Environmental DNA

WILEY

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

ORIGINAL ARTICLE OPEN ACCESS

Environmental DNA Metabarcoding Details the Spatial Structure of a Diverse Tropical Fish Assemblage in a Major East African River System

Asilatu H. Shechonge[1] | Rupert A. Collins[2] | Sophie Ward[3] | Andrew D. Saxon[3] | Alan M. Smith[4] | Patroba Matiku[1] | George F. Turner[5] | Mary A. Kishe[1] | Benjamin P. Ngatunga[1] | Martin J. Genner[3]

[1]Tanzania Fisheries Research Institute (TAFIRI), Dar es Salaam, Tanzania | [2]Department of Life Sciences, The Natural History Museum, London, UK | [3]School of Biological Sciences, University of Bristol, Bristol, UK | [4]School of Natural Sciences, University of Hull, Hull, UK | [5]School of Environmental & Natural Sciences, Bangor University, Bangor, Gwynedd, UK

Correspondence: Asilatu H. Shechonge (ashechonge@yahoo.com) | Martin J. Genner (m.genner@bristol.ac.uk)

Received: 28 March 2024 | **Revised:** 15 July 2024 | **Accepted:** 19 August 2024

Funding: This work was supported by the Royal Society, (LAF\R1\180000), and Royal Society-Leverhulme Trust Africa Awards to MJG, BPN, and GFT (AA100023 and AA130107).

Keywords: aquatic biodiversity | community structure | DNA barcoding | freshwater fishes | phylogeny

ABSTRACT

Management and conservation of species-rich tropical freshwater systems require reliable information on the diversity and distribution of species present. Here, we used environmental DNA metabarcoding to reveal the diversity of the fishes in the Rufiji River catchment of central Tanzania. Across 174 samples from 49 sites, and using a newly developed reference library, we mapped the presence of 66 fish species from an estimated 91 that we are confident are present in the system. We found clear evidence of community structuring of the assemblage linked to key environmental gradients—elevation, temperature, and turbidity. We also identified core distributions of rare or threatened taxa, including migratory species such as the anguillid eels. With a focused analysis of 50 samples collected over a small spatial scale (<2 km) from the Kilombero River, we showed that each single sample can capture an average of 23.1 species, while three samples can capture 39.4 species, from a total of 56 species encountered in the 50 samples. Collectively the results help to identify species vulnerable to ongoing change in the catchment, including dam construction and agricultural intensification. The results clearly demonstrate how eDNA-based metabarcoding can reliably describe the diversity and distributions of riverine fish species across a catchment, providing standardized information that will be valuable for environmental management.

1 | Introduction

Tropical freshwater ecosystems are often characterized by high levels of fish diversity (Oberdorff et al. 2011), yet our knowledge of species distributions is often limited (Lundberg et al. 2000; Jézéquel et al. 2020; Freitas et al. 2021; Schedel et al. 2024)—a concept termed the "Wallacean shortfall" (Hortal et al. 2015). More detailed knowledge of diversity would enable more effective management of these ecosystems, which are challenged by stressors including dam construction, water abstraction, pollution, and climate change (Arthington et al. 2016; Barbarossa et al. 2021). The use of environmental DNA (eDNA) metabarcoding to determine the composition of fish assemblages is now established (Yao et al. 2022), but there are relatively few studies that have applied eDNA-based methods to study tropical freshwater fish communities (e.g., Cilleros et al. 2019; Valdez-Moreno et al. 2019), particularly in Africa (e.g., Doble et al. 2020; von der Heyden 2023; von der Heyden et al. 2023). It is valuable,

therefore, to determine the extent to which eDNA-derived metabarcoding surveys can adequately capture the species richness of tropical ecosystems at the catchment scale, and the extent to which spatial patterns of biodiversity reflect changes in important environmental variables known to structure assemblages, including elevation, water temperature, and dissolved oxygen content (Ibanez et al. 2007; Esselman and Allan 2010; Mondal and Bhat 2020).

One of the key challenges for environmental DNA metabarcoding studies is assignment of derived reads to the correct taxa (Deiner et al. 2017; Bourret et al. 2023), which requires a high-quality reference library of sequences that have been derived from specimens that have been reliably determined to species level. Obtaining reliable assignments is especially challenging to complete in tropical African freshwaters, which can contain upward of 50 species in a single catchment (Hugueny 1989; Darwall et al. 2005), and with the full diversity typically underrepresented in public databases of DNA barcode data (Marques et al. 2021). An additional problem with assignments arises when multiple species share a single haplotype of the mitochondrial DNA markers that are typically used for barcoding (Collins et al. 2021). This can arise in cases where species are recently diverged, and thus have not had time for mutations to accumulate (Hubert et al. 2008; Knebelsberger et al. 2014). It can also arise when species hybridize, introgressing the mitochondrial genome between species, even if they are largely reproductively isolated (e.g., Schmidt, Bart Jr, and Nyingi 2017). A good understanding of the ability of chosen barcodes to discriminate taxa is necessary when designing eDNA-based investigations.

Here, we used an eDNA metabarcoding approach to study the diversity and spatial structure of fishes in a large river system in East Africa, the Rufiji River catchment of Tanzania. This catchment comprises two major tributaries, the Ruaha and Kilombero, which meet to form the Rufiji before flowing into the Indian Ocean (Figure 1). This region was selected due to a known high diversity of fish taxa (Eccles 1992)—that we estimate to be at least 91 species—from 47 genera and 21 families (Table S1). It was also selected because of importance of the river for artisanal fisheries (Mwalyosi 1990), and for the conservation importance of the river system (Seeteram et al. 2019). Of particular note is the construction of the Julius Nyerere Hydropower Station in Stiegler's Gorge on the Rufiji River, which is between our two main sampling areas, and within the boundaries of Nyerere National Park (IUCN 2019; Figure 1).

For this study, our eDNA samples covered a range of habitats across the Ruaha, Kilombero, and Rufiji sections of the catchment, including fast-flowing montane streams, slow-flowing open river and stillwater oxbow lakes (Figure 1). We analyzed these samples using a bespoke reference library that enabled us to assign a high proportion of fish reads to species. The key aims of this study were to (1) develop a comprehensive taxonomic knowledge of the current diversity of the studied river sectors; (2) quantify the number of species that can be detected with increased sampling in a species-rich community; (3) determine if environmental variables associated with spatial patterns of fish community structure derived from eDNA metabarcode data; and (4) determine if eDNA can determine the likely distributions species that may be of conservation interest—particularly given changes to the environment from dam construction and ongoing land-use change in the region.

2 | Materials and Methods

2.1 | Study Region

Our study system is the Rufiji River system, including the upstream Great Ruaha River, the Kilombero River, and the Rufiji River (Figure 1). Our sampling included mountain streams, broad river channels, and shallow lakes. We grouped our Rufiji sampling sites into five geographical categories, namely higher-elevation (>290 m above sea level) tributaries of the Kilombero River, lower-elevation tributaries of the Kilombero River (260–290 m above sea level), the Kilombero River floodplain at Ifakara, the Great Ruaha River at Kidatu, and the Rufiji River (and associated oxbow lakes). We also sampled one location outside the catchment, although in close proximity to the catchment boundary (Lake Mansi), and refer to this as a sixth geographical category. In total, we sampled eDNA during 49 events (locations), between 21 May 2019 and 25 July 2020. Between three and six samples were collected during each event (total of 174 samples) (Figure 1; Table S2). Note that Lake Mansi has a fauna shared with the Rufiji River system.

2.2 | eDNA Sampling

Water samples were collected from the surface with a new (or sterilized with 10% bleach if reused) plastic container and were hand pumped through a 0.22 μm PES membrane Sterivex canister filter (Merck Millipore) using a 50-mL syringe. The volume of water pumped through the filter ranged between 50 and 420 mL (average 202 mL per sample). After water was filtered, air was pumped through the filter to remove residual water. A Combi stopper was then used to seal the lower (luer lock) end of the filter, and the contents were preserved through the additional 0.37 mL of ATL buffer (Qiagen), using a 1-mL syringe. The upper end of the Sterivex was then heat sealed by melting the end with a flame, labeled, and placed in a sealed Whirl-Pak bag (Cole-Parmer). Full details of the eDNA field collection protocol are available (https://shorturl.at/LiBNO; https://doi.org/10.5281/zenodo.4687985). GPS coordinates and elevation were obtained using an eTrex handheld GPS (Garmin), and at each location, environmental variables were recorded including temperature, and data on dissolved oxygen, pH, salinity, conductivity, and turbidity were collected using a CTD profiler (Sea-Bird SBE 19plus V2 SEACAT). We also collected a series of seven field blanks during the sampling, each consisting of ~150 mL of shop-bought bottled drinking water that was processed in the same way as the field samples.

2.3 | Reference Library

DNA samples (fin tissue) from reference specimens were collected and preserved in absolute ethanol. Genomic DNA was extracted using the Wizard kit (Promega), and DNA was quantified using a C40 nanophotometer (Implen). The primers Aa22-12SF

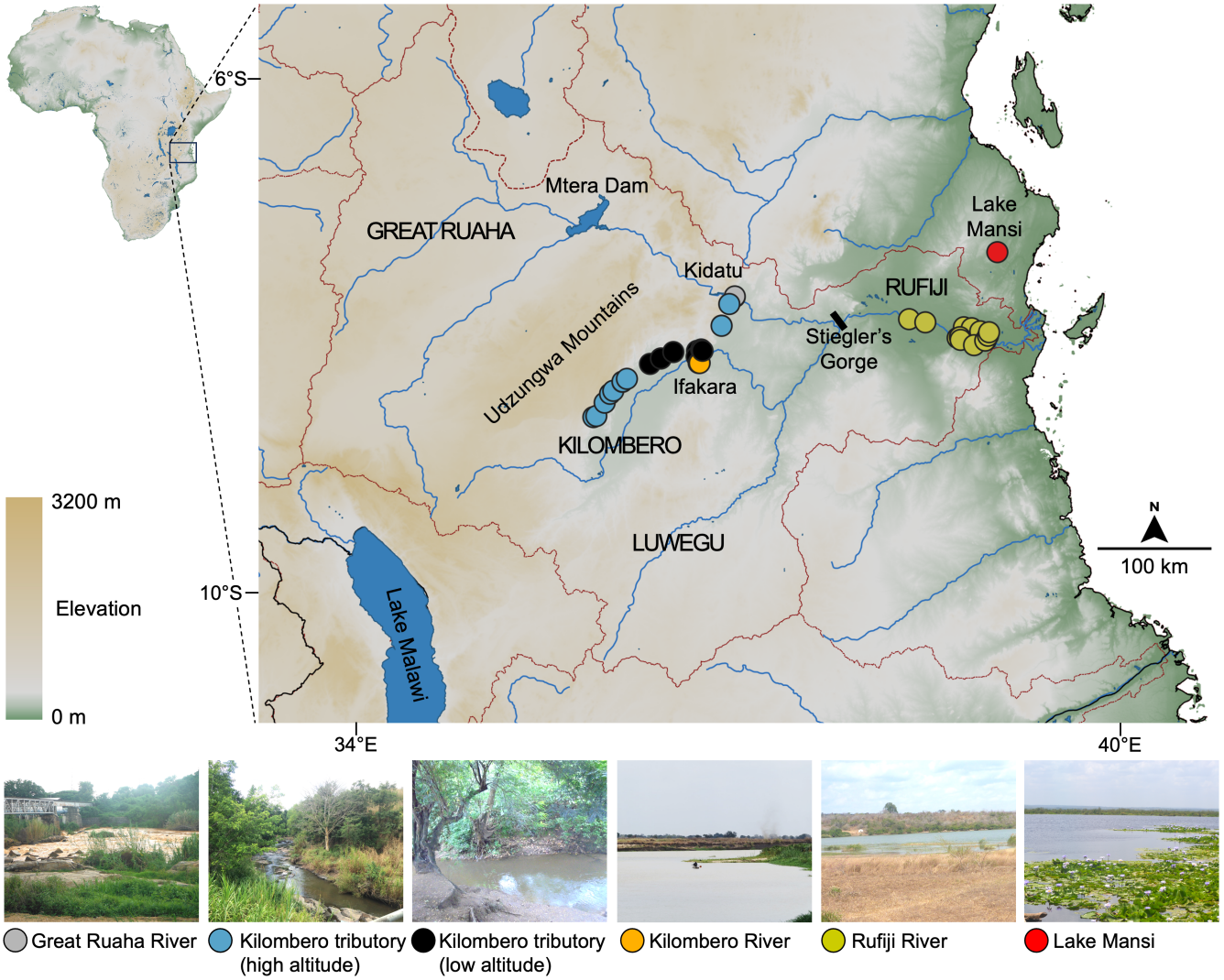

FIGURE 1 | Sampling sites in the Rufiji catchment in Tanzania. In total, 49 locations were sampled, separated into six geographical categories (lower images) indicated by colored circles that overlap in the image. For details of sampling sites, see Table S2.

(5′-AGC ATA ACA CTG AAG ATR YTA RGA-3′) and Aa633-12SR (5′-TTC TAG AAC AGG CTC CTC TAG-3′) were used to amplify a ~567-bp fragment of the mitochondrial 12S rRNA gene from all fish specimens (Collins et al. 2021). The 12S PCR amplifications were performed in 20 μL volumes with 5 μL molecular biology–grade H_2O, 10 μL GoTaq Green MasterMix (Promega), 2 μL of each primer (2 μM), and 1 μL genomic DNA template. PCR conditions for 12S were as follows: an initial denaturation was performed at 95°C for 2 min, followed by 40 cycles of denaturation at 95°C for 30 s, annealing at 53°C for 30 s, and extension at 72°C for 45 s. A final extension step was performed at 72°C for 5 min. PCR success was verified by 1% agarose gel electrophoresis, and PCRs were cleaned using AMPure XP magnetic beads (Beckman Coulter). Sequences were obtained through Sanger sequencing of PCR products by Eurofins Genomics (https://eurofinsgenomics.eu/). In total, our reference library contained 198 newly sequenced individuals from 66 species from the Rufiji–Ruaha–Kilombero, or proximate catchments (Table S3). We found no cases of different species sharing 12S haplotypes. Sequences have GenBank accessions PQ268652 to PQ268849. To augment our reference library, we sourced data from nine species from the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra). Additionally, the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/nucleotide; v259, Jan 2024) was searched using meta-fish-lib (Collins et al. 2021; https://doi.org/10.5281/zenodo.10368996) using a list of genera reported from the catchment (Table S1). The resulting 12S sequence data were then quality controlled using phylogenetic analysis, and mislabeled sequences were removed. We also removed duplicate haplotypes annotated with the same taxonomic name. For preliminary taxonomic assignment to remove nonfish reads, the NCBI RefSeq v208 (https://www.ncbi.nlm.nih.gov/refseq/) database was accessed using refseq-reflib v1.1 (https://doi.org/10.5281/zenodo.5636772).

2.4 | eDNA Sample Processing

Preserved eDNA samples were extracted using a modified version of the in-canister extraction method (Spens et al. 2017) in a dedicated eDNA extraction/pre-PCR laboratory at the University of Bristol. The laboratory was sterilized with 10%

bleach solution and UV light prior to extractions. The extraction protocol involved digestion in proteinase K and digestion for 2 h at 56°C. The supernatant was then removed from the canister and added to buffer AL from the DNeasy Blood & Tissue Kit (Qiagen). Next, the sample was placed and spun in a DNeasy spin column before successive washes with buffers AW1 and AW2 from the DNeasy kit. Finally, the DNA was eluted into AE buffer. Eluted samples were purified using a OneStep PCR Inhibitor Removal Kit (Zymo). Full details of the eDNA laboratory extraction protocol are available (https://shorturl.at/hFIdv; https://doi.org/10.5281/zenodo.4741283).

Pre-PCR handling was undertaken at a dedicated eDNA extraction/pre-PCR laboratory at the University of Bristol. Metabarcode data were obtained by amplifying a ~167 bp fragment of mitochondrial 12S using the Tele02 primer set (Taberlet et al. 2018). Each forward and reverse PCR primer pair was adapted with the same unique 8-mer sample identification tag differing across pairs by at least three nucleotides, and including variable 5′ random heterogeneity of 2–4 bp. PCRs were conducted in triplicate 20 μL reactions of 10 μL AmpliTaqGold 360 Mastermix (ThermoFisher), 0.16 μL BSA (New England Biolabs), 5.84 μL water, 2 μL forward and reverse primer (5 μM), and 2 μL template DNA. Cycling conditions comprised: denaturation at 95°C for 10 min; 40 cycles of 95°C for 30 s, 54°C for 30 s, and 72°C for 60 s; and final extension of 72°C for 7 min.

Post-PCR handling was undertaken in dedicated library preparation space of another molecular laboratory at the University of Bristol. Amplicons for each sample were double size selected with AMPure XP paramagnetic beads (Beckman Coulter) and eluted in EB buffer before being pooled at equal concentrations. Pools were then purified and concentrated with a DNA Clean & Concentrator kit (Zymo). In total, two libraries comprising 174 samples, seven field blanks, and a further five laboratory extraction blank controls were prepared using the PCR-free Kapa HyperPrep kit (Roche) following manufacturer instructions. Vazyme DNA adapters (VAHTS) were triplexed in each library to ensure optimal sequencing. The final libraries were quantified using qPCR and the NEBNext Library Quant Kit (New England Biolabs) and sequenced on a NextSeq 500 instrument (Illumina) using a v2.5 mid-output, 2×150 bp kit, at a concentration of 1.0 pM plus a 10% PhiX spike-in, at the University of Bristol Genomics Facility.

2.5 | Bioinformatic Processing

Bioinformatic processing and taxonomic assignment were conducted using the meta-fish-pipe v1.1 software (https://doi.org/10.5281/zenodo.5636787). The taxonomic assignment method is based on congruence between phylogenetic placement and sequence similarity, allowing for robust assignments to be made with confidence, even in situations where species differ by only single substitutions. The pipeline comprised: (1) sample demultiplexing using cutadapt v3.4 (Martin 2011); (2) sequence denoising and dereplication using dada2 v1.20.0 (Callahan et al. 2016); (3) homology filtering of amplicon sequence variants (ASVs) using hidden Markov models in hmmer v3.3 (http://hmmer.org/); (4) first-pass taxonomic assignment to exclude nonfish ASVs using sintax (Edgar 2016) and NCBI RefSeq v208 (https://www.ncbi.nlm.nih.gov/refseq/); and (5) exhaustive assignment using combined phylogenetic placement (epa-ng v0.3.8; Barbera et al. 2019) and sequence similarity (blastn v2.11.0), using our curated Rufiji River fish reference library. Reference library sequences were used as priors during dada2 denoising to avoid erroneously discarding rare sequences, and an exclusion list of sequences generated from concurrent laboratory projects was used to control for potential cross-contamination. Only reads assigned to freshwater and estuarine fishes were retained for analyses (i.e., reads from marine fishes and non-fish species were removed). We retained estuarine species as several species from our curated species list (Table S1) are known to be primarily estuarine, but range into freshwaters.

2.6 | Data Analysis

The analyses of the eDNA-derived data were conducted in R v4.3.1 (R Core Team 2023). We initially focused on a subset of data from the 50 samples collected from the Kilombero River at Ifakara. Since these were collected from very similar habitats, all within 2 km of one another, this provided an opportunity to determine how increased sampling effort determined the number of species sampled. Species accumulation curves were estimated using the "specaccum" function using the Coleman accumulation method in vegan v2.6.4 (Oksanen et al. 2022).

For the full matrix, and using each sample independently, we generated a matrix comprising Hellinger transformed data, using the "decostand" function in vegan. To ordinate differences among samples, we used principal coordinates analysis (PCOA) implemented with the "pcoa" function in ape v5.0 (Paradis and Schliep 2019), using the resulting primary axes of variation to visualize differences among the geographical categories. To test for differences among sites and the six geographical categories, we used PERMANOVA with the "adonis2" function in vegan, with 10,000 permutations, and nesting samples within site.

Correlations between environmental variables were quantified using Hmisc v5.1-3 (Harrell Jr. 2024). Our environmental data were not complete, therefore we imputed missing data using knnImputation method in DMwR2 v0.0.2 (Torgo 2016). Tests for associations between key environmental variables and major axes of community structure were undertaken using redundancy analyses in vegan v2.6.4, using each sample independently, with associations between community structure and environmental variables quantified using a ANOVA-like permutation tests with 10,000 permutations. Associations between species and the six geographical categories were quantified using indicator value (IndVal) statistics calculated in indicspecies v1.7.14 (De Cáceres and Legendre 2009).

3 | Results

3.1 | Sequencing and Overview of the Diversity Present and Detected in eDNA

In total 62,226,031 reads passed filter and were able to be assigned taxonomic identities from our reference library (Supplementary Table S4). Our aim was to study the distributions of known

species from the catchment using eDNA, and therefore chose not to incorporate reads with uncertain species-level assignments in our analyses. Thus, we further excluded reads assigned to taxa only known from sites outside the catchment [*Labeobarbus intermedius* (Rüppell 1835), *Amphilius* sp. 'CD-2019' and *Marcusenius* sp. 'wami']. We also merged two taxa [*Marcusenius livingstonii* (Boulenger 1899) and *Marcusenius* cf. *livingstonii*]. After this step there were an average of 346,604 assigned reads per sample (±311,338 standard deviation; range 571 to 1,866,218; and median 274,906) and an average of 24.7 fish taxa per sample (±9.5 SD; range 4 to 46) (Table S5). Hereafter these fish taxa are referred to as species. Our seven-field blank controls comprised only 15 reads that were assigned to fish species (mean 2.1 reads per blank ±1.5 SD; Table S6). Our five extraction blank controls comprised only nine reads that were assigned to fish species (mean 1.8 reads per blank ±4.0 SD; Table S6). Thus, we consider our analyses to have minimal contamination. Across all 174 samples, using sampling site as a random factor, we found that the volume of water sampled was not a significant predictor of the $\log_{10}$-transformed total number of assigned reads (linear mixed-effects model, estimate = 0.0007, $t = 0.548$, $p = 0.586$) or the number of species recovered (linear mixed-effects model, estimate = −773.12, $t = -1.216$, $p = 0.230$). However, the total number of species recovered was positively associated with the $\log_{10}$-transformed total number of assigned reads (linear mixed-effects model, estimate = 1.89, $t = 3.029$, $p = 0.003$).

Across all 174 field samples, we confirmed the presence of reads from 66 fish species in the environmental DNA (Table S5). Thus, we identified 73% of the 91 fish species that we estimate to be present in Ruaha–Rufiji catchment, from our own observations and a critical evaluation of published records (Table S1). Eleven sampling sites were in close proximity (<2 km) from one another, at Ifakara on a wide section of the Kilombero River plain, with between three and six samples taken in each, totaling 50 samples (Table S2). Across those 50 samples combined, we identified a total of 56 species. We found an average of 24.69 species per sample (±9.49 SD). Species accumulation curves estimated with the Coleman method showed a single typical sample would yield 23.1 species (±3.3 SD), whereas three samples would typically yield 39.4 species (±2.4 SD). Thus, with one sample we would capture 41% of the species that would be captured with 50 samples, while three samples would provide 70% of those species (Figure 2a).

3.2 | Community Structure in Relation to Environmental Variables

We found highly significant differences in community composition between our six geographical categories (PERMANOVA, $F_{5,125} = 93.01$, $r^2 = 42.83$, $p < 0.001$), with pairwise tests showing significant ($p < 0.05$) differences between all geographical category pairs within the main study catchment (Table S7). Differences in spatial diversity of the assemblage were ordinated using a PCOA, with the primary axis (PCOA1) capturing 27.4% of the observed variation, and the secondary axis (PCOA2) capturing 14.7%. This analysis showed broad separation between communities from most geographical categories, with overlap between communities in the high and low tributaries of the Kilombero River, and between communities in Lake Mansi and the Rufiji River. PCOA1 captured a gradient of community structure from high-elevation tributaries of the Kilombero River at one extreme, to open river and shallow lakes of the Rufiji River at the other (Figure 2b). Our analyses of associations between environmental variables showed that higher-elevation habitats were more likely to be flowing, with relatively low conductivity, high pH, low salinity, and low temperatures (Table S8). Redundancy analyses revealed highly significant associations between community structure and environmental variables; the predictor variables with the greatest explanatory power were elevation and temperature (Figure 2c,d; Table S9).

Although most species were present across multiple geographical categories, calculation of indicator values enabled us to identify geographical categories that species were most closely associated with (Figure 3), based on the probability of occurrence of samples in geographical categories, and relative read abundance (Table S10). High-elevation tributaries were characterized by small-bodied mountain catfish [*Amphilius* sp., *Chiloglanis* sp. 'mbarali river' and *Zaireichthys rotundiceps* (Hilgendorf 1905)], cyprinoids [*Engraulicypris spinifer* Bailey & Matthes 1971, *Enteromius atkinsoni* (Bailey 1969), *Enteromius kerstenii* (Peters 1868) and *Labeo cylindricus* Peters 1852], eels [*Anguilla bengalensis* (Gray 1831), *Anguilla mossambica* (Peters 1852) and *Mastacembelus frenatus* Boulenger 1901] and the kneriid *Parakneria tanzaniae* Poll 1984. Lower elevation tributaries were characterized by catfishes [*Clarias gariepinus* (Burchell 1822), *Clarias theodorae* Weber 1897, *Atopochilus vogti* Pellegrin 1922], cyprinoids [*Enteromius apleurogramma* (Boulenger 1911), *Enteromius luikae* (Ricardo 1939), *Enteromius radiatus* (Peters 1853), *Opsaridium loveridgii* (Norman 1922), *Labeobarbus macrolepis* (Pfeffer 1889), *Labeobarbus oxyrhynchus* (Pfeffer 1889)] and cichlids [*Haplochromis vanheusdeni* (Schedel, Friel, and Schliewen 2014), *Oreochromis leucostictus* (Trewavas 1933)]. The main floodplain of the Kilombero River was dominated by catfishes [*Bagrus orientalis* Boulenger 1902, *Synodontis* sp. 'utete', *Schilbe moebiusii* (Pfeffer 1896), *Pareutropius longifilis* (Steindachner 1914)], mormyrids [*Cyphomyrus discorhynchus* (Peters 1852), *Marcusenius livingstonii* (Boulenger 1899), *Mormyrus longirostris* Peters 1852)], alestids [*Brachyalestes lateralis* (Boulenger 1900), *Rhabdalestes barnardi* (Herre 1936), *Hydrocynus tanzaniae* Brewster 1986] and cyprinoids [*Enteromius lineomaculatus* (Boulenger 1903), *Enteromius macrotaenia* (Worthington 1933)]. The Kilombero River was also characterized by a haplochromine cichlid (*Pseudocrenilabrus* sp. 'kilombero'), an anabantid [*Ctenopoma muriei* (Boulenger 1906)], a poecilid topminnow [*Lacustricola kongoranensis* (Ahl 1924)] and an annual killifish (*Nothobranchius kilomberoensis* Wildekamp, Watters & Sainthouse 2002). The Great Ruaha River was primarily characterized by three undescribed cichlids (*Astatotilapia* sp. "ruaha blue", *A.* sp. "ruaha red cheek" and *O.* sp. 'mtera'), as well as an eel (*Anguilla marmorata* Quoy & Gaimard 1824). The Rufiji River was characterized by relatively large alestids [*Alestes stuhlmannii* Pfeffer 1896, *Brachyalestes affinis* (Günther 1894), *Brachyalestes imberi* (Peters 1852)], a citharinid (*Citharinus congicus* Boulenger 1897), a distichodontid (*Distichodus petersii* Pfeffer 1896), a cyprinoid (*Labeo congoro* Peters 1852), catfish (*Synodontis rufigiensis* Bailey 1968, *Synodontis rukwaensis* Hilgendorf & Pappenheim 1903), a mormyrid (*Petrocephalus steindachneri* Fowler 1958) and an eel (*Anguilla* bicolor McClelland 1844). The Rufiji River also

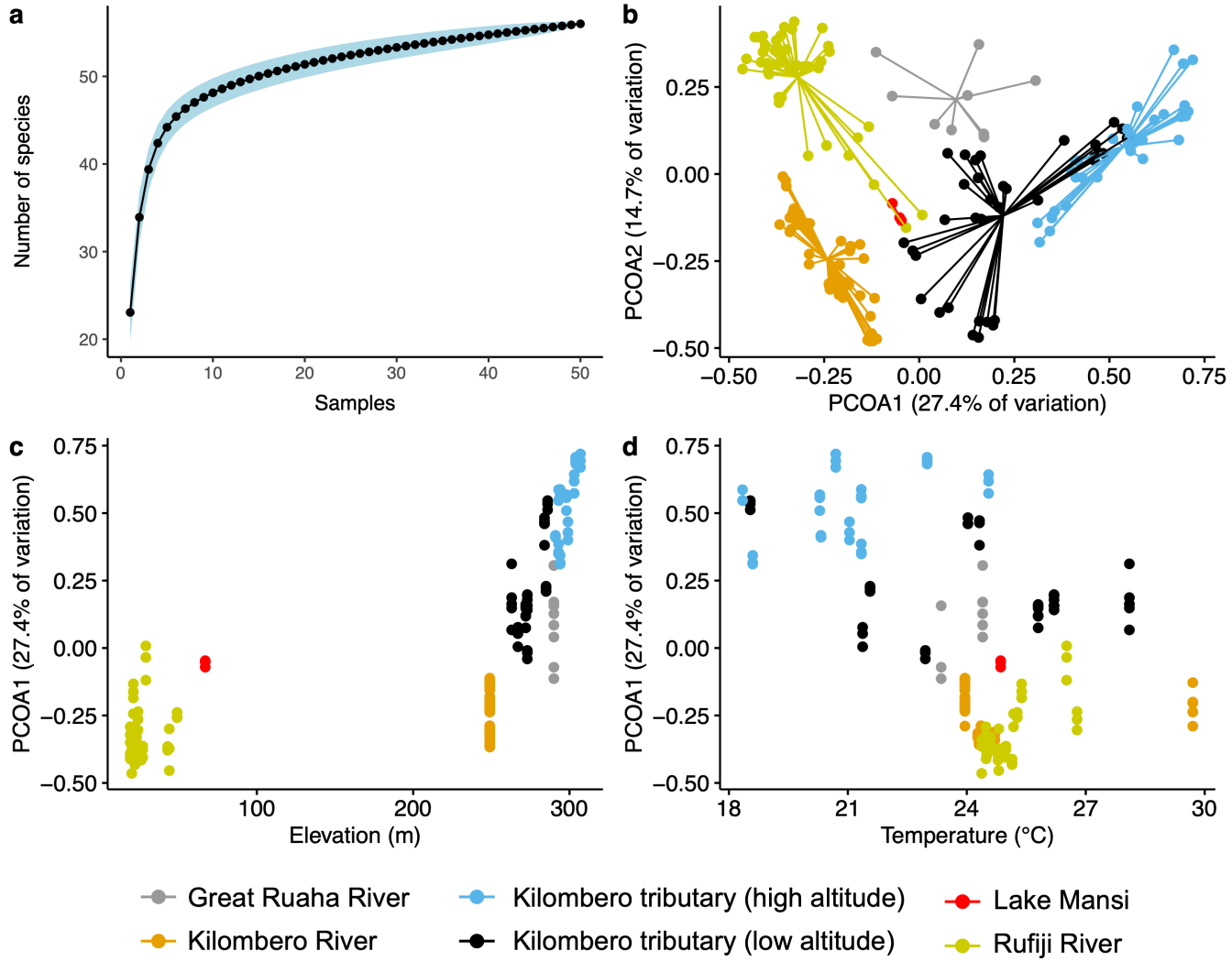

FIGURE 2 | Fish diversity recovered using environmental DNA metabarcoding. (a) Species accumulation with increasing sampling density on the Kilombero River, as estimated using the Coleman method. Adjacent points are joined with lines. The blue shading indicates the standard deviation. (b) Principal coordinate analysis illustrating similarity of assemblages recovered from six geographical categories, across the first two axes of community variation (PCOA1 and PCOA2; Hellinger-transformed read abundance). Each data point represents one sample. (c) Association between elevation and primary axis of community-wide variation (PCOA1). (d) Association between temperature and primary axis of community-wide variation (PCOA1).

was characterised by river-sea migratory species include gobiiforms [*Glossogobius giurus* (Hamilton 1822) and *Eleotris klunzingerii* Pfeffer 1893] and the tarpon *Megalops cyprinoides* (Broussonet 1782). Landlocked Lake Mansi was cichlid dominated, including haplochromines [*Astatotilapia* sp 'rufiji blue', *Astatotilapia gigliolii* (Pfeffer 1896)] and tilapiines [(*Oreochromis urolepis* (Norman 1922), *Oreochromis niloticus* (Linnaeus 1758), *Coptodon rendalli* (Boulenger 1897)]. Also characteristic of Lake Mansi were a mormyrid [*Marcusenius macrolepidotus* (Peters 1852)] and an undescribed characiform herein referred to as *Nannaethiops* sp. 'kilombero'.

3.3 | Resolving the Distribution of Individual Species

Our analyses enabled us to determine geographic spread of individual species within the eDNA data, and to focus on the distributions of key species of conservation concern, including migratory species, invasive species, and species that are endemic to the catchment with very localized distributions. The diadromous Indo-Pacific tarpon *M. cyprinoides* (Figure 4a), the amphidromous tank goby *G. giuris* (Figure 4b), and sleeper *E. klunzingerii* were only found in the lower reaches of the Rufiji section of the system. Reads assigned to three species of catadromous anguillid eels (*A. bengalensis* [Figure 4c]; *A. mossambica* and *A. marmorata*) were primarily found in the higher-altitude regions of the catchment. Reads belonging to the invasive tilapia species *O. niloticus* (Figure 4d) and *O. leucostictus* were identified, with *O. niloticus* reads present in both the main Rufiji system and Lake Mansi (Figure 4), while those assigned to invasive *O. leucostictus* were only recovered from a single sample at Ifakara in the Kilombero River. Although we did not sample upstream locations in the Ruaha where the undescribed catchment endemic *Oreochromis* sp. "mtera" has been recorded, reads assigned to the species were recovered in the Great Ruaha

at Kidatu and the Kilombero River, but not in the downstream sampling locations, or the Rufiji sector of the system. The introduced *C. rendalli* was only recovered in Lake Mansi. We noted some highly localized catchment-endemic species within our sampled region—for example, the chedrin *O. loveridgii* (Figure 4) was primarily recovered in the Kilombero sampling locations (Figure 4e). Two undescribed haplochromine cichlids *Astatotilapia* sp. "ruaha red cheek" and *Astatotilapia* sp. "ruaha blue" were only present at Kidatu (Figure 4f).

4 | Discussion

Our study aimed to determine if environmental DNA metabarcoding could establish the distributions of species within a tropical freshwater species-rich fish assemblage at the catchment scale. Because we had a well-sampled reference library, we chose to use a method for taxonomic classification of reads that was based on congruence between phylogenetic placement (Barbera et al. 2019) and similarity scores, rather than exclusively using similarity scores with default percentage similarity thresholds. Thus, our approach was able to confidently assign species to taxa with very similar but distinct haplotypes, such as the haplochromine cichlids in the genus *Astatotilapia*. In this study we did not generate information on unidentified operation taxonomic units (OTUs), preferring instead to focus on taxa with well-supported identifications. The read data could, however, in future, be used in species discovery projects, with information on the distributions of OTUs being employed to inform field sampling for physical collection of specimens (e.g., Giribet et al. 2023).

In total, using a method that considers phylogenetic placements of reads, our analyses were able to detect 66 species. This represents the majority (73%) of the 91 species that we can be confident are present in the Rufiji–Ruaha system (Table S1). Of these 91 species that we are confident are present in the Rufiji–Ruaha system, 72 had 12S reference sequences available for use in this study (Tables S1), and since we detected 66 of these, then we detected the vast majority (93%) of the confirmed species present that it was possible to detect using our method from the available reference sequences. It is likely that many of the 29 species that are absent from our reference library are sources of the unassigned reads in our data. We also find it likely that six species that were present in our reference library, but absent from our eDNA reads, may be discovered with broader geographic sampling of the catchment. Thus, we suggest that with an enhanced 12S reference library, and extended coverage of our sampling into the sectors of Rufiji–Ruaha system that we did not sample (notably the Great Ruaha and Luwega subcatchments), then a higher number of species would be detected in environmental DNA samples.

4.1 | Species Richness Recovered at Sampling Locations

We investigated how sampling density affected the detection of species at Ifakara, a location where river is wide with extensive flood plains. Here, we collected 50 samples, within 2 km of one another. We found a minimum of three eDNA samples at each location was an effective method of surveying the diversity, enabling us to recover ~70% of the species present in 50 spatially proximate eDNA samples (Figure 2a). Nevertheless, this high-density sampling demonstrates that even with 50 samples, we were still recovering previously missed rare species, as is expected when sampling species-rich natural communities (Magurran and Henderson 2003). Thus, our lower-density sampling elsewhere in the catchment is likely to have overlooked species that were present in the reference library. Based on the recovered pattern from the species accumulation analysis, we suggest that by using more replicate samples from individual sites we would have encountered more species. There is also possibility that larger concentrations of analyzed eDNA filtered from larger water volumes may have resulted in more species detections; however, we did not find any clear evidence that our sampling volumes predicted either the total abundance of assigned reads or number of reads. Similarly, other factors have been implicated in determining the number of species encountered in eDNA metabarcoding runs, including the choice of amplification primers (Zhang, Zhao, and Yao 2020) and the number of PCR replicates (Shirazi, Meyer, and Shapiro 2021). Moreover, sequencing read depth is a key factor influencing the number of species detected (Alberdi et al. 2018). In our study, we averaged 346,604 assigned reads for each of the 174 samples, although the number varied between 571 and 1,866,218. We may, therefore, expect samples with lower sequencing depth to detect fewer species, and this pattern was supported by our analysis.

It has been hypothesized that local biological communities are comprised of core species that are persistent and abundant, and occasional species that are less frequently sampled and with different habitat requirements (Magurran and Henderson 2003). The high-density sampling at Ifakara revealed species in each of these groups. We found a core community of species that characterize this main Kilombero River environment, including a range of commercially fished species such as the bagrid catfish *Bagrus orientalis*, tigerfish *Hydrocynus tanzaniae*, bottle-nosed mormyrid *Mormyrus longirostris*, and the schilbeid catfish *Schilbe moebiusii*, among others (Figure 3). There was, however, a rich diversity of less frequent species in this location. Species present in <20% of samples comprised fluvial catfishes typical of higher elevations (genera *Amphilius*, *Atopochilus*, *Chiloglanis*, and *Zaireichthys*; see Muñoz-Mas et al. 2019), fluvial cyprinoids (*Opsaridium loveridgii*, *Engraulicypris spinifer*, *Labeobarbus macrolepis*, and *Labeobarbus oxyrhynchus*), and the spiny eel *Mastacembelus frenatus*. It seems likely that the presence of infrequently occurring species detected by the high-density sampling of the Kilombero River at Ifakara may be partly explained by vagrant individuals but also it may be attributed to detections of eDNA that have an original source elsewhere and are transported by currents downstream to the sampling location. Notably, we also found eDNA reads belonging to Nile tilapia *Oreochromis niloticus* at Ifakara. This invasive species is present in the broader Ruaha–Rufiji catchment (Shechonge et al. 2019), and has broad habitat requirements but does not yet appear to be widespread in fisheries catches of the Kilombero (MJG, BPN, and AHS pers obs.). Therefore, this species provides an example of a species that is locally present and probably suited to local conditions, but as yet is intrinsically uncommon.

FIGURE 3 | Indicator species for each of the six geographical categories based on eDNA-based reads. Circle sizes indicate the relative indicator value statistic (IndVal) of a species for each particular geographical category. Species are ordered by the geographical category for which they have the greatest indicator value.

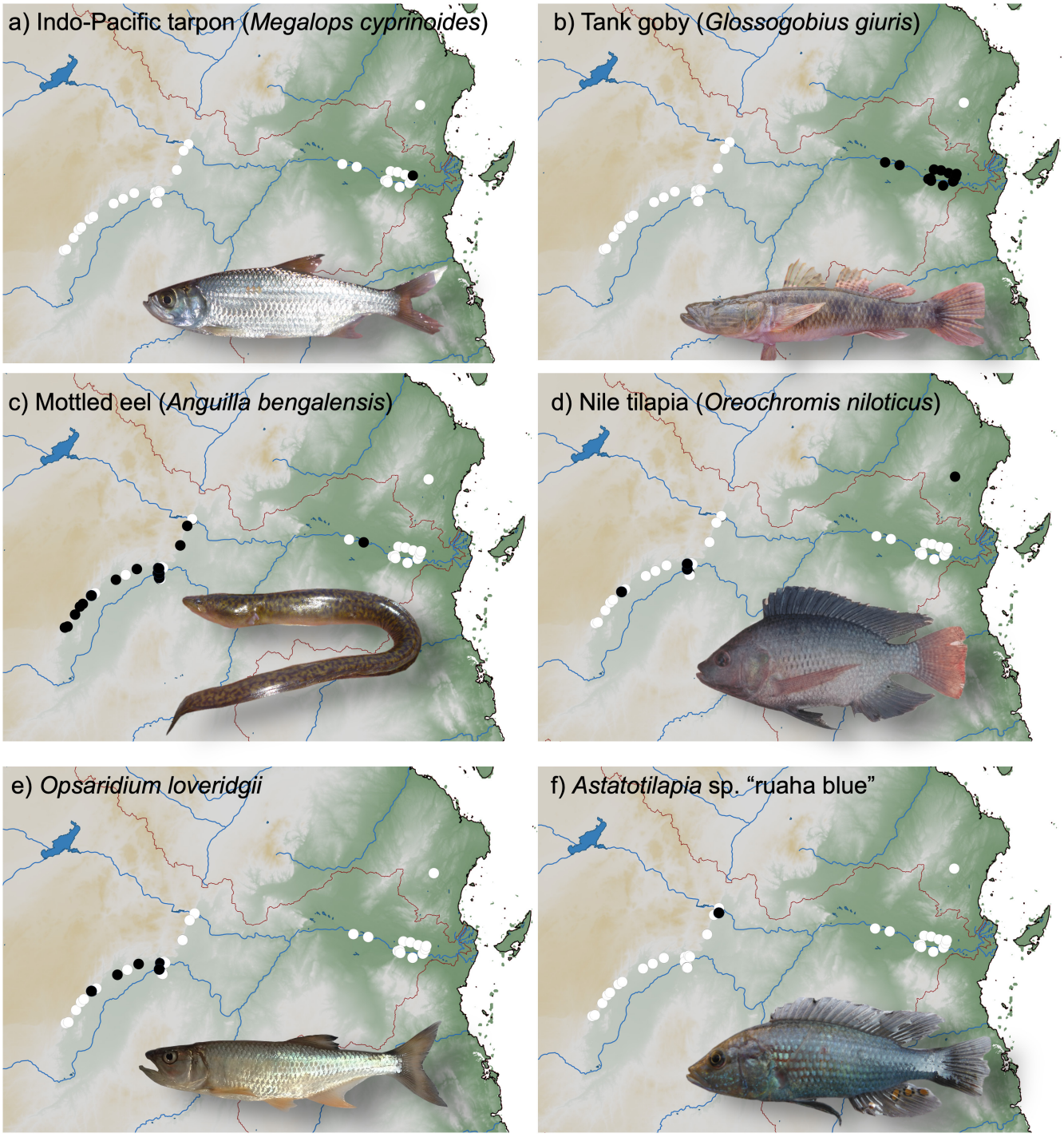

FIGURE 4 | Examples of species distribution information recovered from eDNA reads. Illustrated are presence (black circles) and absence (white circles) for (a) Indo-Pacific tarpon *Megalops cyprinoides*, (b) tank goby *Glossogobius giurus*, (c) mottled eel *Anguilla bengalensis*, (d) the endemic chedrin *Opsaridium loveridgii*, (e) the endemic haplochromine cichlid *Astatotilapia* sp. "ruaha blue," and (f) the introduced Nile tilapia *Oreochromis niloticus*.

4.2 | Factors Affecting eDNA-Based Estimates of Community Structure

A combination of observation and experimental research has helped us to understand the persistence and movement of eDNA in freshwaters. Degradation of eDNA takes place at an exponential rate in aquatic environments (Collins et al. 2019; Andruszkiewicz Allan et al. 2021), thus detection is most likely at locations closest to the source in both space and time. The detectability of eDNA in freshwaters will, however, be dependent on numerous factors, including temperature, water chemistry, flow rate, and substrate composition (Curtis et al. 2021; Joseph et al. 2022; Jo and Yamanaka 2022). In rivers, it has been shown that eDNA can be detected 150 km downstream (Pont et al. 2018), although the highest probability of detection is in much closer proximity to the source site (i.e., <50 km), and a meta-analysis

has indicated that most eDNA particles are transported in rivers <2 km from the source site (Jo and Yamanaka 2022). Moreover, there is evidence of highly localized distributions of eDNA over narrow spatial scales of <1 km (von der Heyden et al. 2023). Thus, the distribution information obtained from the eDNA is very likely to reflect relatively localized occurrence of the distribution of fish species in the region, resulting in the nonhomogeneous patterns of read distributions observed in our data.

Our study demonstrated that the main axis of community structure was between the fish communities in mountain streams, and those of the lower-elevation slow-flowing rivers and oxbow lakes. High-elevation streams typically comprised cooler temperatures with greater dissolved oxygen and lower conductivity (e.g., Kirk, Rahel, and Laughlin 2022). Characteristic species included cooler water streams which include the relatively small-bodied mountain catfishes (*Amphilius* sp., *Chiloglanis* sp.), cyprinoids (*Enteromius kerstenii* and *Opsaridium loveridgii*), and spiny eel (*Mastacembelus frenatus*). By contrast, low-elevation rivers and lakes were associated with warmer temperatures, lower dissolved temperatures, higher conductivity, and higher turbidity. Characteristic species here included catfishes (*Synodontis* sp.), a distichodontid (*Distichodus petersii*), and alestids (*Brachyalestes* sp.). This clear gradient in community structure with elevation is reflective of freshwater fish faunas globally (Jaramillo-Villa, Maldonado-Ocampo, and Escobar 2010; Kirk, Rahel, and Laughlin 2022; Askeyev et al. 2017; Yan et al. 2011), and is associated with multiple metabolic, morphological, and ecological specializations (Cerezer et al. 2023).

4.3 | Geographic Distributions of Exemplar Species

Our results illustrated the ability of eDNA metabarcoding to provide an indication of the distribution of rarer species in our study system. For example, *Opsaridium loveridgii* is a species that we only recorded from the Kilombero system, which matches the collection locality of the type specimen (labeled "Mpanganye, Rufigi," and presumed to be Mpanga, within the upper reaches of the Kilombero system; Loveridge 1960). Our study implies that the stronghold of *O. loveridgii* is within the Kilombero and that the species is not distributed widely across the Rufiji system. Our results also identified the presence of Indian Ocean tarpon (*Megalops cyprinoides*) in the lower Rufiji, a species in which the juveniles will be found in fresh and brackish waters, but as adults typically are found in fully marine coastal or offshore waters (Shen et al. 2009). We also resolved the distribution of the amphidromous tank goby *Glossogobius giuris* in the lower Rufiji, a species that will breed in freshwaters (Skelton 2001), with larvae entering the sea, and then migrating back to freshwaters as they mature.

Importantly, our dataset contained metabarcode reads assigned to four species of *Anguilla* (*A. mossambica*, *A. bengalensis*, *A. marmorata*, and *A. bicolor*) that have catadromous life histories. While all four of these species have been identified from East Africa (Skelton 2001; Hanzen et al. 2019; Hanel et al. 2024), their patterns of riverine habitat use have been unclear. Our results suggest a tendency for *A. mossambica*, *A. bengalensis*, and *A. marmorata* to be most abundant in higher-elevation rivers and streams while being rare or absent in the lower Rufiji sector (Table S3). This matches the upstream distributions from reported catches of these species from Tanzanian rivers (Hanel et al. 2024). Therefore, these species will need to pass through the Julius Nyerere Hydropower Station, constructed across Stiegler's Gorge within the Nyerere National Park (IUCN 2019), if they are to successfully migrate upstream as elvers, and downstream as reproductive adults. By contrast, we only found *A. bicolor* in the lower Rufiji, matching the inferred lower-elevation distribution from catches (Hanel et al. 2024). We suggest that eDNA-based monitoring of these species in the Rufiji will be valuable for studying the impacts of environmental changes, as proposed for other anguillid species (e.g., Cardás et al. 2020).

4.4 | Potential Benefits of eDNA Sampling

Our results indicate that eDNA metabarcoding provides a valuable data source on which to evaluate spatial patterns of fish biodiversity in freshwater catchments. With specific regard to the Rufiji system, until now available information has been based on occurrence records from historic museum collections (Eccles 1992), ad hoc collections (e.g., Turner, Ngatunga, and Genner 2021), formal species descriptions (e.g., Schedel, Friel, and Schliewen 2014), and localized surveys (e.g., Muñoz-Mas et al. 2019). Our study has helped to bring together and enhance knowledge by mapping the likely distributions of species more broadly. However, eDNA-based approaches alone may not be sufficient to study the distributions of all species, for example, eDNA of some species may not readily amplify in metabarcoding assays. Moreover, eDNA cannot inform many attributes of interest, for example, body size distributions of fished stocks (Pikitch 2018). Therefore, eDNA-based approaches can be viewed as complementary to other sampling methods.

Key advantages of using eDNA in the context of tropical aquatic biodiversity surveys are that eDNA collections can be made relatively rapidly and safely, and surveys of water bodies with different depths and flow rates can be undertaken in a standardized manner. This overcomes some of the limitations of standard sampling methods, such as seine nets, gill nets, angling, or electrofishing, that are each strongly and intrinsically different in their selectivity and efficacy depending on habitat characteristics (e.g., Dunn and Paukert 2020; Schwanke and Hubert 2004). Although there are certain steps that can further standardize eDNA sampling protocols and simplify the analysis for users, eDNA-based metabarcoding methods are likely to be effective for future surveys of fish distributions in species-rich river systems, particularly when used alongside traditional sampling methods that can provide specimens to enhance museum collections, barcode reference libraries, and enable detailed systematic study.

4.5 | Reference Libraries of African River Fishes

We used the mitochondrial 12S ribosomal gene for our eDNA-based metabarcoding. Research has shown the marker to have

sufficient variability to enable species-level identifications across a broad range of fish taxa, while also possessing conserved flanking regions enabling reliable primer design that is generally specific to teleosts—providing high proportions of on-target reads in metabarcoding assays (e.g., Collins et al. 2019). However, there is limited availability of 12S sequences of African freshwater fishes (Marques et al. 2021), and if eDNA-based metabarcode assessments are to be used more widely for assessment of African fish biodiversity, then more extensive reference libraries for river catchments or specific geographic regions will be important. Ideally, such libraries would be based on accessible museum reference collections, and supported by genome-wide sequence data for the individuals. Until now large sequencing projects focused on African freshwater fishes have concentrated on charismatic species groups such as the cichlid fish radiations of the rift lakes (Malinsky et al. 2018; Ronco et al. 2021; Meier et al. 2023), or species that represent major food resources in capture fisheries and aquaculture, such as tilapiine cichlids (Ciezarek et al. 2022, 2024). There is scope to expand genome-sequencing efforts to species groups from rivers or smaller lakes with value for conservation or locally important fisheries. In addition to the broad benefits of generating genome sequences for future evolution-focused research, this information can specifically enable assessments of spatial and temporal patterns of species diversity using eDNA-based metabarcoding approaches by (1) providing confidence in the identity of reference specimens from a specific location—particularly if they belong to a cryptic species complex (Sawasawa et al. 2024), and (2) allowing for the characterization of novel metabarcoding markers for both interspecific and intraspecific studies (e.g., Liu et al. 2024).

4.6 | Further Development of eDNA-Based Methods for Surveys of African Riverine Diversity

Our study represents one of the first metabarcoding studies of the fish communities of African freshwaters. Although limited by the absence of a complete reference library, it has been able to successfully identify the spatial distributions of key fish species. There are, however, multiple questions to resolve, including the extent to which our eDNA analyses are affected by shared or similar mtDNA haplotypes among species, and whether the methods could distinguish species that were not well represented in our reference library (most notably *Lacustricola* poeciliids and *Nothobranchius* tooth carps). There is also considerable potential to use eDNA metabarcoding as a method for species discovery, mining sets of unassigned reads to identify potentially new fish biodiversity, which could then be located using targeted sampling with established net- or trap-based survey methods. In summary, these results demonstrate that eDNA methods can provide valuable community-wide information on the distributions of freshwater fish species, including those of conservation concern that may be threatened by infrastructural development.

Author Contributions

Conceptualization: A.H.S. and M.J.G.; funding acquisition: A.H.S. and M.J.G.; investigation: A.H.S., R.A.C., S.W., A.D.S., A.M.S., P.M., G.F.T., M.A.K., B.P.N., and M.J.G.; methodology: R.A.C.; formal analysis: R.A.C. and M.J.G.; visualization: M.J.G.; writing – original draft preparation: M.J.G., A.H.S., and R.A.C.; and writing – review and editing M.J.G., R.A.C., A.H.S., A.M.S., and G.F.T.

Acknowledgments

The work was funded by a Royal Society-Leverhulme Trust Africa Postdoctoral Fellowship to AS and MJG (LAF\R1\180000), Royal Society-Leverhulme Trust Africa Awards to MJG, BPN, and GFT (AA100023 and AA130107), and a UKRI GCRF award that supported MJG and RAC (EP/T015462/1). We thank the Tanzania Commission for Research and Technology (COSTECH) for fieldwork approval and permits, the Ministry of Fisheries and Livestock for export permission, and staff of the Tanzania Fisheries Research Institute for contributions to fieldwork.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The raw metabarcoding data area are accessioned at Sequence Read Archive Bioproject PRJNA1155635. The eDNA sample metadata, the eDNA species read assignment matrices, and the eDNA analysis code are available via GitHub (https://github.com/genner-lab/rufiji-kilombero-biodiversity, https://zenodo.org/doi/10.5281/zenodo.13622930). Barcode 12S and COI sequences for reference specimens have GenBank accession numbers PQ268652 – PQ268849.

References

Alberdi, A., O. Aizpurua, M. T. P. Gilbert, and K. Bohmann. 2018. "Scrutinizing Key Steps for Reliable Metabarcoding of Environmental Samples." *Methods in Ecology and Evolution* 9: 134–147.

Andruszkiewicz Allan, E., W. G. Zhang, A. C. Lavery, and A. F. Govindarajan. 2021. "Environmental DNA Shedding and Decay Rates From Diverse Animal Forms and Thermal Regimes." *Environmental DNA* 3: 492–514.

Arthington, A. H., N. K. Dulvy, W. Gladstone, and I. J. Winfield. 2016. "Fish Conservation in Freshwater and Marine Realms: Status, Threats and Management." *Aquatic Conservation: Marine and Freshwater Ecosystems* 26: 838–857.

Askeyev, A., O. Askeyev, N. Yanybaev, et al. 2017. "River Fish Assemblages Along an Elevation Gradient in the Eastern Extremity of Europe." *Environmental Biology of Fishes* 100: 585–596.

Barbarossa, V., J. Bosmans, N. Wanders, et al. 2021. "Threats of Global Warming to the World's Freshwater Fishes." *Nature Communications* 12: 1701.

Barbera, P., A. M. Kozlov, L. Czech, et al. 2019. "EPA-Ng: Massively Parallel Evolutionary Placement of Genetic Sequences." *Systematic Biology* 68: 365–369.

Bourret, A., C. Nozères, E. Parent, and G. J. Parent. 2023. "Maximizing the Reliability and the Number of Species Assignments in Metabarcoding Studies Using a Curated Regional Library and a Public Repository." *Metabarcoding and Metagenomics* 7: e98539.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. "DADA2: High-Resolution Sample Inference From Illumina Amplicon Data." *Nature Methods* 13: 581–583.

Cardás, J. B., D. Deconinck, I. Márquez, P. P. Torre, E. Garcia-Vazquez, and G. Machado-Schiaffino. 2020. "New eDNA Based Tool Applied to the Specific Detection and Monitoring of the Endangered European Eel." *Biological Conservation* 250: 108750.

Cerezer, F. O., C. S. Dambros, M. T. Coelho, et al. 2023. "Accelerated Body Size Evolution in Upland Environments Is Correlated With Recent Speciation in South American Freshwater Fishes." *Nature Communications* 14: 6070.

Ciezarek, A., A. G. Ford, G. J. Etherington, et al. 2022. "Whole Genome Resequencing Data Enables a Targeted SNP Panel for Conservation and Aquaculture of *Oreochromis* Cichlid Fishes." *Aquaculture* 548: 737637.

Ciezarek, A. G., T. K. Mehta, A. Man, et al. 2024. "Ancient and Ongoing Hybridization in the *Oreochromis* Cichlid Fishes." *Molecular Biology and Evoloution* 41: msae116.

Cilleros, K., A. Valentini, L. Allard, et al. 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.

Collins, R. A., J. Bakker, O. S. Wangensteen, et al. 2019. "Non-Specific Amplification Compromises Environmental DNA Metabarcoding With COI." *Methods in Ecology and Evolution* 10: 1985–2001.

Collins, R. A., G. Trauzzi, K. M. Maltby, et al. 2021. "Meta-Fish-Lib: A Generalised, Dynamic DNA Reference Library Pipeline for Metabarcoding of Fishes." *Journal of Fish Biology* 99: 1446–1454.

Curtis, A. N., J. S. Tiemann, S. A. Douglass, M. A. Davis, and E. R. Larson. 2021. "High Stream Flows Dilute Environmental DNA (eDNA) Concentrations and Reduce Detectability." *Diversity and Distributions* 27: 1918–1931.

Darwall, W., K. Smith, T. Lowe, and J. C. Vié. 2005. "The Status and Distribution of Freshwater Biodiversity in Eastern Africa. IUCN SSC Freshwater Biodiversity Assessment Programme." IUCN, Gland, Switzerland and Cambridge, UK.

De Cáceres, M., and P. Legendre. 2009. "Associations Between Species and Groups of Sites: Indices and Statistical Inference." *Ecology* 90: 3566–3574.

Deiner, K., H. M. Bik, E. Mächler, et al. 2017. "Environmental DNA Metabarcoding: Transforming How We Survey Animal and Plant Communities." *Molecular Ecology* 26: 5872–5895.

Doble, C. J., H. Hipperson, W. Salzburger, et al. 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.

Dunn, C. G., and C. P. Paukert. 2020. "A Flexible Survey Design for Monitoring Spatiotemporal Fish Richness in Nonwadeable Rivers: Optimizing Efficiency by Integrating Gears." *Canadian Journal of Fisheries and Aquatic Sciences* 77: 978–990.

Eccles, D. H. 1992. *Field Guide to the Freshwater Fishes of Tanzania*. FAO Species Identification Sheets for Fishery Purposes. Rome, Italy: FAO.

Edgar, R. C. 2016. "SINTAX: A Simple Non-Bayesian Taxonomy Classifier for 16S and ITS Sequences." *bioRxiv*: 074161. https://doi.org/10.1101/074161.

Esselman, P. C., and J. D. Allan. 2010. "Relative Influences of Catchment- and Reach-Scale Abiotic Factors on Freshwater Fish Communities in Rivers of Northeastern Mesoamerica." *Ecology of Freshwater Fish* 19: 439–454.

Freitas, T. M. D. S., J. Stropp, B. B. Calegari, et al. 2021. "Quantifying Shortfalls in the Knowledge on Neotropical Auchenipteridae Fishes." *Fish and Fisheries* 22: 87–104.

Giribet, G., O. S. Wangensteen, S. Garcés-Pastor, P. R. Møller, and K. Worsaae. 2023. "Using eDNA to Find Micrognathozoa." *Current Biology* 33: R756–R757.

Hanel, R., L. Gaspare, C. Mwakosya, et al. 2024. "Assessing the Biodiversity of Eels of Tanzania—Promoting Sustainable Fisheries Through Environmental Monitoring and Capacity Building (BIOEELS)." Project Brief Thünen Inst 2024/02a. Bremerhaven, Germany: Thünen Institute of Fisheries Ecology. https://literatur.thuenen.de/digbib_extern/dn067508.pdf.

Hanzen, C., O. Weyl, M. Lucas, K. Brink, C. Downs, and G. O'Brien. 2019. "Distribution, Ecology and Status of Anguillid Eels in East Africa and the Western Indian Ocean." In *Eels—Biology, Monitoring, Management, Culture and Exploitation. Proceedings of the First International Eel Science Symposium*, edited by A. Don and P. Coulson, 33–58. Sheffield, UK: 5M Publishing.

Harrell, F., Jr. 2024. "Hmisc: Harrell Miscellaneous." R Package Version 5.1-3. https://CRAN.R-project.org/package=Hmisc.

Hortal, J., F. de Bello, J. A. F. Diniz-Filho, T. M. Lewinsohn, J. M. Lobo, and R. J. Ladle. 2015. "Seven Shortfalls That Beset Large-Scale Knowledge of Biodiversity." *Annual Review of Ecology, Evolution, and Systematics* 46: 523–549.

Hubert, N., R. Hanner, E. Holm, et al. 2008. "Identifying Canadian Freshwater Fishes Through DNA Barcodes." *PLoS One* 3: e2490.

Hugueny, B. 1989. "West African Rivers as Biogeographic Islands: Species Richness of Fish Communities." *Oecologia* 79: 236–243.

Ibanez, C., T. Oberdorff, G. Teugels, et al. 2007. "Fish Assemblages Structure and Function Along Environmental Gradients in Rivers of Gabon (Africa)." *Ecology of Freshwater Fish* 16: 315–334.

IUCN. 2019. *Technical Review of the Environmental Impact Assessment for the Rufiji Hydropower Project in Selous Game Reserve, Tanzania*. Gland, Switzerland: IUCN.

Jaramillo-Villa, U., J. A. Maldonado-Ocampo, and F. Escobar. 2010. "Altitudinal Variation in Fish Assemblage Diversity in Streams of the Central Andes of Colombia." *Journal of Fish Biology* 76: 2401–2417.

Jézéquel, C., P. A. Tedesco, R. Bigorne, et al. 2020. "A Database of Freshwater Fish Species of the Amazon Basin." *Scientific Data* 7: 96.

Jo, T., and H. Yamanaka. 2022. "Meta-Analyses of Environmental DNA Downstream Transport and Deposition in Relation to Hydrogeography in Riverine Environments." *Freshwater Biology* 67: 1333–1343.

Joseph, C., M. E. Faiq, Z. Li, and G. Chen. 2022. "Persistence and Degradation Dynamics of eDNA Affected by Environmental Factors in Aquatic Ecosystems." *Hydrobiologia* 849: 4119–4133.

Kirk, M. A., F. J. Rahel, and D. C. Laughlin. 2022. "Environmental Filters of Freshwater Fish Community Assembly Along Elevation and Latitudinal Gradients." *Global Ecology and Biogeography* 31: 470–485.

Knebelsberger, T., M. Landi, H. Neumann, et al. 2014. "A Reliable DNA Barcode Reference Library for the Identification of the North European Shelf Fish Fauna." *Molecular Ecology Resources* 14: 1060–1071.

Liu, Z., M. A. Kishe, N. P. Gabagambi, et al. 2024. "Nuclear Environmental DNA Resolves Fine-Scale Population Genetic Structure in an Aquatic Habitat." *IScience* 27: 108669.

Loveridge, A. 1960. "Status of New Vertebrates Described or Collected by Loveridge." *Journal of the East Africa Natural History Society* 23: 250–280.

Lundberg, J. G., M. Kottelat, G. R. Smith, M. L. Stiassny, and A. C. Gill. 2000. "So Many Fishes, So Little Time: An Overview of Recent Ichthyological Discovery in Continental Waters." *Annals of the Missouri Botanical Garden* 87: 26–62.

Magurran, A. E., and P. A. Henderson. 2003. "Explaining the Excess of Rare Species in Natural Species Abundance Distributions." *Nature* 422: 714–716.

Malinsky, M., H. Svardal, A. M. Tyers, et al. 2018. "Whole-Genome Sequences of Malawi Cichlids Reveal Multiple Radiations Interconnected by Gene Flow." *Nature Ecology & Evolution* 2: 1940–1955.

Marques, V., T. Milhau, C. Albouy, et al. 2021. "GAPeDNA: Assessing and Mapping Global Species Gaps in Genetic Databases for eDNA Metabarcoding." *Diversity and Distributions* 27: 1880–1892.

Martin, M. 2011. "Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads." *EMBnet Journal* 17: 10–12.

Meier, J. I., M. D. McGee, D. A. Marques, et al. 2023. "Cycles of Fusion and Fission Enabled Rapid Parallel Adaptive Radiations in African Cichlids." *Science* 381: eade2833.

Mondal, R., and A. Bhat. 2020. "Temporal and Environmental Drivers of Fish-Community Structure in Tropical Streams From Two Contrasting Regions in India." *PLoS One* 15: e0227354.

Muñoz-Mas, R., J. Sánchez-Hernández, F. Martínez-Capel, et al. 2019. "Microhabitat Preferences of Fish Assemblages in the Udzungwa Mountains (Eastern Africa)." *Ecology of Freshwater Fish* 28: 473–484.

Mwalyosi, R. B. 1990. "Resource Potentials of the Rufiji River Basin, Tanzania." *Ambio* 19: 16–20.

Oberdorff, T., P. A. Tedesco, B. Hugueny, et al. 2011. "Global and Regional Patterns in Riverine Fish Species Richness: A Review." *International Journal of Ecology* 12, no. 3: 967631.

Oksanen, J., G. Simpson, F. Blanchet, et al. 2022. "vegan: Community Ecology Package." R Package Version 2.6-4. https://CRAN.R-project.org/package=vegan.

Paradis, E., and K. Schliep. 2019. "Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R." *Bioinformatics* 35: 526–528.

Pikitch, E. K. 2018. "A Tool for Finding Rare Marine Species." *Science* 360: 1180–1182.

Pont, D., M. Rocle, A. Valentini, et al. 2018. "Environmental DNA Reveals Quantitative Patterns of Fish Biodiversity in Large Rivers Despite its Downstream Transportation." *Scientific Reports* 8: 10361.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*_. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ronco, F., M. Matschiner, A. Böhne, et al. 2021. "Drivers and Dynamics of a Massive Adaptive Radiation in Cichlid Fishes." *Nature* 589: 76–81.

Sawasawa, W., A. H. van Huysduynen, S. Gresham, et al. 2024. "Population Genomic Analysis Reveals Cryptic Population Structure in the Commercially Important Lake Malawi Cichlid *Copadichromis mloto*." *bioRxiv*: 2024.04.26.591058. https://doi.org/10.1101/2024.04.26.591058.

Schedel, F. D., J. P. Friel, and U. K. Schliewen. 2014. "*Haplochromis vanheusdeni* a New Haplochromine Cichlid Species From the Great Ruaha River Drainage, Rufiji Basin, Tanzania." *Spixiana* 37: 135–149.

Schedel, F. D., C. E. Huyghe, N. Boileau, et al. 2024. "Surrounded by Great Lakes: Assessing the Ichthyological Diversity of Northern Zambian Drainage Systems." *Journal of Great Lakes Research* 50: 102245.

Schmidt, R. C., H. L. Bart Jr., and W. D. Nyingi. 2017. "Multi-Locus Phylogeny Reveals Instances of Mitochondrial Introgression and Unrecognized Diversity in Kenyan Barbs (Cypriniae: Smiliogastrini)." *Molecular Phylogenetics and Evolution* 111: 35–43.

Schwanke, C. J., and W. A. Hubert. 2004. "Evaluation of Three Gears for Sampling Spawning Populations of Rainbow Trout in a Large Alaskan River." *North American Journal of Fisheries Management* 24: 1078–1082.

Seeteram, N. A., P. T. Hyera, L. T. Kaaya, M. C. Lalika, and E. P. Anderson. 2019. "Conserving Rivers and Their Biodiversity in Tanzania." *Watermark* 11: 2612.

Shechonge, A., B. P. Ngatunga, S. J. Bradbeer, et al. 2019. "Widespread Colonisation of Tanzanian Catchments by Introduced *Oreochromis* Tilapia Fishes: The Legacy From Decades of Deliberate Introduction." *Hydrobiologia* 832: 235–253.

Shen, K. N., C. W. Chang, Y. Iizuka, and W. N. Tzeng. 2009. "Facultative Habitat Selection in Pacific Tarpon *Megalops cyprinoides* as Revealed by Otolith Sr:Ca Ratios." *Marine Ecology Progress Series* 387: 255–263.

Shirazi, S., R. S. Meyer, and B. Shapiro. 2021. "Revisiting the Effect of PCR Replication and Sequencing Depth on Biodiversity Metrics in Environmental DNA Metabarcoding." *Ecology and Evolution* 11: 15766–15779.

Skelton, P. H. 2001. *A Complete Guide to the Freshwater Fishes of Southern Africa*. 2nd ed. Cape Town, South Africa: Struik.

Spens, J., A. R. Evans, D. Halfmaerten, et al. 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.

Taberlet, P., A. Bonin, L. Zinger, and E. Coissac. 2018. *Environmental DNA: For Biodiversity Research and Monitoring*. Oxford, UK: Oxford University Press.

Torgo, L. 2016. *Data Mining With R, Learning With Case Studies*. 2nd ed. Boca Raton, FL: Chapman and Hall/CRC. http://ltorgo.github.io/DMwR2.

Turner, G. F., B. P. Ngatunga, and M. J. Genner. 2021. "*Astatotilapia* Species (Teleostei, Cichlidae) From Malawi, Mozambique and Tanzania, Excluding the Basin of Lake Victoria." *EcoEvoRxiv*. https://doi.org/10.32942/osf.io/eu6rx.

Valdez-Moreno, M., N. V. Ivanova, M. Elias-Gutierrez, S. L. Pedersen, K. Bessonov, and P. D. Hebert. 2019. "Using eDNA to Biomonitor the Fish Community in a Tropical Oligotrophic Lake." *PLoS One* 14: e0215505.

von der Heyden, S. 2023. "Environmental DNA Surveys of African Biodiversity: State of Knowledge, Challenges, and Opportunities." *Environmental DNA* 5: 12–17.

von der Heyden, S., G. Neef, T. Grevesse, et al. 2023. "Environmental DNA Biomonitoring in Biodiversity Hotspots: A Case Study of Fishes of the Okavango Delta." *Environmental DNA* 5: 1720–1731.

Yan, Y., X. Xiang, L. Chu, Y. Zhan, and C. Fu. 2011. "Influences of Local Habitat and Stream Spatial Position on Fish Assemblages in a Dammed Watershed, the Qingyi Stream, China." *Ecology of Freshwater Fish* 20: 199–208.

Yao, M., S. Zhang, Q. Lu, et al. 2022. "Fishing for Fish Environmental DNA: Ecological Applications, Methodological Considerations, Surveying Designs, and Ways Forward." *Molecular Ecology* 31: 5132–5164.

Zhang, S., J. Zhao, and M. Yao. 2020. "A Comprehensive and Comparative Evaluation of Primers for Metabarcoding eDNA From Fish." *Methods in Ecology and Evolution* 11: 1609–1625.

Supporting Information

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