Received: 22 April 2019 | Revised: 27 September 2019 | Accepted: 2 October 2019

DOI: 10.1002/edn3.43

ORIGINAL ARTICLE

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

WILEY

Testing the performance of environmental DNA metabarcoding for surveying highly diverse tropical fish communities: A case study from Lake Tanganyika

Christopher J. Doble[1,2,3] | Helen Hipperson[3] | Walter Salzburger[4] | Gavin J. Horsburgh[3] | Chacha Mwita[5] | David J. Murrell[1,2] | Julia J. Day[1,2]

[1]Department of Genetics, Evolution and Environment, University College London, London, UK

[2]Centre for Biodiversity and Environment Research, University College London, London, UK

[3]Department of Animal and Plant Sciences, NERC Biomolecular Analysis Facility, University of Sheffield, Sheffield, UK

[4]Department of Environmental Sciences, Zoological Institute, University of Basel, Basel, Switzerland

[5]Department of Aquatic Sciences and Fisheries, University of Dar es Salaam, Dar es Salaam, Tanzania

Correspondence
Christopher J. Doble, Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK.
Email: christopher.doble.13@ucl.ac.uk
Julia J. Day, Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK.
Email: j.day@ucl.ac.uk

Funding information
Natural Environment Research Council, Grant/Award Number: NE/L002485/1; Genetics Society Heredity Fieldwork Grant; Systematics Research Fund; Fisheries Society of the British Isle Small Research Grant; Percy Sladen Memorial Trust Fund; NERC Bimolecular Analysis Facility Sheffield; European Research Council; Swiss National Science Foundation

Abstract

Background and Aims: Environmental DNA (eDNA) metabarcoding provides a highly sensitive method of surveying freshwater fish communities, although studies to date have largely been restricted to temperate ecosystems. Due to limited reference sequence availability and challenges identifying closely related and rare species in diverse tropical ecosystems, the effectiveness of metabarcoding methods for surveying tropical fish communities from eDNA samples remains uncertain. To address this, we applied an eDNA metabarcoding approach to survey Lake Tanganyika's (LT) species-rich littoral fish communities.

Materials and Methods: As this system contains many closely related species, particularly cichlid fishes, we used four primer sets including a cichlid-specific primer set (Cichlid_CR). A reference database was built for the 12s, 16s, and control region for 358 fish species including over 93% of known cichlids.

Results and Discussion: In silico and in situ results demonstrated wide variability in the taxonomic resolution of assignments by each primer with the cichlid-specific marker (Cichlid_CR) enabling greater species-level assignments for this highly diverse family. A greater number of non-cichlid teleost species were detected at sites compared to the visual survey data. For cichlid species however, sequencing depth substantially influenced species richness estimates obtained from eDNA samples, with increased depths producing estimates comparable to that obtained from the visual survey data.

Conclusions: Our study highlights the importance of sequencing depth and local reference databases when undertaking metabarcoding studies within diverse ecosystems, as well as demonstrating the potential of eDNA metabarcoding for surveying diverse tropical fish communities, even those containing closely related species within evolutionary radiations.

KEYWORDS
cichlids, environmental DNA, Lake Tanganyika, metabarcoding, reference database, tropical freshwaters

1 | INTRODUCTION

Freshwaters globally represent highly productive and biologically diverse ecosystems, with much of this diversity centered in the tropics (Collen et al., 2014). Aquatic habitats in South America, Central and Eastern Africa, and South-East Asia contain the highest species richness and endemicity across all major freshwater taxonomic groups (excluding crayfish), highlighting the importance of these regions for global freshwater diversity (Tisseuil et al., 2013). As well as hot spots of diversity, tropical freshwaters have also been focal points of development and as a result face a broad range of stressors (Strayer & Dudgeon, 2010). This has resulted in rates of biodiversity decline in freshwaters surpassing that in both terrestrial and marine ecosystems, with extinction rates of freshwater fishes in the twentieth century exceeding that of all other vertebrate groups (Burkhead, 2012; Collen et al., 2014). The recent Freshwater Living Planet Index reports average recorded population declines since 1970 in the Neo- and Afrotropics of 94% and 75%, respectively (Grooten & Almond, 2018). This exceeds terrestrial declines and highlights the substantial pressures on species within tropical freshwater ecosystems.

Species richness and evenness measures underpin our understanding of biological diversity and our ability to monitor their responses to anthropogenic stressors. These measures are reliant on the accurate detections of species as they assume survey data are representative of the community sampled (Buckland, Studeny, Magurran, & Newson, 2011). Therefore, variation in the detectability of species due to differing behaviors or across habitat types can lead to inaccuracies in diversity measures (Gotelli & Colwell, 2011). Traditional methods of surveying freshwater ecosystems all impose biases on datasets often relating to the size and activity of individuals (Jackson & Harvey, 1997). As a result, the survey method used has been shown to significantly influence the diversity values obtained from sites (Deacon et al., 2017; Jackson & Harvey, 1997; Oliveira, Gomes, Latini, & Agostinho, 2012). To help overcome these individual biases, applying multiple methods for surveying fish communities has been widely advocated (Kubecka et al., 2009).

Recently, environmental DNA (eDNA) metabarcoding has been shown to be a sensitive and cost-effective method of surveying freshwater communities (Evans et al., 2017; Hänfling et al., 2016; Valentini et al., 2016). Comparisons with other survey methods have demonstrated how eDNA metabarcoding data can complement traditional approaches often leading to increased detection of species and improved species richness estimates (Deiner et al., 2017). Despite this, current applications of eDNA metabarcoding within freshwaters have largely been restricted to temperate ecosystems, with few published studies within tropical freshwaters (Cantera et al., 2019; Cilleros et al., 2019; Lopes et al., 2017). Potential variability in the effectiveness of these methods across temperate and tropical ecosystems could result from differences in the ecology of eDNA (Eichmiller, Best, & Sorensen, 2016; Strickler, Fremier, & Goldberg, 2015), variation in the detectability of rare species within highly diverse tropical fish communities, the existence of taxonomic problems for some fish taxa (Decru et al., 2016), and the ability of markers to distinguish between closely related species in tropical ecosystems (Breman, Loix, Jordaens, Snoeks, & Van Steenberge, 2016; Pereira, Hanner, Foresti, & Oliveira, 2013). More diverse systems also require higher sampling depths to obtain accurate species richness estimates (Gotelli & Colwell, 2011). As a result, current eDNA metabarcoding designs may require modifications (e.g., increased sequencing depth) to effectively survey the diverse fish communities found in many tropical freshwater ecosystems. A recent application of eDNA metabarcoding within tropical South American streams highlighted the potential of this method for surveying diverse fish communities, while also demonstrating some current limitations in the detection of species compared to studies within temperate ecosystems (Cilleros et al., 2019). Still in its infancy as a method, there remains a need to test eDNA metabarcoding methods across a wider range of complex tropical ecosystems to better understand the potential of these approaches for surveying freshwater fish diversity globally.

Limiting factors preventing the application of eDNA metabarcoding within tropical systems include difficulties collecting and preserving samples within remote locations, incomplete taxonomic descriptions for many fish groups resulting in cryptic biodiversity (Decru et al., 2016; Sales et al., 2018), and the limited availability of sequences within public databases for tropical fish species, particularly for the commonly used 12s and 16s mitochondrial gene regions. Due to the extensive molecular and taxonomic work undertaken on Lake Tanganyika's (LT) fish fauna (Salzburger, Van Bocxlaer, & Cohen, 2014; and refs therein), most fish groups in this system are well studied and there is also a good availability of DNA samples and sequences from museum and research group collections. As a result, LT provides an ideal tropical system with which to test eDNA metabarcoding methods.

Lake Tanganyika contains an exceptional fish diversity with over 400 species, most of which are endemic to the basin (Salzburger et al., 2014). A key feature of LT's fish fauna is that much of its diversity emerged through in situ evolutionary radiations, including radiations of cichlid fishes with at least 241 known species (Day, Cotton, & Barraclough, 2008; Muschick, Indermaur, & Salzburger, 2012; Ronco, Büscher, Indermaur, & Salzburger, 2019), catfishes (Day and Wilkinson 2006, Peart et al. 2014), and mastacembelid spiny eels (Brown et al., 2010), with the latter noncichlid groups containing far fewer species. The high levels of sequence similarity between closely related and young species emerging from rapid evolutionary radiations can make accurate barcode identifications challenging (Salzburger, 2018). For example, a recent study of LT cichlid fishes showed the taxonomic resolution of the traditional COI barcoding region was limited in some cases to species complexes and genera (Breman et al., 2016). As with other large freshwater lakes, much of LT's fish diversity is found within the littoral zone (Vadeboncoeur, McIntyre, & Vander Zanden, 2011). The local species richness of fish communities is particularly high within littoral rocky habitats where as many as 60 fish species, including 54 cichlids, have been identified within a multiyear 10 × 40 m quadrat survey (Takeuchi, Ochi, Kohda, Sinyinza, & Hori, 2010), and 49 cichlid species identified at one site in Mahale National Park

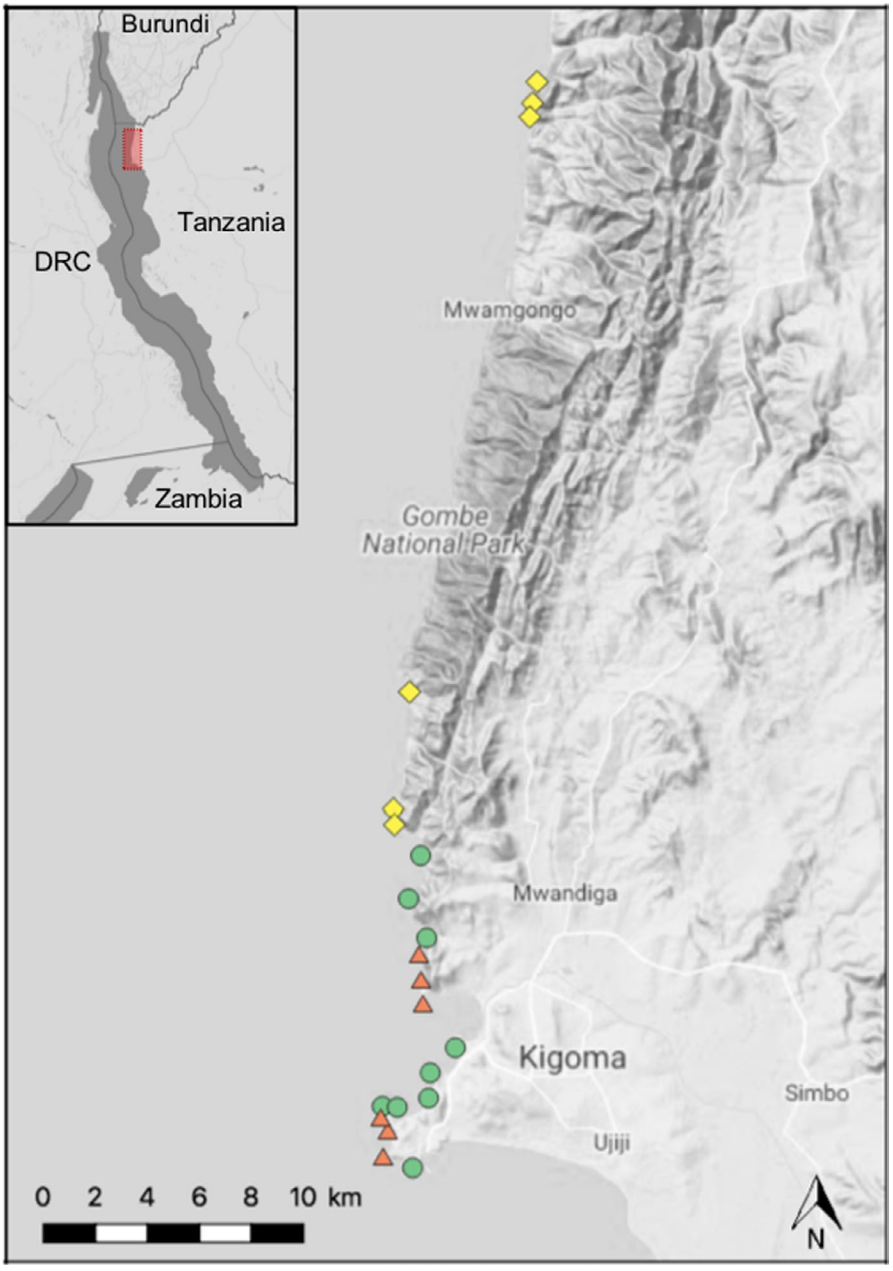

FIGURE 1 Map of Lake Tanganyika sampling locations for both field seasons in 2016 and 2017. Site numbers increase from 1 to 21 in a northwards direction. Green circles are sites surveyed in 2016 only, yellow diamonds show sites surveyed in 2017 only, and red triangles are sites surveyed in both years. Inset: Lake Tanganyika, with the red box highlighting the study area. Maps were created with QGIS v3.2. The base map is Google™ Terrain map

(Britton et al., 2017). As such, accurate and consistent surveying of species within this habitat poses significant challenges, with methods often reliant on labor-intensive SCUBA visual surveys or gill netting, although see Widmer et al. (2019) who utilized video technology for surveying LT cichlid fishes.

The high local diversity and large number of closely related and young species within LT's littoral habitat poses a number of challenges to eDNA metabarcoding methods, many of which will be common across other tropical ecosystems. Here, we develop an extensive reference sequence database across key barcoding regions and collect eDNA samples alongside visual survey data to address a number of these challenges. Specifically, we asked: (a) Can eDNA metabarcoding methods accurately identify LT cichlid fishes to species level? (b) How effective is eDNA metabarcoding at detecting noncichlid teleost species within the LT littoral habitat? (c) Are eDNA species richness estimates and detection rates comparable to visual survey data?

2 | MATERIALS AND METHODS

2.1 | Site locations

Sampling was undertaken within the Kigoma region of LT in Tanzania (Figure 1), representing a well-studied and easily reachable area of the lake (GPS locations, Table S5). The fish communities along this section of coastline have received substantial research, with surveys of the littoral fish communities having been undertaken since 2015 (Britton et al., 2017). Two field seasons were undertaken to LT in September–October 2016 and May–June 2017 during which visual

SCUBA survey data and eDNA samples were collected across 21 sites along a 42 km stretch of coastline.

2.2 | Sampling

Surveys were focused along areas of rocky habitat, as this contains the most diverse littoral fish fauna (Hori, Gashagaza, Nshombo, & Kawanabe, 1993). The structure of littoral communities has been shown to differ across very fine spatial scales, particularly between depths of 1–15 m (Takeuchi et al., 2010). To capture this, a nested design survey following Britton et al. (2017) was adopted at each site in which a series of ten stationary visual surveys were undertaken at depths of 5 and 10 m (Figure S1) (Bohnsack & Bannerot, 1986). At each depth, five surveys were undertaken positioned at the central eDNA collection point and at 30 and 60 m along the coastline in either direction. For further details of the survey design, see Appendix S2.

Prior to undertaking the visual surveys, a water sample was collected from the mid-survey point at a depth of 5 m following a design similar to Port et al. (2016). Containers used for collecting each eDNA sample were rinsed with 50% bleach solution (sodium hypochlorite concentration unknown) followed by lake water at the collection site. Divers remained 1 m off the bottom to prevent kicking up sediment that could alter the eDNA within the water column. Nitrile gloves were also worn to reduce the potential of contaminating samples with their own DNA. A single 10 L water sample was collected within a collapsible container, transferred to a more solid container, and stored on ice within a cooler box, while visual surveys were undertaken. Water samples were subsequently filtered once back onshore within 12 hr of collection.

Nitrile gloves were worn throughout the filtering process, and prior to filtering, the work surface was cleaned with 50% commercial bleach solution. eDNA samples consisted of 1.5-L water filters (subsamples of the 10 L water sample), with one eDNA sample collected per site in 2016 and three filter replicates collected per site in 2017. Each 1.5 L sample was vacuum-filtered onto sterile 47 mm diameter, 0.45 µm pore size cellulose nitrate filter paper contained within 250 ml disposable Thermo Scientific™ Nalgene™ Analytical Test Filter Funnels. Filters were then folded inwards three times and placed within a 2-ml Eppendorf tube. In 2016, this tube was stored immediately at −18°C, transported back to the UK at >−80°C using a dry shipper, and then stored at −70°C until extraction. In 2017, samples were fully submerged in 95% molecular grade ethanol at room temperature until returning to the UK after which they were stored at −20°C. A filtration blank was collected between the filtration of every five samples to monitor for potential contamination. This involved filtering 1.5 L of commercial bottled water (Kilimanjaro brand) following the same procedure as above.

2.3 | Reference database

A multimarker approach was adopted to overcome low divergences between closely related species for individual markers. A total fish list for species within the LT basin was developed based on FishBase (www.fishbase.org), Brichard (1989), and Ronco et al. (2019) (Table S1). Cichlid 12s, 16s, and control region sequences for 250 species were extracted from available mitogenome alignments derived from whole genome assemblies. Separate species with identical marker sequences within the reference database were grouped into species complexes, similar to Breman et al. (2016). Separate species complex groupings were undertaken for each primer set based on marker resolution (Table S2). For each region, one sequence per species was included in the reference database, except for *Oreochromis tanganicae*, where a second sequence was included from an individual collected from a fish farm within LT, in case the farmed populations differed from wild ones. *Oreochromis tanganicae* fish farm, *Oreochromis malagarasi*, and *Oreochromis niloticus* sequences were obtained separately through Sanger sequencing (Appendix S2), and an *Aulonocranus dewindti* control region sequence from NCBI was included in the reference database as mitogenome extract sequences were not available for these species.

Noncichlid fish sequences were obtained using samples from research group collections (Day and Salzburger labs), the South African Institute for Aquatic Biodiversity (SAIAB) and the American Museum of Natural History (AMNH). Some samples for species within the reference database, largely inhabiting river catchments, were collected outside of the LT basin. As the taxonomies for some of these fish groups remain unresolved, samples collected outside of this range may represent separate species yet to be described. Nevertheless, they would be close relatives to those species found within the LT basin. Eight samples collected within the LT basin were identified to genus level for these fish groups. These samples were also sequenced and included in the database to help overcome the potential issue of poorly described taxonomies (e.g., *Amphilius*, *Clarias*, *Enteromius*, *Kneria*, *Leobarbus*, and *Opsaridium*). For the laboratory methods undertaken to obtain reference database sequences, see Appendix S2.

2.4 | Primer design, in silico, and in vitro testing

Reference database sequences for each region were used to design new primers targeting the lake's cichlids and test these in silico along with previously published universal fish primer sets. ecoPrimers was used to search these regions for suitable variable barcode locations (Riaz et al., 2011). This identified a highly variable portion of the control region enabling improved taxonomic resolution. Primers flanking this region were designed using sequence alignments and Primer3 v0.4.0 (Untergasser et al., 2012). This primer set amplifies a 307 bp barcode fragment, with the forward primer located within the same region as the commonly used L-Pro-F forward primer (Meyer, Morrissey, & Scharti, 1994). The 16s_Teleo primer set was designed following a similar method targeting the same variable region as Ve16s, while amplifying a shorter barcode of 275 bp compared to 310 bp (Evans et al., 2016). The newly designed Cichlid_CR and 16s_Teleo primers along with previously published 12S-V5 and MiFish-U primer sets were selected for in silico testing (Miya et al., 2015; Riaz et al., 2011).

TABLE 1 Primers used for the first round of amplifications

Name	Region	Forward primer 12 (5'–3')	Reverse primer (5'–3')	Barcode length (bp)	Reference
MiFish-U	12s	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG	170	Miya et al. (2015)
12S-V5	12s	ACTGGGATTAGATACCCC	TAGAACAGGCTCCTCTAG	106	Riaz et al. (2011)
16s_Teleo	16s	GACGAGAAGACCCTDTGGAG	GTCCTGATCCAACATCGAG	278	This publication
Cichlid_CR	Control Region	CCTACCCCTAGCTCCCAAAG	ACTGATGGTGGGCTCTTACTACA	307	This publication

Primer specificity was evaluated using PrimerMiner v0.18 with threshold scores ranging between 10 and 300 (Bylemans, Gleeson, Hardy, & Furlan, 2018; Elbrecht & Leese, 2017). Kimura-2-Parameter (K2P) distances were calculated for each marker using ape v5.1 to investigate the genetic divergences between species within the reference databases (Kimura, 1980; Paradis, Claude, & Strimmer, 2004). DNA extracts for 17 fish species found within LT were amplified with each primer set to test their consistency of amplification across taxa (Table S8). 1.5 L eDNA samples, filtered following the same methods as for the field samples, were also collected from the LT tank at Zoological Society of London (ZSL). This aquarium contained five cichlid species endemic to the lake (*Julidochromis* sp., *Haplotaxodon microlepis*, *Altolamprologus calvus*, *Neolamprologus longior*, and *Lepidiolamprologus kendalli*) with eDNA samples sequenced following the same metabarcoding methods detailed below. The four selected primers are shown in Table 1 with their locations in each gene region highlighted in Figure S2.

2.5 | DNA extraction and PCR amplifications

Further details of the DNA extractions and PCR amplifications are provided within the Appendix S2. Briefly, sample filters were extracted using the Qiagen DNeasy® Blood & Tissue Kit following a modified protocol in combination with a Qiagen QIAshredder based on the methods of Goldberg et al. (2011) and Lacoursière-roussel et al. (2016). DNA was amplified using a two-step PCR protocol, with barcoded Illumina adapters added in a second amplification (PCR2). Amplicon PCRs (PCR1) were replicated four times for each sample and pooled to minimize PCR bias. All filter and extraction controls were included in the PCRs.

2.6 | Library quantification and original sequencing run

Following PCR2, 2 µl of product from each reaction was quantified using a FLUOstar Optima (Promega). Based on these results, samples were normalized to equal concentrations, pooled into groups of 8, and cleaned with AmPure XP beads. The Illumina-tagged DNA concentration of each pool was quantified using the KAPA Library Quantification Kit run on a QuantStudio 12K (Applied Biosystems) and DNA fragment size identified with an Agilent 2100 Analyzer. As this identified likely primer dimer in some sample pools, these were size-selected using a BluePippin (Sage Science) and re-run on an Agilent 2100 Analyser. Final pools were quantified using both the KAPA Library Quantification Kit and a QUBIT 3.0 using the dsDNA HS assay.

Libraries were sequenced on an Illumina MiSeq platform at the Sheffield Children's Hospital Next Generation Sequencing Facility. The MiFish and 12S-V5 pool were sequenced on a 150 bp Paired-End sequencing run, and a 250 bp Paired-End run was used for the 16s_Teleo and Cichlid_CR pool. A 10% PhiX spike-in was included on both runs to increase the sequence complexity. In total, 75 samples were sequenced for each primer set, comprising 51 field samples, 5 aquarium samples, 11 filter negative controls, and 8 extraction negative controls.

2.7 | Bioinformatic analyses

Analyses were run on the High Performance Computing Cluster at the University of Sheffield, with full details on software and parameters used provided in the Appendix S2. Briefly, reads were quality checked with FastQC (Andrews, 2010) and trimmed based on read quality with the removal of Illumina sequencing adaptors using Trimmomatic v0.36 (Bolger, Lohse, & Usadel, 2014). Quality filtered reads were aligned with FLASH v1.2.11 (Magoč & Salzberg, 2011) and primers trimmed allowing for one mismatch with Mothur v1.37.1 (Schloss et al., 2009). Sequences were dereplicated with USEARCH v9.2.64 (Edgar, 2010) and clustered into high resolution MOTUs with Swarm v2 (d = 1) (Mahé, Rognes, Quince, De Vargas, & Dunthorn, 2015). MOTUs with a read count less than 3 were removed (as in Hänfling et al., 2016).

Blast+ searches were undertaken against the local reference database for each marker. Following this, taxonomies were assigned with MEGAN v6 based on primer-specific identity thresholds of 97/98%, using default parameters apart for a minimum score of 200, a minimum e-value of 10^{-10}, and a top percent of 2. Remaining unassigned MOTUs were removed.

To investigate the likely identity of unassigned MOTUs, a secondary blast search was undertaken against the NCBI nucleotide database with taxa assigned using the same methods as for the local reference database. Taxa assigned at this stage to previously unassigned reads were not included within the final dataset from which species richness estimates were derived.

2.8 | Error filtering and final eDNA matrix assignment

To reduce the impact of false positives, the maximum read count for each MOTU found in any of the filter or extraction negative controls

was subtracted from the read counts for the respective MOTUs within the LT and aquarium samples. MOTUs were then grouped together by their assigned taxonomy. Finally, taxonomic assignments with a sample read count below 0.15% were removed. This was based on read counts within the aquarium samples, removing any assignments to species not found within the aquarium, except potential misidentifications to close relatives by the MiFish marker. Taxonomic assignments from each primer set were aggregated together to form a final matrix. Those at a higher level than species or species complex were removed.

2.9 | Increased sequencing depth

Following the analysis of the initial results, a subset of samples (*N* = 24) collected in 2017 from eight sites (3, 4, 10, 11, 16, 17, 20, & 21) were re-sequenced on a MiSeq run with the Cichlid_CR marker to investigate the impact of sequencing depth on the species richness estimates obtained from the metabarcoding data. Six PCR replicates were undertaken per a sample, producing a total of 18 technical replicates per a site. Library preparation, sequencing, and bioinformatic analysis steps were consistent with those for the original sequencing runs detailed above (for further information, see Appendix S2).

2.10 | Statistical analysis

All statistical analysis was undertaken in RStudio v1.1.453 (RStudio Team, 2015). Total matrices were converted to presence–absence for comparisons between year, filter replicates, and survey methods. Visual survey species accumulation curves were produced using iNEXT v2.0.17 (Hsieh, Ma, & Chao, 2016). Accumulation curves for the eDNA samples and site species richness values were calculated using Vegan v2.5.4. eDNA accumulation curves were calculated for 2017 sites only as these contained three filter replicates per site. To investigate scales of detection, Sørenson dissimilarity values were calculated between the eDNA site species richness estimates and those derived from the visual survey data at five different scales; (a) site scale where there is a maximum distance of 60 m between surveys and eDNA samples; (b) the central three survey points (at both depths) with a maximum of 30 m between surveys and eDNA samples; (c) the central survey points where eDNA samples were collected; (d) surveys at 5 m; and (e) surveys at 10 m depth. Sørenson dissimilarity values were calculated with betapart v1.5.0 (Baselga & Orme, 2012; Oksanen et al., 2012).

3 | RESULTS

3.1 | Visual surveys

A total of 945 detections representing 63 species were made across the visual surveys, including 55 cichlid and 8 noncichlid species (Table S4). Interpolated and extrapolated sampling curves showed clear plateauing at the majority of sites, demonstrating sufficient sampling completeness in both field seasons (Figures S3 and S4).

3.2 | Reference database and in silico testing

A total of 431 fish species from 22 families were identified as occurring in LT and its broader catchment area (Tables S1 and S6). This includes 272 cichlid species of which 213 are described with the remainder currently either undescribed or putative (Ronco et al., 2019). Reference database sequences were obtained for 358 fish species (including eight taxa only identified to genus level) representing 83% of species and 254 cichlid species representing 93% from this family. In silico results demonstrate the MiFish, 12S-V5, and 16s_Teleo primer sets are highly conserved across the lake's fishes, except for a first base mismatch against the *Synodontis* catfishes (~11 spp.) for the 12S-V5 primer set (Figure S5). The Cichlid_CR primer set is highly conserved across the lake's cichlid fishes with no mismatches within the first seven bases of either primer. These results were supported by the consistent amplification of DNA across 17 fish species in the lake by the three universal fish primers and 10 cichlid fishes by Cichlid_CR (Figure S6 and Table S8). Further testing of the newly designed 16s_Teleo and Cichlid_CR primers against the MitoFish database and cichlid mitogenomes within NCBI, respectively, demonstrated these primers are largely conserved across fish species and cichlids globally (Table S7).

The percentage of species with unique barcodes in the reference database ranged between 36.1% for the 12S-V5 marker, 68.6% for MiFish, 80.3% for 16s_Teleo, and 96.8% for Cichlid_CR (Figure S7). Species with identical sequences were grouped into marker-specific species complexes (Table S2). The genetic distances of species to their closest neighbor had a mean of 2.0% for the MiFish marker, 1.0% for 12S-V5, 2.5% for 16s_Teleo, and 4.6% for Cichlid_CR (Figure S9). Closest neighbor genetic distances for the cichlid fishes were largely below 2% for most markers with 9.8% of MiFish, 2.0% of 12-V5, 18.0% of 16s_Teleo, and 72.5% of Cichlid_CR barcodes having divergence values greater than 2%. This increased for the noncichlid fishes across the three universal fish primers with 59.6%, 37.0%, and 57.0% of MiFish, 12S-V5, and 16s_Teleo barcodes, respectively, having closest neighbor divergence values higher than 2%.

Mean within-genus genetic distances also varied considerably between markers (Figure S8) ranging between 2.31% and 12.13% (Table S9). Within-genus K2P distances for the cichlid fishes were three to four times higher for Cichlid_CR compared to the other three markers. The distribution of genetic divergence is also greater for Cichlid_CR compared to the other three markers ranging between 0% and 28.5% (Figure S8). The increased interspecific variability within the Cichlid_CR barcode for the cichlid fishes suggests a likely improved taxonomic resolution compared to the other three markers, with the potential to identify many cichlids down to species level.

3.3 | Original sequence data

In total, 11.5 million and 8.6 million paired-end reads were obtained from the 2 × 150 bp and 2 × 250 bp MiSeq runs, respectively. Following the bioinformatic filtering steps (shown in Table S10), a

total of 9.1 million reads remained across the four primer sets. 5.1 million reads were assigned to the MiFish primer set with the other three primers ranging between 952 thousand and 1.7 million reads. Low read counts were identified for some samples with each primer set (MiFish = 6, 12S-V5 = 7, 16s_Teleo = 2 and Cichlid_CR = 2) that were ultimately removed. These were not consistent across the samples apart from replicates for sites 2 and 18 in 2017 that had very low extract DNA concentrations (<2 ng/μl). Following filtering (Table S10), mean sequencing depths per a site were 155,208 in 2016 (one filter replicate) and 531,267 in 2017 (with three filter replicates). Mean LT sample depths for each marker were 108,362 reads for MiFish, 36,880 for 12S-V5, 25,787 for 16s_Teleo, and 19,198 for Cichlid_CR. Species accumulation curves for the 2017 eDNA samples showed little plateauing demonstrating limited sampling completeness and suggesting sequencing depth may not be sufficient at some sites (Figure S10). Of the 19 filter and extraction controls sequenced, no contamination was identified within the negative controls for any of the species within the local reference database. The secondary NCBI blast search did identify human assigned 12S-V5 (*N* = 5) and 16s_Teleo (*N* = 3) reads within some of the field filter controls likely resulting from the collecting or filtering of eDNA samples in the field as these were not present in any of the DNA extraction negative controls.

Taxa from 12 fish families within the reference database were assigned to MOTUs (Figure 2). Cichlidae dominated both the MOTU and read counts of all four primer sets, largely reflecting their abundance within the littoral habitat. The MiFish primer set showed a high specificity to fishes with 96.9% of reads assigned to sequences within the reference database. For the 12S-V5, 16s_Teleo, and Cichlid_CR markers, 47.2%, 49.4%, and 32.5% of reads, respectively, remained unassigned. The secondary blast search against the NCBI nt database demonstrated the majority of these reads were assigned to Vertebrata, primarily Hominidae, for the 12S-V5 and 16s_Teleo primer sets (Figure 2). For Cichlid_CR, 86% of unassigned reads matched to LT cichlid sequences not included within the local reference database. This likely reflects intraspecific variability within the Cichlid_CR barcode not accounted for within the reference database for which there is currently one sequence per species. Species-level assignments to MOTUs made by the local reference and NCBI databases were found to differ in 34.5% of cases with species in different genera assigned to 3.5% of MOTUs. To ensure the accuracy of identifications, MOTUs with only NCBI assignments were therefore not included within the final dataset.

3.4 | Aquarium samples

All four markers consistently identified *N. longior*, which is by far the most abundant fish species within the ZSL tank (Table S11). *Julidochromis dickfeldi* species-level assignments were made by the MiFish and Cichlid_CR markers for the *Julidochromis* sp. within the tank. The remaining assignments, however, were limited to genus, tribe, or family level, with no read assignments corresponding to *A. calvus* or *H. microlepis* for the Cichlid_CR marker. The MiFish marker also made two erroneous assignments to close relatives of species found within the aquarium shown below the dashed line in Table S11. The aquarium samples highlight the challenges identifying taxa down to species level with varying taxonomic resolutions achieved across all four primer sets.

3.5 | Lake Tanganyika samples

The majority of MOTUs and reads were assigned to taxonomic levels above species and species complex for the three universal fish primers (Figures S11 and S12). 37.1%, 40.9%, and 34.2% of total reads were identified to species level by the MiFish, 12S-V5, and 16s_Teleo markers, respectively. 39.2% of 12S-V5 reads were assigned to family level, greater than all three other markers (MiFish = 27.6%, 16s_Teleo = 6.7%, and Cichlid_CR = 0.01%). A much higher proportion of MOTUs and reads were assigned to species level by Cichlid_CR with 81.2% of MOTUs and 76.6% of reads assigned to species level, respectively.

Within the final eDNA dataset, 645 detections of 109 species or species complexes (*N* = 8) were made across the eDNA samples in both years, including 84 cichlids and 25 noncichlids (Table S3). No species-level assignments were made for *Clarias* catfishes, so only the genus assignment was included. Of the 84 cichlid species identified, 19 were unlikely to be found along the surveyed range, based on previous coastline survey data and Konings (2015) (Table S3). These species represented 7.6% of site occurrences within the eDNA dataset with 16 species occurring three or less times. Only three of the mis-assigned species had nearest neighbor K2P distances greater than 2% with a mean of 1.37% within their identifying markers. This demonstrates they all had close genetic relatives within the reference database. Similar to the aquarium samples, these were considered to be erroneous assignments likely representing false positives. As a result, they were removed from further analysis for comparisons with the visual survey data.

Of the 65 remaining cichlid identifications, 31 (48%) were made by one marker, 19 (29%) by two, 9 (14%) by three, and 6 (9%) by four markers (Figure 3a). A total of 44 cichlid species were detected by Cichlid_CR, 33 by 16s_Teleo, 29 by MiFish, and 14 by 12S-V5. The Cichlid_CR primer set also detected the most unique cichlid species with 17 independent identifications. Of the 25 noncichlid identifications, eight were made by all three markers, with 16s_Teleo (*N* = 19) and MiFish (*N* = 18) detecting more species than 12S-V5 (*N* = 11) (Figure 3b). Only two of the 19 erroneous species identifications were made by more than one marker (Figure 3c). Of the remaining identifications, three were made by MiFish, two by 12S-V5, seven by 16s_Teleo, and five by Cichlid_CR.

3.6 | eDNA sample comparison

No relationship was identified between species richness estimate and the standardized read counts of samples (Spearman rho = −.14,

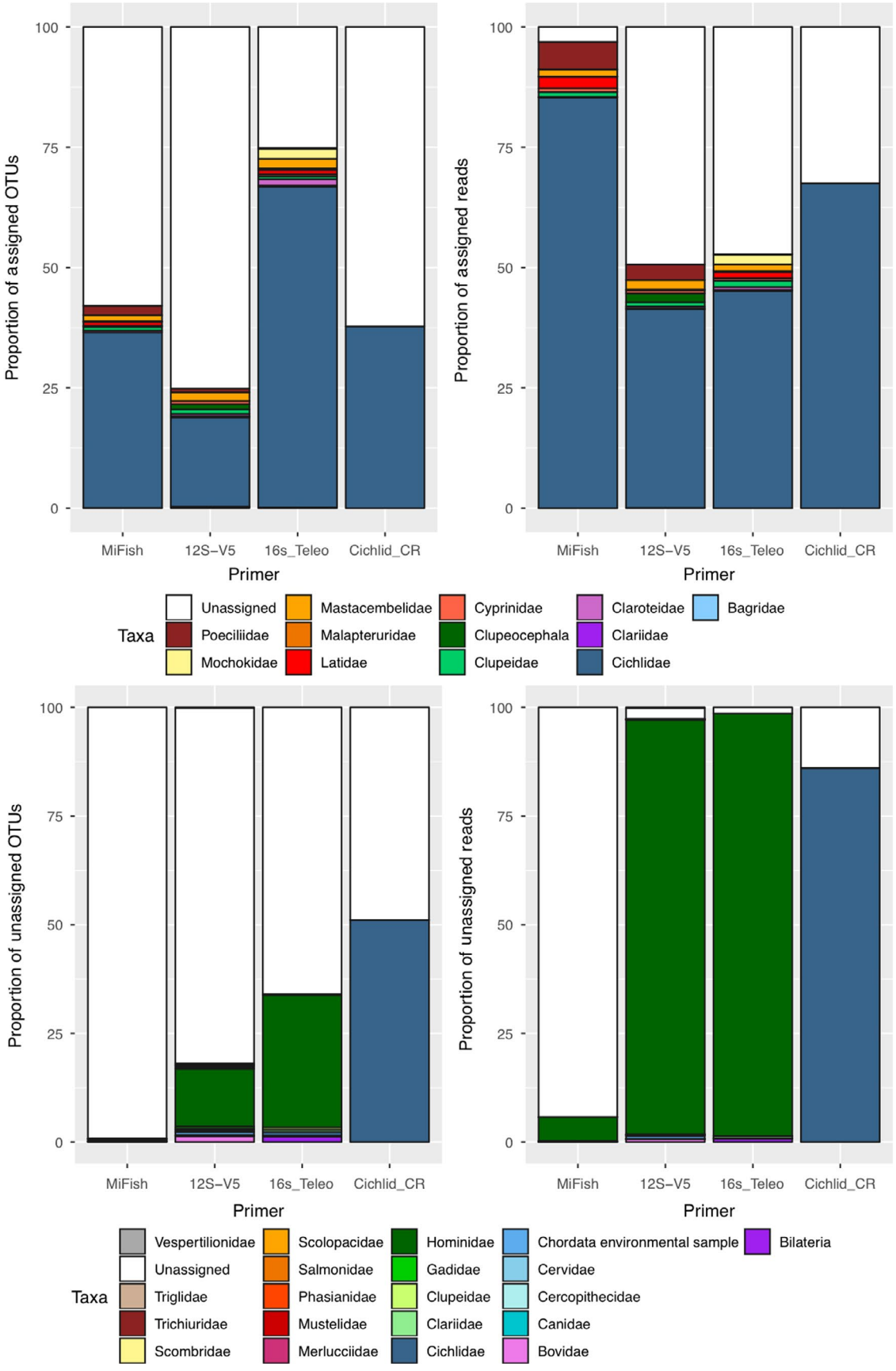

FIGURE 2 Families to which MOTUs and sequence reads were assigned (top) and families to which unassigned MOTUs and sequences were assigned with a secondary blast search against the NCBI nt database (bottom)

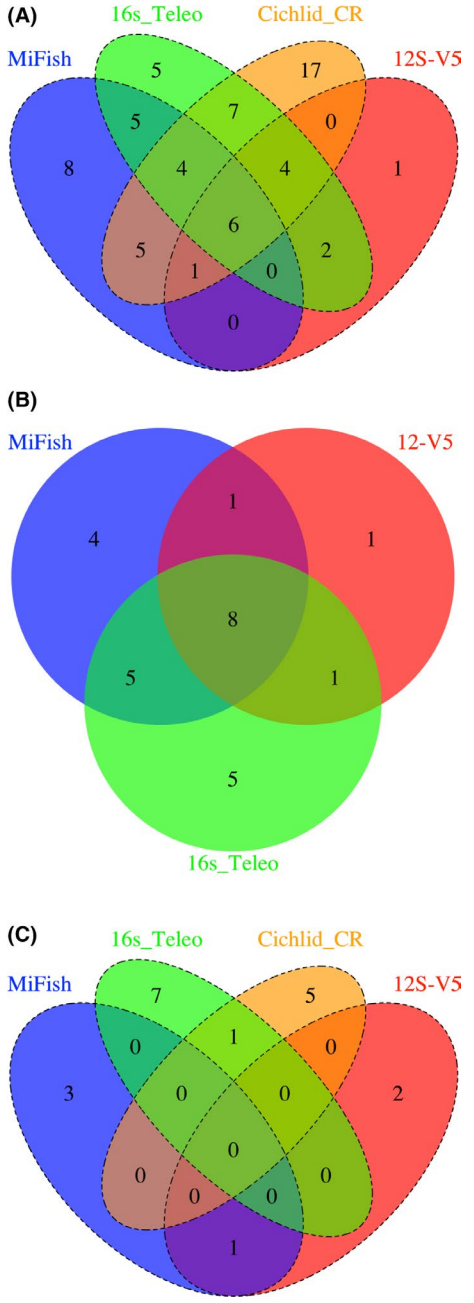

FIGURE 3 Primer detections of fish species within the eDNA dataset. Identifications are split into cichlid fishes only (a), noncichlid fishes (b) and likely false positives (c)

p = .058, N = 180) or sites (Spearman rho = .19, p = .35, N = 27). Filter replicates collected at sites in 2017 showed limited similarity in species detected (Figure 4). In total, 53% of species detections were made by one filter, 23% by two, and 24% by three. A significant negative correlation was identified between total site species richness and the percent of species identified in only one biological replicate (Spearman rho = −.70, p = .015, N = 12), with a significant positive correlation also identified between site species richness and the percent of species identified in three biological replicates (Spearman rho = .80, p = .002, N = 12).

3.7 | eDNA field season comparison

The use of multiple filter replicates in the 2017 field season resulted in increased eDNA species richness estimates for each site compared to 2016. Mean eDNA species richness estimates across all sites were 17.1 in 2016 and 28.3 in 2017, while the six sites surveyed in both years had a mean species richness estimate of 13.0 in 2016 and 27.7 in 2017. Sørenson dissimilarity comparisons of the 2016 and 2017 species richness estimates derived from each site therefore showed limited similarity ranging between 0.37 and 0.85 (Table S12). The nestedness component dominated at four sites compared with species turnover as a result of the lower species richness estimates derived from the 2016 samples compared to 2017.

3.8 | Comparison of eDNA and visual survey site diversity estimates

Based on the initial sequencing run across all surveys, a total of 103 species were detected, with 50 species (43 cichlids and 7 noncichlids) identified by both methods; 40 (22 cichlids and 18 noncichlids) by the eDNA method only; and 13 (12 cichlids and 1 noncichlid) by the visual surveys only. Visual survey species richness estimates were consistently higher compared with those derived from the eDNA samples in both years (Figure 5). This was largely due to a reduced detection of cichlids at each site, with a number of commonly observed species missing from or underrepresented within the eDNA dataset. For example, *Lamprologus callipterus*, *Lepidiolamprologus attenuates*, and *Perrisodus microlepis* that had 26, 26, and 27 site occurrences within the visual survey data, respectively, were not present within the eDNA dataset. A reduced number of detections per species were also consistently observed within the eDNA data compared to the visual survey dataset (Figure S14).

The eDNA dataset consistently detected a greater number of noncichlid species at each site compared with the visual survey data particularly in 2017 (Figure 5). Across the 2016 and 2017 field seasons, a mean of 3.9 and 8.3 noncichlid species, respectively, was detected per site within the eDNA dataset compared to 2.0 and 3.3, respectively, from the visual surveys. In total, a greater number of noncichlid species detected within the eDNA samples (N = 25) compared to the visual surveys (N = 8). This includes an increased number of detected species within the *Mastacembelus* spiny eel, *Synodontis*, and claroteid catfish radiations, as well as of other catfishes (e.g., *Malapterurus tanganyikaensis*, *Tanganikallabes mortiauxi*), *Lates* species, the lake's two freshwater herring species (*Limnothrissa miodon* and *Stolothrissa tanganicae*) and *Acapoeta tanganicae*. Observed *Barbus* sp. (possibly a misidentification of *A. tanganicae*) assigned to genus level at sites in 2017 was the only noncichlid detected by the visual surveys not included in the eDNA dataset.

Species detections for the eDNA and visual survey methods showed limited similarity at each site (Table 2). In 2016, an average of 22% of species detected at each site were found within the eDNA and visual survey datasets, 21% were in the eDNA data only, and 57%

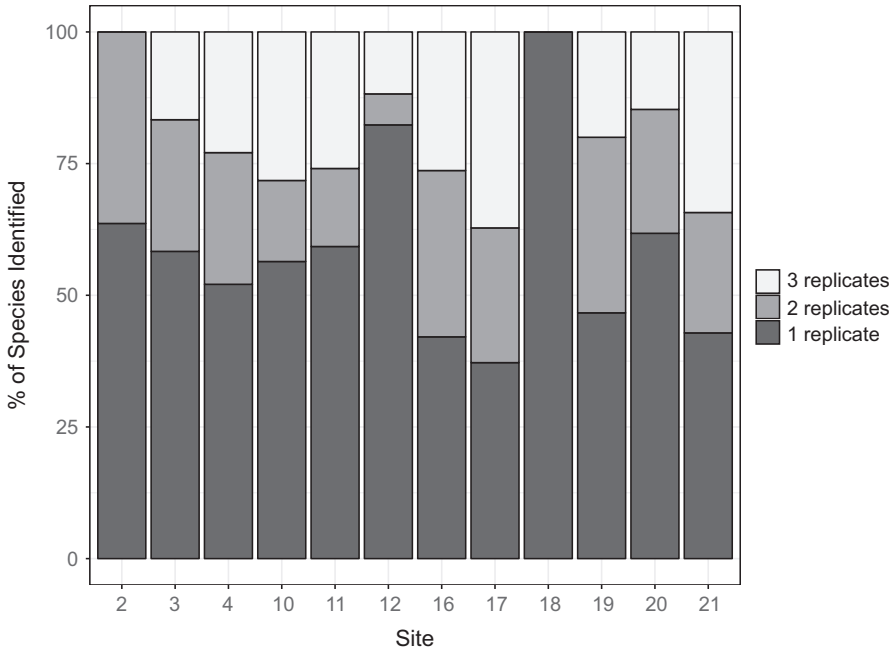

FIGURE 4 Percentage of species identified in each of the three filter replicates collected at sites in 2017

in the visual survey data only. In 2017, the number of species detected by both survey methods at each site increased to 27% with 26% detected by the eDNA data only and 46% by the visual survey only. As a result, mean Sørenson dissimilarity values of 0.66 and 0.59 were detected across 2016 and 2017, respectively, demonstrating this difference in composition. At most sites, dissimilarity was largely driven by species turnover (variation caused by the replacement of one species by a different species) due to the high proportion of species identified by only one survey method. For sites where the eDNA species richness estimate was much lower than the visual survey data, a greater proportion of variance resulted from species nestedness as the eDNA species list represented a subset of the more diverse visual survey. Subsets of the visual survey site data were also analyzed against the eDNA data to investigate whether these better reflected the eDNA species richness estimates. Little difference in Sørenson dissimilarity values was observed across spatial scales, with the total site species richness estimate on average showing the greatest similarity to eDNA samples (Figure S13). As a result, the eDNA best reflects the visual survey data at the site scale (within 60 m) supporting comparisons being made between these methods at this scale.

3.9 | Impact of increased sequencing depth on cichlid species richness estimates

The additional sequencing run for the Cichlid_CR marker detected a total of 61 cichlid species across the eight re-sequenced sites (Table S13). A mean cichlid species richness estimate of 35.0 was calculated across all sites, making it comparable to the mean estimate of 35.1 from the visual survey data. In comparison, estimates from the original run were much lower with a mean of 25.4 with all four primers, and 13.1 with the Cichlid_CR marker only (Figure 6). As a result, site community compositions differed between sequencing runs with mean site Sørenson pairwise dissimilarity values between the additional run and the original run of 0.40 (turnover = 0.28, nestedness = 0.12) for all primers and 0.50 (turnover = 0.06, nestedness = 0.44) for the Cichlid_CR marker only. Overall, there was an average sequencing depth of 167,687 reads per sample and 503,061 reads per site for the additional run postfiltering. Species accumulation plots for each site demonstrate substantial plateauing at most sites, highlighting improved sampling completeness within the additional sequence data with site estimates closer to species saturation (Figure S15).

4 | DISCUSSION

This study has demonstrated the potential of eDNA metabarcoding for surveying diverse and complex fish communities as well as detecting closely related species within evolutionary radiations. It also highlights the importance of sequencing depth and reference database completeness when designing eDNA metabarcoding studies for surveying diverse fish communities with recommendations for further improvements. Finally, through establishing a novel reference database for LT's fish communities, information on the interspecific genetic divergence across multiple markers is provided as a future resource that can be built upon for future metabarcoding work within this system.

4.1 | Genetic divergence and resolution of assignments

Analysis of the reference database showed limited interspecific genetic divergence between species for the 12s and 16s primer sets due to the large number of closely related species within the lake. Similar reduced interspecific genetic distances have been reported within the COI region for diverse neotropical fish communities that also include genera containing multiple species (Pereira et al., 2013). While

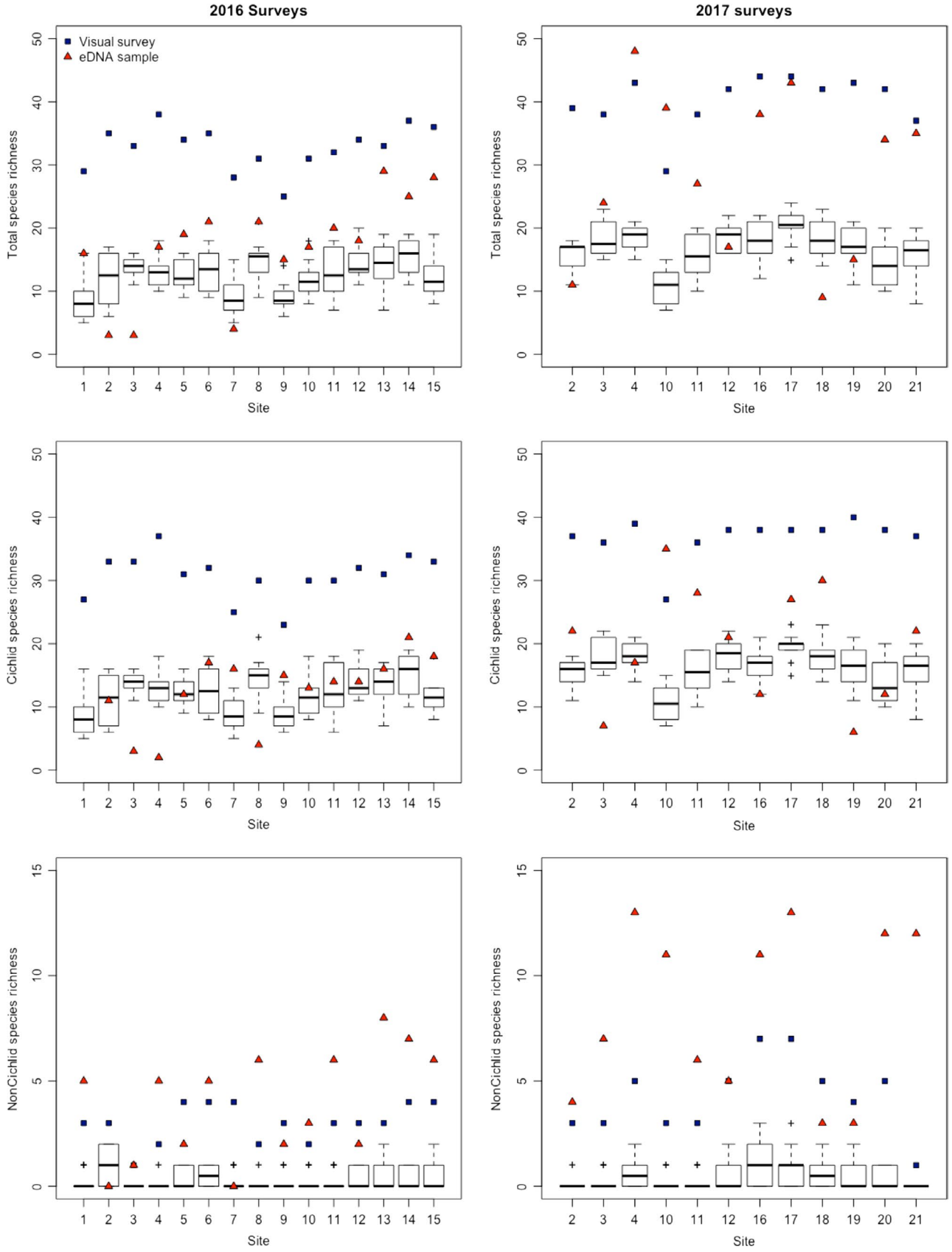

FIGURE 5 Species richness estimates from the individual surveys (boxplots), total site estimates from the ten visual surveys (blue squares) and from the eDNA samples (red triangles). The top, middle, and bottom rows show total species richness for all fish, cichlid species and noncichlid species, respectively

TABLE 2 Comparisons of site species richness estimates from the eDNA and visual surveys

Site	Total SR	Shared species	eDNA unique	Visual unique	Beta Sor.	Beta Sim.	Beta Sne.
S1_16	37	7	9	21	0.68	0.56	0.12
S2_16	35	2	1	32	0.89	0.33	0.56
S3_16	34	1	2	31	0.94	0.67	0.28
S4_16	48	6	11	31	0.78	0.65	0.13
S5_16	43	9	10	24	0.65	0.53	0.13
S6_16	44	11	10	23	0.6	0.48	0.12
S7_16	30	1	3	26	0.94	0.75	0.19
S8_16	40	11	10	19	0.57	0.48	0.09
S9_16	32	7	8	17	0.64	0.53	0.11
S10_16	40	7	10	23	0.7	0.59	0.11
S11_16	40	11	9	20	0.57	0.45	0.12
S12_16	37	14	4	19	0.45	0.22	0.23
S13_16	47	14	15	18	0.54	0.52	0.02
S14_16	46	15	10	21	0.51	0.4	0.11
S15_16	47	16	12	19	0.49	0.43	0.06
S2_17	43	6	5	32	0.76	0.45	0.3
S3_17	49	12	12	25	0.61	0.5	0.11
S4_17	64	26	22	16	0.42	0.38	0.04
S10_17	52	14	24	14	0.58	0.5	0.08
S11_17	51	13	14	24	0.59	0.52	0.08
S12_17	46	12	5	29	0.59	0.29	0.29
S16_17	62	19	19	24	0.53	0.5	0.03
S17_17	62	24	19	19	0.44	0.44	0
S18_17	44	6	3	35	0.76	0.33	0.43
S19_17	46	11	4	31	0.61	0.27	0.35
S20_17	57	18	16	23	0.52	0.47	0.05
S21_17	58	13	22	23	0.63	0.63	0.01

Note: Beta.Sor is the Sørenson dissimilarity between estimates, Beta Sim. is the Simpson pairwise dissimilarity measuring species turnover, and beta Sne. is the dissimilarity accounting for species nestedness. Total SR is the combined species richness estimate from both survey methods. Site descriptions state the site number followed by the survey year (16 = 2016; 17 = 2017).

interspecific divergence values for these barcodes often fell below the traditional 2% cutoff for species delimitation, the ability to correctly distinguish between species with variation below this threshold has been demonstrated (Breman et al., 2016; Pereira et al., 2013).

A number of species were found to have identical barcodes across some of the markers used, and these were ultimately grouped into species complexes. The taxa included in species complexes were not consistent across primers, resulting in all species containing a unique sequence within at least one marker apart from *Benthochromis tricoti/Benthochromis* sp. "horii mahale" and *Cyprichromis coloratus/Cyprichromis* sp. "jumbo" complexes that both likely represent geographical variants, rather than distinct species (Ronco et al., 2019). This is largely due to the increased interspecific variation observed within the Cichlid_CR barcode where genetic distances were three to four times higher for the cichlid fishes compared to the other three markers. The increased substitution rate within this region can improve the taxonomic resolution of barcodes, with the control region previously shown to be a more robust marker for species-level identifications across a number of fish genera compared to COI (Cawthorn, Duncan, Kastern, Francis, & Hoffman, 2015; Pedrosa-Gerasmio, Babaran, & Santos, 2012; Shum et al., 2017). Substantial intraspecific variability is also likely to exist within this highly variable region, with population structuring identified for a number of cichlid species (Sefc, Baric, Salzburger, & Sturmbauer, 2007; Wagner & McCune, 2009), and recently a catfish species (Peart, Dasmahapatra, & Day, 2018) inhabiting the lake's rocky littoral habitat. Although the strict conditions for the inclusion of specimens within our reference database limits error due to misidentifications, it has restricted the number of sequences per species to one for this study. Including multiple sequences per species will enable the assessment of intraspecific variability within this barcode and the quantification of any barcoding gap.

The improved interspecific variability within the Cichlid_CR marker led to an increased proportion of LT eDNA MOTUs and

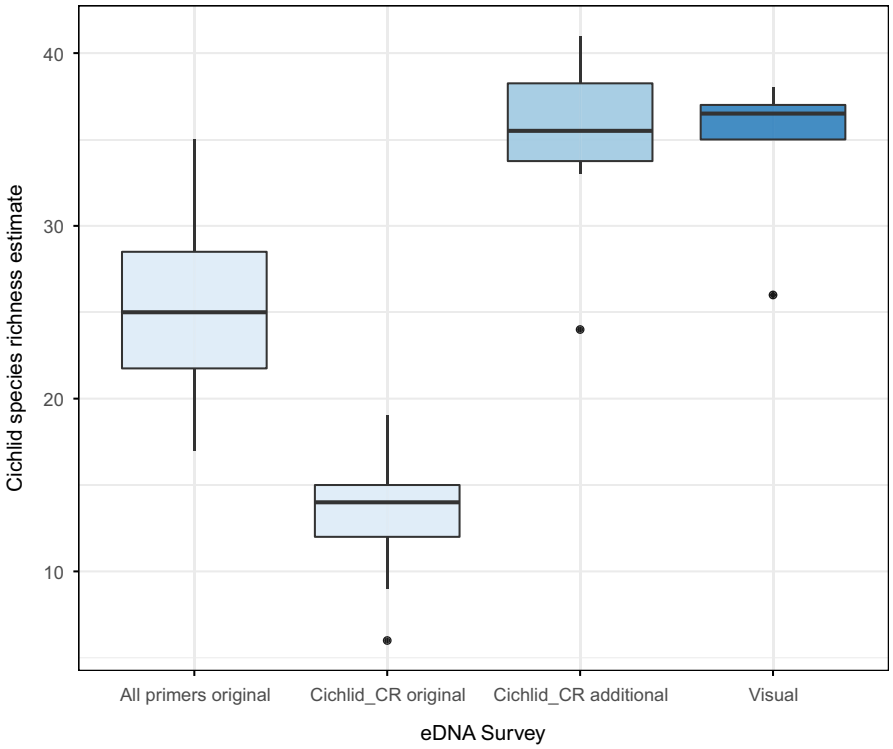

FIGURE 6 Cichlid site species richness estimates obtained from the original sequencing run, additional sequencing run, and visual survey data. Only species richness estimates for the eight re-sequenced sites are included. "All Primers Original" represents species richness estimates derived from the four primers sequenced, while "Cichlid_CR Original" shows estimates obtained from the Cichlid_CR marker only with the original sequence data

reads identified to species level compared to the other three markers that showed a limited taxonomic resolution with over 50% of reads assigned to genus level or above. Due to the large number of genera containing multiple species within LT, species-level identifications are largely required to be ecologically informative. As a result, reads assigned to higher taxonomic levels represent lost information, limiting the number of detections from these primer sets.

4.2 | Species-level identifications

Nevertheless, the eDNA samples resulted in a greater number of species identifications compared to the visual surveys, with 90 species identified in total (excluding false positives). There is a strong depth gradient in the community structure of LT's littoral fish communities and as a result significant changes in species composition can occur over small spatial scales (Takeuchi et al., 2010). A number of the cichlid species identified only in the eDNA dataset are more commonly found in either the wave-washed habitat at shallower depths than those surveyed, such as *Pseudosimochromis curvifrons*, *Tanganicodus irsacae*, and *Spathodus erythrodon*, as well as from deepwater habitats at depths below the visual surveys, including *Benthochromis horii*, *Xenotilapia caudafasciata*, and two *Trematocara* species (Konings, 2015). Furthermore, longitudinal variation in cichlid community composition is heavily influenced by substrate type (e.g., rocky, sandy, muddy) (Widmer et al., 2019). While all eDNA samples were collected in rocky habitats, some additional species more commonly found within sandy habitats including *Cardiopharynx schoutedeni*, *Lestradea stappersii*, and a number of *Xenotilapia* species were detected. The lake's two freshwater herring species, *L. miodon* and *Stolothrissa tanganyicae*, which are pelagic were also identified within the samples.

These examples suggest the spatial scales of detection within the eDNA samples may extend beyond the local littoral habitat surveyed. eDNA studies in similar coastal marine and lentic freshwater systems have shown fine scaled detection for fish communities, with longer barcodes such as the Cichlid_CR primer potentially reducing the spatial scales of detection further (Andruszkiewicz, Starks, et al., 2017; Hänfling et al., 2016; Port et al., 2016). This is likely to vary across ecosystems, however, with processes including lake mixing and stratification theoretically influencing the scales of detection (Deiner et al. 2017). For example, seasonal upwelling common in LT could result in the transportation of deepwater eDNA into the littoral habitat. Annual surface water temperatures between 25.9°C and 27.7°C have been recorded within the surveyed region (Kimirei & Mgaya, 2007). Warm temperatures such as this reduce eDNA persistence within the water column (Andruszkiewicz, Sassoubre, et al., 2017), with Eichmiller et al. (2016) detecting exponential DNA degradation rates in lake water at 25°C with a half-life of only 6.9 hr (Collins et al., 2018). While high degradation rates would suggest finer spatio-temporal scales of detection, the unique nature of LT's ecosystem and its fish communities means both degradation rates and scales of detection need to be investigated in future studies. The latter could be achieved through sampling across depths and habitat boundaries with marked shifts in fish community composition.

Of the 19 detected species considered to be likely false positives, 17 were identified by individual primers, of which 10 occurred at only one site and six at two or three sites. *Neolamprologus caudopunctatus* and *Xenotilapia* sp. "papilio sunflower" assignments were made

by two markers with occurrences at 11 and 8 sites, respectively, therefore not adhering to the lower confidence levels expected from false-positive assignments. Due to the large number of closely related species with low interspecific genetic distances, particularly within the 12s and 16s markers, the potential for mis-assignments is increased. This is highlighted by the low nearest neighbor genetic divergences for each mis-assigned species within their identifying markers. For example, *Cyphotilapia gibberosa* likely represents a mis-assignment by MiFish to *Cyphotilapia frontosa*, a species common along the surveyed coastline. Similarly, *Xenochromis hecqui* identified by 16s_Teleo has only a 1.4% genetic divergence within this marker from *H. microlepis* as well as being closely related to *Perissodus microlepis*, both of which were frequently observed within the visual surveys. In these cases, it is possible the limited interspecific genetic distances for some markers increase the likelihood of erroneous assignments resulting in false-positive identifications. Erroneous assignments to close relatives were also identified by Cilleros et al. (2019) within tropical South American streams. These findings demonstrate the potential for assignment errors within complex tropical communities containing closely related taxa often with limited sequence availability in reference databases.

The eDNA samples showed higher detection of noncichlid species at sites compared to the visual surveys, with the use of filter replicates in 2017 resulting in an increase in the number of species detected. The improved detections of species within the catfish and mastacembelid spiny eel radiations demonstrated the effectiveness of eDNA metabarcoding for distinguishing between closely related species within these groups. While an increased number of detections were made by all three primer sets, a number of species were still identified by one or two markers with *Synodontis* species only assigned to MOTUs by 16s_Teleo. Building on the findings of earlier eDNA metabarcoding studies surveying fish communities (Evans et al., 2017; Stat et al., 2017), the use of multiple primer sets enables the improved detection of both cichlid and noncichlid species.

Many of the noncichlid species only detected in the eDNA samples are likely to be underrepresented within the visual surveys due to their behavioral habits. For example, the claroteine catfishes (e.g., *Chrysichthys sianenna*, *Lophiobragrus cyclurus*, and *Bathybagrus tetranema*) are largely nocturnal (Peart et al., 2014) and many *Mastacembelus* species live in the substrate or within the complex rocky environment (Brown et al., 2011). In comparison, the territorial nature of many cichlid fishes means they are less shy toward SCUBA divers, a behavior that has been shown to positively bias the detection of fish species within visual survey data (Bozec et al., 2011). These behaviors are therefore likely to result in the positive bias of cichlids and underrepresentation of many noncichlids within visual surveys. Many of the species present only in the eDNA dataset are also wider ranging with lower local abundances than the majority of cichlid species in the littoral habitat and are therefore much less likely to be consistently recorded by the stationary visual surveys. Similarly species that commonly exist in schools, such as *Lamprichthys tanganicanus*, are likely to have biased detection rates from visual surveys (Pais & Cabral, 2018), explaining why this species was more commonly detected within the eDNA data. As the eDNA survey method is less influenced by species behavior, its combined use alongside visual surveys holds the potential to help overcome some of these survey biases particularly for the often more genetically distinct noncichlid teleost species. This could therefore represent an immediate benefit of incorporating an eDNA metabarcoding approach within survey methodologies of LT's fish communities.

4.3 | Filter replicate similarity

Limited similarity between filter replicates at each site in 2017 was observed, as has been previously reported (Andruszkiewicz, Starks, et al., 2017). Sites with a more diverse species richness estimate had an increased percentage of species identified in three filter replicates and a lower percentage of species detected in one filter replicate. Similarity between replicates is therefore greater at sites with a higher species richness. The limited similarity at low diversity sites likely results from inconsistencies surrounding the preservation, amplification, or sequencing of one or more filter replicates at these sites, leading to variable species detections. The optimization of methods would likely lead to an improved similarity between replicates as observed at sites that detected a greater number of species (e.g., sites 17 and 21). For example, while the storage of cellulose nitrate filter papers in ethanol at −20°C has been shown to be effective for eDNA preservation (Hinlo, Gleeson, Lintermans, & Furlan, 2017), recent research has demonstrated the use of lysis buffer or drying in silica gel can give more consistent community composition estimates from lake eDNA samples (Majaneva et al., 2018) and therefore likely more consistent filter replicates.

4.4 | Species richness estimates

While detecting more species overall, the eDNA samples consistently produced lower site species richness estimates compared to the visual surveys. Similarly, the community composition of species richness estimates from the eDNA and visual survey methods was also found to largely differ. Dissimilar fish assemblage patterns were also detected between eDNA metabarcoding and traditional survey approaches across diverse tropical streams in French Guiana (Cilleros et al., 2019). Much higher similarities between eDNA and traditional methods at sites have been reported within temperate ecosystems (Pont et al., 2018), largely due to the high detection sensitivity of eDNA metabarcoding methods within these ecosystems (Evans et al., 2017; Hänfling et al., 2016; Valentini et al., 2016). In LT, this difference results from consistently observed cichlid species either not being detected or being underrepresented within the eDNA dataset, with a lower number of detections per species overall compared to the visual surveys.

While limitations in eDNA detections could be caused by PCR bias or the taxonomic resolution of individual markers, in this instance it is likely derived from under sampling due to the sequencing depth used as well as limitations in the reference database, particularly for the Cichlid_CR marker. Despite the sample sequencing depth

being comparable to that used in other eDNA metabarcoding studies (Li et al., 2019; Yamamoto et al., 2017), the exceptional diversity of LT's littoral habitat compared to these systems means a larger sequencing depth is likely required to get closer to saturation of species detection (Figure S10). This is highlighted by the improved species richness estimates obtained for sites from the additional sequencing run that greatly exceed estimates from the original sequence data (Figure S15). This higher sequencing depth combined with the greater resolution of detections with the Cichid_CR marker compared to the other three primers, resulted in improved cichlid species detections comparable to those obtained from the visual survey data (Figure 6). Two eDNA metabarcoding studies recently published focusing on Guianese tropical streams had sequencing depths of over 400,000 reads per sample prefiltering (Cantera et al., 2019; Cilleros et al., 2019). As the diversity of LT is twice that identified in Guianese streams, high sequencing depths such as that used in the additional run and these two studies are required to obtain sufficient sampling completeness and accurate species richness estimates.

Species richness estimates derived from the Cichlid_CR marker in the original run represent subsets of those from the additional run with Sørenson dissimilarity values dominated by species nestedness. Turnover represented a larger proportion of the dissimilarity when site species richness estimates from the additional run were compared with those from the original sequence data with all four primer sets. This is because species detected by some of the other three markers still remained undetected by the Cichlid_CR marker despite the improved sequencing depth used due to current reference database limitations. For example, common *Altolamprologus compressiceps*, *Neolamprologus brichardi*, and *Neolamprologus mondabu* species remained undetected. Further expansions of the control region reference database could help overcome this, improving species richness estimates derived from this data further through improving the detections of species currently missing or underrepresented within the eDNA data.

While publicly available control region sequences could potentially enable this, the observed discrepancies in taxonomic assignments between the NCBI and local reference databases highlight the challenges of using public databases. These contain verified and unverified sequences that can lead to the presence of ambiguous assignments (Shum et al., 2017). Similarly, due to a number of recent taxonomic changes to LT's fishes within recent years, the NCBI taxonomy is not up to date for many sequences (Ronco et al., 2019). These issues can be overcome by providing a comprehensive dataset of verified NCBI sequences with reliable references as demonstrated by Shum et al. (2015). The inclusion of verified NCBI sequences and obtaining further sequences from sample collections will help to reduce the number of unassigned MOTUs, likely improving the resolution and number of detections from the eDNA samples. There is also the potential to investigate the biodiversity of sites using a taxonomy-free approach more focused on the diversity of MOTUs. This could be challenging, however, due to the likely intraspecific variability within the markers for cichlid species (particularly Cichlid_CR). There would be the risk of over splitting species with substantial population structuring if it was assumed MOTUs reflected true species, or indeed under splitting for some markers with insufficient taxonomic resolution.

5 | CONCLUSION

To better understand the potential for eDNA metabarcoding approaches to survey freshwater fish communities globally, there is a need to apply these methods across a broad range of ecosystems and communities. This study provides a first application of these methods within one of the world's most diverse freshwater ecosystems. Our findings demonstrate the potential and limitations of eDNA metabarcoding for identifying taxa to species level, and thereby contributing to diversity estimates for fish communities. Wide variation in the resolution of markers highlights the importance of primer selection, with the use of a family-specific cichlid control region marker improving the taxonomic resolution of identifications within this species-rich group. Using multiple markers also improved species detections across the cichlid and noncichlid fishes. A number of false positives were identified in this study, likely reflecting current limitations in the resolution of the 12s and 16s markers as well as the reference database, particularly for the control region.

Increasing the sequencing depth substantially improved site species richness estimates from the eDNA samples, resulting in estimates much more comparable to that obtained from the visual survey data. While inconsistencies in the detections of some cichlid species remain, further reference database expansions, particularly for the Cichlid_CR marker, would likely further improve species richness estimates from the eDNA samples. These advancements combined with the improved detection of noncichlid species highlight the benefits of including eDNA metabarcoding methods within survey designs for LT's fishes alongside traditional methods. This study has highlighted the potential for eDNA metabarcoding to survey even highly diverse tropical communities and closely related taxa within evolutionary radiations, demonstrating the contributions this method could make toward surveying freshwater fish communities within tropical systems in the future.

ACKNOWLEDGMENTS

This project was conducted with the permission of the Tanzania Commission for Science and Technology (COSTECH) under permit 2016-322-NA-2014-267. This work was supported by the Natural Environment Research Council (award NE/L002485/1); Genetics Society Heredity Fieldwork Grant (CJD); Systematics Research Fund (CJD); Fisheries Society of the British Isle Small Research Grant (CJD); Percy Sladen Memorial Trust Fund (JJD); the NERC Bimolecular Analysis Facility Sheffield (CJD and JJD); European Research Council (ERC; CoG CICHLID ~ X) (WS); and the Swiss National Science Foundation (WS). We thank Melanie Stiassny (AMNH) and Roger Bills (SAIAB) for providing additional fish samples for the establishment of the reference database; George Kazumbe for field assistance, and Deborah Dawson for her substantial input with the design and

management of laboratory work at the NERC Biomolecular Analysis Facility Sheffield. We thank Adrian Indermaur for the taxonomic identification of reference database cichlid species as well as Fabrizia Ronco and Michael Matschiner for the preparation of the cichlid reference database sequences (all University of Basel). We are also grateful to the ZSL Aquarium, particularly Samantha Guillaume, for facilitating collection of eDNA samples from their LT tank. Finally, we thank the reviewers and associate editor Dr Kristy Deiner for their valuable comments that greatly improved the manuscript.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

CJD, JJD, and DJM designed the study. CJD, JJD, and CM planned and organized the fieldwork, CJD collected the data, CJD, HH, and GH designed the laboratory work, CJD and GH undertook the laboratory work, CJD undertook the bioinformatic analysis based on pipelines developed by HH., CJD, JJD, and WS collected and developed the reference database sequences. All authors contributed to the analysis and writing of the manuscript.

ORCID

Christopher J. Doble https://orcid.org/0000-0002-1358-9577
Walter Salzburger https://orcid.org/0000-0002-9988-1674
David J. Murrell https://orcid.org/0000-0002-4830-8966
Julia J. Day https://orcid.org/0000-0002-6765-7782

DATA AVAILABILITY STATEMENT

The visual survey community matrices (https://doi.org/10.5522/04/9879614), reference database sequence files (https://doi.org/10.5522/04/9861461) and Illumina raw sequence data from the three Illumina MiSeq runs (https://doi.org/10.5522/04/9874451) are publicly available via the UCL research data depository. Noncichlid reference database 12s and 16s sequences produced for this study have been uploaded to NCBI (accession numbers: MN255556–MN255727). Cichlid genomes will be uploaded to NCBI by WS.

REFERENCES

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Andruszkiewicz, E. A., Sassoubre, L. M., & Boehm, A. B. (2017). Persistence of marine fish environmental DNA and the influence of sunlight. *PLoS ONE*, *12*(9), 1819–1827. https://doi.org/10.1371/journal.pone.0185043

Andruszkiewicz, E. A., Starks, H. A., Chavez, F. P., Sassoubre, L. M., Block, B. A., & Boehm, A. B. (2017). Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *Plos ONE*, *12*(4), e0176343. https://doi.org/10.1371/journal.pone.0176343

Baselga, A., & Orme, C. D. L. (2012). Betapart: An R package for the study of beta diversity. *Methods in Ecology and Evolution*, *3*(5), 808–812. https://doi.org/10.1111/j.2041-210X.2012.00224.x

Bohnsack, J. A., & Bannerot, S. P. (1986). A stationary visual census technique for quantitatively assessing community structure of coral reef fishes. In *NOAA technical report NMFS* (Vol. *41*).

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bozec, Y.-M., Kulbicki, M., Lahoe, F., Mou-Tham, G., & Gascuel, D. (2011). Factors affecting the detection distances of reef fish: Implications for visual counts. *Marine Biology*, *158*, 969–981. https://doi.org/10.1007/s00227-011-1623-9

Breman, F. C., Loix, S., Jordaens, K., Snoeks, J., & Van Steenberge, M. (2016). Testing the potential of DNA barcoding in vertebrate radiations: The case of the littoral cichlids (Pisces, Perciformes, Cichlidae) from Lake Tanganyika. *Molecular Ecology Resources*, *16*(6), 1455–1464. https://doi.org/10.1111/1755-0998.12523

Brichard, P. (1989). *Cichlids and all the other fishes of Lake Tanganyika*. Neptune City, NJ: T.F.H Publications.

Britton, A. W., Day, J. J., Doble, C. J., Ngatunga, B. P., Kemp, K. M., Carbone, C., & Murrell, D. J. (2017). Terrestrial-focused protected areas are effective for conservation of freshwater fish diversity in Lake Tanganyika. *Biological Conservation*, *212*, 120–129. https://doi.org/10.1016/j.biocon.2017.06.001

Brown, K. J., Britz, R., Bills, R., Rüber, L., & Day, J. J. (2011). Pectoral fin loss in the Mastacembelidae: A new species from Lake Tanganyika. *Journal of Zoology*, *284*(4), 286–293. https://doi.org/10.1111/j.1469-7998.2011.00804.x

Brown, K. J., Rüber, L., Bills, R., & Day, J. J. (2010). Mastacembelid eels support Lake Tanganyika as an evolutionary hotspot of diversification. *BMC Evolutionary Biology*, *10*, 188. https://doi.org/10.1186/1471-2148-10-188

Buckland, S. T., Studeny, A. C., Magurran, A. E., & Newson, S. E. (2011). Biodiversity monitoring: The relevance of detectability. In A. E. Magurran, & B. J. McGill (Eds.), *Biological Diversity: Frontiers in measurement and assessment* (pp. 25–38). Oxford, UK: Oxford University Press.

Burkhead, N. M. (2012). Extinction rates in North American freshwater fishes, 1900–2010. *BioScience*, *62*(9), 798–808. https://doi.org/10.1525/bio.2012.62.9.5

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*(17), 8697–8712. https://doi.org/10.1002/ece3.4387

Cantera, I., Cilleros, K., Valentini, A., Cerdan, A., Dejean, T., Iribar, A., … Brosse, S. (2019). Optimizing environmental DNA sampling effort for fish inventories in tropical streams and rivers. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-39399-5

Cawthorn, D. M., Duncan, J., Kastern, C., Francis, J., & Hoffman, L. C. (2015). Fish species substitution and misnaming in South Africa: An economic, safety and sustainability conundrum revisited. *Food Chemistry*, *185*, 165–181. https://doi.org/10.1016/j.foodchem.2015.03.113

Cilleros, K., Valentini, A., Allard, L., Dejean, T., Etienne, R., Grenouillet, G., … Brosse, S. (2019). Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): A test with Guianese freshwater fishes. *Molecular Ecology Resources*, *19*, 27–46. https://doi.org/10.1111/1755-0998.12900

Collen, B., Whitton, F., Dyer, E. E., Baillie, J. E. M., Cumberlidge, N., Darwall, W. R. T., … Böhm, M. (2014). Global patterns of freshwater species

diversity, threat and endemism. *Global Ecology and Biogeography*, *23*(1), 40–51. https://doi.org/10.1111/geb.12096

Collins, R. A., Wangensteen, O. S., O'Gorman, E. J., Mariani, S., Sims, D. W., & Genner, M. J. (2018). Persistence of environmental DNA in marine systems. *Communications Biology*, *1*, 185. https://doi.org/10.1038/s42003-018-0192-6

Day, J. J., Cotton, J. A., & Barraclough, T. G. (2008). Tempo and mode of diversification of Lake Tanganyika cichlid fishes. *PloS One*, *3*(3), e1730. https://doi.org/10.1371/journal.pone.0001730

Day, J. J., & Wilkinson, M. (2006). On the origin of the Synodontis catfish species flock from Lake Tanganyika. *Biology Letters*, *2*(4), 548–552. https://doi.org/10.1098/rsbl.2006.0532

Deacon, A. E., Mahabir, R., Inderlall, D., Ramnarine, I. W., & Magurran, A. E. (2017). Evaluating detectability of freshwater fish assemblages in tropical streams: Is hand-seining sufficient? *Environmental Biology of Fishes*, *100*(7), 839–849. https://doi.org/10.1007/s10641-017-0610-5

Decru, E., Moelants, T., De Gelas, K., Vreven, E., Verheyen, E., & Snoeks, J. (2016). Taxonomic challenges in freshwater fishes: A mismatch between morphology and DNA barcoding in fish of the north-eastern part of the Congo basin. *Molecular Ecology Resources*, *16*(1), 342–352. https://doi.org/10.1111/1755-0998.12445

Deiner, K., Bik, H. M., Machler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., … 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

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Eichmiller, J. J., Best, S. E., & Sorensen, P. W. (2016). Effects of temperature and trophic state on degradation of environmental DNA in lake water. *Environmental Science and Technology*, *50*(4), 1859–1867. https://doi.org/10.1021/acs.est.5b05672

Elbrecht, V., & Leese, F. (2017). PrimerMiner: An R package for development and in silico validation of DNA metabarcoding primers. *Methods in Ecology and Evolution*, *8*, 622–626. https://doi.org/10.1111/2041-210X.12687

Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., … 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

Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., … Lodge, D. M. (2016). Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 29–41. https://doi.org/10.1111/1755-0998.12433

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*(7), https://doi.org/10.1371/journal.pone.0022746

Gotelli, N. J., & Colwell, R. K. (2011). Estimating species richness. In A. E. Magurran, & B. J. McGill (Eds.), *Biological diversity: Frontiers in measurement and assessment* (pp. 39–54). Oxford, UK: Oxford University Press.

Grooten, M., & Almond, R. E. A. (2018). *Living Planet report – 2018: Aiming higher*. Gland, Switzerland: WWF.

Hänfling, B., Lawson Handley, L., Read, D. S., Hahn, C., Li, J., Nichols, P., … 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

Hinlo, R., Gleeson, D., Lintermans, M., & Furlan, E. (2017). Methods to maximise recovery of DNA from water samples. *PLoS ONE*, *12*(6), 1–22. https://doi.org/10.1371/journal.pone.0179251

Hori, M., Gashagaza, M. M., Nshombo, M., & Kawanabe, H. (1993). Littoral fish communities in Lake Tanganyika: Irreplaceable diversity supported by intricate interactions among species. *Conservation Biology*, *7*(3), 657–666. https://doi.org/10.1046/j.1523-1739.1993.07030657.x

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

Jackson, D. A., & Harvey, H. H. (1997). Qualitative and quantitative sampling of lake fish communities. *Canadian Journal of Fisheries and Aquatic Sciences*, *54*, 2807–2813. https://doi.org/10.1139/f97-182

Kimirei, I., & Mgaya, Y. (2007). Influence of environmental factors on seasonal changes in clupeid catches in the Kigoma area of Lake Tanganyika. *African Journal of Aquatic Science*, *32*(3), 291–298. https://doi.org/10.2989/AJAS.2007.32.3.9.308

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120. https://doi.org/10.1007/BF01731581

Konings, A. (2015). *Tanganyika cichlids in their natural habitat* (3rd ed.). El Paso, TX: Cichlid Press.

Kubecka, J., Hohausová, E., Matena, J., Peterka, J., Amarasinghe, U. S., Bonar, S. A., … Winfield, I. J. (2009). The true picture of a lake or reservoir fish stock: A review of needs and progress. *Fisheries Research*, *96*, 1–5. https://doi.org/10.1016/j.fishres.2008.09.021

Lacoursière-Roussel, A., Dubois, Y., Normandeau, E., & Bernatchez, L. (2016). Improving herpetological surveys in eastern North America using the environmental DNA method1. *Genome*, *59*(11), 991–1007. https://doi.org/10.1139/gen-2015-0218

Li, J., Hatton-Ellis, T. W., Lawson Handley, L. J., Kimbell, H. S., Benucci, M., Peirson, G., & Hänfling, B. (2019). Ground-truthing of a fish-based environmental DNA metabarcoding method for assessing the quality of lakes. *Journal of Applied Ecology*, *56*, 1232–1244. https://doi.org/10.1111/1365-2664.13352

Lopes, C. M., Sasso, T., Valentini, A., Dejean, T., Martins, M., Zamudio, K. R., & Haddad, C. F. B. (2017). eDNA metabarcoding: A promising method for anuran surveys in highly diverse tropical forests. *Molecular Ecology Resources*, *17*(5), 904–914. https://doi.org/10.1111/1755-0998.12643

Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

Mahé, F., Rognes, T., Quince, C., De Vargas, C., & Dunthorn, M. (2015). Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ*, *3*, e1420. https://doi.org/10.7717/peerj.1420

Majaneva, M., Diserud, O. H., Eagle, S. H. C., Boström, E., Hajibabaei, M., & Ekrem, T. (2018). Environmental DNA filtration techniques affect recovered biodiversity. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-23052-8

Meyer, A., Morrissey, J. M., & Scharti, M. (1994). Recurrent origin of a sexually selected trait in *Xiphophorus* fishes inferred from a molecular phylogeny. *Nature*, *368*, 539–541. https://doi.org/10.1038/368539a0

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., … 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*(7), 150088. https://doi.org/10.1098/rsos.150088

Muschick, M., Indermaur, A., & Salzburger, W. (2012). Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology*, *22*(24), 2362–2368. https://doi.org/10.1016/j.cub.2012.10.048

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., … Oksanen, M. J. (2012). *Community ecology package: Vegan. Version 2.4-0*. Retrieved from https://cran.r-project.org

Oliveira, A. G., Gomes, L. C., Latini, J. D., & Agostinho, A. A. (2012). Implications of using a variety of fishing strategies and sampling techniques across different biotopes to determine fish species

composition and diversity. *Natureza e Conservacao*, *12*(2), 112–117. https://doi.org/10.1016/j.ncon.2014.08.004

Pais, M. P., & Cabral, H. N. (2018). Effect of underwater visual survey methodology on bias and precision of fish counts: A simulation approach. *PeerJ*, *6*, e5378. https://doi.org/10.7717/peerj.5378

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

Peart, C. R., Bills, R., Wilkinson, M., & Day, J. J. (2014). Nocturnal claroteine catfishes reveal dual colonisation but a single radiation in Lake Tanganyika. *Molecular Phylogenetics and Evolution*, *73*(1), 119–128. https://doi.org/10.1016/j.ympev.2014.01.013

Peart, C. R., Dasmahapatra, K. K., & Day, J. J. (2018). Contrasting geographic structure in evolutionarily divergent Lake Tanganyika catfishes. *Ecology and Evolution*, *8*(5), 2688–2697. https://doi.org/10.1002/ece3.3860

Pedrosa-Gerasmio, I. R., Babaran, R. P., & Santos, M. D. (2012). Discrimination of juvenile yellowfin (*Thunnus albacares*) and bigeye (*T. obesus*) tunas using mitochondrial DNA control region and liver morphology. *PLoS ONE*, *7*(4), e35604. https://doi.org/10.1371/journal.pone.0035604

Pereira, L. H. G., Hanner, R., Foresti, F., & Oliveira, C. (2013). Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? *BMC Genetics*, *14*, 20. https://doi.org/10.1186/1471-2156-14-20

Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A., … Dejean, T. (2018). Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific Reports*, *8*(1), 1–13. https://doi.org/10.1038/s41598-018-28424-8

Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., … Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, *25*(2), 527–541. https://doi.org/10.1111/mec.13481

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*(21), 1–11. https://doi.org/10.1093/nar/gkr732

Ronco, F., Büscher, H. H., Indermaur, A., & Salzburger, W. (2019). The taxonomic diversity of the cichlid fish fauna of ancient Lake Tanganyika, East Africa. *Journal of Great Lakes Research*, in press. https://doi.org/10.1016/j.jglr.2019.05.009

RStudio Team (2015). *RStudio: Integrated development for R*. Retrieved from http://www.rstudio.com/

Sales, N. G., Mariani, S., Salvador, G. N., Pessali, T. C., & Carvalho, D. C. (2018). Hidden diversity hampers conservation efforts in a highly impacted neotropical river system. *Frontiers in Genetics*, *9*, 1–11. https://doi.org/10.3389/fgene.2018.00271

Salzburger, W. (2018). Understanding explosive diversification through cichlid fish genomics. *Nature Reviews Genetics*, *19*, 705–717. https://doi.org/10.1038/s41576-018-0043-9

Salzburger, W., Van Bocxlaer, B., & Cohen, A. S. (2014). Ecology and evolution of the African Great Lakes and their faunas. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 519–545. https://doi.org/10.1146/annurev-ecolsys-120213-091804

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

Sefc, K. M., Baric, S., Salzburger, W., & Sturmbauer, C. (2007). Species-specific population structure in rock-specialized sympatric cichlid species in Lake Tanganyika, East Africa. *Journal of Molecular Evolution*, *64*(1), 33–49. https://doi.org/10.1007/s00239-006-0011-4

Shum, P., Moore, L., Pampoulie, C., Di Muri, C., Vandamme, S., & Mariani, S. (2017). Harnessing mtDNA variation to resolve ambiguity in 'Redfish' sold in Europe. *PeerJ*, *5*, e3746. https://doi.org/10.7717/peerj.3746

Shum, P., Pampoulie, C., Kristinsson, K., & Mariani, S. (2015). Three-dimensional post-glacial expansion and diversification of an exploited oceanic fish. *Molecular Ecology*, *24*(14), 3652–3667. https://doi.org/10.1111/mec.13262

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., … Bunce, M. (2017). Ecosystem biomonitoring with eDNA: Metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-12501-5

Strayer, D. L., & Dudgeon, D. (2010). Freshwater biodiversity conservation: Recent progress and future challenges. *Journal of the North American Benthological Society*, *29*(1), 344–358. https://doi.org/10.1899/08-171.1

Strickler, K. M., Fremier, A. K., & Goldberg, C. S. (2015). Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biological Conservation*, *183*, 85–92. https://doi.org/10.1016/j.biocon.2014.11.038

Takeuchi, Y., Ochi, H., Kohda, M., Sinyinza, D., & Hori, M. (2010). A 20-year census of a rocky littoral fish community in Lake Tanganyika. *Ecology of Freshwater Fish*, *19*(2), 239–248. https://doi.org/10.1111/j.1600-0633.2010.00408.x

Tisseuil, C., Cornu, J.-F., Beauchard, O., Brosse, S., Darwall, W., Holland, R., … Oberdorff, T. (2013). Global diversity patterns and cross-taxa convergence in freshwater systems. *Journal of Animal Ecology*, *82*(2), 365–376. https://doi.org/10.1111/1365-2656.12018

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), 1–12. https://doi.org/10.1093/nar/gks596

Vadeboncoeur, Y., McIntyre, P. B., & Vander Zanden, M. J. (2011). Borders of biodiversity: Life at the edge of the World's large lakes. *Bioscience*, *61*(7), 526–537. https://doi.org/10.1525/bio.2011.61.7.7

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., … Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, *25*(4), 929–942. https://doi.org/10.1111/mec.13428

Wagner, C. E., & McCune, A. R. (2009). Contrasting patterns of spatial genetic structure in sympatric rock-dwelling cichlid fishes. *Evolution*, *63*(5), 1312–1326. https://doi.org/10.1111/j.1558-5646.2009.00612.x

Widmer, L., Ronco, F., Heule, E., Indermaur, A., Colombo, M., Salzburger, W., & Rueegg, A. (2019). Point-combination transect (PCT): Incorporation of small underwater cameras to study fish communities. *Methods in Ecology and Evolution*, *10*(6), 1–11. https://doi.org/10.1111/2041-210x.13163

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., … Miya, M. (2017). Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific Reports*, *7*, 40368. https://doi.org/10.1038/srep40368

SUPPORTING INFORMATION

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

How to cite this article: Doble CJ, Hipperson H, Salzburger W, et al. Testing the performance of environmental DNA metabarcoding for surveying highly diverse tropical fish communities: A case study from Lake Tanganyika. *Environmental DNA*. 2020;2:24–41. https://doi.org/10.1002/edn3.43