Received: 1 March 2021 | Revised: 15 June 2021 | Accepted: 30 June 2021

DOI: 10.1002/edn3.238

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

WILEY

SPECIAL ISSUE ARTICLE

Novel insights into marine fish biodiversity across a pronounced environmental gradient using replicated environmental DNA analyses

Molly Victoria Czachur[1] | Mathew Seymour[2] | Simon Creer[3] | Sophie von der Heyden[1]

[1]Evolutionary Genomics Group, Department of Botany and Zoology, University of Stellenbosch, Matieland, South Africa

[2]Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden

[3]MEFGL, School of Natural Sciences, Bangor University, Bangor, UK

Correspondence
Molly Victoria Czachur, Evolutionary Genomics Group, Department of Botany and Zoology, University of Stellenbosch, Private Bag X1, 7602, Matieland, South Africa.
Email: zoologymolly@gmail.com

Present address
Molly Victoria Czachur, Nature Metrics Ltd, Egham, UK

Funding information
National Research Foundation, Grant/Award Number: 105842; Royal Society, Grant/Award Number: NI70129

Abstract

Safeguarding marine ecosystems is essential for maintaining ecosystem function and biodiversity, but effective monitoring of marine habitats can be logistically challenging, costly, and difficult to regularly implement. Environmental DNA-based biomonitoring is a rapidly growing tool that is non-destructive, cost-effective, and reliable. However, discrepancies in eDNA sampling protocols and methodology persist, which can greatly impact the interpretations of biomonitoring results, particularly across highly diverse ecosystems with historically elevated biodiversity. The South African coastal system is a unique and highly diverse ecoregion consisting of two ocean boundary currents creating one of the most diverse biological regions on the planet. Here, we present the first eDNA-based metabarcoding assessment of South African coastal fishes while also providing key management insights into study and sample design. We observed strong ecological associations with fish species richness across the extent of the South African coast, along with weaker associations with seasonality. We detected 466 operational taxonomic units across 112 of the 270 families described previously from the region, with greater species richness on the eastern subtropical coast compared to the western coast, which follows expected species richness patterns. Additionally, we provide evidence that biological replication is necessary to detect intra-site fish diversity and that three biological replicates are sufficient for capturing species diversity dynamics. Our work highlights the value of eDNA biomonitoring across space and time enabling biodiversity characterizations for the management of a gradient of coastal marine environments.

KEYWORDS
biological replication, biomonitoring, environmental DNA, metabarcoding

1 | INTRODUCTION

Coastal ecosystems provide an array of vital ecosystem services, and it is well established that biodiversity is essential for healthy marine and coastal systems. However, anthropogenic pressures, including climate change, constantly threaten natural systems globally, requiring rapid and reliable means to actively monitor marine systems and to respond to anthropogenic pressures (Gilbey et al., 2021; Häder et al., 2020; Levin et al., 2014). One of the largest obstacles in the conservation and sustainable utilization of marine species and their environments is a lack of data on spatial biodiversity patterns that could be used to support management actions (Canonico et al., 2019; Underwood et al., 2018). The lack of spatially resolved marine biodiversity data is due in part to the high logistic cost of current biodiversity monitoring approaches, which are usually expensive, often miss small, cryptic, or rare species, require special permits, are time-consuming, can be destructive, and rely on expert taxonomic identification that is not always available (Kumar et al., 2020; Miya et al., 2020; Shu et al., 2020). Distribution data for fishes, which are among the most well-studied marine species, are often based on catch data from commercial fishing practices or ecological studies, which can be biased toward certain species and highly dependent on sampling strategy and timing (Boenish et al., 2020; Pope et al., 2010). As such, traditional methods can limit biodiversity surveys both temporally and spatially, where data cannot be systematically collected and compared (McElroy et al., 2020; Seymour et al., 2021).

Environmental DNA (eDNA) metabarcoding is gaining momentum for fish surveys globally to overcome some of the traditional challenges of biomonitoring. Fishes have been assessed in a large body of eDNA-based research for validating eDNA methodology (Miya et al., 2015; Valentini et al., 2016), developing quantitative methods (Hänfling et al., 2016), and for fisheries management (Hansen et al., 2018; Jerde, 2019; Miya et al., 2020). The potential applications of eDNA metabarcoding are well documented and have been applied successfully in freshwater, estuarine, and marine environments, ranging from single species detection to the characterization of whole communities (Creer et al., 2016; Deiner et al., 2017; Garlapati et al., 2019; Hansen et al., 2018; Miya et al., 2015; Ruppert et al., 2019; Valentini et al., 2016), and across various spatio-temporal scales (Miya et al., 2020; Seymour et al., 2021; Sigsgaard et al., 2017). Fish eDNA studies have described species distributions (Lacoursière-Roussel et al., 2018), the distribution of fishes vertically in the water column (Littlefair et al. 2020), and for detecting invasive species (Adrian-Kalchhauser & Burkhardt-Holm, 2016). The detectability of fish populations in some settings has been shown to be more sensitive and reliable using eDNA-based biomonitoring (Hänfling et al., 2016; Miya et al., 2015); however, gaps in global assessment of marine eDNA still persist along with a more general understanding of the sensitivity of eDNA sampling methodology.

One of the major considerations when collecting eDNA is the number of biological replicate samples collected at each sampling point (Goldberg et al., 2016). Replication is clearly important within the context of improving observed levels of biodiversity across eDNA studies (Deiner et al., 2017; Minamoto et al., 2016; Yamamoto et al., 2017). What remains unclear is the level of replication needed to ensure adequate sampling of the biological community, which may be study-specific (Bessey et al., 2020; Koziol et al., 2018), particularly for systems where diversity may have strong spatio-temporal patterns. There is a cost trade-off associated with the number of replicates chosen (higher assessment cost per site) to adequately assess the biodiversity of a given site (Smart et al., 2016). Effective biomonitoring assessment relies on maximizing spatial and temporal sampling, and funding is generally limited. Increasing eDNA sampling replication per site may become less cost-effective compared to increasing the number of sampling locations or time points. Taking these trade-offs into account, it is therefore appropriate to optimize the eDNA sampling protocol with a replication level that is appropriate for the project goals, whether for detecting a single species or for diversity estimates. For example, for detecting subarctic deep-sea fishes in Greenland, Thomsen et al. (2016) found that metabarcoding data from a single 2 L water sample sufficiently represented known fish catch data from trawling. Study areas with greater biodiversity and environmental complexity, however, may require different levels of biological replication for efficient eDNA-based biomonitoring.

South Africa lies at the confluence of two major boundary currents and ocean basins, the Benguela Current in the Atlantic and the Agulhas Current in the Indian Ocean. This creates a strong gradient of environmental variation, including temperature (Smit et al., 2013) and productivity, shaping a unique assemblage of coastal biodiversity, with high levels of endemism (Griffiths et al., 2010) and differing spatial levels of species diversity and phylogeographic structure (Teske et al., 2011). There have been extensive concerted efforts to characterize the biodiversity of the region, with over 13,000 species recorded to date (Griffiths et al., 2010). It is likely that many species remain unaccounted for, and estimates suggest that endemic fish biodiversity is vastly underestimated, which likely applies to many other taxonomic groups (von der Heyden, 2011). Further, the spatio-temporal distribution of the majority of species is poorly known, which is problematic given that evidence suggests contemporary species range shifts of coastal marine species (Bolton et al., 2012; James et al., 2008) and that this lack of knowledge may limit successful conservation and management actions. An approach based on eDNA may well provide novel insights into the spatio-temporal patterns of biodiversity and provide much-needed data to support ongoing monitoring, not only of native, but also alien and invasive species.

Environmental DNA biomonitoring has seen very limited application in coastal South Africa, and as such, there remains a need to better understand the methodological aspects of its application. The unique gradient of biodiversity, with lower species richness on the west coast, increasing along the south and east coasts (Awad et al., 2002; Griffiths et al., 2010) may require a differential approach to sample design, as it may take more intensive sampling to recover representative species richness in areas with elevated biodiversity compared to those with fewer species. In addition, fishes are some

of the best-studied taxa in South Africa, with ~2000 described species (Heemstra & Heemstra, 2004) and their distribution, particularly of key commercially exploited species, such as the Sparids, is well characterized. Therefore, the first aim of this study was to gain a practical understanding of how eDNA replicate sampling effort affected the detection of South African coastal fish diversity across different sampling areas. Secondly, this study aimed to provide a preliminary eDNA biodiversity assessment of fishes across the marine biogeographic gradient of South Africa.

2 | MATERIALS AND METHODS

2.1 | Data collection and experimental design

Two eDNA datasets were independently collected and processed for (1) a biological replication experiment, where seven biological replicates were collected from three different sites in the Western Cape of South Africa, and from one exhibit in the Two Oceans Aquarium, Cape Town, and (2) for an extended baseline survey where triplicate samples were collected from seven natural sites across the South African biogeographical gradient (Figure 1).

The biological replication experiment included sample collection in an aquarium tank and three natural sites. Seven biological replicate water samples were collected from the Two Oceans aquarium, from a saltwater tank maintained at ambient temperature, salinity, and pH, and exposed to a natural photoperiod as experienced by the captive species in the wild. No UV light treatment is used in the tank. For the natural sites, seven biological replicate water samples were collected from three coastal sites ($n = 28$) in South Africa (Figure 1). The natural sites were selected for representing three different coastal types and included Langebaan, a sandy site in close proximity to seagrass, Jacobsbaai, a rocky, sheltered, kelp-dominated shore, and Cape Agulhas, an exposed rocky shore with low algal cover.

For the large-scale baselining experiment, independently of the previously mentioned samples, three biological replicate samples were collected from each of seven coastal sites in South Africa, spanning ~2400km and a strong environmental gradient (Figure 1). All sites were sampled once in October 2017 ($n = 21$) and once in May 2018 ($n = 21$), resulting in 42 samples collected for this experiment.

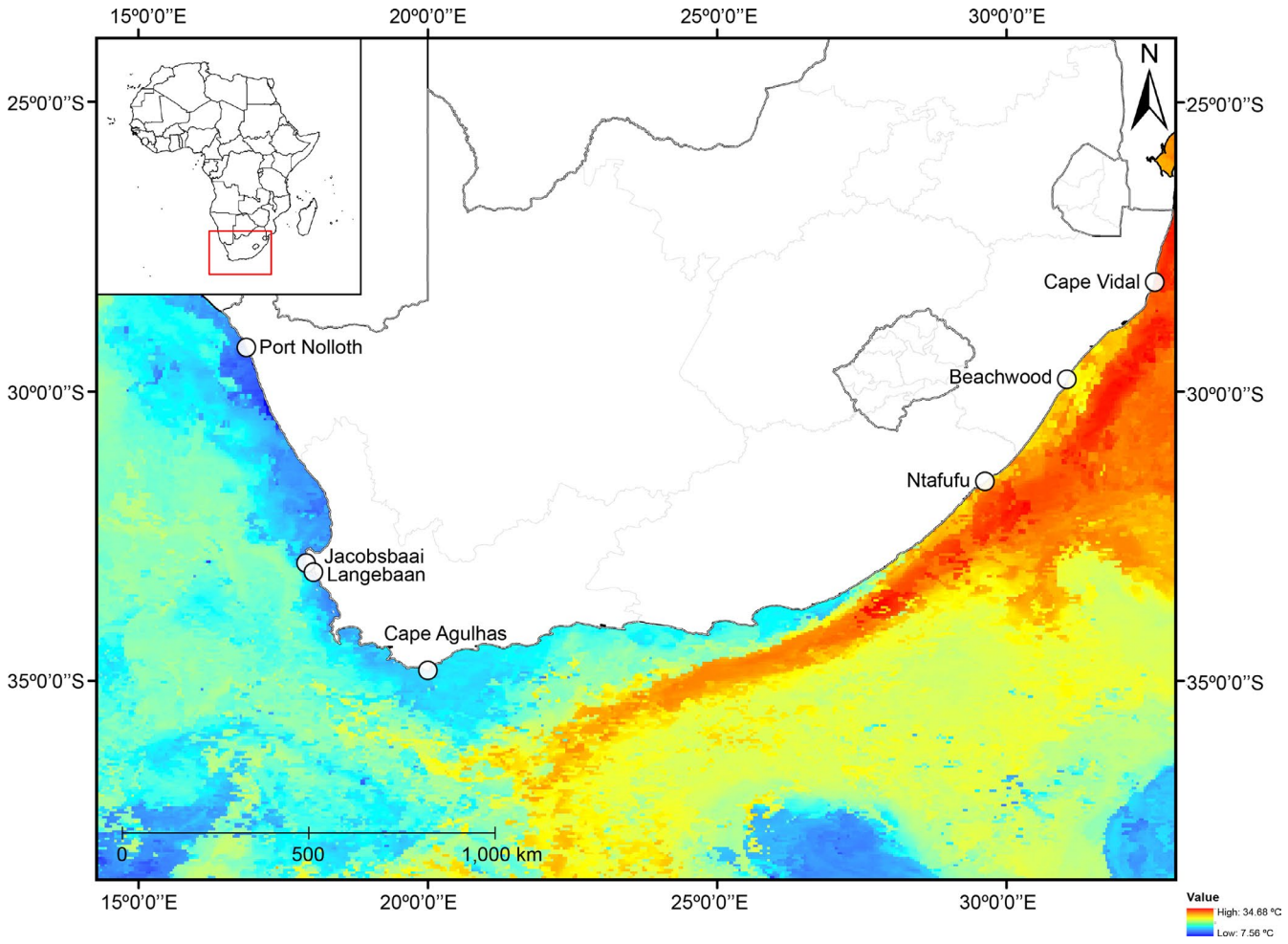

FIGURE 1 Field sampling region in South Africa across South Africa's coastal environmental variation. White circles indicate sampling locations. Jacobsbaai, Langebaan, and Cape Agulhas were sampled at an additional sampling event to collect samples for the replication experiment. Sea surface temperature (SST) data source: NASA OBPG (2020)

2.2 | Water collection and filtering

For each biological replicate, a 2 L water sample was collected from the surface water (within 1 m of the water surface) of the tank or natural site using a 2-L container with a wide opening. Once ashore, the water sample was immediately drawn into a sterile 50-ml luer lock syringe, which was then connected to a Sterivex™ 0.22 μm filter, and the water repeatedly pushed through the filter unit in 50 ml subsamples. This was repeated for every biological replicate, using one filter unit per 2 L sample, except where the sample could not be passed through the filter (e.g., in highly sedimented samples). For the latter, once the filter reached its capacity, additional Sterivex filter units were used for samples where <250 ml was achieved, or/and if >250 ml was achieved, the total volume was recorded, and the filtering stage considered complete. After filtering each sample, an air-filled syringe was used to empty the remaining seawater through the filter outlet. Each filter unit was subsequently filled with 2 ml of a tissue lysis buffer (ATL, Qiagen) using a sterile 2–ml syringe, and then, the filter unit was sealed with Helapet Combi-Caps at both ends and stored at room temperature until DNA extraction as soon as possible, or within 5 months. Regular negative controls were prepared by processing 1 L of sterile distilled water through a Sterivex filter unit, and otherwise treating it equally to a seawater sample for the rest of the workflow.

2.3 | DNA extraction

All pre-PCR laboratory procedures were carried out in a disinfected, PCR-free area that was thoroughly cleaned using 10% sodium hypochlorite and exposure to UV light for 30 min before and after PCR setup. DNA was extracted from the filter units using the DNeasy Blood and Tissue kit (Qiagen) with a modified protocol, involving adding 70 μl proteinase K directly to the Sterivex filter unit and incubating at 56°C overnight before removing the sample from the filter unit and into Eppendorf tubes. Once the sample was in Eppendorf tubes, the standard spin-column extraction protocol was performed, and the final elution volume was 50 μl. Each extraction round contained blank controls of sterile, UV-treated Milli-Q water to monitor potential contamination from the laboratory setup. Post-extraction, all DNA extracts were stored at −20°C and were thawed and spun in a centrifuge before being used in the PCR-based library preparations.

2.4 | PCR-based library preparations

A two-step PCR-based library preparation protocol was performed after initially optimizing the PCR methods, based on the two-step adapter system outlined by Miya et al., 2015 and others (e.g., Brennan et al., 2019; Holman et al., 2021; Seymour et al., 2021). In summary, each biological replicate was prepared as a single library; thus, three libraries resulted from each site for each sampling event. The DNA templates amplified in this study included DNA extracted from Sterivex filter units (including samples and negative field controls), DNA extracted from fish tissue (pooled into a mock community and treated as a positive control), and distilled water (negative control). The mock community contained DNA extracts of eleven fish species that were normalized to 10 ng/ml and pooled. All first-round PCRs were performed in 25 μl reaction volume that contained 1 μl of the aforementioned DNA template, 0.75 μl of each 10 μM primer (MiFish-U-F, MiFish-U-R, MiFish-E-R and MiFish-E-R by Miya et al., 2015 with an added adaptor region), 12.5 μl of KAPA HiFi HotStart ReadyMix (KAPA Biosystems), 1 μl of 5 mg/ml bovine serum album (BSA) solution, and the remaining volume made up with PCR-free, molecular-grade water. The PCR thermal cycle was performed with an initial denaturation at 95°C for 3 min, followed by 35 cycles of (1) denaturation at 98°C for 20 s, (2) annealing at 65°C for 15 s, and (3) extension at 72°C for 10 s, then followed by one cycle of final extension at 72°C for 5 min and cooling at 4°C for 5 min, then held at 4°C and transferred to the refrigerator until used for further stages.

PCR1 reactions were performed in triplicate and pooled after PCR1 to minimize the processing costs of the subsequent stages, after which they were purified. Purification involved bead cleaning using a two-step protocol of 0.5× ratio of MagBio HighPrep PCR beads, followed by a 0.8× ratio, which was eluted from the beads in 48 μl of AE buffer (Qiagen).

A second round of PCR was then performed to attach Illumina-associated P5/P7 index adapter ends (purchased from IDT) using the same methods as described in PCR1, except (1) the DNA template used in each reaction was the PCR1 product, (2) the number of PCR cycles was reduced to 15 cycles instead of 35 cycles, and (3) PCR2 products were bead cleaned following the same two-step MagBio bead cleaning protocol as described for PCR1, then each bead cleaned PCR product visualized by gel electrophoresis, and its concentration quantified with Qubit. All PCR2 products were normalized to 4 nM, mixed, and sequenced on an Illumina MiSeq with the 2x150bp kit at the Institute for Microbial Biotechnology and Metagenomics (IMBM) at the University of the Western Cape.

2.5 | Bioinformatics

The data were demultiplexed using the Illumina MiSeq software. Thereafter, data were processed with the most recently available version of the MiFish pipeline (version 2.40, Miya et al., 2020) at the time of analysis. MiSeq data files were added into a predefined directory and the reference database folder was populated with the MiFish ver.40 reference database, which was based on the MitoFish database (Iwasaki et al., 2013). The nomenclature in the database was not curated, so species-level taxonomic assignments were manually reviewed using taxonomic records from the World

Register of Marine Species (WoRMS) database, and nomenclature was updated where necessary. The MiFish pipeline Perl scripts were run locally via Terminal on a Mac without any external processing power.

2.6 | Data processing, statistical analysis, and data visualizations

Data were processed using R Statistics version 4.0.3 (R Core Team, 2020) and Microsoft Excel, with data converted to presence/absence. The data were analyzed at OTU level throughout as taxonomically assigned by the MiFish pipeline, including "low," "medium," and "high" assignments for the purpose of demonstrating the diversity of OTU-derived taxa recovered from the environmental samples. All OTUs were checked against available literature on fish diversity (Heemstra & Heemstra, 2004). Diversity estimates for Shannon–Wiener Diversity Index were generated using the "diversity" function in the "vegan" package in R. Diversity estimates were log transformed for normality of distribution, and data were compared between samples, between sites, and between seasons using analysis of variance (ANOVA) using the "aov" function in R. The ANOVA assumptions were checked using the "shapiro.test" function in the "dplyr" package and "leveneTest" in the "car" package for homogeneity of variances, and a diagnostic plot of residuals was used to assess the assumption of normality of distribution. Where significant differences were revealed by ANOVA, a pairwise comparison *t* test was performed between factors using the "pairwise.*t*.test" function in R. Line and bar charts were generated with the "ggplot" package and rarefaction curve with "rarefy" package, both using R Statistics. The read abundance data were based on OTU-level data and visualized by heatmaps, but presented with a reduced taxonomic resolution (family level) due to the large volume of data recovered. Mapping was performed in QGIS version 3.16 (QGIS.org, 2021).

3 | RESULTS

3.1 | Biological replication experiment

From the four sampling locations in the biological replication experiment, a total of 818,524 sequences were retrieved after bioinformatic processing, belonging to 100 operational taxonomic units (OTUs). The replication experiment detected between 10 and 37 OTUs across samples, capturing 8 to 17 families (Figure S1). For samples taken from the Two Oceans aquarium, 33 OTUs were detected from 13 families, and the mean Shannon–Wiener index value (SW) was 3.22 (±SE 0.05) (see Figure S2 for SW of the three natural sites). All OTUs were assigned to teleost fishes, with no elasmobranchs detected (Figure 2).

Species richness between sites and between replicates varied. For the seven aquarium samples, 33 OTUs were detected in total, and the average detection was 23.7 (±SE 1.20). For Jacobsbaai, a total of 32 OTUs were detected, with average species richness per replicate of 21.43 (±SE 2.14). A total of 28 OTUs were detected in Langebaan, with an average species richness of 13.57 (±SE 0.69). Cape Agulhas samples contained the highest species richness and highest variation between replicates, with 53 OTUs detected and average species richness of 31.43 (±SE 2.28). In the Two Oceans aquarium, the majority of OTUs were taxonomically assigned to species of Sparidae (sparids) and Engraulidae (anchovies), whereas in the natural sites (Jacobsbaai, Langebaan, and Cape Agulhas), the majority of OTUs were assigned to species of Blenniidae (blennies), Gobiidae (gobies), Mugilidae (mullets), and Opistognathidae (jawfishes) (Figure 2).

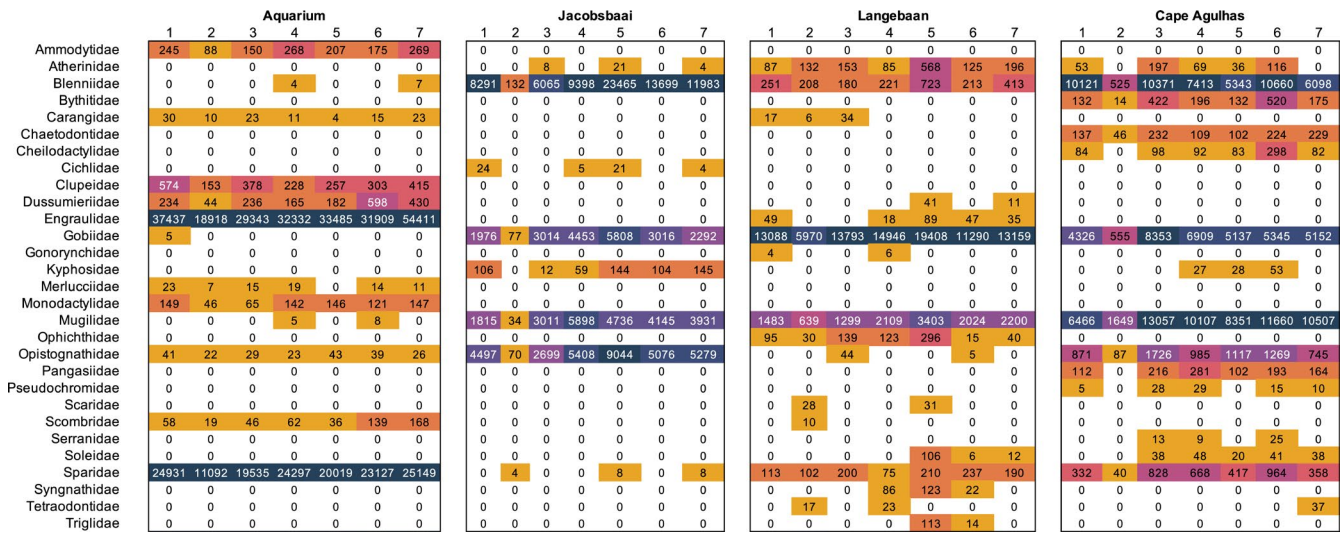

FIGURE 2 Read abundance of each taxonomic group (family) detected in the replication experiment for the aquarium and three natural sites (Jacobsbaai, Langebaan, and Cape Agulhas)

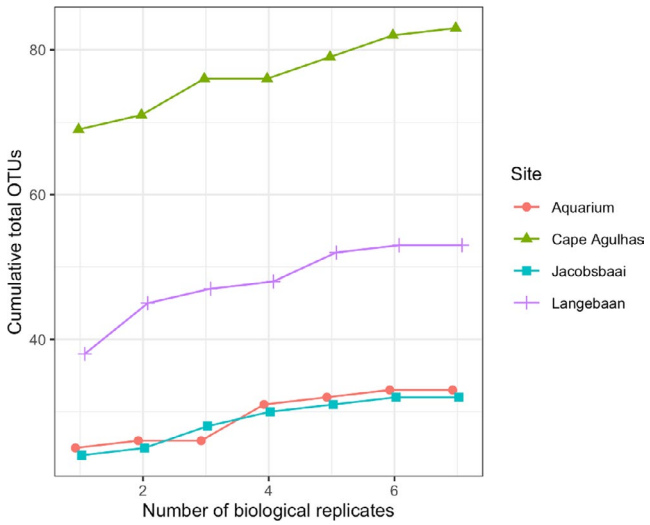

FIGURE 3 Cumulative total number of OTUs detected in the replication experiment, where the color of the line and icon indicates the site (red circles = aquarium, green triangles = Cape Agulhas, blue squares = Jacobsbaai, and purple crosses = Langebaan)

The replication experiment showed that OTUs with high read abundance were consistently recovered across all of the seven replicates, but that some taxonomic groups with lower read numbers were not detected in multiple replicates (Figure 2). There was no statistically significant difference between samples ($F(1,26) = 0.453$, $p = 0.507$) or sites ($F(1,26) = 0.201$, $p = 0.657$). The cumulative detection curves for each site suggest that, for the sites tested, there is an increase in detection of OTUs with increasing biological replicate sampling, and the majority of OTUs are detected within the first three replicates (Figure 3 and Figure S5).

Two positive controls (mock communities) detected 16 and 17 OTUs, respectively. One of the twelve blank controls contained low LOD (log of odds, assigned during the MiFish pipeline) sequences and in future, with the use of an improved reference database, these sequences should be taxonomically assigned and removed from the dataset proportionately.

3.2 | Large-scale baseline

From the seven sampling locations in the large-scale baseline, a total of 649,721 sequences were retrieved after bioinformatic processing, belonging to 449 taxonomic units. All except one eDNA sample in the large-scale baseline detected at least one target fish species, and the largest number of fish OTUs detected in a single replicate sample was 132 OTUs from Cape Vidal (Figure 4, Table S1 and Figures S3 and S4). The seasonal datasets suggest that similar species richness was recovered from sites between seasons, yet with novel detections specific to season (Figure 5). There was a strong spatial trend of increased species richness on the subtropical east coast of South Africa when compared with the temperate west and south coast field sites (Figure 5, Table S1).

4 | DISCUSSION

We detected 466 distinct OTUs across 112 of the 270 known families of fish in the first eDNA-based assessment of South African coastal fisheries and thus demonstrating the potential of eDNA in a South African coastal setting. Using a partial fragment of the 12S rRNA gene (Miya et al., 2015), we specifically targeted fishes given the relatively advanced global understanding of fish-targeted eDNA metabarcoding (e.g., see Miya et al., 2020). The diversity of OTUs captured in these datasets signifies an advancement in fish biomonitoring techniques regionally, and the data resolution trade-offs of reducing biological eDNA replication are likely outweighed by the breadth of data achieved even with lower numbers of biological replicates.

4.1 | Biological replication is necessary for accurate eDNA-based biomonitoring

Given the distinct patterns of biodiversity in South Africa, which increases consistently across taxonomic groups, from west to east (Awad et al., 2002; Griffiths et al., 2010; Turpie et al., 2000), this study set out to determine the role of biological replication (i.e., the number of samples taken per site) in detecting fish taxa. This is particularly important in the context of species detection versus cost, which increases with additional sampling and downstream DNA extraction, library preparation, and sequencing.

Our results show that biological replication within sites is essential for accurately measuring biological diversity. For sites with lower diversity, high read abundant OTUs were consistently detected across replicate samples. However, lower abundant or rarer reads were inconsistently detected across replicates, regardless of the number of replicates used, especially for high-diversity sites. There were also variations in detectability of fishes between replicates despite no significant differences in diversity indices at different replication levels. For example, increasing replication from one to two samples improved the number of families detected by an average of 7.5% across sites with an additional increase of 4.6% with three replicates, clearly showing the importance of replicate sampling for assigning OTUs to family level (Figure 3). While additional samples did increase detectability, the detection return for increased sampling was substantially lower than additional site sampling may provide. Therefore, the cost-benefit of reducing biological replication is that increased spatial or temporal sampling can be performed. To achieve wider spatial sampling, another approach is to subsample and pool water before processing. For example, Hänfling et al. (2016) collected five 400 ml subsamples from 100 m around a central sampling point and pooled into a 2 L sample for

	Port Nolloth		Jacobsbaai		Langebaan		Cape Agulhas		Ntafufu		Beachwood		Cape Vidal	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Acanthuridae	0	0	0	0	0	0	0	29	0	0	0	0	0	0
Apogonidae	0	0	0	0	9	0	1940	0	0	1318	0	0	8	357
Blenniidae	0	0	0	0	0	0	0	0	0	0	0	0	10	51
Caesionidae	0	0	0	0	0	0	0	0	0	0	5	0	0	46
Callionymidae	0	0	0	0	0	0	0	0	0	0	0	0	57	7
Carangidae	0	0	0	0	0	0	0	0	0	0	0	0	4	0
Chaetodontidae	0	0	0	54	0	0	0	0	0	5	0	0	0	15
Champsodontidae	0	0	0	0	0	0	0	0	0	4	0	0	0	0
Cheilodactylidae	0	0	0	0	0	0	0	4	0	0	0	0	0	0
Cirrhitidae	0	0	0	0	0	0	494	265	0	0	0	0	0	0
Clupeidae	0	0	0	0	0	0	0	0	73	1685	70	23	792	985
Gobiidae	0	0	0	0	0	0	0	0	5	0	0	0	20	63
Haemulidae	0	0	0	0	0	0	0	0	0	0	0	0	22	55
Labridae	0	0	5	0	63	101	0	0	0	17	0	0	47	79
Lethrinidae	0	0	0	0	0	0	0	0	0	0	0	0	25	45
Lutjanidae	0	0	0	0	0	0	0	0	0	9	86	218	138	306
Malacanthidae	0	0	0	0	0	0	0	0	0	0	0	0	8	0
Melanocetidae	0	0	0	0	0	0	205	70	0	0	0	0	0	0
Mugilidae	3826	855	0	0	0	0	198	808	0	0	0	0	0	0
Mullidae	5	119	23	17	0	0	2586	642	0	89	0	0	0	0
Muraenidae	624	830	1349	510	0	37	1056	443	0	0	0	0	0	0
Myctophidae	0	0	0	0	0	0	149	280	0	0	0	0	0	0
Ogcocephalidae	0	0	0	0	0	0	0	55	0	0	0	0	0	0
Oplegnathidae	0	0	0	0	0	0	13	0	0	0	0	0	0	0
Phosichthyidae	56	0	0	0	0	0	3452	3402	0	0	0	0	0	0
Platycephalidae	5145	8919	2707	6152	0	0	0	0	0	0	0	0	0	0
Scaridae	0	0	0	0	0	0	1250	488	22	253	0	0	267	594
Schindleriidae	1215	1375	67	270	0	0	5	0	0	0	0	0	0	0
Scombridae	0	0	0	0	0	0	0	0	11	457	0	0	0	0
Scorpaenidae	138	305	0	20	0	0	17	68	0	0	0	0	0	0
Serranidae	0	0	0	0	0	0	110	40	0	0	0	0	0	0
Sparidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Stomiidae	0	0	156	154	7	0	137	102	0	0	0	0	0	0
Tetraodontidae	3614	827	690	107	0	0	735	1945	0	0	0	0	0	0
Trichiuridae	0	0	10	0	0	0	346	738	0	0	0	0	0	0
Tripterygiidae	0	0	356	5	0	0	0	0	0	0	0	0	0	0

FIGURE 4 Read abundance (total per family) in seven coastal baselining sites for two seasons (S1 and S2)

their large-scale freshwater fish analysis, which suitably reflected known fish patterns in the region.

Here, we found that three replicates revealed fish communities that are broadly reflective of known national biodiversity trends (Turpie et al., 2000), encompassing 270 unique families. In conclusion, three biological replicates is a suitable level of biological replication for the study of large-scale South African coastal fish biodiversity. However, when expecting elevated level biodiversity, such as for invertebrate communities, it may be essential to consider additional replication to ensure sufficient assessment.

4.2 | eDNA provides novel insights into fish biodiversity patterns across a strong environmental gradient

There are 270 families of fishes known to inhabit South African waters (Heemstra & Heemstra, 2004), of which 112 were detected in this baseline study. Our results mirror known patterns of coastal species diversity by detecting a trend of increasing numbers of families and OTUs on the east coast, described for coastal fishes by Turpie et al. (2000), invertebrate communities (Awad et al., 2002), and interestingly patterns recovered for metazoan, protist, and bacterial communities using a metabarcoding approach (Holman et al., 2021). In fact, community composition differs markedly between west, south, and east coast sites (Figure 4), with fish fauna more dominated by Indo-Pacific fish fauna along the eastern coastal margins. The high diversity in families was also reflected in the high number of OTUs recovered from three replicate samples; for Cape Vidal along the subtropical/tropical boundary, 268 OTUs were detected, compared to a range of 21 to 37 OTUs on the west coast. However, Cape Agulhas, at the southernmost tip of Africa, had the largest diversity of families, with 19, perhaps as a result of the overlap of the Benguela and Agulhas Currents along the south coast, which drives a unique fish faunal composition (Turpie et al., 2000).

Many of the families detected using eDNA are well described and relatively common and include clupeids, butterflyfishes, mullets, and pufferfishes (Figures S1 and S3). Interestingly, we also identified OTUs that are rare or show diurnal movement and would otherwise not be accounted for in surveys. For example, for both seasons at Cape Agulhas, reads matching those of Devil-anglers (Melanocetidae) were recovered likely due to the presence of their

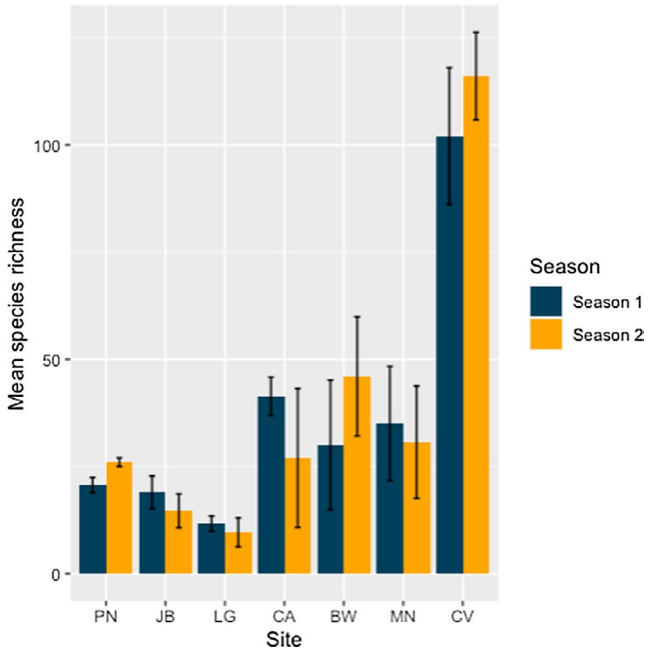

FIGURE 5 Species richness, based on species-level taxonomic assignments, of the seven coastal baseline sites in two sampling seasons (BW, Beachwood; CA, Cape Agulhas; CV, Cape Vidal; JB, Jacobsbaai; LG, Langebaan; MN, Ntafufu; PN, Port Nolloth)

larval stages, given that juvenile and adult fishes are generally in the mesopelagic zones and beyond (Heemstra & Heemstra, 2004). As the ecology of larval stages is generally poorly understood and because larvae are difficult to identify (Steinke et al., 2016), our eDNA barcoding approach provides a powerful additional tool for surveying fish diversity beyond easy to recognize adult or juvenile stages, thus providing a more complete biodiversity inventory for the region.

4.3 | A future for eDNA monitoring of coastal biodiversity in South Africa

There are several considerations when considering future eDNA studies in South Africa and beyond. Environmental DNA is essentially an indirect sampling method with compounding effects of primer selection influencing the level of detectable diversity. When taxonomic assignments were reviewed to species level, positive controls containing a mock community of 11 species of fish tissue were characterized by 17 separate OTUs. This overestimation also occurred in the Two Oceans aquarium samples, where only four families were known to be present (Clinidae, Sparidae, Haemulidae, and Monodactylidae), yet 13 families were detected, potentially due to fish feed or cross-tank contamination by divers using gear unwashed between different tanks or simply moving between tanks in close proximity themselves. This highlights the sensitivity of the eDNA approach, but such overestimating of OTUs needs to be investigated further. For example, the overestimation of OTUs in positive controls may be attributed to multiple OTUs representing single taxa, or a lack of South African reference sequences in many taxonomic groups leading to ambiguous taxonomic assignments. As such, a more thorough regional reference database review is required before assigning taxonomy to species level. There is a global challenge with lack of reference sequences for some groups, which is often a barrier for interpreting metabarcoding data, and regional-specific databases are often required for higher resolution taxonomic assignments (e.g., Gold et al., 2020 for Californian fishes).

The eDNA workflow employed in this study utilized a publicly available pipeline using available resources and generated a large amount of data with relatively low in-field effort. Overall, the results of this field experiment broadly suggest that eDNA is a valuable tool for monitoring coastal fish biodiversity in South Africa. There remain large gaps in the 12S barcode database for fishes, which need to be addressed in order to provide better resolution and power to species assignments. Additionally, a more general assessment of the MiFish primers for detecting groups with high numbers of endemic species and those of commercial interest, such as the Clindae and Sparidae respectively, which were not found in our preliminary surveys. Further, closer examination of the ability to detect elasmobranch species would be highly beneficial given that ~25% of the 185 elasmobranch species recorded from South Africa are endemic and many species are threatened through over-exploitation (Bester-van der Merwe & Gledhill, 2015; Compagno et al., 2005; Miya et al., 2015). Nevertheless, eDNA biomonitoring of coastal fish communities undoubtedly provides a novel approach with increased spatial and temporal sampling capabilities needed to more effectively manage valuable and dwindling natural resources.

ACKNOWLEDGEMENTS

The Two Oceans Aquarium contributed access to their saltwater tanks and aquarium team for collecting eDNA samples. Luke Holman, Thomas Grevesse, and Alicia Dalongeville contributed to the eDNA sample collection. The Central Analytical Facility at Stellenbosch University supported this study with Qubit services, and the Institute for Microbial Biotechnology and Metagenomics (IMBM) at the University of the Western Cape provided Illumina MiSeq sequencing services. This project was funded by the National Research Foundation, South Africa (Grant number 105842) to SvdH and partially through a Royal Society-Newton mobility grant (Grant number NI70129) awarded to SvdH and SC. MVC was supported through a NRF grantholder-linked bursary, as well as a NRF extension bursary.

CONFLICT OF INTEREST

No conflict of interest has been declared by the authors.

AUTHOR CONTRIBUTIONS

MVC, MS, SC, and SVDH contributed to the concept and design of the study and the writing of the manuscript. MVC, MS, and SVDH contributed to the acquisition, analysis, and interpretation of the data.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Molly Victoria Czachur https://orcid.org/0000-0001-5688-9353
Sophie von der Heyden https://orcid.org/0000-0001-9166-976X

REFERENCES

Adrian-Kalchhauser, I., & Burkhardt-Holm, P. (2016). An eDNA assay to monitor a globally invasive fish species from flowing freshwater. *PLoS One*, *11*(1), e0147558. https://doi.org/10.1371/journal.pone.0147558

Awad, A. A., Griffiths, C. L., & Turpie, J. K. (2002). Distribution of South African marine benthic invertebrates applied to the selection of priority conservation areas. *Diversity Distributions*, *8*(3), 129–145. https://doi.org/10.1046/j.1472-4642.2002.00132.x

Bessey, C., Jarman, S. N., Berry, O., Olsen, Y. S., Bunce, M., Simpson, T., Power, M., McLaughlin, J., Edgar, G. J., & Keesing, J. (2020). Maximizing fish detection with eDNA metabarcoding. *Environmental DNA*, *2*(4), 493–504. https://doi.org/10.1002/edn3.74

Bester-van der Merwe, A. E., & Gledhill, K. S. (2015). Molecular species identification and population genetics of chondrichthyans in South Africa: current challenges, priorities and progress. *African Zoology*, *50*(3), 205–217. https://doi.org/10.1080/15627020.2015.1063408

Boenish, R., Willard, D., Kritzer, J. P., & Reardon, K. (2020). Fisheries monitoring: perspectives from the United States. *Aquaculture and Fisheries*, *5*(3), 131–138. https://doi.org/10.1016/j.aaf.2019.10.002

Bolton, J. J., Anderson, R. J., Smit, A. J., & Rothman, M. (2012). South African kelp moving eastwards: the discobery of *Ecklonia maxima* (Osbeck) Papenfuss at De Hoop Nature Reserve on the coast of South Africa. *African Journal of Marine Science*, *34*(1), 147-151. https://doi.org/10.2989/1814232X.2012.675125

Brennan, G. L., Potter, C., de Vere, N., Griffith, G. W., Skjøth, C. A., Osborne, N. J., Wheeler, B. W., McInnes, R. N., Clewlow, Y., Barber, A., Hanlon, H. M., Hegarty, M., Jones, L., Kurganskiy, A., Rowney, F. M., Armitage, C., Adams-Groom, B., Ford, C. R., Petch, G. M., & Creer, S. (2019). Temperate airborne grass pollen defined by spatio-temporal shifts in community composition. *Nature Ecology & Evolution*, *3*(5), 750–754. https://doi.org/10.1038/s41559-019-0849-7

Canonico, G., Buttigieg, P. L., Montes, E., Muller-Karger, F. E., Stepien, C., Wright, D., Benson, A., Helmuth, B., Costello, M., Sousa-Pinto, I., Saeedi, H., Newton, J., Appeltans, W., Bednaršek, N., Bodrossy, L., Best, B. D., Brandt, A., Goodwin, K. D., Iken, K., … Murton, B. (2019). Global Observational Needs and Resources for Marine Biodiversity. *Frontiers in Marine Science*, *6*, 1. https://doi.org/10.3389/fmars.2019.00367

Compagno, L., Dando, M., & Fowler, S. (2005). *Sharks of the World (Princeton Field Guides)* (Illustrated. ed.). Princeton University Press.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., Potter, C., & Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, *7*(9), 1008–1018. https://doi.org/10.1111/2041-210x.12574

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895. https://doi.org/10.1111/mec.14350

Garlapati, D., Charankumar, B., Ramu, K., Madeswaran, P., & Ramana Murthy, M. V. (2019). A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. *Reviews in Environmental Science and Bio/Technology*, *18*(3), 389–411. https://doi.org/10.1007/s11157-019-09501-4

Gilbey, J., Carvalho, G., Castilho, R., Coscia, I., Coulson, M. W., Dahle, G., Derycke, S., Francisco, S. M., Helyar, S. J., Johansen, T., Junge, C., Layton, K. K. S., Martinsohn, J., Matejusova, I., Robalo, J. I., Rodríguez-Ezpeleta, N., Silva, G., Strammer, I., Vasemägi, A., & Volckaert, F. A. M. (2021). Life in a drop: Sampling environmental DNA for marine fishery management and ecosystem monitoring. *Marine Policy*, *124*, 104331. https://doi.org/10.1016/j.marpol.2020.10433

Gold, Z., Choi, E., Kacev, D., Frable, B., Burton, R., Goodwin, K., Thompson, A., & Barber, P. (2020). *FishCARD: Fish 12S California current specific reference database for enhanced metabarcoding efforts*. Authorea, 1. https://doi.org/10.22541/au.159136805.55528691

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., Spear, S. F., McKee, A., Oyler-McCance, S. J., Cornman, R. S., Laramie, M. B., Mahon, A. R., Lance, R. F., Pilliod, D. S., Strickler, K. M., Waits, L. P., Fremier, A. K., Takahara, T., Herder, J. E., & Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, *7*(11), 1299–1307. https://doi.org/10.1111/2041-210x.12595

Griffiths, C. L., Robinson, T. B., Lange, L., & Mead, A. (2010). Marine biodiversity in South Africa: An evaluation of current states of knowledge. *PLoS One*, *5*(8), e12008. https://doi.org/10.1371/journal.pone.0012008

Häder, D. P., Banaszak, A. T., Villafañe, V. E., Narvarte, M. A., González, R. A., & Helbling, E. W. (2020). Anthropogenic pollution of aquatic ecosystems: Emerging problems with global implications. *Science of the Total Environment*, *713*, 136586. https://doi.org/10.1016/j.scitotenv.2020.136586

Hänfling, B., Lawson Handley, L., Read, D. S., Hahn, C., Li, J., Nichols, P., Blackman, R. C., Oliver, A., & Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology*, *25*(13), 3101–3119. https://doi.org/10.1111/mec.13660

Hansen, B. K., Bekkevold, D., Clausen, L. W., & Nielsen, E. E. (2018). The sceptical optimist: Challenges and perspectives for the application of environmental DNA in marine fisheries. *Fish and Fisheries*, *19*(5), 751–768. https://doi.org/10.1111/faf.12286

Heemstra, P., & Heemstra, E. (2004). *Coastal Fishes of Southern Africa by Phil Heemstra, Elaine Heemstra (2004) Paperback*. NISC.

Holman, L. E., de Bruyn, M., Creer, S., Carvalho, G., Robidart, J., & Rius, M. (2021). Animals, protists and bacteria share marine biogeographic patterns. *Nature Ecology & Evolution*, *5*(6), 738–746. https://doi.org/10.1038/s41559-021-01439-7

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., Sado, T., Mabuchi, K., Takeshima, H., Miya, M., & Nishida, M. (2013). MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Molecular Biology and Evolution*, *30*(11), 2531–2540. https://doi.org/10.1093/molbev/mst141

James, N. C., Whitfield, A. K., & Cowley, P. D. (2008). Preliminary indications of climate-induced change in a warm-temperate South African estuarine fish community. *Journal of Fish Biology*, *72*(7), 1855–1863. https://doi.org/10.1111/j.1095-8649.2008.01850.x

Jerde, C. L. (2019). Can we manage fisheries with the inherent uncertainty from eDNA? *Journal of Fish Biology*, *98*(2), 341–353. https://doi.org/10.1111/jfb.14218

Koziol, A., Stat, M., Simpson, T., Jarman, S., DiBattista, J. D., Harvey, E. S., Marnane, M., McDonald, J., & Bunce, M. (2018). Environmental DNA metabarcoding studies are critically affected by substrate selection. *Molecular Ecology Resources*, *19*(2), 366–376. https://doi.org/10.1111/1755-0998.12971

Kumar, G., Eble, J. E., & Gaither, M. R. (2020). A practical guide to sample preservation and pre-PCR processing of aquatic environmental DNA. *Molecular Ecology Resources*, *20*(1), 29–39.

Lacoursière-Roussel, A., Howland, K., Normandeau, E., Grey, E. K., Archambault, P., Deiner, K., Lodge, D. M., Hernandez, C., Leduc, N., & Bernatchez, L. (2018). eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. *Ecology and Evolution*, *8*(16), 7763–7777. https://doi.org/10.1002/ece3.4213

Levin, N., Coll, M., Fraschetti, S., Gal, G., Giakoumi, S., Göke, C., Heymans, J., Katsanevakis, S., Mazor, T., Öztürk, B., Rilov, G., Gajewski, J., Steenbeek, J., & Kark, S. (2014). Biodiversity data requirements for systematic conservation planning in the Mediterranean Sea. *Marine Ecology Progress Series*, *508*, 261–281. https://doi.org/10.3354/meps10857

McElroy, M. E., Dressler, T. L., Titcomb, G. C., Wilson, E. A., Deiner, K., Dudley, T. L., Eliason, E. J., Evans, N. T., Gaines, S. D., Lafferty, K. D., Lamberti, G. A., Li, Y., Lodge, D. M., Love, M. S., Mahon, A. R., Pfrender, M. E., Renshaw, M. A., Selkoe, K. A., & Jerde, C. L. (2020). Calibrating environmental DNA metabarcoding to conventional surveys for measuring fish species richness. *Frontiers in Ecology and Evolution*, *8*, 276. https://doi.org/10.3389/fevo.2020.00276

Minamoto, T., Naka, T., Moji, K., & Maruyama, A. (2016). Techniques for the practical collection of environmental DNA: filter selection, preservation, and extraction. *Limnology*, *17*(1), 23–32. https://doi.org/10.1007/s10201-015-0457-4

Miya, M., Gotoh, R. O., & Sado, T. (2020). MiFish metabarcoding: a high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. *Fisheries Science*, *86*(6), 939–970. https://doi.org/10.1007/s12562-020-01461-x

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science*, *2*(7), 150088. https://doi.org/10.1098/rsos.150088

NASA OBPG (2020). *MODIS Aqua Global Level 3 Mapped SST. Ver. 2019.0*. PO.DAAC, CA, USA. Dataset accessed [2021-02-27] at https://doi.org/10.5067/MODSA-1D4D9

Pope, K. L., Lochmann, S. E., & Young, M. K. (2010). Methods for Assessing Fish Populations. In M. C. Quist, & W. A. Hubert (Eds.), *Inland fisheries management in North America* (3rd ed.). American Fisheries Society. https://core.ac.uk/download/pdf/188140002.pdf

QGIS.org (2021). *QGIS geographic information system*. QGIS Association. http://www.qgis.org

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.r-project.org

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, *17*, e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Seymour, M., Edwards, F. K., Cosby, B. J., Bista, I., Scarlett, P. M., Brailsford, F. L., Glanville, H. C., de Bruyn, M., Carvalho, G. R., & Creer, S. (2021). Environmental DNA provides higher resolution assessment of riverine biodiversity and ecosystem function via spatio-temporal nestedness and turnover partitioning. *Communications Biology*, *4*(1), 512. https://doi.org/10.1038/s42003-021-02031-2

Shu, L., Ludwig, A., & Peng, Z. (2020). Standards for methods utilizing environmental DNA for detection of fish species. *Genes*, *2020*(11), 296.

Sigsgaard, E. E., Nielsen, I. B., Carl, H., Krag, M. A., Knudsen, S. W., Xing, Y., Holm-Hansen, T. H., Møller, P. R., & Thomsen, P. F. (2017). Seawater environmental DNA reflects seasonality of a coastal fish community. *Marine Biology*, *164*(6), https://doi.org/10.1007/s00227-017-3147-4

Smart, A. S., Weeks, A. R., Rooyen, A. R., Moore, A., McCarthy, M. A., & Tingley, R. (2016). Assessing the cost-efficiency of environmental DNA sampling. *Methods in Ecology and Evolution*, *7*(11), 1291–1298. https://doi.org/10.1111/2041-210x.12598

Smit, A. J., Roberts, M., Anderson, R. J., Dufois, F., Dudley, S. F. J., Bornman, T. G., Olbers, J., & Bolton, J. J. (2013). A coastal seawater temperature dataset for biogeographical studies: large biases between in situ and remotely-sensed data sets around the coast of South Africa. *PLoS One*, *8*(12), e81944. https://doi.org/10.1371/journal.pone.0081944

Steinke, D., Connell, A. D., & Hebert, P. D. N. (2016). Linking adults and immatures of South African marine fishes. *Genome*, *59*(11), 959–967. https://doi.org/10.1139/gen-2015-0212

Teske, P. R., Von der Heyden, S., McQuaid, C. D., & Barker, N. P. (2011). A review of marine phylogeography in southern Africa. *South African Journal of Science*, *107*(5/6), 43–53. https://doi.org/10.4102/sajs.v107i5/6.514

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., & Willerslev, E. (2016). Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PLOS ONE*, *11*(11), e0165252. https://doi.org/10.1371/journal.pone.0165252

Turpie, J. K., Beckley, L. E., & Katua, S. M. (2000). Biogeography and the selection of priority areas for conservation of South African coastal fishes. *Biological Conservation*, *92*(1), 59–72. https://doi.org/10.1016/s0006-3207(99)00063-4

Underwood, E., Taylor, K., & Tucker, G. (2018). The use of biodiversity data in spatial planning and impact assessment in Europe. *Research Ideas and Outcomes*, *4*, e28045. https://doi.org/10.3897/rio.4.e28045

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G. H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.-M., … 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

Von Der Heyden, S. (2011). 'Carry on sampling!'- assessing marine fish biodiversity and discovery rates in southern Africa. *Diversity and Distributions*, *17*(1), 81–92. https://doi.org/10.1111/j.1472-4642.2010.00712.x

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., Minamoto, T., & 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: Czachur, M. V., Seymour, M., Creer, S., & von der Heyden, S. (2022). Novel insights into marine fish biodiversity across a pronounced environmental gradient using replicated environmental DNA analyses. *Environmental DNA*, 4, 181–190. https://doi.org/10.1002/edn3.238