and 16S: 5.7% and 52.4%) for *Couesius plumbeus* in lakes Blair, Chavoy, and Sept-Frères, *Semotilus atromaculatus* (from 2.%7 to 66.0% vs. 12S: 1.5% to 27.7% and 16S: 1.6% and 25.2%) in lakes Chavoy, Herman, and Tape, *Ameiurus nebulosus* (from 4.5% to 15.0% vs. 12S: 0.4% to 1.9% and 16S: 2.1% and 8.7%) in lakes Blair, Brompton, Papineau, and Sept-Frères and *Lepomis macrochirus*, *Notemigonus crysoleucas*), *Notropis heterolepis*, *Percopsis omiscomaycus*, and *Coregonus artedi* in a few lakes.

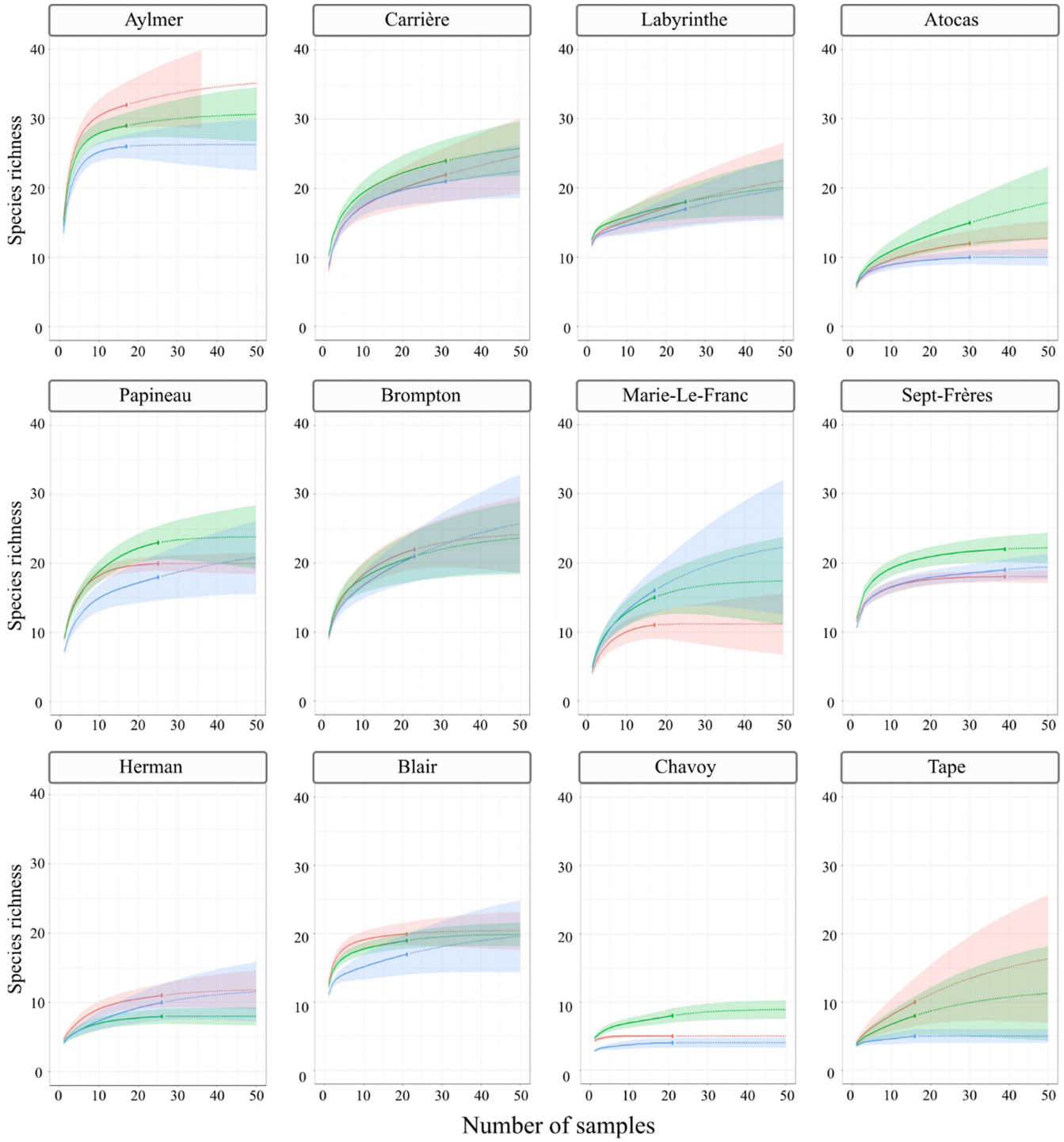

FIGURE 5 Rarefied species accumulation curves for the three eDNA metabarcoding primer sets for each lake surveyed. The solid lines depict the rarefaction curves for 12S (red), 16S (green), cytb (blue), and their respective 95% confidence intervals (shaded areas) obtained after 1000 bootstrap replications. The dots indicated the observed richness, and the dashed line represents the species richness obtained by extrapolation (see Hsieh et al., 2016).

4 | DISCUSSION

4.1 | Sequencing depth

The analysis of the number of sites in which species were detected as a function of species minimum read proportion thresholds and sequencing depth revealed that the number of sampling sites decreases for all species as we increased the threshold but that this effect was stronger in lakes with lower richness. Before we applied any minimum read threshold, sequencing depth strongly impacted the number of sites occupied. A high percentage (47%) of species disappeared in more than half of sites in low richness

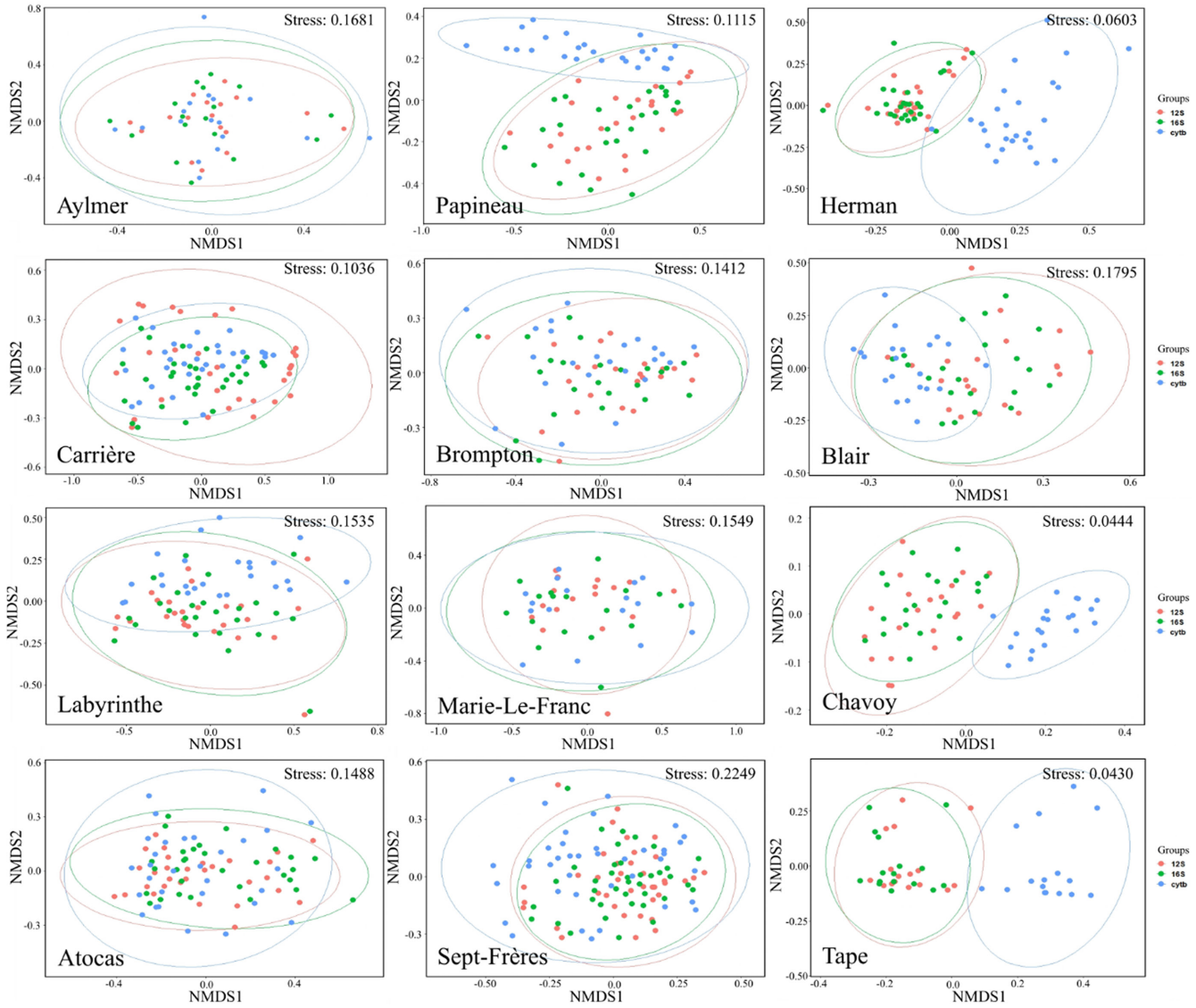

FIGURE 6 Nonmetric multidimensional scaling plots based on Hellinger distance matrices showing sample ordination for each lake and primer set, along with their 95% confidence ellipses.

lakes when sequencing depth was reduced below 10,000 reads per sample, compared with only 19% of species in the high richness lakes for the same sequencing depth. We observed more low-frequency false positives in low richness than in high richness lakes but, ultimately, this effect is linked to the presence of very low abundance species.

Once the read threshold was applied to remove false positives, all MOTUS were detected across lakes, regardless of the species richness level and sequencing depth. This suggests that, at this species minimum read proportion threshold, the increase in sequencing depth did not improve species detection in our study lakes. This is in line with Bylemans, Gleeson, Lintermans, et al. (2018), who found that for low species-richness communities (14 species), increasing sequencing depth only increased species detection moderately. On the other hand, this is in contrast with habitats with exceptional diversities where a higher sequencing depth seems necessary to approach saturation of species (Doble et al., 2020). However, increasing sequencing depth above the minimal number of reads needed for species detection may improve the accuracy of relative read abundance estimates (Shirazi et al., 2021).

4.2 | Species minimum read proportion stringency

The 0.001 species minimum read proportion threshold we used here was adequate to remove most false positives, which represented an important fraction of low-frequency species detected at each lake, as suggested by the historical and more recent record of species obtained with traditional fishing methods. This threshold is in the same range of the 0.001 applied to the 12S sequences in a metabarcoding analysis of lake fish communities in England (Hänfling et al., 2016) and <0.0015 applied to analyze hyperdiverse fish communities of Lake Tanganyika (Doble et al., 2020).

Comparing high and low richness lakes revealed that the number of low-frequency false positives was higher in low compared with high richness lakes. We hypothesize that the proportion of false positives observed resulted from some kind of sequencing-derived cross-contamination on multiplexed samples during library preparation (Kircher et al., 2012; Rodriguez-Martinez et al., 2023; van der Valk et al., 2020; Yin et al., 2019). This well-known effect may artificially increase estimates of species richness (Rodriguez-Martinez et al., 2023). This could explain why we observed a higher impact of false positives on low species diversity samples. Finally, as we expected, it is noteworthy that the minimum read threshold of 0.001 also resulted in removing a few true positives in several lakes, an anticipated and difficult to circumvent consequence of applying reads threshold in metabarcoding analysis (Shirazi et al., 2021; Tsuji et al., 2020). However, moderate to highly abundant species were always recovered in contrast to rare species, which tend to be detected more stochastically, as previously reported (Shirazi et al., 2021).

4.3 | Comparisons among primer sets

Compared with 12S and cytb, the 16S primer set produced about three times fewer fish sequence reads despite equal DNA concentrations during library preparation. This impacted our subsequent analyses on primer sets comparisons since we had to reduce the sequencing depth to 20,000 reads/sample, as well as the number of samples per lake. Reduced sequencing efficiency for 16S resulted from the amplification of invertebrate and human DNA. Similar observations have been previously reported for different 16S primer sets that also showed reduced sequenced efficiency compared with 12S primers (Vences et al., 2016; Zhang et al., 2020).

No primer set alone was optimal to unambiguously identify all species detected across the 12 lakes but, overall, 12S and 16S surpassed cytb in most comparisons. The higher correlation between the 12S and 16S primer sets revealed that both markers returned a highly consistent relative number of reads assigned to each species, a similar spatial detection of species across samples, and that both were in better agreement about what species were detected across lakes relative to cytb. The 12S and 16S (median = 18.5) primer sets also slightly outperformed cytb (median = 17) among and within lakes.

By equalizing sequencing depths, our results show that the expected increase of species richness when using multiple primer sets was modest.

Bylemans, Gleeson, Lintermans, et al. (2018) found no considerable differences in the number of freshwater fishes recovered at different sequencing depths in the Murrumbidgee River in Australia. However, rare species could be missed at 10,000 reads per site. Alberdi et al. (2018) compared read depths from 2500 to 25,000 reads per replicate which showed that diversity increased with sequencing depth. With our filtration criteria, we found that, with very few exceptions, the same species richness was detected with either 20,000 or 6667 reads per sample. Thus, for the same sampling effort in the field, using two or three primer sets could provide a slightly more complete picture of fish biodiversity at the cost of doubling or tripling the cost of library preparation and sequencing. The question then becomes: how important is it in a given biomonitoring survey to ensure that a few rare species are not missed in relative to the increased cost? Yet, we suggest using a higher sequencing depth (ex., 50,000 reads per site) to survey temperate lakes, a value commonly applied in metabarcoding projects (Bruce et al., 2021). In particular, although evaluating the effect of the sequencing depth on species' relative abundances of reads was beyond the scope of this study, such information is important for biodiversity monitoring and higher sequencing depth should improve the estimation of relative abundance, especially for rare species (Lamb et al., 2019).

Within lakes, the comparisons among species richness medians and the species richness rarefaction-based curves revealed variation among primer sets in the number of species detected. Thus, 12S and 16S outperformed cytb with higher species richness estimates in 25% of the lakes, producing similar values or taking turns showing the highest value of number of species. Also, the rarefied species accumulation curves were higher for 12S in four and for 16S in five lakes, respectively. A similar trend using 12S (Riaz et al., 2011) and a modified version of 16S primers (Deagle et al., 2009) was observed when constructing rarefaction species accumulation curves from fish communities from eight lakes in Michigan (Sard et al., 2019). This variability regarding which marker (12S or 16S) performs best across lakes might result from the stochasticity in the amplification of rare species DNA, which could persist despite using five pooled PCR replicates in our study. In particular, performing more PCR replicates independently sequenced to detect rare species is sometimes proposed (Dopheide et al., 2019), but this choice also depends on the filtering parameters, the number of samples, and the sequencing depth, parameters that should all be considered depending on the goal of a given study (Shirazi et al., 2021). Here, the comparison between the number of species observed and inferred (Chao2 bias-corrected estimates) revealed that both estimates were identical for 58.3% of the lakes for 12S and 33.3% of the lakes for 16S. It also revealed that between one and ten species were missed per lake and primer set in those cases where the inferred species number was higher. These values must be considered cautiously since we reduced the number of samples per lake to equate sequencing effort among primers sets, in addition to the reduction of the number of sites species are detected after filtering for false positives. Nonetheless, our results suggest that for the same number of field samples, 12S was in general more efficient in estimating the expected number of species within lakes than either 16S or cytb.

Our results concur with Kelly et al. (2019) who showed that primer choice determines conclusions pertaining to community-wide diversity because differences in amplification efficiency lead to variation in read abundance. We identified several species that contributed significantly to the observed differences between the 12S and 16S fish communities and the ones observed with cytb. These species had lower read count correlation coefficient values

between cytb vs. 12S, and cytb vs. 16S compared to 12S vs. 16S. This also suggests that despite relatively high correlation values in the proportion of reads detected per species by the different primer sets, differences in amplification efficiency are at the origin of the discrepancies among markers, as previously reported (Jusino et al., 2019; Skelton et al., 2022). In our case, however, we can discard the effect of amplicon size as we selected the markers to produce similar size products.

Species richness values from 12S and 16S were always higher than those obtained with cytb, as was observed using 12S and cytb in Lake Windermere, England (Hänfling et al., 2016; Lawson Handley et al., 2019). Similar results were also reported for freshwater and marine waters in England and Australia (Bylemans, Gleeson, Hardy, et al., 2018; Collins et al., 2019). Zhang et al. (2020) showed that the seven 12S primer sets tested (including 12S MiFish) consistently detected a high number of species followed by the 16S primer sets that showed variable results, and cytb that detected fewer species. Zhang et al. (2020) also showed that, compared with different 16S and cytb primer sets, 12S primers exhibited the greatest performance in terms of species discrimination, higher universality, and specificity in fish eDNA. A recent study by Yang et al. (2023) provided further support for higher efficiency of 12S primers relative to 16S by evaluating new and modified variants of published primer sets targeting 12S (Teleo2 Taberlet et al., 2018) and 16S. In summary, although not a general rule, there is a trend emerging from the literature that for freshwater fish in temperate ecosystems at least, 12S primer sets tend to outperform or perform like 16S primer sets and that both outperform cytb primer sets.

5 | CONCLUSIONS AND RECOMMENDATIONS

Based on the results of this study, we propose the following recommendations toward improving the monitoring protocols of fish communities in temperate lakes using eDNA metabarcoding. First, a minimum read threshold of 0.001 to consider a species as valid in a sample is adequate to limit the presence of false positive identifications, while ensuring the true detection of species present, sometimes with few exceptions. As sequencing depth (down to 1000 sequence reads per sample) has limited effect on eDNA detection of rare species it should be adjusted to the objective of each survey (e.g., number of samples, local expected species richness, marker used). Nonetheless, we suggest using higher sequencing depths (e.g., 50,000 reads per site) to increase both the detectability of rare species and improve the estimates of relative read abundance. Among the primer sets we tested, the 12S MiFish-U primer set appears as the best choice to survey freshwater fish communities in northern temperate lakes that would have a comparable level of species diversity as in the lakes surveyed in this study. This primer set accurately discriminated a high number of species across lakes and, on average, performed similarly to 16S within lakes when comparing an equal number of reads per sample. However, as mentioned above, a much lower number of fish sequence reads were obtained with 16S for a same sequencing effort because of non-specific amplification, making 12S a better choice. Although less efficiently sequenced in our study, surveys using the 16S primer set in combination with the 12S could reduce the probability of missing rare species although this would come at the expanse of increasing sequencing and analysis costs. Finally, our results highlight the usefulness of conducting pilot studies to determine the best experimental design for reducing biases and ensure optimal monitoring of fish communities in north temperate lakes using eDNA metabarcoding.

AUTHOR CONTRIBUTIONS

EG-M made major contributions to the acquisition of the data, analysis, interpretation of the data, and writing the manuscript. EN made major contributions to the analysis and interpretation of the data and writing the manuscript. GC made major contributions to the acquisition of the data and writing the manuscript. LB made major contributions to the conception of the study, interpretation of the data, and writing the manuscript.

ACKNOWLEDGMENTS

Funding for this project was provided by the Ministère de l'Environnement, de la Lutte contre les changements climatiques, de la Faune et des Parcs, the Société des Établissements de Plein Air du Québec (SÉPAQ), and the Strategic Partnership Grants for Projects from the Natural Sciences and Engineering Research Council of Canada (NSERC) to LB. The project is also part of the Ressources Aquatiques Québec (RAQ) research program. The authors are sincerely grateful to Amélie Gilbert, Sarah Aubé, Pierre Alexis Drolet, Sébastien Massé, Emilie, Isabeau Caza-Allard, Félix-Antoine Deschênes, Maude Sevigny, and Sann Delaive for field support. The authors also thank Cecilia Hernández, Charles Babin, Alysse Perreault-Payette, and Bérénice Bougas for their precious laboratory assistance, Gabriela Ulmo-Díaz for precious help with R, and Christopher Jerde and two anonymous reviewers for their valuable suggestions to improve the manuscript. The authors thank the Centre d'expertise et de services Génome Québec genomic sequencing platform for their assistance in NovaSeq sequencing.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Dryad at [https://datadryad.org/stash/dataset/doi:10.5061/dryad.k6djh9wc5], reference number [doi:10.5061/dryad.k6djh9wc5].

ORCID

Erik García-Machado https://orcid.org/0000-0001-5720-1733
Louis Bernatchez https://orcid.org/0000-0002-8085-9709

REFERENCES

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*, 134–147. https://doi.org/10.1111/2041-210x.12849

April, J., Mayden, R. L., Hanner, R. H., & Bernatchez, L. (2011). Genetic calibration of species diversity among North America's freshwater fishes. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 10602–10607. https://doi.org/10.1073/pnas.1016437108

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological*, *57*, 289–300.

Berger, C. S., Hernandez, C., Laporte, M., Côté, G., Paradis, Y., Kameni, T. D. W., Normandeau, E., & Bernatchez, L. (2020). Fine-scale environmental heterogeneity shapes fluvial fish communities as revealed by eDNA metabarcoding. *Environmental DNA*, *2*, 647–666. https://doi.org/10.1002/edn3.129

Bernatchez, L., & Giroux, M. (2012). Les poissons d'eau douce du Québec. Broquet. pp. 348.

Bista, I., Carvalho, G. R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., Christmas, M., & Creer, S. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, *8*, 14087. https://doi.org/10.1038/ncomms14087

Blackman, R. C., Ling, K. K. S., Harper, L. R., Shum, P., Hänfling, B., & Lawson-Handley, L. (2020). Targeted and passive environmental DNA approaches outperform established methods for detection of quagga mussels, *Dreissena rostriformis bugensis* in flowing water. *Ecology and Evolution*, *10*, 13248–13259. https://doi.org/10.1002/ece3.6921

Boivin-Delisle, D., Laporte, M., Burton, F., Dion, R., Normandeau, E., & Bernatchez, L. (2021). Using environmental DNA for biomonitoring of freshwater fish communities: Comparison with established gill-net surveys in a boreal hydroelectric impoundment. *Environmental DNA*, *3*, 105–120. https://doi.org/10.1002/edn3.135

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bruce, K., Blackman, R. C., Bourlat, S. J., Hellström, M., Bakker, J., Bista, I., Bohmann, K., Bouchez, A., Brys, R., Clark, K., Elbrecht, V., Fazi, S., Fonseca, V. G., Hänfling, B., Leese, F., Mächler, E., Mahon, A. R., Meissner, K., Panksep, K., … Deiner, K. (2021). *A practical guide to DNA-based methods for biodiversity assessment*. Pensoft Publishers. https://doi.org/10.3897/ab.e68634

Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., Martius, C., Zlinszky, A., Calvignac-Spencer, S., Cobbold, C. A., Dawson, T. P., Emerson, B. C., Ferrier, S., Gilbert, M. T. P., Herold, M., Jones, L., Leendertz, F. H., Matthews, L., Millington, J. D. A., … Yu, D. W. (2017). Connecting earth observation to high-throughput biodiversity data. *Nature Ecology and Evolution*, *1*, 0176. https://doi.org/10.1038/s41559-017-0176

Bylemans, J., Gleeson, D. M., Hardy, C. M., & Furlan, E. (2018). Toward an ecoregion scale evaluation of eDNA metabarcoding primers: A case study for the freshwater fish biodiversity of the Murray–Darling Basin (Australia). *Ecology and Evolution*, *8*, 8697–8712. https://doi.org/10.1002/ece3.4387

Bylemans, J., Gleeson, D. M., Lintermans, M., Hardy, C. M., Beitzel, M., Gilligan, D. M., & Furlan, E. M. (2018). Monitoring riverine fish communities through eDNA metabarcoding: Determining optimal sampling strategies along an altitudinal and biodiversity gradient. *Metabarcoding Metagenomics*, *2*, e30457. https://doi.org/10.3897/mbmg.2.30457

Carvalho, C. S., de Oliveira, M. E., Rodriguez-Castro, K. G., Saranholi, B. H., & Galetti, P. M., Jr. (2022). Efficiency of eDNA and iDNA in assessing vertebrate diversity and its abundance. *Molecular Ecology Resources*, *22*, 1262–1273. https://doi.org/10.1111/1755-0998.13543

Chao, A. (2005). Species estimation and applications. In S. Kotz, N. Balakrishnan, C. B. Read, & B. Vidakovic (Eds.), *Encyclopedia of statistical sciences* (2nd ed., pp. 7907–7916. Wiley.

Chao, A., & Jost, L. (2012). Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology*, *93*, 2533–2547. https://doi.org/10.1890/11-1952.1

Chao, A., Ma, K. H., Hsieh, T. C., & Chiu, C.-H. (2016). *SpadeR (species-richness prediction and diversity estimation in R): An R package in CRAN*. Program and User's Guide.

Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J.-C., Bonin, A., Taberlet, P., & Pont, D. (2016). Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. *PLoS One*, *11*, e0157366. https://doi.org/10.1371/journal.pone.0157366

Cole, V. J., Harasti, D., Lines, R., & Stat, M. (2022). Estuarine fishes associated with intertidal oyster reefs characterized using environmental DNA and baited remote underwater video. *Environmental DNA*, *4*, 50–62. https://doi.org/10.1002/edn3.190

Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., Genner, M. J., & Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution*, *10*, 1985–2001. https://doi.org/10.1111/2041-210x.13276

Colwell, R. K., Chao, A., Gotelli, N. J., Lin, S.-Y., Mao, C. X., Chazdon, R. L., & Longino, J. T. (2012). Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology*, *5*, 3–21. https://doi.org/10.1093/jpe/rtr044

De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular Ecology Resources*, *14*, 306–323. https://doi.org/10.1111/1755-0998.12188

de Bello, F., Lavorel, S., Gerhold, P., Reier, Ü., & Pärtel, M. (2010). A biodiversity monitoring framework for practical conservation of grasslands and shrublands. *Biological Conservation*, *143*, 9–17. https://doi.org/10.1016/j.biocon.2009.04.022

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

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., de Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*, 5872–5895. https://doi.org/10.1111/mec.14350

Deiner, K., Yamanaka, H., & Bernatchez, L. (2021). The future of biodiversity monitoring and conservation utilizing environmental DNA. *Environmental DNA*, *3*, 3–7. https://doi.org/10.1002/edn3.178

Dickie, I. A., Boyer, S., Buckley, H. L., Duncan, R. P., Gardner, P. P., Hogg, I. D., Holdaway, R. J., Lear, G., Makiola, A., Morales, S. E., Powell, J. R., & Weaver, L. (2018). Towards robust and repeatable sampling methods in eDNA-based studies. *Molecular Ecology Resources*, *18*, 940–952. https://doi.org/10.1111/1755-0998.12907

Doble, C. J., Hipperson, H., Salzburger, W., Horsburgh, G. J., Mwita, C., Murrell, D. J., & Day, J. J. (2020). Testing the performance of environmental DNA metabarcoding for surveying highly diverse tropical fish communities: A case study from Lake Tanganyika. *Environmental DNA*, *2*, 24–41. https://doi.org/10.1002/edn3.43

Dopheide, A., Xie, D., Buckley, T. R., Drummond, A. J., & Newcomb, R. D. (2019). Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. *Methods in Ecology and Evolution*, *10*, 120–133. https://doi.org/10.1111/2041-210X.13086

Duke, E. M., & Burton, R. S. (2020). Efficacy of metabarcoding for identification of fish eggs evaluated with mock communities. *Ecology and Evolution*, *10*, 3463–3476. https://doi.org/10.1002/ece3.6144

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—Sequence relationships with an innovative metabarcoding protocol. *PLoS One*, *10*, e0130324. https://doi.org/10.1371/journal.pone.0130324

Euclide, P. T., Lor, Y., Spear, M. J., Tajjioui, T., Vander Zanden, J., Larson, W. A., & Amberg, J. J. (2021). Environmental DNA metabarcoding as a tool for biodiversity assessment and monitoring: Reconstructing established fish communities of north-temperate lakes and rivers. *Diversity and Distributions*, *27*, 1966–1980. https://doi.org/10.1111/ddi.13253

Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., Jerde, C. L., Lodge, D. M., Lamberti, G. A., & Pfrender, M. E. (2017). Fish community assessment with eDNA metabarcoding: Effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, *74*, 1362–1374. https://doi.org/10.1139/cjfas-2016-0306

García-Machado, E., Laporte, M., Normandeau, E., Hernández, C., Côté, G., Paradis, Y., Mingelbier, M., & Bernatchez, L. (2021). Fish community shifts along a strong fluvial environmental gradient revealed by eDNA metabarcoding. *Environmental DNA*, *4*, 117–134. https://doi.org/10.1002/edn3.221

Gehri, R. R., Larson, W. A., Gruenthal, K., Sard, N. M., & Shi, Y. (2021). eDNA metabarcoding outperforms traditional fisheries sampling and reveals fine-scale heterogeneity in a temperate freshwater lake. *Environmental DNA*, *3*, 912–929. https://doi.org/10.1002/edn3.197

Goldberg, C. S., Pilliod, D. S., Arkle, R. S., & Waits, L. P. (2011). Molecular detection of vertebrates in stream water: A demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. *PLoS One*, *6*, e22746. https://doi.org/10.1371/journal.pone.0022746

Grey, E. K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K. L., Lacoursière-Roussel, A., Leong, S. C. Y., Li, Y., Olds, B., Pfrender, M. E., Prowse, T. A. A., Renshaw, M. A., & Lodge, D. M. (2018). Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Scientific Reports*, *8*, 8843. https://doi.org/10.1038/s41598-018-27048-2

Hajibabaei, M., Porter, T. M., Robinson, C. V., Baird, D. J., Shokralla, S., & Wright, M. T. G. (2019). Watered-down biodiversity? A comparison of metabarcoding results from DNA extracted from matched water and bulk tissue biomonitoring samples. *PLoS One*, *14*, e0225409. https://doi.org/10.1371/journal.pone.0225409

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

Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: An R package for rarefaction and extrapolation of species diversity (hill numbers). *Methods in Ecology and Evolution*, *7*, 1451–1456. https://doi.org/10.1111/2041-210X.12613

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., Sado, T., Mabuchi, K., Takeshima, H., Miya, M., & Nishida, M. (2013). MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Molecular Biology and Evolution*, *30*, 2531–2540. https://doi.org/10.1093/molbev/mst141

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

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Woodcock, P., Edwards, F. A., Larsen, T. H., Hsu, W. W., Benedick, S., Hamer, K. C., Wilcove, D. S., Bruce, C., Wang, X., Levi, T., Lott, M., … Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*, 1245–1257. https://doi.org/10.1111/ele.12162

Jusino, M. A., Banik, M. T., Palmer, J. M., Wray, A. K., Xiao, L., Pelton, E., Barber, J. R., Kawahara, A. Y., Gratton, C., Peery, M. Z., & Lindner, D. L. (2019). An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Molecular Ecology Resources*, *19*, 176–190. https://doi.org/10.1111/1755-0998.12951

Kelly, R. P., Shelton, A. O., & Gallego, R. (2019). Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. *Scientific Reports*, *9*, 12133. https://doi.org/10.1038/s41598-019-48546-x

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*, e3. https://doi.org/10.1093/nar/gkr771

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*, 1547–1549. https://doi.org/10.1093/molbev/msy096

Lacoursière-Roussel, A., Rosabal, M., & Bernatchez, L. (2016). Estimating fish abundance and biomass from eDNA concentrations: Variability among capture methods and environmental conditions. *Molecular Ecology Resources*, *16*, 1401–1414. https://doi.org/10.1111/1755-0998.12522

Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., & Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology*, *28*, 420–430. https://doi.org/10.1111/mec.14920

Laporte, M., Reny-Nolin, E., Chouinard, V., Hernandez, C., Normandeau, E., Bougas, B., Côté, C., Behmel, S., & Bernatchez, L. (2021). Proper environmental DNA metabarcoding data transformation reveals temporal stability of fish communities in a dendritic river system. *Environmental DNA*, *3*, 1007–1022. https://doi.org/10.1002/edn3.224

Lawson Handley, L., Read, D. S., Winfield, I. J., Kimbell, H., Johnson, H., Li, J., Hahn, C., Blackman, R., Wilcox, R., Donnelly, R., Szitenberg, A., & Hänfling, B. (2019). Temporal and spatial variation in distribution of fish environmental DNA in England's largest lake. *Environmental DNA*, *1*, 26–39. https://doi.org/10.1002/edn3.5

Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, *129*, 271–280.

Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*, 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

Mathon, L., Valentini, A., Guérin, P.-E., Normandeau, E., Noel, C., Lionnet, C., Boulanger, E., Thuiller, W., Bernatchez, L., Mouillot, D., Dejean, T., & Manel, S. (2021). Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification. *Molecular Ecology Resources*, *21*, 2565–2579. https://doi.org/10.1111/1755-0998.13430

Minamoto, T., Miya, M., Sado, T., Seino, S., Doi, H., Kondoh, M., Nakamura, K., Takahara, T., Yamamoto, S., Yamanaka, H., Araki, H., Iwasaki, W., Kasai, A., Masuda, R., & Uchii, K. (2021). An illustrated manual for environmental DNA research: Water sampling guidelines and experimental protocols. *Environmental DNA*, *3*, 8–13. https://doi.org/10.1002/edn3.121

Miya, M., Gotoh, R. O., & Sado, T. (2020). MiFish metabarcoding: A high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. *Fisheries Science*, *86*, 939–970. https://doi.org/10.1007/s12562-020-01461-x

Miya, M., Sato, K., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M.,

& Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science*, *2*, 150088. https://doi.org/10.1098/rsos.150088

Oksanen, J., Simpson, G., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P., Hara, R., Solymos, P., Stevens, H., Szöcs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Cáceres, M., Durand, S., & Weedon, J. (2022). *vegan community ecology package version 2.6-2 April 2022*.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). ecoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, *39*, e145. https://doi.org/10.1093/nar/gkr732

Rivière, T., Arvisais, M., Banville, D., & Couillard, M. A. (2018). *Rapport sur la situation de l'omble chevalier oquassa (*Salvelinus alpinus oquassa*) au Québec, ministère des Forêts, de la Faune et des Parcs, Direction de la gestion de la faune et des habitats*, 50 p.

Rodriguez-Martinez, S., Klaminder, J., Morlock, M. A., Dalén, L., & Huang, D. Y.-T. (2023). The topological nature of tag jumping in environmental DNA metabarcoding studies. *Molecular Ecology Resources*, *23*, 621–631. https://doi.org/10.1111/1755-0998.13745

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

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

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated – Reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, *15*, 1289–1303. https://doi.org/10.1111/1755-0998.12402

Shirazi, S., Meyer, R. S., & Shapiro, B. (2021). Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. *Ecology and Evolution*, *11*, 15766–15779. https://doi.org/10.1002/ece3.8239

Skelton, J., Cauvin, A., & Hunter, M. E. (2022). Environmental DNA metabarcoding read numbers and their variability predict species abundance, but weakly in non-dominant species. *Environmental DNA*, 1–13. https://doi.org/10.1002/edn3.355

Smith, D. P., & Peay, K. G. (2014). Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PLoS One*, *9*, e90234. https://doi.org/10.1371/journal.pone.0090234

Smith, M. M., & Goldberg, C. S. (2020). Occupancy in dynamic systems: Accounting for multiple scales and false positives using environmental DNA to inform monitoring. *Ecography*, *43*, 376–386. https://doi.org/10.1111/ecog.04743

Spens, J., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S. T., Sigsgaard, E. E., & Hellström, M. (2017). Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: Advantage of enclosed filter. *Methods in Ecology and Evolution*, *8*, 635–645. https://doi.org/10.1111/2041-210X.12683

Stauffer, S., Jucker, M., Keggin, T., Marques, V., Andrello, M., Bessudo, S., Cheutin, M.-C., Borrero-Pérez, G. H., Richards, E., Dejean, T., Hocdé, R., Juhel, J.-B., Ladino, F., Letessier, T. B., Loiseau, N., Maire, E., Mouillot, D., Mutis Martinezguerra, M., Manel, S., … Waldock, C. (2021). How many replicates to accurately estimate fish biodiversity using environmental DNA on coral reefs? *Ecology and Evolution*, *11*, 14630–14643. https://doi.org/10.1002/ece3.8150

Su, G., Logez, M., Xu, J., Tao, S., Villéger, S., & Brosse, S. (2021). Human impacts on global freshwater fish biodiversity. *Science*, *371*, 835–838. https://doi.org/10.1126/science.abd3369

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For biodiversity research and monitoring*. Oxford University Press. https://doi.org/10.1093/oso/9780198767220.001.0001

Tsuji, S., Miya, M., Ushio, M., Sato, H., Minamoto, T., & Yamanaka, H. (2020). Evaluating intraspecific genetic diversity using environmental DNA and denoising approach: A case study using tank water. *Environmental DNA*, *2*, 42–52. https://doi.org/10.1002/edn3.44

van der Valk, T., Vezzi, F., Ormestad, M., Dalén, L., & Guschanski, K. (2020). Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. *Molecular Ecology Resources*, *20*, 1171–1181. https://doi.org/10.1111/1755-0998.13009

Vences, M., Lyra, M. L., Perl, R. G. B., Bletz, M. C., Stanković, D., Lopes, C. M., Jarek, M., Bhuju, S., Geffers, R., Haddad, C. F. B., & Steinfartz, S. (2016). Freshwater vertebrate metabarcoding on Illumina platforms using double-indexed primers of the mitochondrial 16S rRNA gene. *Conservation Genetics Resources*, *8*, 323–327. https://doi.org/10.1007/s12686-016-0550-y

Yang, J., Zhang, L., Mu, Y., & Zhang, X. (2023). Small changes make big progress: A more efficient eDNA monitoring method for freshwater fish. *Environmental DNA*, *5*, 363–374. https://doi.org/10.1002/edn3.387

Yin, C., Liu, Y., Guo, X., Li, D., Fang, W., Yang, J., Zhou, F., Niu, W., Jia, Y., Yang, H., & Xing, J. (2019). An effective strategy to eliminate inherent cross-contamination in mtDNA next-generation sequencing of multiple samples. *The Journal of Molecular Diagnostics*, *21*, 593–601. https://doi.org/10.1016/j.jmoldx.2019.02.006

Zhang, H., Yoshizawa, S., Iwasaki, W., & Xian, W. (2019). Seasonal fish assemblage structure using environmental DNA in the Yangtze estuary and its adjacent waters. *Frontiers in Marine Science*, *6*, 1–10. https://doi.org/10.3389/fmars.2019.00515

Zhang, S., Zhao, J., & Yao, M. (2020). A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. *Methods in Ecology and Evolution*, *11*, 1609–1625. https://doi.org/10.1111/2041-210X.13485

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: García-Machado, E., Normandeau, E., Côté, G., & Bernatchez, L. (2023). How eDNA data filtration, sequence coverage, and primer selection influence assessment of fish communities in northern temperate lakes. *Environmental DNA*, *00*, 1–18. https://doi.org/10.1002/edn3.444