

RESEARCH

Open Access



Development of environmental DNA metabarcoding primers for marine mollusks and comparison with published primers

Xiaojing Shi¹, Yihui Jiang¹, Ling Cao² and Cong Zeng^{1,2*}

Abstract

Monitoring mollusk biodiversity is a great challenge due to their large diversity and broad distribution. Environmental DNA (eDNA) technology is increasingly applied for biodiversity monitoring, but relevant studies on marine mollusks are still limited. Although previous studies have developed several pairs of primers for mollusk eDNA analyses, most of them targeted only a small group of mollusks. In this study, seven primers were designed for the mollusk community and validated and compared with eight pairs of published primers to select the best candidates. After *in silico* test, MollCOI154 and MollCOI255 primers showed non-specific amplification, and same results were also obtained in published primers (COI204, Sepi, and veneroida). Moll12S100, Moll12S195 and Moll16S primers failed to amplify across all genomic DNA from selected mollusk. Except Moll16S, all developed and two published (unionoida and veneroida) primers were successfully amplified on four eDNA samples from Yangtze River estuary. After annotation of the amplified sequences, MollCOI253 showed higher annotation of the amplification results than the other primers. In conclusion, MollCOI253 had better performance in terms of amplification success and specificity, and can provide technical support for eDNA-based research, which will be beneficial for molluscan biodiversity investigation and conservation.

Keywords eDNA, Primer development, Mollusk, Metabarcoding, Mitochondrial genes

Introduction

Mollusks are the second largest group of invertebrates and are widely distributed in freshwater and marine ecosystems, playing important roles in the ecosystem, such as modifying the sediment and purifying water [1]. Currently, nearly 100,000 species of mollusks have been reported in the world, and they are considered as the key taxa for marine biodiversity in different marine areas

[2]. Additionally, mollusks are important components of global fisheries products, and the world production of major molluscan species has shown a steady increase since 1950 [3]. On the other side, due to their poor mobility and sensitivity to environmental changes, mollusks are more susceptible to the impacts of global warming and human disturbances compared to other marine organisms [4], which makes it urgent to monitor and protect their biodiversity.

Traditional biodiversity monitoring is mainly based on obtaining data on community composition through trawling, diving and the use of underwater cameras, which are costly, time-consuming and highly dependent on the involvement of taxonomic experts, who are currently becoming increasingly scarce [5]. Therefore,

*Correspondence:

Cong Zeng

congzung@sytu.edu.cn

¹School of Oceanography, Shanghai Jiao Tong University, Shanghai 200030, China

²State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361104, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

these limitations of traditional survey pose significant challenges to biodiversity research [6]. In recent years, environmental DNA (eDNA) technology has emerged as a novel approach for studying community compositions and species detection in ecological and biodiversity research [7]. Unlike traditional methods, eDNA technology is non-invasive, which does not damage the target species and the ecosystems [8]. Besides, its high specificity and sensitivity make eDNA particularly effective for studying hard-to-capture, invasive, and rare species [9]. With an increasing concern on the marine conservation and sustainable fisheries, eDNA technology is particularly important for the study of aquatic organisms [10]. As the cost of high-through sequencing decreased, the metabarcoding, a combination of DNA barcoding and high-through sequencing, provided solutions for the simultaneous detection of multiple species as well as relative abundance, which is more suitable for community analysis. Environmental DNA-based metabarcoding (eDNA metabarcoding) technology was expected to further improve the detection efficiency of aquatic organisms and contribute to the establishment of a standardized monitoring technology system for freshwater and marine ecosystems.

While in recent years eDNA technology has been applied to mollusk studies, existing primers target specific groups, such as Unionida and Venerida in freshwater and Cephalopoda in marine ecosystems [10]. A total of eight primer pairs were developed in previous publications for molluscan eDNA detection or diversity analysis, of which four primers (NZMS, NADH, COI204, and Sepi) were used for single-species detection, and four primers (16 S rRNA, unionoida, veneroida, and Ceph18S) were used for diversity analysis. Among the single-species target primers, *Cytochrome b* (*Cytb*) [11], *NADH* [12] and *Cytochrome c oxidase* subunit 1 (*COI*) [13, 14] gene were the candidate regions for primer design. For diversity analysis, metabarcoding primers (16 S rRNA, unionoida, veneroida and Ceph18S) were reported to be able to be used in a wide range of molluscan community analyses [15–17], and these primers were mainly targeting *16 S* and *18 S* rRNA regions. However, *16 S* rRNA, unionoida and veneroida primers were applied in the analyses of freshwater bivalves but not in marine mollusks, and Ceph18S was only designed for the biodiversity study of cephalopod species [18]. Since most mollusks inhabit the marine environment, further development of eDNA primers for mollusks is still required to include more species to meet the urgent needs of mollusk biodiversity surveys.

The aim of this study was to develop new environmental DNA primers and evaluate their effectiveness with those already developed in previous studies. For this, Chinese offshore mollusks were selected as candidates

for primer design and testing, and the mitochondrial genes *COI*, *12s* and *16 S* were selected as target gene regions. The amplification performance of designed primers from different gene regions, as well as published primers, were compared by *in silico* PCR, genomic DNA amplification and eDNA amplification. The results are expected to identify suitable primers for environmental DNA-based mollusk biodiversity surveys, contributing to future marine biodiversity conservation efforts.

Materials and methods

Sequence collection for primer design

Given the vast diversity and uneven distribution of mollusks, the specificity (number of non-target species amplified), universality (number of target species amplified) and robustness (amplification success rate) of the developed primers were mainly tested by mollusks from the coastal areas of China. Candidate species were selected based on the *Atlas of Marine Mollusks in China* [3], and further checked the existence of genetic sequences of these species in NCBI database (<https://www.ncbi.nlm.nih.gov/>). To ensure sufficient sequence length for primer region screening, only species with complete mitochondrial genomes in the NCBI database were chosen. After screening, a total of 213 molluscan species were obtained and their sequences were downloaded from NCBI using Geneious (versionR11) (<https://www.geneious.com/>). Based on previous studies and initial alignments, the full-length gene regions of *COI*, *12 S* and *16 S* were selected as candidate regions for primer design. Multiple alignments for these three genes across the 213 species were performed using MAFFT alignment in Geneious. Alignments were also further visualized using Geneious [18], and the visualized alignments were then used to identify primer binding regions and variant sites.

Primer development

Primer regions were identified considering three crucial requirements: (1) because most eDNA is easily degraded and the sequencing platform (mainly Illumina) usually requires the sequencing fragments to be shorter than 500 bp [19], and further shortened due to the additional adapters at both ends of the amplified fragments required during sequencing and library construction. As such and based on previous studies, in which the amplification products of environmental DNA usually did not exceed 300 bp in length and to obtain a sufficient number of variable sites within the sequence variants, the length of the target region was set to be between 100 and 300 bp [19]; (2) to provide a good taxonomical resolution, the target regions should include sufficient inter-specific DNA variation for all target species [20]; and (3) to successful conduct PCR amplification, conservative regions

for binding PCR primers (more than 18 bp in length) across all target species should be located at both ends of the targeted regions [21]. Primers were designed using Geneious accounting for G/C contents (40–60%), melting temperature (T_m : 50–60 °C), primer length (18–27 bp) and product size (100–300 bp).

Analysis of *in silico* PCR

The mollusk metabarcoding developed primers were evaluated using *in silico* PCR implemented in the Primer-BLAST program from NCBI website (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome). Specific settings included: 1) PCR product size within 100–500 bp; 2) primer melting temperatures between 57°C–63°C; 3) a maximum T_m difference of 2°C between forward and reverse primers; 4) a maximum of 3 mismatches between each primer and the target sequence; and no mismatches in the last two nucleotides at the 3' end of the primer [22]. The Primer-BLAST search was conducted against the nr nucleotide database in NCBI. Specificity, universality and robustness results of *in silico* PCR were compared among developed and published primers.

Tests in genomic DNA amplification

To further confirm the universality of developed primers, genomic DNA was extracted from 24 species across three mollusk classes: 6 Gastropoda (*Haliotis discus*, *Nodilitorina pyramidalis*, *Babylonia lutosa*, *Neptunea cumingi*, *Rapana bezoar*, *Rapana rapiformis*), 16 Bivalvia (*Mactra quadrangularis*, *Mactra veneriformis*, *Mactra antiquata*, *Mactra chinensis*, *Tegillarca granosa*, *Scapharca subcrenata*, *Paphia undulata*, *Meretrix meretrix*, *Grassostrea gigas*, *Azumapecten farreri*, *Mytilus edulis*, *Perna viridis*, *Musculus senhousi*, *Trichomya hissutus*, *Sinonovacula constricta*, *Solen strictus*) and 2 Cephalopoda (*Octopus vulgaris*, *Loligo chinensis*). In addition, two decapod species, *Penaeus monodon* and *Ocypoda ceratophthalma*, from the East China Sea, were also selected to verify the specificity of the mollusk primers through PCR amplification. Also, according to previous studies, 16SrRNA primer had a better performance [15] and it was also selected to test in all specimens. Total genomic DNA was extracted from each species, preserved in 95% ethanol, using the MolPure Cell/Tissue DNA Kit (Yeasen Biotechnology, China). The extracted DNA was stored in -20 °C.

PCR was carried out in a 50.0 µl reaction volume containing 20.0 µl sterile distilled water, 25.0 µl 2 × Gflex PCR buffer, 2.0 µl of each primer, and 1.0 µl DNA template. The thermal cycle profile after an initial 5 min denaturation at 94°C was as follows: denaturation at 94°C for 30 s, annealing at 53°C for 30 s and extension at 72°C for 30s with the final extension at 72°C for 10 min. The PCR products were checked by 1.80% agarose gel

electrophoresis to check whether the amplification was successful. The amplified products were further sequenced for verification by Tsingke Biotechnology Co. Ltd. (Beijing, China).

Environmental DNA amplification tests

To test *in-situ* sample amplification, four seawater samples were collected at Yangtze Estuary in the East China Sea, where the presence of mollusks is well documented. Sampling was conducted at four sites (30.70°N 123.21°E, 31.32°N 121.75°E, 29.17°N 122.77°E, 28.75°N 122.47°E) covering Yangtze estuary onboard of R/V “Zheyuke2” and “Runjiang1” implementing the open research cruise NORC2021-03. Water sampling was collected from October to November 2021.

For environmental DNA collection, 1 L water were collected at each site from the bottom layer (1–3 m above the seabed) and stored at 4 °C. Water samples were first filtered through a 1 µm glass microfiber filter (Whatman, UK) to remove the large particles and organisms, and then filtered through a 0.45 µm mixed-fiber filter (Jinteng, China) within 10 h after sampling. The filter papers were preserved in ethanol at -20 °C. The eDNA extraction, library construction and pair-end 2×250 bp sequencing were completed in Shanghai Personal Biotechnology Co., Ltd. (Shanghai, China).

Bioinformatics analyses were mainly performed with QIIME 2 (2023.2) [23]. Briefly, raw sequence data were demultiplexed using the demux plugin followed by primers cutting with cutadapt plugin [24]. Sequences were then merged, filtered and dereplicated using functions of fastq_mergepairs, fastq_filter, and derep_full-length in Vsearch. All the unique sequences were then clustered at 98% (via cluster_size) followed by chimera removing (via uchime_denovo). Finally, the non-chimera sequences were re-clustered at 97% to generate Molecular Operational Taxonomic Unit (MOTU) representative sequences and MOTU frequency table, and the MOTU clustering procedure following the Vsearch (v2.13.4) [25]. Taxonomy was assigned to MOTUs using BROCC [26] against the mollusk sequences from mollusk nr Database in NCBI (accessed in March 2023).

Results

Based on the selection criteria, a total of 213 mollusk species were kept for primer design, including 125 bivalves, 27 cephalopods, and 61 gastropods. Three gene sequences, *COI*, *12 S* and *16 S*, were selected from the mitochondrial genomes of these mollusks, and the sequences of these three genes were separately aligned to identify conservative regions. After alignment, seven pairs of primers were designed from these regions, including three pairs from *COI*, three pairs from *12 S* and one pair from *16 S*. To find the most effective primers,

these seven primer pairs, as well as eight previously published primer, were analyzed to compare their amplification performance.

New developed primers.

Seven primer pairs were identified, each targeting a conservative region flanking a hyper-variable region within the respective genes. In the *COI* region, three primers, MollCOI253, MollCOI154 and MollCOI255, were developed, and all these primers shared a conservative region (Fig. 1a; Table 1). All three pairs exhibited an annealing temperature of approximately 57°C, with PCR product lengths ranging from 154 to 255 bp. For the *I2 S* region, Moll12S150, Moll12S195 and Moll12S100 were designed, and all these primers did not share binding regions (Fig. 1b; Table 1). The annealing temperature and PCR product lengths for these primers varied from 56 to 59 °C and 100–195 bp, respectively. The primer Moll16S was the only primer designed targeting the *I6 S* gene, and the annealing temperature and PCR product lengths were 60 °C and 228 bp, respectively (Fig. 1c; Table 1). These seven primer pairs basically met the eDNA amplification requirements in terms of length and annealing temperature.

Tests of *in silico* PCR amplification

Numbers of amplified species from *in silico* PCR amplification were presented in Table S1 for both newly developed and published primers, and the results showed that robustness, specificity and universality varied considerably among the primers assessed. Among the newly developed primer pairs, MollCOI255 had the highest robustness, while Moll12S150 had the lowest (Fig. 2). Both MollCOI154 and MollCOI255 amplified non-molluscan taxa with a not insignificant percentage of 12.63% and 17.68%, respectively, indicating that these two pairs of primers had less specificity (Table S1). Furthermore, distinct amplification preferences were observed for different primers, except for MollCOI253. The MollCOI255 primers and Moll12S100 amplified more gastropods (69.86% and 96.46%, respectively) (Table S1), while the other primers favored bivalves. Among these primers, only MollCOI154 and MollCOI253 were successful in amplifying all classes of mollusks (Table S1).

Among reported primers, NADH (targeted *Lampsilis siliquoidea*), 16SrRNA (targeted Bivalvia) and unionida (targeted Unionida) exclusively amplified bivalves, while Ceph18s (targeted Cephalopoda) exclusively amplified cephalopods, suggesting that these primers had a high specificity but low universality for mollusks (Table S1). COI204 (targeted *Octopus vulgaris*), Sepi (targeted Cephalopoda) and veneroida (targeted Veneroida) primers demonstrated broader mollusk coverage but they also amplified non-molluscan taxa, particularly COI204 (25.41% non-mollusks). NZMS, although it exhibited

good specificity, had the lowest species coverage among all examined primers (Fig. 2). Due to their lack of specificity in *in silico* PCR test, COI204, MollCOI154 and MollCOI255 were excluded from genomic and environmental DNA amplification assays.

Testing developed primers using genomic DNA

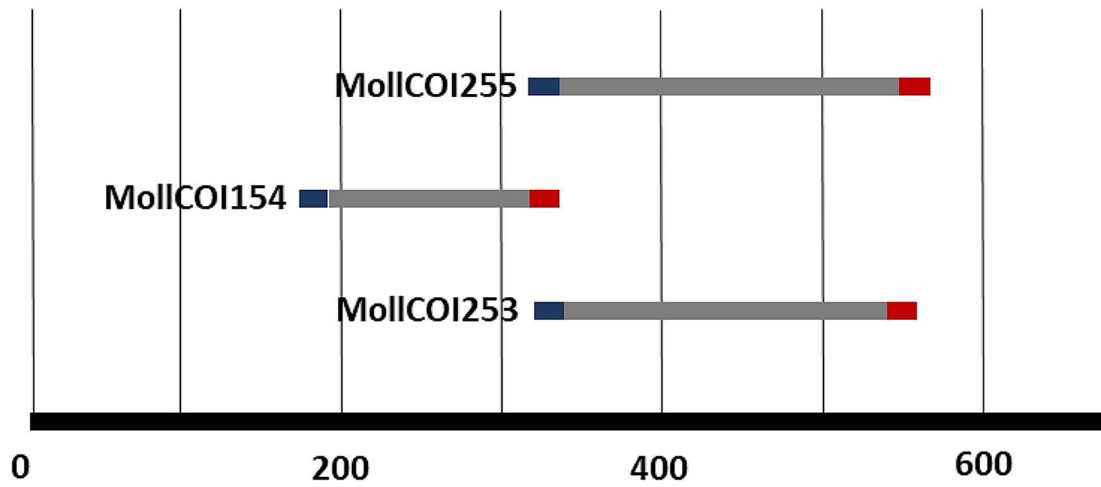
After PCR amplification with genomic DNA, 16SrRNA was successfully amplified in all tested species, indicating the capability and suitability of DNA templates for assessing primer amplification (Table 2; Table S2). Apart from 16SrRNA, Moll12S150 and MollCOI253 primers successfully amplified all 24 mollusks from 15 families, and Moll12S100 was the least successful (45.83% success rate). Moll12S195 amplified 66.67% and 75.00% of the genomic DNA successfully for Gastropoda and Bivalvia, respectively, whereas Moll16S amplified these two Classes less successfully (50.00% and 43.75%). In addition, neither primer pair was able to amplify cephalopods. In contrast, Moll12S100 was able to amplify the all sample of gastropods and cephalopods, while bivalves were amplified very inefficiently (18.75%). These differences reflected amplification bias between primers. None of the examined primers amplified decapods except 16SrRNA, which further implied that these primers had a high specificity (Table 2). Sanger sequencing confirmed that the PCR product lengths of all tested primers closely matched the expected lengths, and further BLAST of the sequenced PCR products against the nr database did not reveal any non-target regions of amplification.

Testing developed primers using environmental DNA

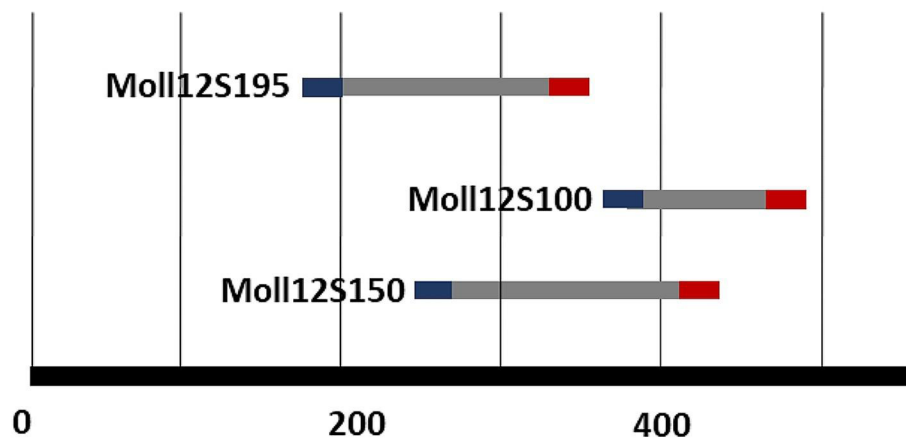
After extraction of the eDNA, five designed primers (MollCOI253, Moll12S150, Moll12S100, Moll12S195, Moll16S) were tested using eDNA PCR amplification. To better compare the effect of the newly developed primers, two pairs of published metabarcoding primers (unionoida and veneroida) were also used for amplification and high-throughput sequencing. Moll16S failed to amplify and perform the sequencing library construction, while the rest of the primers successfully passed quality control for high-throughput sequencing. In addition, the results of two samples of the high-throughput sequencing for Moll12S195, Moll12S100, unionoida and veneroida had very few reads, so the results for these two samples were not included in the analysis as well.

The high-throughput sequencing data for MollCOI253, Moll12S150, Moll12S195 and Moll12S100 amplicons comprised a total of 116,750, 351,475, 567,992 and 384,919 raw reads, respectively, which generated 485, 8031, 952 and 6 MOTUs. After annotation using the nr database, 13,749 reads of MollCOI253 amplicon were mollusk (11.78%), and 12 species were detected, of which the largest proportion was *Sulcospira paludiformis*

(a) *COI*



(b) *12S*



(c) *16S*

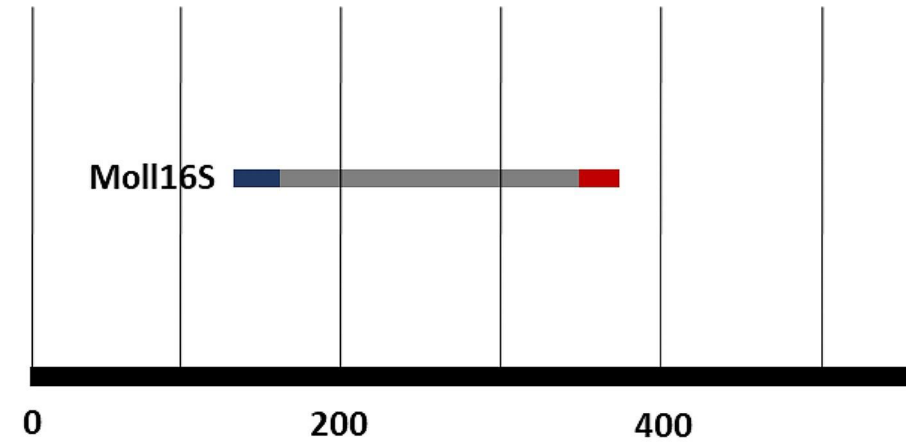


Fig. 1 Target amplification regions of new primers for *COI*(a), *12S*(b) and *16S*(c). The black lines are the three genes and the numbers below the lines are the sequence lengths. The gray line is the target amplification region, and the blue and red fragments at the ends of the gray line are forward and reverse primers

Table 1 Primers tested in this study

Target gene	Primer name	Primers (5'~3')	Product size(bp)	Annealing Temperature	GC%	Source
COI	MollCOI154	F-TGGGGGTTTTGGTAATTGGT	154	57°C	45.00	Present study
		R-TCAACCAGTACCAGCTCCA			52.63	
	MollCOI253	F-GGAGTAGGAACTGGTTGGAC	253	57°C	55.00	
		R-CAGCTGCTAACACAGGCA			55.56	
	MollCOI255	F-TGGAGCTGGTACTGGTTGA	255	57°C	52.63	
		R-GCCCCAGCTAAAACAGGTAT			50.00	
12S	Moll12S100	F-TTGATACCGTCGTCGTCAG	100	57°C	50.00	
		R-AGCTGCACCTTGATCTGAC			52.63	
	Moll12S150	F-TATGCTTGCCGGGCAACT	150	59°C	55.56	
		R-ACGGCCATACACCAACTGA			52.63	
	Moll12S195	F-GTATTGCCGTTGTCAGCTT	195	56°C	47.37	
		R-CCTTACTCCTAAGTTCACCTTC			45.45	
16S	Moll16S	F-GTCTGTGAATGGTTGACGAG	228	60°C	50.00	
		R-TTGTGCCCCAGCCAAAAC			57.89	
COI	COI204	F-TGTTACAGCTCACGCATTTGTT	204	60°C	40.91	[13]
		R-CCGGTACCTGCACCTCTTTC			60.00	
	Sepi	F-CACCAGACATAGCCTTCC	155	54°C	55.56	[14]
		R-GCCAGCATGAGATAGATTAC			45.00	
NADH	NADH	F-TCGAGCCATAGCTCAAACCA	147	59°C	50.00	[12]
		R-GCGAGTGGTAGTGAAAGAGT			50.00	
Cytb	NZMS	F-TGTTTCAAGTGTGCTGGTTT	92	56°C	40.00	[11]
		R-CAAATGGGCTAGTTGATTCTTT			36.36	
16S	16S rRNA	F-TGAGCGTGCTAAGGTAGC	360	60°C	55.56	[15]
		R-AGCCAACATCGAGGTCGC			61.11	
	unionoida	F-GCTGTTATCCCCGGGTAR	170	58°C	61.11	[16]
		R-AAGACGAAAAGACCCCGC			55.56	
	veneroida	F-CSCTGTTATCCCYRCGGTA	159	56°C	52.63	[16]
		R-TTDTAAAAGCCGAGAAGACCC			42.86	
18S	Ceph18S	F-CGCGGGCGCTACATATTAGAC	173–235	61°C	55.00	[17]
		R-GCACTTAACCGACCGTCGAC			60.00	

(3.93% of total abundance, including MOTUs that could not be annotated), followed by *Mopalia spectabilis* (3.54%) and *Choanomphalus hyaliniiformis* (2.42%) (Table 3, Table S2). For the Moll12S150 amplicon, 418 reads were annotated as mollusks (0.12%), and 3 species were found, including *Crassostrea ariakensis* (0.11%), *Radix acuminata* (0.003%) and *Doryteuthis opalescens* (0.002%) (Table S3). However, Moll12S195 amplicon had only 9 reads identified as mollusks, belonging to Lucinidae. Moll12S100 amplicon did not yield any annotated mollusk species across all MOTUS in the four samples (Table 3). Although most MOTUs from developed primers could not be annotated to specific mollusk species, those that were annotated largely represent mollusk species that have been previously reported from coastal areas of China.

When comparing the annotation effect of all the primer amplicons, it was found that none of the amplification products were found to contain Aplacophora, which could mean that none of these primers were able to amplify Aplacophora or that the sampling point region did not have Aplacophora. In addition, the results

of amplification showed that most of examined primers also amplified Cephalopoda and Polyplacophora poorly, with only one pair of primer products detecting them (Table 4). Unlike the developed primers, even though amplifications of unionoida and veneroida was successful and relatively good sequencing results were obtained, the majority of MOTUs were annotated as microorganisms (mainly bacteria). With the exception of Moll12S100, all developed and published primers had a relatively large proportion of products that could not be annotated (42–93%), implying that the use of the nr database as a reference data for mollusks remained insufficient (Table 4).

Based on the results of eDNA amplification and annotation, MollCOI253 had a relatively high amplification success rate (all samples succeeded), and produced raw reads and MOTUs similar to those reported in other studies. Although the percentage of non-annotatable MOTUs was still not low, similar to the other primers, it could be improved by constructing specific annotation database. Since none of the examined primers provided a perfect solution, it can be assumed that MollCOI253 is

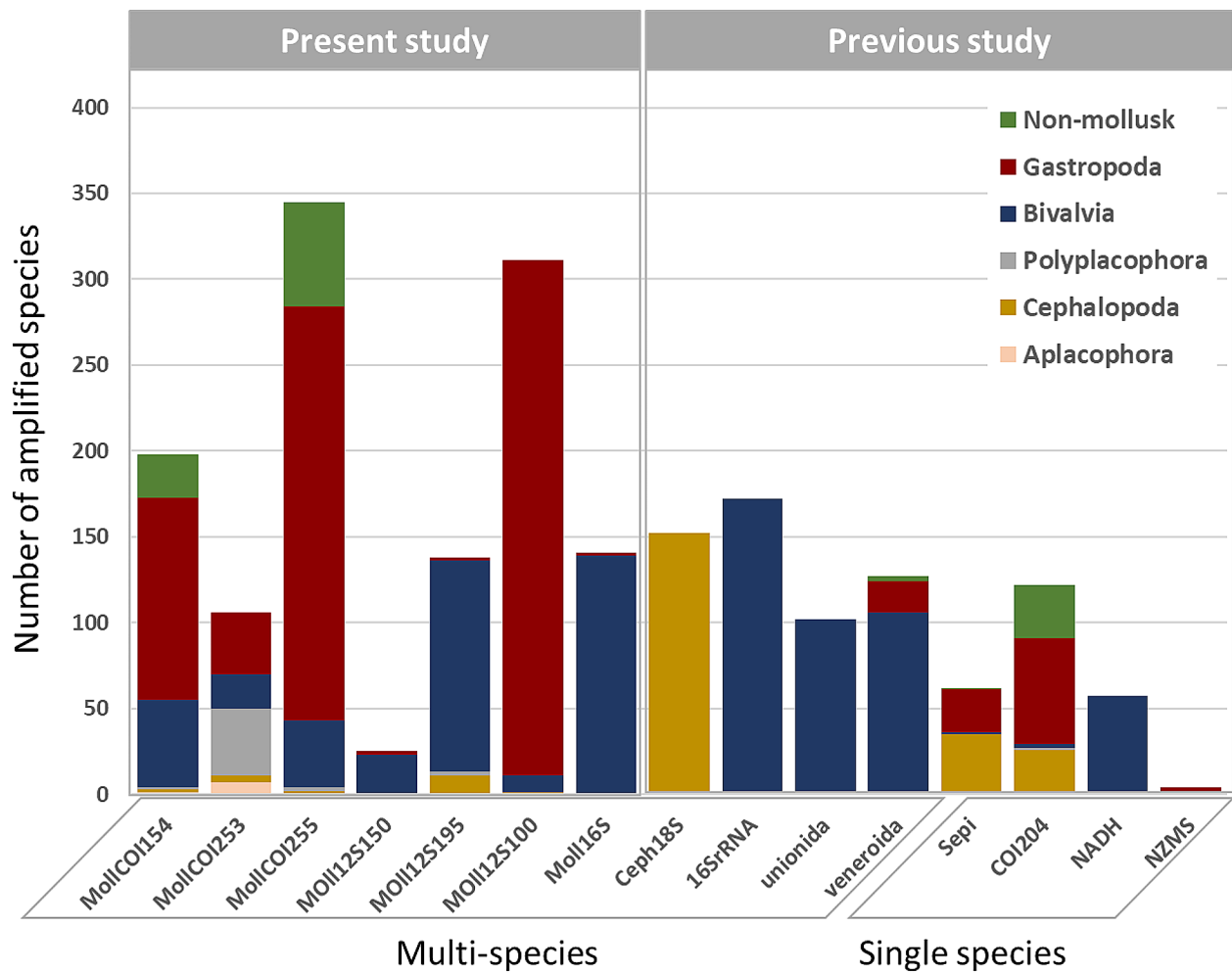


Fig. 2 Comparison of new primers and previous primers *in silico* PCR amplification results

Table 2 Primer validation results based on genomic DNA amplification

Primer name	Bivalvia	Cephalopoda	Gastropoda	Non-mollusk
MollCOI253	16/16 (100%)	2/2 (100%)	6/6 (100%)	0
Moll12S150	16/16 (100%)	2/2 (100%)	6/6 (100%)	0
Moll12S195	12/16 (75%)	0	4/6 (75%)	0
Moll12S100	3/16 (18.75%)	2/2 (100%)	6/6 (100%)	0
Moll16S	7/16 (43.75%)	0	3/6 (50%)	0
16SrRNA	16/16 (100%)	2/2 (100%)	6/6 (100%)	2/2 (100%)

a relatively good candidate primer for molluscan eDNA studies.

Discussion

Given the critical importance of metabarcoding primers in eDNA based biodiversity detection, a pair of primer with good performance is essential for environmental DNA research [27]. However, despite the increasingly wide application of eDNA metabarcoding in mollusk community surveys across diverse ecosystems, the effective identification of all mollusk species is still poorly known. In this study, seven pairs of primers were developed and tested in *in silico* PCR and PCR amplification with genomic and environmental DNA. Newly developed primers were compared with published primers to identify the most effective candidate primer for marine mollusk biodiversity surveys. Among the newly-developed primers, MollCOI253 was found to be the best

Table 3 High-throughput sequencing results based on environmental DNA amplification

Primer name	Samples	Raw reads Total (range)	MOTUs	Mollusks %	No. Annotated Mollusks	Source
MollCOI253	4	116,750 (23,747 – 35,831)	4,85	11.78%	12	Present study
Moll12S150	4	351,475 (83,344 – 92,908)	8,031	0.12%	3	
Moll12S195	2	567,992 (198,732 – 369,260)	952	0.001%	1	
Moll12S100	2	384,919 (159,907 – 225,012)	6	0	0	
unionoida	2	139,868 (68,264 – 71,604)	347	0	0	
veneroida	2	225,221 (109,582 – 115,639)	1,521	0	0	
16SrRNA	6	461,185	-	26.69%	8	[15]
Ceph18S	19	9,936,758 (23,311 – 50,066)	-	-	15	[40]
Unionoida + veneroida	15	-	-	-	42	[16]

Note: - meant information was not available in the publications

Table 4 Amplicon annotation results for the examined primers after pooling all samples

Taxon	MollCOI253	Moll12S150	Moll12S195	Moll12S100	unionoida	veneroida
Aplacophora	0	0	0	0	0	0
Cephalopoda	0	1 (7, 0.00%)	0	0	0	0
Polyplacophora	1 (4,138, 3.54%)	0	0	0	0	0
Bivalvia	2 (611, 0.52%)	14 (400, 0.11%)	1 (9, 0.00%)	0	0	0
Gastropoda	15 (9,000, 7.71%)	1 (11, 0.00%)	0	0	0	0
Non-mollusk	70 (3,746, 3.21%)	1,036 (82,214, 23.39%)	101 (40,203, 7.08%)	6 (384,919, 100%)	324 (14,057, 10.05%)	406 (131,022, 58.17%)
Unassigned	397 (99,255, 85.02%)	6,979 (268,843, 76.49%)	850 (527,780, 92.92%)	0	23 (125,811, 89.95%)	1115 (94,199, 41.83%)

Note: Numbers in table were expressed as number of MOTUs (Abundance of these MOTUs, proportion of all MOTUs)

candidate, demonstrating good specificity and universality for marine mollusks. This primer is expected to provide valuable technical support for mollusk biodiversity surveys, and it would be also beneficial for their conservation in the range of Chinese sea.

Comparison of different methods for assessing primer effects

Although the *in silico* method allowed for rapid and high-volume screening of amplified species [28], the primer binding was tolerant of certain mismatches, especially in the 5' end region [29], which could lead to discrepancies between actual amplification and *in silico* PCR results. This was further confirmed in the present study. For example, 16SrRNA showed strong specificity in the *in silico* PCR analysis, but non-specific amplification was found in genomic amplification. In contrast, Moll12S100 and Moll12S195 primers amplified a higher number of species than MollCOI253 in the *in silico* PCR analysis, while the former two amplified fewer species than the latter in the genomic amplification. Furthermore, the amplifications of eDNA showed that the performance of primers was also different from *in silico* PCR tests, with the primer that amplified the most MOTUs being Moll12S150, which was consistent with the genomic amplification results but not with the *in silico* PCR results.

Despite some differences in the assessment of specificity, universality and robustness among the three

methods, all methods confirmed the presence of amplification preferences for most tested primers. For example, Moll12S100 and Moll12S195 favored gastropods and bivalves in both *in silico* PCR and genomic DNA amplification tests, whereas MollCOI253 amplifies more Polyplacophora in both *in silico* tests and environmental DNA amplification compared to the other primers. Therefore, the disagreements and agreements between these methods suggested that eDNA primer performance assays may need to be evaluated comprehensively by a combination of multiple methods.

Comparison of designed and published primers

The primer requirements for environmental DNA metabarcoding techniques usually depend on their amplification specificity and universality [30]. In the present study, MollCOI154, MollCOI255, COI204, Sepi, and veneroida primers exhibited non-target amplifications in *in silico* PCR amplifications. Additionally, 16 S rRNA primer was further excluded from genomic DNA amplification due to cross-amplification in non-molluscan species. This suggests that these primers lacked specificity and may result in Type I errors, amplifying non-target DNA from samples that contain no eDNA of the target species [30], thus impacting the accuracy of biodiversity estimates.

On the other hand, when considering universality, Moll12S150 and NZMS amplified a much lower number of species than the other primers. Also, the primers NADH, Moll12S150, 16 S rRNA, Ceph18S and veneroida,

did not perform well in terms of universality and only amplified a certain taxon. Genomic DNA amplification revealed that Moll12S195, Moll12S100 and Moll16S failed to amplify across all the tested samples. These limited coverages of amplified species had the potential to bias biodiversity estimates due to primer binding preferences [30]. After different assessments, MollCOI253 was superior to the other primers considering both specificity and species amplification coverage (Table 5).

Whether a barcode region contains an appropriate conservative sequence region directly determines whether primers based on that region can amplify all taxa. If there is no conservative region within the region, a proportion of taxa may remain unidentified following amplicon-based metabarcoding [31]. Although previous studies reported that “universal” *COI* primers that amplify barcoding regions anneal to primer-binding sites that were poorly conservative across gastropods [32], the primers designed in this study based on the *COI* region showed relatively good coverage for *in silico* amplification and genomic DNA amplification in gastropods, suggesting that the *COI* region remains a viable candidate for metabarcoding in mollusks. Although it could not be excluded that it is related to the conservativeness of the region, the relatively low number of reads amplified by MollCOI253 compared to other primers may be related to the length and size of the DNA fragments it amplifies [19]. MollCOI253 amplified a longer DNA fragment compared to the amplification products of the other primers, and therefore may yield relatively fewer reads at similar sequencing volumes. Although longer DNA fragments were easier to degradation in nature, it is clear from the results that MollCOI253 amplifies fewer but more efficient reads under the same conditions [33]. Degenerate primers containing degenerate base pairs in

the primer sequences were expected to increase the species of amplification and improve primer performance [41], but this may require the selection of non-conservative regions for primer re-design in the future.

Annotation effects of MollCOI253 on eDNA samples

The highest percentage of annotations was from the amplification of MollCOI253, which accounted for only 11.78% of the reads. Compared with 16SrRNA in previous studies, the ratio of annotation of mollusks by MollCOI253 was lower than 26.69% of 16SrRNA [15]. This small proportion of annotated MOTUs could be explained by the incompleteness of reference databases [34]. Currently, the reference sequences used in most mollusk metabarcoding analysis were from the NCBI database [35], and insufficient sequence data could result in low efficiency of mollusk classification [36]. Therefore, to use eDNA technology for more accurate and effective studies of mollusk biodiversity, further efforts are needed to establish a more complete reference sequence database of mollusks.

In this study, some common East China Sea mollusk species were undetected in the eDNA samples (e.g., *Rapana bezoar*, *Coelomactra antiquate*, *Octopus vulgaris*), and the reason for this may be related to the fact that eDNA of these mollusks were not collected, failed to be amplified, or were not annotated [5]. Therefore, further improvements in eDNA extraction [37], amplifications [38], and annotated reference database [36] should be made in the future. For instance, adding proteinase K could significantly improve DNA extraction [39], and the use of multiple primers for amplification could also improve the accuracy and detection rate of eDNA technology [38].

Table 5 Summary of all primers' faults in application of environmental DNA

Primer name	in silico PCR	Genomic DNA PCR	Environmental DNA PCR	High-throughput sequencing
MollCOI154	Cross amplification	-	-	-
MollCOI253				
MollCOI255	Cross amplification	-	-	-
Moll12S150	Limited coverage			
Moll12S195		Partial amplification		
Moll12S100		Partial amplification		Annotation failure
Moll16S		Partial amplification	Amplification failure	-
COI204	Cross amplification	-	-	-
NADH	Limited coverage	-	-	-
NZMS	Poor coverage	-	-	-
Sepi	Cross amplification	-	-	-
16 S rRNA	Limited coverage	Cross amplification	-	-
Ceph18S	Limited coverage	-	-	-
Unionoida	Limited coverage	-		Annotation failure
Veneroida	Cross amplification	-		Annotation failure

Note: - meant analyses were not conducted in this study

Among the annotated results, two exotic species, *Powelliphanta patrickensis* and *Bursa scrobilator*, were found, both of which have never been reported offshore China [3]. Given their extremely low abundance (0.003–0.037%), it was possible that they were artificially transported to the sampling area (e.g., ballast water), or that they are recently introduced alien species. Also, it could not be excluded that it was the result of errors in the sequencing and assembly process. Although further investigation of MOTUs was needed for these species, these very low abundance or occasional MOTUs were often excluded from community diversity analyses [5]. Therefore, these unreported low abundance species did not have a significant impact on the results of community diversity analyses, especially beta diversity. However, if these species needed to be detected, it would be recommended that more specific primers should be designed for better detection.

Conclusions

Seven primer pairs were designed based on the *COI*, *12 S* and *16 S* genes in this study, of which MollCOI253, as the best candidate, outperformed other newly designed and published primers for molluscan eDNA studies in terms of versatility, specificity and robustness. The results of this study not only provided technical support for molluscan eDNA investigations, but also provide a reference for the selection of multiple molecular markers for future molluscan eDNA studies. The results of this study will be helpful for more effective investigations of molluscan biodiversity in the future.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-024-02265-8>.

Supplementary Material 1

Acknowledgements

We are grateful to Yue Liu, Shanghai Jiao Tong University, for his advice and the collection of mollusks. We would also like to thank the anonymous reviewers and editors for their comments and suggestions that improved our manuscript.

Author contributions

Conceptualization: Cong Zeng, Ling Cao, Xiaojing Shi; Methodology: Xiaojing Shi, Yihui Jiang; Formal analysis and investigation: Xiaojing Shi, Yihui Jiang; Writing - original draft preparation: Xiaojing Shi; Writing - review and editing: Cong Zeng, Ling Cao; Funding acquisition: Cong Zeng, Ling Cao; Supervision: Cong Zeng.

Funding

This study was supported by the Ministry of Science and Technology of China (Grant No. 2022YFC3102400), National Natural Science Foundation of China (Grant No. 42206082), Oceanic Interdisciplinary Program of Shanghai Jiao Tong University (Grant No. SL2022PT110), New Faculty Start-Up Program of Shanghai Jiao Tong University (22X010503822), Blue Planet Found (OPF & WWF) (Grant No. PORO001426).

Data availability

Raw sequencing data from the MollCOI253 (SRR27338740–41, SRR27338732–33), Moll12S150 (SRR27338728–31), Moll12S195 (SRR27338726–27), Moll12S100 (SRR27338738–39), unionoida (SRR27338736–37) and veneroida (SRR27338734–35) sequencing are available in the NCBI SRA database under bioproject (PRJNA1049577).

Declarations

Ethics approval

All collection and use of the mollusk and decapod samples for this study were performed following local legislation and institutional and/or national guidelines for the care and use of animals and were approved by the Institutional Animal Care and Use Committee of Shanghai Jiao Tong University.

Ethics declarations

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 8 December 2023 / Accepted: 24 May 2024

Published online: 31 May 2024

References

- Dillon RT. The ecology of freshwater molluscs. Cambridge: Cambridge University Press; 2000. pp. 22–6. <https://doi.org/10.1017/CBO9780511542008>.
- Zhang JL, Zhang SQ, Jiao YY. Progress on marine molluscan biodiversity in the Indo-Pacific convergence region. *Oceanologia et Limnologia Sinica*. 2021;52:487–97. <https://doi.org/10.11693/hyh20200700212>.
- Zhang SP. Atlas of marine mollusks in China. Beijing: China Ocean; 2008.
- Bai ZY, Wen PC, Yuan L, Wang GL, Li JL. Germplasm resources situation and protection & utilization of freshwater molluscs in China. *J Fisheries China*. 2022;46:149–57. <https://doi.org/10.11964/jfc.20210212638>.
- Komai T, Gotoh RO, Sado T, Miya M. Development of a new set of PCR primers for eDNA metabarcoding decapod crustaceans. *Metabarcoding Metagenomics*. 2019;3:1–19. <https://doi.org/10.3897/mbmg.3.33835>.
- Lim NKM, Tay YC, Srivathsan A, Tan JWT, Kwik JTB, Baloğlu B, Meier R, Yeo DCJ. Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals species-rich and reservoir-specific communities. *Royal Soc open Sci*. 2016;3:160635. <https://doi.org/10.1098/rsos.160635>.
- Thomsen PF, Kielgast J, Iversen LL, Peter RM, 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>.
- Willerslev E. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*. 2003;300:791–5. <https://doi.org/10.1126/science.1084114>.
- Chen L, Wu L, Liu Y, Xu HG. Application of environmental DNA metabarcoding in ecology. *Acta Ecol Sin*. 2016;36:4573–82. <https://doi.org/10.5846/stxb201501150125>.
- Chen JP, Li XH, Zhou CH. Application of environmental DNA technology in the study of molluscs. *Anhui Agric Sci*. 2021;49:22–4. <https://doi.org/10.3969/j.issn.0517-6611.2021.14.006>.
- Goldberg CS, Sepulveda A, Ray A, Baumgardt J, Waits LP. Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshw Sci*. 2013;32:792–800. <https://doi.org/10.1899/13-046.1>.
- Sansom BJ, Sassoubre LM. Environmental DNA (eDNA) shedding and decay rates to model freshwater mussel eDNA transport in a river. *Environ Sci Technol*. 2017;51:14244–53. <https://doi.org/10.1021/acs.est.7b05199>.
- Mauvisseau Q, Parrondo M, Fernández MP, García L, Martínez JL, García-Vázquez E, Borrell YJ. On the way for detecting and quantifying elusive species in the sea: the Octopus vulgaris case study. *Fish Res*. 2017;191:41–8. <https://doi.org/10.1016/j.fishres.2017.02.023>.

14. Chen Z, Song N, Toshifumi M, Wu QQ, Gao TX. The eDNA collection method of Zhoushan coastal waters. *Acta Hydrobiol Sin.* 2020;44:50–8. <https://doi.org/10.7541/2020.007>.
15. Chen JP, Zhou CH, Shan OY, Huang XC, Wu XP. Universal primers screening and verification for environmental DNA metabarcoding of freshwater mussels in the Lake Poyang Basin. *Lake Sci.* 2021;33:1254–64. <https://doi.org/10.18307/2021.0425>.
16. Prie' V, Valentini A, Lopes-Lima M, Froufe E, Rocle M, Poulet N, Taberlet P, Dejean T. Environmental DNA metabarcoding for freshwater bivalves biodiversity assessment: methods and results for the western palearctic (European sub-region). *Hydrobiologia.* 2021;848:2931–50. <https://doi.org/10.1007/s10750-020-04260-8>.
17. de Jonge D, Merten V, Bayer T, Puebla O, Reusch TBH, Hoving HJT. A novel metabarcoding primer pair for environmental DNA analysis of Cephalopoda (Mollusca) targeting the nuclear 18S rRNA region. *Royal Soc Open Sci.* 2021;8:1–16. <https://doi.org/10.1098/rsos.201388>.
18. Elbrecht V, Leese F. PRIMERMINER: an R package for development and in silico validation of DNA metabarcoding primers. *Methods Ecol Evol.* 2017;8:622–6. <https://doi.org/10.1111/2041-210X.12687>.
19. Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Res.* 2011;39:e145. <https://doi.org/10.1093/nar/gkr732>.
20. Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, Kondoh M, Iwasaki W. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Soc Open Sci.* 2015;2:150088. <https://doi.org/10.1098/rsos.150088>.
21. Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bel-Jemain E, Besnard AL, Coissac E, Boyer F, Gaboriaud C, Poulet PJN, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin JM, Peroux T, Crivelli AJ, Olivier A, Acqueberge M, Le Brun M, Møller PR, Willerslev E, Dejean T. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol Ecol.* 2016;25:929–42. <https://doi.org/10.1111/mec.13428>.
22. Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kausserud H. ITS as an environmental DNA barcode for fungi: an in-silico approach reveals potential PCR biases. *BMC Microbiol.* 2010;10:189. <https://doi.org/10.1186/1471-2180-10-189>.
23. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodriguez AM, Chase, Caporaso JG. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* 2019;37:852–7. <https://doi.org/10.1038/s41587-019-0209-9>.
24. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.* 2011;17:1–10. <https://doi.org/10.14806/EJ.17.1.200>.
25. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ.* 2016;4:e2584. <https://doi.org/10.7717/peerj.2584>.
26. Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson KH, Koljalg U. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS ONE.* 2006;1:e59. <https://doi.org/10.1371/journal.pone.0000059>.
27. Mills LS, Pilgrim KL, Schwartz MK, McKelvey K. Identifying Lynx and other north American felids based on eDNA analysis. *Conserv Genet.* 2000;1:285–8. <https://doi.org/10.1023/A:1011574209558>.
28. Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessièrè J, Taberlet P, Pompanon F. An in silico approach for the evaluation of DNA barcodes. *BMC Genomics.* 2010;11:434. <https://doi.org/10.1186/1471-2164-11-434>.
29. Wright ES, Yilmaz LS, Ram S, Gasser JM, Harrington GW, Noguera DR. Exploiting extension bias in polymerase chain reaction to improve primer specificity in ensembles of nearly identical DNA templates. *Environ Microbiol.* 2014;16:1354–65. <https://doi.org/10.1111/1462-2920.12259>.
30. Freeland JR. The importance of molecular markers and primer design when characterizing biodiversity from environmental DNA. *Genome.* 2017;60:358–74. <https://doi.org/10.1139/gen-2016-0100>.
31. Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biol Lett.* 2014;10. <https://doi.org/10.1098/rsbl.2014.0562>.
32. Meyer CP. Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. *Biol J Linn Soc.* 2003;79:401–59. <https://doi.org/10.1046/j.1095-8312.2003.00197.x>.
33. Mächler E, Osathanunkul M, Altermatt F. Shedding light on eDNA: neither natural levels of UV radiation nor the presence of a filter feeder affect eDNA-based detection of aquatic organisms. *PLoS ONE.* 2018;13:e0195529. <https://doi.org/10.1371/journal.pone.0195529>.
34. Leray M, Knowlton N. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proc Natl Acad Sci USA.* 2015;112:2076–81. <https://doi.org/10.1073/pnas.1424997112>.
35. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank Nucleic Acids Res. 2007;35:D21–5. <https://doi.org/10.1093/nar/gkt1030>.
36. Locatelli NS, McIntyre PB, Therkildsen NO, Baetscher DS. GenBank's reliability is uncertain for biodiversity researchers seeking species-level assignment for eDNA. *Proc Natl Acad Sci USA.* 2020;117:32211–2. <https://doi.org/10.1073/pnas.2007421117>.
37. Majaneva M, Diserud OH, Eagle SHC, Hajibabaei M, Ekrem T. Choice of DNA extraction method affects DNA metabarcoding of unsorted invertebrate bulk samples. *Metabarcoding Metagenomics.* 2018;2:1–12. <https://doi.org/10.3897/mbmg.2.26664>.
38. Zhang GK, Chain FJJ, Abbott CL, Cristescu ME. Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. *Evol Appl.* 2018;11:1901–14. <https://doi.org/10.1111/eva.12694>.
39. Tsuji S, Yamanaka H, Minamoto T. Effects of water pH and proteinase K treatment on the yield of environmental DNA from water samples. *Limnology.* 2016;18:1–7. <https://doi.org/10.1007/s10201-016-0483-x>.
40. Véronique M, Oscar P, Till B, Thorsten BHR, Janina F, Julia S, Katja M, Julian BS, Henk-Jan H. Arctic Nekton uncovered by eDNA metabarcoding: diversity, potential range expansions, and pelagic-benthic coupling. *Environ DNA.* 2023;00:1–16. <https://doi.org/10.1002/edn3.403>.
41. Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly-related sequences. *Nucleic Acids Res.* 1998;26:1628–35. <https://doi.org/10.1093/nar/26.7.1628>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.