Short Communication

Evaluation of Fish Species Detection in the Northwestern Pacific using eDNA Metabarcoding: A Mock Community Approach

Sergei V. Turanov[1,2,*,†] , Olesia A. Rutenko[3,4,†]

[1]Laboratory of Deep Sea Research, A.V. Zhirmunsky National Scientific Center of Marine Biology, Far Eastern Branch, Russian Academy of Sciences, 690041 Vladivostok, Russia

[2]Chair of Water Biological Resources and Aquaculture, Far Eastern State Technical Fisheries University, 690087 Vladivostok, Russia

[3]Laboratory of Ecology and Evolutionary Biology of Aquatic Organisms, Far Eastern Federal University, 690922 Vladivostok, Russia

[4]Laboratory of Molecular Systematics, A.V. Zhirmunsky National Scientific Center of Marine Biology, Far Eastern Branch, Russian Academy of Sciences, 690041 Vladivostok, Russia

*Correspondence: sturcoal@mail.ru (Sergei V. Turanov)

[†]These authors contributed equally.

Academic Editor: Monica Butnariu

Submitted: 23 August 2024   Revised: 18 November 2024   Accepted: 26 November 2024   Published: 24 March 2025

Abstract

Background: Metabarcoding of environmental DNA (eDNA), a technique using high-throughput sequencing, has transformed biodiversity monitoring by identifying organisms from DNA fragments present in the environment. This method, particularly useful for aquatic ecosystems, allows for non-invasive species monitoring, helping to provide insight into ecosystem composition and taxonomic diversity. The objective of this study was to assess the efficacy of eDNA metabarcoding for fish species identification in a model community from the northeast Pacific Ocean using *12S* ribosomal RNA (*12S* rRNA) marker. **Methods**: Water samples were collected from the tank of the Primorsky Aquarium, which contains fish species from the Sea of Japan, Sea of Okhotsk, and Bering Sea. DNA was extracted using syringe filters and enriched with polymerase chain reaction (PCR) of mitochondrial *12S* rRNA fragment, followed by sequencing on Illumina platform. The resulting reads were processed using the bayesian generalized uncertainty modeling (BEGUM) pipeline and their taxonomic diversity was assessed by basic local alignment search tool (BLAST) search. Using *in silico* PCR, we also assessed the possible association of detection failures of some species with the presence of primer-to-target sequence mismatches. **Results**: From a fish community of only 20 species in the tank, we identified 56 operational taxonomic units (OTUs) corresponding to 28 genera. Among these OTUs, 20 species were unambiguously classified by BLAST-based analysis, though only 9 of them corresponded to the species actually present in the tank. Significant problems included inconsistent reference data and marker biases that affected the accuracy of species identification. In addition to DNA contamination from feed, contamination from the water source may have introduced extraneous DNA into the samples. Also, using *in silico* PCR analysis with a small number of available reference sequences, we demonstrated a significantly higher number of primer mismatches for species that were not identified. **Conclusions**: This study highlights the relative efficacy of eDNA metabarcoding for fish species identification, but also highlights the need to improve reference databases and minimise contamination, searching for references and primers to improve accuracy. Further research should focus on optimising marker selection and controlling methodological bias to ensure robust biodiversity estimates.

Keywords: *12S* rRNA; eDNA metabarcoding; fish tank; mock community; north-east pacific fish

1. Introduction

Metabarcoding, leveraging high-throughput sequencing, has revolutionized biodiversity assessment by allowing the identification of organisms through their DNA sequences. This technique, when paired with environmental DNA (eDNA) analysis, captures and enriches specific regions of extracellular DNA molecules released into the environment by various organisms. eDNA analysis is especially valuable for non-invasive monitoring of biodiversity, offering a comprehensive snapshot of ecosystem composition without the need for direct specimen collection [1–3]. This method not only allows researchers to trace organisms through environmental samples, but also enhances our understanding into taxonomic diversity [4,5], estimating fish biomass [6–9], and evaluating genetic diversity via amplified sequence variants [10–13].

In order to validate these methods mock community analysis has been used where bunch of specimens with known taxonomic composition are treated and analyzed as a proxy of natural community [14]. In aquatic environments, mock community analyses using eDNA metabarcoding have commonly employed the species composition of fish housed in aquariums [15,16]. As a result, it was found that eDNA can effectively detect the majority of bony fish species present in a controlled mesocosm environment, but facing challenges in detecting cartilaginous fishes and sea turtles, due to biases in primer specificity [15] and may be highly influenced by marker selection [16]. Furthermore, analysis of DNA from known fish egg ratios has been explored to evaluate metabarcoding efficacy, revealing that the accuracy of species identification is significantly dependent on the quality of the reference database and the specific

Publisher's Note: IMR Press stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

molecular markers used [17]. Similar outcomes were highlighted in the research on neotropical fish mock communities which are influenced by factors such as DNA concentration, community composition, and marker choice, thus emphasizing the importance of using multiple markers and conducting controlled experiments for accurate biodiversity assessments [18]. This overall still makes it crucial to assess potential limitations of the approach and generate the necessary methodological recommendations before launching the analysis of natural samples of the local fauna by the modern technique of DNA metabarcoding.

The northwestern Pacific, including the Sea of Japan, Sea of Okhotsk, and the Bering Sea, maintains significant fish diversity, with over 1300 species inhabiting these waters [19,20]. The unified methods, including the construction of a reference library of DNA fragments, facilitate accurate species identification [21–23], providing a robust framework for ongoing research and conservation efforts. Recent systematic revisions, aided by molecular genetic approaches, have further refined the understanding of this diverse marine life [24–29]. In this study, we present findings on the evaluation of marine fish composition in the northwestern Pacific area, specifically in the mock communities settled in Primorsky Aquarium (Russia, Vladivostok), using eDNA metabarcoding techniques based on 12S ribosomal RNA (*12S* rRNA) mitochondrial marker. This study also aimed a comparison between the *in silico* polymerase chain reaction (PCR) approach and traditional eDNA metabarcoding, focusing on the use of a standard set of MiFish *12S* rRNA primers [30] to simulate PCR amplification of available reference sequences from databases. By simulating PCR amplification *in silico*, we aimed to evaluate how well the selected primers perform in amplifying DNA sequences from various species in a controlled environment.

2. Materials and Methods

2.1 eDNA Metabarcoding Analysis of the Mock Community

Water from a tank (volume 1.68 m^3) at the Primorsky Aquarium (Russky Island, Vladivostok, Primorsky Krai, Russia) was taken on December 12, 2019. Water comes to the tank directly from the bay after a small amount of purification. This tank was holding 20 species spanning 11 families and 19 genera of the coast of the Far Eastern seas area (Sea of Japan, Sea of Okhotsk, and the Bering Sea). This selection includes species from both benthic and pelagic habitats, with varying ecological roles. The water temperature was maintained in the range of 10–13 °C according to the sea temperature at that time of year at that location. Taxonomic information and the number of fishes held in the tank is provided in the Table 1. *Pseudopleuronectes* sp. among them could not be identified to species level due to small size. Fishes in this tank are fed with minced fillet of pink (*Oncorhynchus gorbuscha* (Walbaum, 1792)) and chum salmons (*O. keta* (Walbaum,

1792)), greenlings (*Hexagrammos* sp. Tilesius, 1810), chub mackerel (*Scomber japonicus* Linnaeus, 1758), squids (*Todarodes pacificus* (Steenstrup, 1880), *Berryteuthis magister* (Berry, 1913)), and shrimp (*Pandalus latirostris* Rathbun, 1902). Three replicate samples of water (about 450 mL each) were collected from the tank with a syringe (150 mL capacity, luer lock type). The full volume of water from each replicate was filtered through a single syringe filter (25 mm diameter, 0.45 µm pore size, PES material).

To preserve the DNA on the filter, 1 mL of Longmire's buffer [31] was passed through it, after which the inlet and outlet ports were sealed with combi-stopper plugs. The fixed filter was then stored at –20 °C. DNA was isolated from the preserved filter using the M-Sorb-OOM kit (Sintol, Moscow, Russia). The manufacturer's protocol was modified such that the lysis buffer was warmed to 65 °C, pushed in the reverse direction of filtration and poured into a clean test tube (backflushing method, after [32]). The isolated DNA was stored at –20 °C. A mitochondrial fragment —*12S* rRNA with a length of ~170 bp [30]—was used to retrieve the information on the taxonomic diversity from the eDNA samples. A primer pair, each carrying a unique 7-nucleotide tag (doubly-tagged primers) developed in ecotag (Alpine Ecology Laboratory, Grenoble, France) [33] was used to amplify the fragment for each sample. Negative PCR controls were also performed using separate pairs of tagged primers. The PCR for each sample was performed using three replicates. The reaction mixture consisted of 10 µL of AmpliTaq Gold 360 Master Mix (Applied Biosystems, Foster City, CA, USA), 0.5 µL of each forward and reverse primer (10 µM), 0.16 µL of bovine serum albumin, 10 ng of DNA, and deionized water to reach a final volume of 20 µL. The thermal cycling protocols are provided in Table 2.

The amplification products were assessed by running the fragments on a 1% agarose gel, staining with ethidium bromide, and visualizing them under ultraviolet light. The amplicons were purified using the Cleanup S-Cap kit (Evrogen, Moscow, Russia) and normalized (refer to [34]) prior to pooling. The amount of control reactions was taken as the average from the resulting volume of the normalized samples before pooling. The normalized amplicons were then combined with PCR controls and sequenced at Novogene (Tianjin, China). The library was prepared with the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) and sequenced on an Illumina NovaSeq 6000 platform employing a 250 bp paired-end sequencing approach.

The obtained reads were processed according to the bayesian generalized uncertainty modeling (BEGUM) metabarcoding pipeline (open-source software) [35,36]. After adapter removal and preliminary evaluation of the read quality with fastqc (Babraham Institute, Cambridge, UK), possible read errors were corrected in Spades (Center for Algorithmic Biotechnology, St. Petersburg State University, St. Petersburg, Russia) [37]. Subsequently, paired-

Table 1. Fish species composition in the tank of the Primorsky Aquarium and their detection status.

No.	Species Name	Family	No. of Individuals	Detected by eDNA Metabarcoding	No. of Reads	Detected by *In Silico* PCR
1	*Agonomalus proboscidalis*	Agonidae	1	No	0	No
2	*Alcichthys elongatus*	Cottidae	3	No	0	No
3	*Bathymaster derjugini*	Bathymasteridae	1	Yes	163	No
4	*Bero elegans*	Cottidae	3	No	0	No
5	*Blepsias cirrhosus*	Agonidae	1	Yes	79	Yes
6	*Brachyopsis segaliensis*	Agonidae	2	No	0	Yes
7	*Chirolophis japonicus*	Stichaeidae	3	No	0	Yes
8	*Chirolophis saitone*	Stichaeidae	2	No	0	No
9	*Gymnocanthus herzensteini*	Cottidae	5	Yes	32	Yes
10	*Gymnogobius heptacanthus*	Gobiidae	5	Yes	59	Yes
11	*Hypomesus japonicus*	Osmeridae	3	Yes	3741	Yes
12	*Hypsagonus jordani*	Agonidae	2	Yes	1932	No
13	*Hypoptychus dybowskii*	Hypoptychidae	5	Yes	78	Yes
14	*Lumpenus sagitta*	Lumpenidae	2	No	0	Yes
15	*Opisthocentrus ocellatus*	Opisthocentridae	3	Yes	622	Yes
16	*Pholidapus dybowskii*	Opisthocentridae	10	No	0	Yes
17	*Podothecus gilberti*	Agonidae	3	No	0	No*
18	*Pseudopleuronectes* sp.	Pleuronectidae	2	No	0	Yes
19	*Pungitius pungitius*	Gasterosteidae	6	Yes	317	Yes
20	*Stichaeopsis epallax*	Stichaeidae	5	No	0	No

*, not present in GenBank. eDNA, environmental DNA.

Table 2. Thermal cycling protocol for 12S ribosomal RNA (*12S* rRNA) marker amplification.

Preheating	Cycles	Denaturation	Annealing	Elongation	Final elongation
95 °C for 3 min.	35	98 °C for 20 sec.	60 °C for 15 sec.	72 °C for 15 sec.	72 °C for 5 min.

end reads were assembled into consensus sequences using PandaSeq (Department of Biology, University of Waterloo, Waterloo, Ontario, Canada) [38]. We then used the BE-GUM program to demultiplex and filter the reads for each sample. Clustering was done by sumaclust (Institut de Biologie Computationnelle, Montpellier, France) [39] with parameters -t 0.98 and -R 0.85. The taxonomic assignment for the generated operational taxonomic units (OTUs) was performed using basic local alignment search tool (BLAST) (National Center for Biotechnology Information, Bethesda, MD, USA) [40] implemented in Ubuntu command line interface. The taxonomic information was then summarized in a single table based on the output from the Metagenome Analyzer (MEGAN) community edition program (University of Tübingen, Tübingen, Germany) [41]. The top percent of lowest common ancestor (LCA) parameters was set to 2.0 implementing the naïve LCA algorithm with 80 percent to cover. The number of OTUs after taxonomic referencing was corrected using the LULU package (Center for GeoGenetics, University of Copenhagen, Copenhagen, Denmark) [42] with the following parameters: minimum match –95%, minimum relative co-occurrence –0.97.

2.2 In silico PCR Based on 12S rRNA Sequences

In addition, an *in silico* approach was used to test the assumption of the presence of more mismatches in the sequences of those species that were not identified by eDNA metabarcoding. For this purpose, *12S* rRNA sequences available for the taxa kept in the tank were retrieved from GenBank (**Supplementary Table 1**). Additionally, 5 sequences of *Oncorhynchus gorbuscha* and 24 sequences of *O. Keta* were included in the analysis to assess the potential impact of feed-derived eDNA on our results. For *Pseudopleuronectes* sp. the sequences of all species of this genus that may occur in the far eastern seas area were searched. For this purpose, a search based on the query was performed. Next, *in silico* PCR with MiFish primers [30] was performed using the retrieved sequences in the ecoPCR tool of the OBITools package (Laboratoire d'Écologie Alpine, Grenoble, France) [33,43,44]. PCR parameters allowed up to 3 mismatches relative to the target sequence per primer. Melting temperature (*Tm*) was estimated based on [45]. Salt concentration used for estimating *Tm* was set 0.05. Based on the data obtained, aggregation by species was performed so that each species corresponded to one observation (**Supplementary Table 2**). To test the hypothesis whether there is a statistically significant difference

between detected and undetected species in the number of primers mismatches detected, Welch's *t*-test [46] was performed in the stats R package (R Foundation for Statistical Computing, Vienna, Austria) [47]. Sequences that were not amplified *in silico* were not included in the analysis. Visualisation of the analysis was performed in the ggplot2 R package (R Foundation for Statistical Computing, Vienna, Austria) [48]. All figures were labeled and finalized using Inkscape (Inkscape Project Inc., Silver Spring, MD, USA, https://inkscape.org/).

3. Results and Discussion

After demultiplexing, the *12S* rRNA fragment had 42,260 reads, 2827 of which were unique. The control identified 60 reads. After filtering the reads and deleting the chimeric sequences, there were 38,441 sequences for *12S* rRNA. Sequences clustered into 56 OTUs, four of which were assigned to bacteria with 57 reads in total. The remaining 52 OTUs were related to ray-finned fishes, classified into seven orders, 15 families and 28 genera (Fig. 1). Species-level taxonomy was unambiguously assigned to 20 OTUs (39%). There were 15 OTUs (26.8%) with taxonomic contradictions at the species level; at the genus level, there were nine OTUs (16%), and one OTU at the family level. The conflict in this case is defined as equal (or undetectable) probability of assigning the same OTU to different taxa according to BLAST results.

This may indicate either erroneous taxonomic identification of one or several sequences that make up the reference library, or that the marker is highly conserved hence divergence is not sufficient enough to distinguish two or more species. These and other limitations of the approach have been mentioned earlier when estimating fish diversity [2,3,49,50]. On the other hand, despite the data correction with the LULU R package (version 0.1.0), some species names, specifically *Hypomesus japonicus*, *Oncorhynchus keta*, and *Opisthocentrus ocellatus*, were assigned to three different OTUs simultaneously (see **Supplementary Table 3**). Selecting a standard delineation threshold for clustering based on hierarchical approaches can distort the true number of OTUs. Using a fixed similarity threshold in clustering can sometimes misrepresent the true diversity of OTUs, either by overestimating or underestimating the number of clusters. Clustering approaches that adapt to the inherent structure of the data—without applying a strict cut-off threshold—such as unsupervised Bayesian clustering [51] and the Swarm algorithm [52], may provide more accurate and biologically meaningful results. More than 1000 reads were accounted for each of the clusters among *Hypomesus japonicus* McAllister, 1963, *Hypsagonus jordani* Jordan & Starks, 1904 and Stichaeidae sp.—26 OTUs had 20 to 1200 reads, whereas each of the other 23 OTUs had less than 20 reads.

The largest number of reads came from two species of the genus Oncorhynchus (whose fillets are used for feeding, see above)—more than 13,000 for each species. Hence,

it is encouraged to account for this when planning sufficient sequencing depth for mesocosm experiments [15,16] to provide adequate coverage for the detection of DNA for targeted organisms with low amounts of eDNA. An option would be to use blocking primers for the forage organisms, similar to those used in diet assessment. However, it is necessary to consider all the advantages and limitations of this approach [53]. Primer bias should not be excluded either [54,55], so we decided to investigate its possible impact.

A total of 155 sequences from 24 fish species were available in GenBank for the *12S* rRNA fragment (**Supplementary Table 1**). One sequence of *Oncorhynchus gorbuscha* and 5 sequences of *O. keta* were successfully amplified *in silico*, with 1 mismatch per forward primer detected. The species *G. herzensteini* had the highest number of sequences (30). Only one *12S* rRNA sequence each was found for the species *A. proboscidalis*, *B. derjugini*, *C. saitone*, *H. jordani* and *S. epallax*. The mean number of *12S rRNA* sequences per species was 6.46 ± 7.42 (mean ± standard deviation). 17 of 24 species were successfully amplified using *in silico* PCR (had at least one successfully amplified sequence). The highest number of primer mismatches (3) was attributed to the species *H. japonicus*, while *A. elongatus* and *B. elegans* had 2 mismatches each. The rest of the *in silico* amplified showed 1 mismatch each. At the same time, it should be noted that all mismatches of *H. japonicus* are concentrated on one ('forward') primer, whereas in *A. elongatus* and *B. elegans 1* mismatch per each (forward and reverse) primer was found. Mismatches near the 3′ end of the primer can significantly reduce amplification efficiency, whereas mismatches near the 5′ end have less effect [56,57]. It is possible that the mismatches in *H. japonicus* sequences occurred in less critical regions of the primer binding sites, allowing for successful amplification, while the mismatches in *A. elongatus* and *B. elegans* may have been located in positions that more adversely affected *in silico* amplification. Furthermore, differences in species-specific behaviors and physiology can influence the amount of eDNA shed into the environment which may increase the probability of detecting certain species. *Hypomesus japonicus* is known to be an active swimmer, often forming schools [58]. This increased activity can lead to higher eDNA shedding rates [59]. In contrast, *Alcichthys elongatus* and *Bero elegans* may exhibit less active behaviors, potentially resulting in lower eDNA contributions despite their presence in the tank.

Although we had the same number of individuals (three) for each species in the tank, their behaviors and physiological differences could have led to varying levels of eDNA shedding. While *A. elongatus* is larger in size— approximately three times larger than *H. japonicus* and *B. elegans*—we lack precise biomass measurements to quantify their contributions accurately.

Despite the small representation of the group of species not detected under mock community conditions but detected by *in silico* PCR, Welch's *t*-test showed (Fig. 2) a
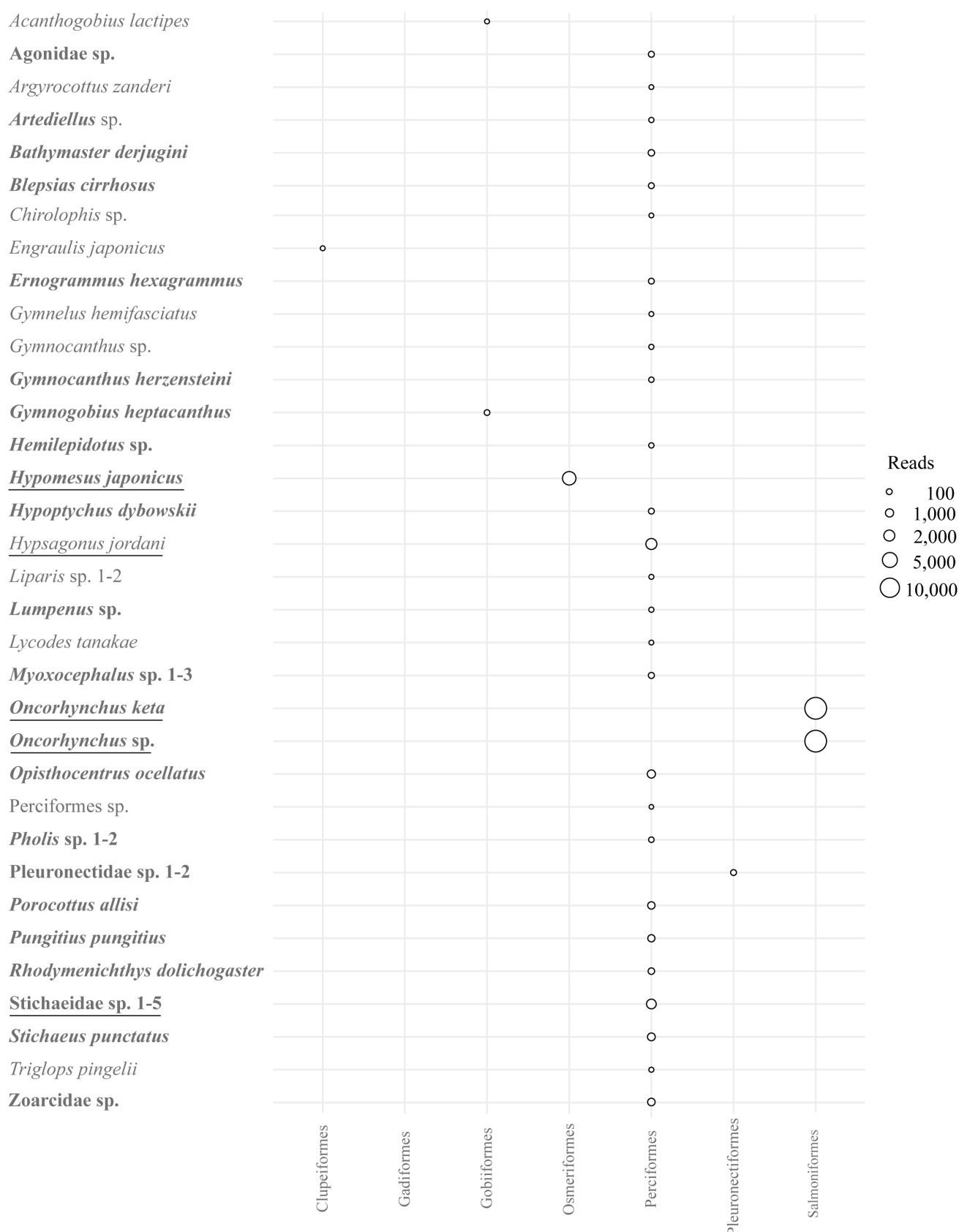
Fig. 1. Bubble graph showing the number of reads for different operational taxonomic units (OTUs) obtained for *12S* rRNA, amplified based on the water eDNA samples from a fish tank at the Primorsky Aquarium. The taxonomic orders are presented horizontally and the OTUs information are placed vertically. The size of circles demonstrates the number of reads assigned to the different OTUs. OTU names in bold have no less than 20 reads. Each underlined OTU accounts for more than 1000 reads.

significantly higher number of mismatches in these species compared to successfully detected species by both methods. The success of in silico PCR (see **Supplementary Table 2**) varied widely between species. The analyses still allow speculation on the primer bias identified above. Although one of the species successfully detected by eDNA had a greater number mismatches, both species not detected by eDNA mock community also exhibited multiple primers mismatches *in silico*. This approach, which to our knowledge is not very popular in the field of eDNA metabarcoding, can be used at the preliminary stage of research, between the collection and validation of a local reference library and the selection of optimal primers for the target genome regions.

Fig. 2. Comparison of the mean primer mismatches between species detected and not detected via eDNA metabarcoding. The boxplot shows two classes of variable (0, not detected via eDNA metabarcoding, 1, detected) significantly discriminated by Welch's *t*-test ($t = 6.5$, *df* = 14, *p*-value = 0.00001402). Vertical lines indicate median values.

We unambiguously identified nine out of 20 taxa (45%) in the tank. Two taxa present (*Alcichthys elongatus* (Steindachner 1881), *Bero elegans* (Steindachner, 1881)) were not identified, i.e., there is no OTU that might be related to these species. Another 11 taxa (55%) may somehow be concealed behind OTUs with conflicting taxonomic reference (see above), such as *Chirolophis* sp., *Lumpenus*

sp., *Pleuronectidae* sp. (Fig. 1, **Supplementary Table 3**). The indication of all these taxa may come from environmental DNA originating from the water storage reservoir, whose intake is located next to the Primorsky Aquarium building. Therefore, these OTUs likely result from external environmental DNA introduced via the water supply rather than solely from the fish present in the tank. The presence of such external DNA is expected and, regretfully, is not unusual [15,16]. In addition to reducing the probability of contamination, the development and verification of a reference library for fish identification should be considered as an important way to reduce uncertainty and improve the reliability of high-throughput monitoring methods based on aquatic eDNA [3,49].

4. Conclusions

Our study highlights the potential and challenges of using eDNA metabarcoding for monitoring fish biodiversity in controlled environments. Despite the potential of this method, our results showed that only 9 out of the 20 species present in the tank were unambiguously identified, and there were 25 instances of incorrect species identifications, resulting in a low accuracy rate. These findings indicate a substantial potential for misidentification, suggesting that the method, as applied in our study, should be used with extreme caution and awareness of the its limitations.

The high rate of misidentification may be attributed to several factors, including inconsistent and incomplete reference databases, primer biases affecting amplification efficiency, and contamination from external DNA sources such as feed and water supply. These limitations underscore the critical need for comprehensive and accurate reference libraries, improved primer design, and stringent contamination controls to enhance the accuracy and reliability of eDNA metabarcoding. By acknowledging these limitations, we aim to contribute to the advancement of eDNA metabarcoding methods and encourage further studies to overcome the obstacles identified in our research. Improving these methodologies is essential before eDNA metabarcoding can be confidently applied in practical biodiversity monitoring and conservation efforts.

Availability of Data and Materials

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

Author Contributions

OR performed the collection of water from tank and DNA isolation from syringe filters. ST designed the research, performed genetic analysis and analyzed the data. ST wrote the manuscript. Both authors contributed to editorial changes in the manuscript. Both authors read and approved the final manuscript. Both authors have participated sufficiently in the work to take public responsibility

for appropriate portions of the content and agreed to be accountable for all aspects of the work in ensuring that questions related to its accuracy or integrity.

Ethics Approval and Consent to Participate

Acknowledgment

Funding

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/FBS26247.

References

[1] McGee KM, Robinson CV, Hajibabaei M. Gaps in DNA-Based Biomonitoring Across the Globe. Frontiers in Ecology and Evolution. 2019; 7: 337. https://doi.org/10.3389/fevo.2019.00337.

[2] Schenekar T, Schletterer M, Lecaudey LA, Weiss SJ. Reference databases, primer choice, and assay sensitivity for environmental metabarcoding: Lessons learnt from a re-evaluation of an eDNA fish assessment in the Volga headwaters. River Research and Applications. 2020; 36: 1004–1013. http://doi.org/10.1002/rra.3610.

[3] Weigand H, Beermann AJ, Čiampor F, Costa FO, Csabai Z, Duarte S, et al. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. The Science of the Total Environment. 2019; 678: 499–524. https://doi.org/10.1016/j.scitotenv.2019.04.247.

[4] Thomsen PF, Kielgast J, Iversen LL, Møller PR, Rasmussen M, Willerslev E. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLoS ONE. 2012; 7: e41732. https://doi.org/10.1371/journal.pone.0041732.

[5] Stauffer S, Jucker M, Keggin T, Marques V, Andrello M, Bessudo S, et al. How many replicates to accurately estimate fish biodiversity using environmental DNA on coral reefs? Ecology and Evolution. 2021; 11: 14630–14643. https://doi.org/10.1002/ece3.8150.

[6] Di Muri C, Handley L, Bean CW, Li J, Peirson G, Sellers GS, et al. Read counts from environmental DNA (eDNA) metabarcoding reflect fish abundance and biomass in drained ponds. Metabarcoding Metagenomics. 2020; 4: e56959. http://doi.org/10.3897/mbmg.4.56959.

[7] Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, et al. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. Molecular Ecology Resources. 2016; 16: 29–41. https://doi.org/10.1111/1755-0998.12433.

[8] Salter I, Joensen M, Kristiansen R, Steingrund P, Vestergaard P. Environmental DNA concentrations are correlated with regional biomass of Atlantic cod in oceanic waters. Communications Biology. 2019; 2: 461. https://doi.org/10.1038/s42003-019-0696-8.

[9] Takahara T, Minamoto T, Yamanaka H, Doi H, Kawabata Z. Estimation of fish biomass using environmental DNA. PloS One. 2012; 7: e35868. https://doi.org/10.1371/journal.pone.0035868.

[10] Tsuji S, Shibata N, Sawada H, Ushio M. Quantitative evaluation of intraspecific genetic diversity in a natural fish population using environmental DNA analysis. Molecular Ecology Resources. 2020; 20: 1323–1332. https://doi.org/10.1111/1755-0998.13200.

[11] Adams CIM, Knapp M, Gemmell NJ, Jeunen GJ, Bunce M, Lamare MD, et al. Beyond Biodiversity: Can Environmental DNA (eDNA) Cut It as a Population Genetics Tool? Genes. 2019; 10: 192. https://doi.org/10.3390/genes10030192.

[12] Antich A, Palacin C, Wangensteen OS, Turon X. To denoise or to cluster, that is not the question: optimizing pipelines for COI metabarcoding and metaphylogeography. BMC Bioinformatics. 2021; 22: 1–24. http://doi.org/10.1186/s12859-021-04115-6.

[13] Turanov SV, Koltsova MA, Rutenko OA. Experimental evaluation of genetic variability based on DNA metabarcoding from the aquatic environment: Insights from the Leray COI fragment. Ecology and Evolution. 2024; 14: e11631. https://doi.org/10.1002/ece3.11631.

[14] Braukmann TWA, Ivanova NV, Prosser SWJ, Elbrecht V, Steinke D, Ratnasingham S, et al. Metabarcoding a diverse arthropod mock community. Molecular Ecology Resources. 2019; 19: 711–727. https://doi.org/10.1111/1755-0998.13008.

[15] Kelly RP, Port JA, Yamahara KM, Crowder LB. Using environmental DNA to census marine fishes in a large mesocosm. PLoS ONE. 2014; 9: e86175. https://doi.org/10.1371/journal.pone.0086175.

[16] Morey KC, Bartley TJ, Hanner RH. Validating environmental DNA metabarcoding for marine fishes in diverse ecosystems using a public aquarium. Environmental DNA. 2020; 2: 330–342. http://doi.org/10.1002/edn3.76.

[17] Duke EM, Burton RS. Efficacy of metabarcoding for identification of fish eggs evaluated with mock communities. Ecology and Evolution. 2020; 10: 3463–3476. https://doi.org/10.1002/ece3.6144.

[18] Hilário HO, Mendes IS, Guimarães SN, Carvalho DC. DNA metabarcoding of mock communities highlights potential biases when assessing Neotropical fish diversity. Environmental DNA. 2023; 5: 1351–1361. http://doi.org/10.1002/edn3.456.

[19] Orlov AM, Tokranov AM. Checklist of deep-sea fishes of the Russian northwestern Pacific Ocean found at depths below 1000 m. Progress in Oceanography. 2019; 176: 102143. http://doi.org/10.1016/j.pocean.2019.102143.

[20] Parin NV, Evseenko SA, Vasil'eva ED. Fishes of the Rusian Seas: Annotated Catalogue. KMK Scientific Press: Moscow. 2014.

[21] Mecklenburg CW, Møller PR, Steinke D. Biodiversity of arctic marine fishes: taxonomy and zoogeography. Marine Biodiversity. 2011; 41: 109–140. http://doi.org/10.1007/s12526-010-0070-z.

[22] Turanov SV, Kartavtsev YP, Lipinsky VV, Zemnukhov VV, Balanov AA, Lee YH, *et al*. DNA-barcoding of perch-like fishes (Actinopterygii: Perciformes) from far-eastern seas of Russia with taxonomic remarks for some groups. Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis. 2016; 27: 1188–1209. https://doi.org/10.3109/19401736.2014.945525.

[23] Zhang JB, Hanner R. DNA barcoding is a useful tool for the identification of marine fishes from Japan. Biochemical Systematics and Ecology. 2011; 39: 31–42. http://doi.org/10.1016/j.bse.2010.12.017.

[24] Balakirev ES, Kravchenko AY, Semenchenko AA. Genetic Evidence for a Mixed Composition of the Genus *Myoxocephalus* (Cottoidei: Cottidae) Necessitates Generic Realignment. Genes. 2020; 11: 1071. https://doi.org/10.3390/genes11091071.

[25] Moreva IN, Radchenko OA, Petrovskaya AV. Karyological and Molecular Genetic Differentiation of Fringed Sculpins of the Genus Porocottus Gill, 1859 (Cottidae: Myoxocephalinae). Russian Journal of Marine Biology. 2019; 45: 128–136. http://doi.org/10.1134/S1063074019020081.

[26] Moreva I, Radchenko O, Petrovskaya A, Borisenko S. Molecular genetic and karyological analysis of antlered sculpins of Enophrys diceraus group (Cottidae). Russian Journal of Genetics. 2017; 53: 1030–1041. http://doi.org/10.1134/S1022795417090113.

[27] Orlova SY, Volkov AA, Shcepetov DM, Maznikova OA, Chernova NV, Chikurova EA, *et al*. Inter- and Intra-Species Relationships of Greenland Halibut Reinhardtius hippoglossoides (Pleuronectidae) Based on the Analysis of Nuclear and Mitochondrial Genetic Markers. Journal of Ichthyology. 2019; 59: 65–77. http://doi.org/10.1134/S0032945219010119.

[28] Radchenko OA. The System of the Suborder Zoarcoidei (Pisces, Perciformes) as Inferred from Molecular Genetic Data. Russian Journal of Genetics. 2015; 51: 1096–1112. http://doi.org/10.1134/S1022795415100130.

[29] Radchenko OA, Moreva IN, Petrovskaya AV. Karyological and Molecular Genetic Divergence of Sculpins Myoxocephalus Gill, 1859 (Cottidae). Russian Journal of Genetics. 2020; 56: 1212–1223. http://doi.org/10.1134/S1022795420100117.

[30] Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, *et al*. 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. 2015; 2: 150088. https://doi.org/10.1098/rsos.150088.

[31] Longmire JL, Maltbie M, Baker RJ, Texas Tech University Museum. Use of "Lysis Buffer" in DNA isolation and its implication for museum collections. Museum of Texas Tech University: Lubbock, TX. 1997. http://doi.org/10.5962/bhl.title.143318.

[32] Kesberg AI, Schleheck D. Improved protocol for recovery of bacterial DNA from water filters: Sonication and backflushing of commercial syringe filters. Journal of Microbiological Methods. 2013; 93: 55–57. https://doi.org/10.1016/j.mimet.2013.02.001.

[33] Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E. obitools: a unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources. 2016; 16: 176–182. https://doi.org/10.1111/1755-0998.12428.

[34] Elbrecht V, Steinke D. Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. Freshwater Biology. 2019; 64: 380–387. http://doi.org/10.1111/fwb.13220.

[35] Yang C, Bohmann K, Wang X, Cai W, Wales N, Ding Z, *et al*. Biodiversity Soup II: A bulk-sample metabarcoding pipeline emphasizing error reduction. Methods Ecology and Evolution. 2021; 12: 1252–1264. http://doi.org/10.1111/2041-210X.13602.

[36] Zepeda-Mendoza ML, Bohmann K, Carmona Baez A, Gilbert MTP. DAMe: a toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses. BMC Research Notes. 2016; 9: 255. https://doi.org/10.1186/s13104-016-2064-9.

[37] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology. 2012; 19: 455–477. https://doi.org/10.1089/cmb.2012.0021.

[38] Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics. 2012; 13: 31. https://doi.org/10.1186/1471-2105-13-31.

[39] Mercier C, Boyer F, Bonin A, Coissac E. SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. In Programs and Abstracts of the SeqBio 2013 workshop. Abstract (pp. 27–29). 2013.

[40] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, *et al*. BLAST+: architecture and applications. BMC Bioinformatics. 2009; 10: 421. https://doi.org/10.1186/1471-2105-10-421.

[41] Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, *et al*. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Computational Biology. 2016; 12: e1004957. https://doi.org/10.1371/journal.pcbi.1004957.

[42] Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, *et al*. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications. 2017; 8: 1188. https://doi.org/10.1038/s41467-017-01312-x.

[43] Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiology. 2010; 10: 189. https://doi.org/10.1186/1471-2180-10-189.

[44] Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessière J, *et al*. An in silico approach for the evaluation of DNA barcodes. BMC Genomics. 2010; 11: 434. https://doi.org/10.1186/1471-2164-11-434.

[45] SantaLucia J, Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proceedings of the National Academy of Sciences of the United States of America. 1998; 95: 1460–1465. https://doi.org/10.1073/pnas.95.4.1460.

[46] WELCH BL. The generalisation of student's problems when several different population variances are involved. Biometrika. 1947; 34: 28–35. https://doi.org/10.1093/biomet/34.1-2.28.

[47] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2021. Available at: https://www.R-project.org/ (Accessed: 3 June 2021).

[48] Wickham H. ggplot2. Wiley interdisciplinary reviews: computational statistics. 2011; 3: 180–185. http://doi.org/10.1002/wics.147.

[49] Collins RA, Bakker J, Wangensteen OS, Soto AZ, Corrigan L, Sims DW, *et al*. Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods Ecology and Evolution. 2019; 10: 1985–2001. http://doi.org/10.1111/2041-210X.13276.

[50] Stoeckle MY, Das Mishu M, Charlop-Powers Z. Improved Environmental DNA Reference Library Detects Overlooked Marine Fishes in New Jersey, United States. Frontiers in Marine Science. 2020; 7: 226. http://doi.org/10.3389/fmars.2020.00226.

[51] Hao X, Jiang R, Chen T. Clustering 16S rRNA for OTU prediction: a method of unsupervised Bayesian clustering. Bioinformatics. 2011; 27: 611–618. https://doi.org/10.1093/bioinformatics/btq725.

[52] Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. Swarm v2: highly-scalable and high-resolution amplicon clustering.

PeerJ. 2015; 3: e1420. https://doi.org/10.7717/peerj.1420.

[53] Piñol J, Mir G, Gomez-Polo P, Agustí N. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Molecular Ecology Resources. 2015; 15: 819–830. https://doi.org/10.1111/1755-0998.12355.

[54] Clarke LJ, Soubrier J, Weyrich LS, Cooper A. Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. Molecular Ecology Resources. 2014; 14: 1160–1170. https://doi.org/10.1111/1755-0998.12265.

[55] Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, *et al*. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology. 2017; 26: 5872–5895. https://doi.org/10.1111/mec.14350.

[56] Bru D, Martin-Laurent F, Philippot L. Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. Applied and Environmental Microbiology. 2008; 74: 1660–1663. https://doi.org/10.1128/AEM.02403-07.

[57] Stadhouders R, Pas SD, Anber J, Voermans J, Mes THM, Schutten M. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay. The Journal of Molecular Diagnostics. 2010; 12: 109–117. https://doi.org/10.2353/jmoldx.2010.090035.

[58] Hamada K. Taxonomic and ecological studies of the genus Hypomesus of Japan. Memoirs of the Faculty of Fisheries Hokkaido University. 1961; 9: 1–55.

[59] Maruyama A, Nakamura K, Yamanaka H, Kondoh M, Minamoto T. The release rate of environmental DNA from juvenile and adult fish. PLoS ONE. 2014; 9: e114639. https://doi.org/10.1371/journal.pone.0114639.