Experimental evaluation of species and genetic variability based on DNA metabarcoding from the aquatic environment: extra OTUs formed by NUMTS may reduce the diversity of ASVs

Sergei Turanov[1,2], Olesia Rutenko[1,3], and Marina Koltsova[3]

[1]A.V. Zhirmunsky National Scientific Center of Marine Biology, Far Eastern Branch, Russian Academy of Sciences (NSCMB FEB RAS), Vladivostok, Russia, 690041
[2]Chair of Water Biological Resources and Aquaculture, Far Eastern State Technical Fisheries University, Vladivostok, Russia, 69008
[3]Chair of Biodiversity and Marine Bioresources, Far Eastern Federal University, Vladivostok, Russia, 690090

November 4, 2022

Abstract

The data on the intraspecific genetic variation for monitoring and conservation of wild populations is an important link for the assessment of the organisms resistance to changing environmental conditions and anthropogenic pressures. The metabarcoding of DNA from the aquatic environment provides a gradual transition to non-invasive methods of biodiversity research, including within-species level. However, the degradation of DNA under UV light in the aquatic environment limits the choice of markers in favor of short standardized regions. Hence, the consequences of information loss when shifting from barcode to metabarcode are not entirely clear. The efforts on approbation and calibration at the intraspecies level under experimental conditions are limited to molecular genetic markers designed for target species. In this study, we aimed to address these challenges: to assess the intraspecific variation in different taxa based on the *COI* barcode reduced to Leray region (˜313bp), accessible from the GenBank, as well as experimentally evaluate the possibility to identify Operational Taxonomic Units (OTUs) and Amplicon Sequence Variant (ASVs) in marine eDNA among abundant species of the *Zostera* sp. community in the northern Sea of Japan: *Hexagrammos octogrammus*, *Pholidapus dybowskii* (Teleostei: Perciformes), and *Pandalus latirostris* (Arthropoda: Decapoda). The three abovementioned species were collected at two distant locations in the Great Bay of the Japan Sea and placed into a separate 150-liter aquaria to produce both – individual and mock communities eDNA samples. Then all individuals were euthanized and genotyped individually for 650 bp and 313 bp *COI* gene regions. The *COI* Leray region was amplified based on the eDNA of mock communities and individual specimens. The resulting amplicons were sequenced on the Illumina 250 bp pair-end platform and processed based on the Begum pipeline. Along with the OTUs based on both global and local references we tried to retrieve individual haplotypes from the obtained reads. We found that eDNA samples from the experiment when blasting on local reference produce additional OTUS which we consider to be NUMTS. Surprisingly, the presence of NUMTS in the eDNA samples reduces the detection of ASVs, which may be related both to the low sequencing coverage in the experiment and probably to the natural competition of pseudogenes for primer binding sites during amplification. Perhaps a PCR-free, metagenomic approach, despite poor accessibility, might solve these difficulties. In addition, we have gathered and analyzed natural water samples from one of the sample locations of *Zostera* sp. community with a little sequence coverage and failed to retrieve any reliable information about OTUs and ASVs of taxa in mock communities, which may indicate much higher biomass of non-target organisms in the studied community.

A total of 90 sequence data sets were collected for some common groups of multicellular organisms (Mollusca, Echinodermata, Crustacea, Polychaeta and Actinopterygii) through the search on the mitochondrial *COI* gene in the popset database of the NCBI. The separate sets of sequences of Leray region were generated. Then, the values of haplotypic variability, as well as the number of population clusters of the same dataset were calculated for the region of original length and Leray region. The produced results reflect the decrease of population diversity by 1 cluster in average while switching from barcode to metabarcode. In addition we found that the length of the Leray fragment can vary in the Echinoderms.

Introduction

Studying biodiversity is rather challenging. Especially when it comes to assessing intraspecific variability at the DNA level. Data on intraspecific genetic variation for monitoring and conservation of wild populations is an important link in assessing the resistance of organisms to changing environmental conditions and anthropogenic pressures (Hilborn et al., 2003; Schindler et al., 2010). In recent years, data from whole-genome sequencing allow not only estimating the population structure but also revealing features of population demography, gene flow, selection, and introgressive hybridization of individual valuable species with high accuracy (Leitwein et al., 2020). At the same time, simple measures of the genetic diversity for natural populations based on haplotypic variation of individual markers using high-throughput monitoring would help in theory to form preliminary information about the structure of natural populations and provide preliminary recommendations for a more elaborate multilocus analysis.

Despite the obvious advantages of classic (direct, invasive) monitoring methods for obtaining information on the ecological condition of wildlife, their applicability for some species (e.g. rare and endangered species or species with low population densities) is limited. They can also produce biased results because of the direct interference of humans and their technologies in the research process (Vucetich, Nelson 2007; Minteer et al. 2014; Field et al. 2019). In addition, these methods, despite their long tradition, are time-consuming and labor-intensive (Zemanova, 2020). Therefore, gradual development and transition to alternative, noninvasive methods is highly needed.

Non-invasive methods for monitoring biodiversity in aquatic environments (Li et al., 2019) include the hydroacoustic technique (Egerton et al., 2018; Wang et al, 2022), the image recognition of aquatic organisms using trained neural networks (Siddiqui et al., 2018; Alemu, 2021), and the use of nucleic acid molecules from the environment (Jerde et al., 2011; Hering et al., 2018; Veilleux et al., 2021). The first 2 methods help to make a real-time assessments, while DNA from the aquatic environment is being introduced worldwide as an additional tool that can provide insights into the presence of aquatic animals in a particular location even when their density is low and inaccessible to other approaches (Rees et al., 2014; Li et al., 2019; Jerde et al., 2019; Veilleux et al., 2021; Nester et al., 2022). This method, in contrast to hydroacoustics and neural networks which are mostly restricted to the species level, also represents a promising tool for population genetics and phylogeography (Elbrecht et al., 2018; Adams et al., 2019; Tsuji et al., 2020a,b; Sigsgaard et al., 2020; Turon et al., 2020; Andres et al., 2021).

At the same time, studies assessing intraspecific genetic variability in high-throughput monitoring based on the environmental DNA are largely individualized in a methodological way (Sigsgaard et al., 2016; Elbrecht et al., 2018; Tsuji et al., 2020a,b; Andres et al., 2021; Adams et al., 2022). Validation and calibration under experimental conditions have not been performed on standardized molecular genetic markers, only on individual, taxon-specific ones (Tsuji et al., 2020a,b; Adams et al., 2022), making it difficult to extend species identification methods to high-throughput approaches for evaluating the population structure of species in communities. Thus, there have been excellent experimental data (Tsuji et al., 2020a,b) showing the possibility of extracting genetic diversity from hydrobionts through the use of their environmental DNA. These are useful and ready to apply data when it comes to target species. At the same time, another adventurous question arises regarding the possibility of extracting genetic diversity information using standardized markers to assess OTUs (Elbrecht et al., 2018), thereby, in theory, facilitating a rapid primary screening the diversity in abundant aquatic species.

Two different approaches are used when assessing intraspecific sequence diversity: noise reduction (ZOTUs or ASVs) and clustering (OTUs), which, however, are recommended to be used in combination (Antich et al., 2021).

One of the most commonly used markers in metabarcoding is now mitochondrial *COI*, a Leray fragment. Localized within the barcode (Folmer et al., 1994), its length is 313 bp (Geller et al. 2013; Leray et al., 2013). As a first approximation, the use of such a short fragment to assess not only species but also genetic variability of organisms is not reasonable. In fact, intuitively, the longer the nucleotide sequence of the marker, the

more information on genetic variation it can provide and the more accurate the estimate and prediction will be. In this case, it would be reasonable to use longer markers for eDNA-based rapid monitoring, followed by the sequencing on a 3rd generation platforms. However, this, in particular, does not work well with DNA from aquatic environments as it is subject to fairly rapid degradation on daylight (Murakami et al., 2019). On the one hand, this shows a significant disadvantage of aquatic DNA, on the other hand, it provides a fundamental opportunity to conduct biodiversity monitoring in dynamics.

Accordingly, to consider the possibility of noninvasive rapid assessment of genetic diversity among abundant aquatic species using DNA from aquatic environments based on metabarcoding of the Leray marker *COI* region, we designed an experiment based on the two artificial communities consisting of abundant, relatively large hydrobiont species that inhabit the *Zostera* sp. belt comminities in the northwestern part of the Japan Sea. Peter the Great Bay is the largest bay of the Japan Sea off the Russian coast. It is located between two climatic zones, where waters of cold Primorsky and warm North-Korean currents meet, the bay is characterized by high species diversity and abundance of fish resources (Kalchugin, 2021). The objects for the experiment had to meet a number of criteria: relatively small size, suitable for keeping animals in common aquariums, abundance in both collection sites (Vostok and Vityaz bays), dietary unpretentiousness, ability to sustain transportation and long-term keeping in aquarium. In accordance with these criteria, three common species inhabiting seagrass belts of *Zostera* sp. were selected: the greenling *Hexagrammos octogrammus* Pallas, 1814, the shrimp *Pandalus latirostris* Rathbun, 1902 and the prickleback *Pholidapus dybowskii* (Steindachner, 1880).

In the present work, with the use of metabarcoding of aquatic DNA from experimental conditions and the calculation of the genetic variability of the standardized *COI* region based on a large volume of published data, we intended to evaluate the applicability of this region for high-throughput monitoring of the genetic diversity of wild populations of aquatic organisms.

Materials and Methods

1. The assessment of intraspecific genetic diversity based on aquatic DNA metabarcoding

1.1. Collecting hydrobionts and DNA from the aquatic environment

The animals were collected from two locations in the Peter the Great Bay of the Japan Sea (Figure 1): Vostok Bay (3 specimens of masked greenling *Hexagrammos octogrammus*, 8 specimens of prickleback *Pholidapus dybowskii* and 16 specimens of shrimp *Pandalus latirostris*) and Vityaz Bay (4, 6 and 22 specimens, respectively) in September 2020 using a fish fry net. In addition, in September 2021 we sampled water from the *Zostera* sp. community for environmental DNA analysis of the Vostok Bay. Water in a volume of 900 ml was sampled twice using a 150 ml syringe. The entire volume of water was passed through a syringe filter cap with a diameter of 33 mm and a pore size of 0.45 μm (the material is PES). Then, DNA on the filter was preserved by passing 1 mL of Longmayer's buffer through the syringe tip, and the inlet and outlet ports were closed with combi-stopper plugs. The filter was stored at 20 0C below zero until isolation.

1.2 The experiment with aquatic animals

The collected hydrobionts were settled in two separate aquariums of the NNCMB FEB RAS with the volume of 150 liters each according to the site of capture - "Vostok" and "Vityaz". Temperature was maintained at 15 0C throughout the experiment. The hydrobionts were fed with crushed squid fillets of *Todarodes pacificus*. Shortly after the introduction to the tanks, one of the greenlings in "Vostok" aquarium died and was eaten by shrimps. Similarly, some of the shrimps were lost in both aquariums. Thus, only single shrimp remained in the aquarium "Vityaz", and 8 shrimps left in the aquarium "Vostok". The number of pricklebacks did not change. The circulation and filtration of water in the tanks were turned off 1 hour before sampling.

Environmental DNA was collected from both aquariums according to the method described above, in a volume of 900 ml, in triplicate. In addition, we performed the filtration of water from the marine water

storage reservoir as a control sample. The aquatic DNA was then collected from the animals individually. Each hydrobiont was settled into a separate 1.2 liter aquarium. Prior to the transfer of the individuals each aquarium was cleaned with a 10% bleach solution. Then the tank was rinsed with water from the storage reservoir. After the hydrobiont was placed, it was rinsed with three volumes (˜3.6 liters) of water followed by filling the individual aquarium. After some time (from 30 minutes to 1 hour), the water was sampled and filtered (900 ml per aminal), and each animal was placed into a container with a 10% urethane solution for sedation. After sedation, each animal was measured, weighed, and then material was taken for genetic analysis (a piece of skeletal muscle was cut off from the back of the fish body, and one of the legs in shrimp was taken; then the tissue was preserved in 95% ethanol). A total of 29 individual aquatic eDNA and tissue samples were collected.

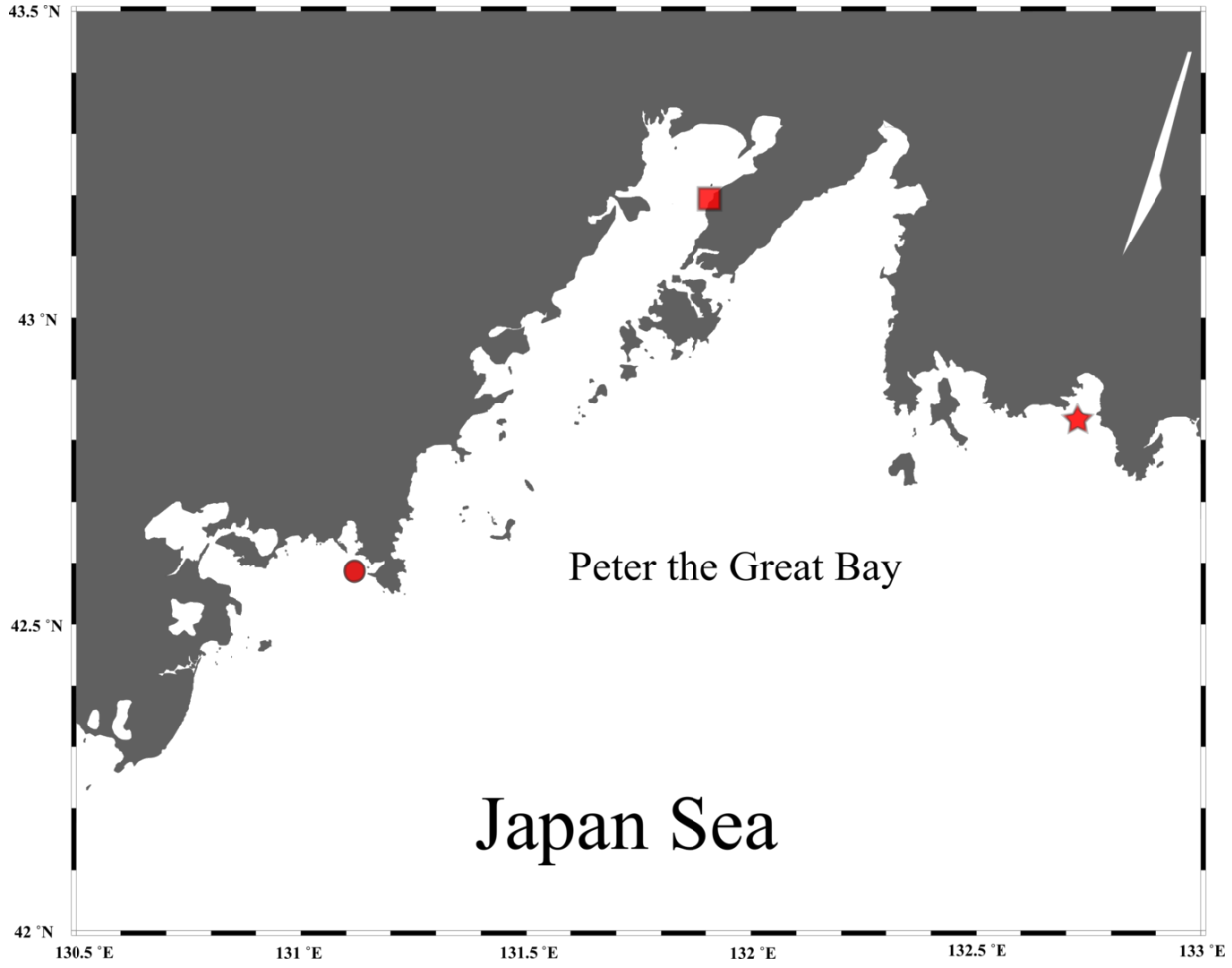


Figure 1. Map of the area where the animals were collected and the experimental work were carried out. The locality of Vityaz Bay is marked with a circle. The asterisk indicates the Vostok Bay. The square marks the location where the animals were placed in the aquarium to form the mock communities.

1.3 The DNA isolation and individual genotyping of the hydrobionts

The extraction of total DNA from the fixed tissue was performed using K-SORB-100 kit (Syntol, Russia). The isolated DNA was then used to amplify a 313 bp (Geller et al. 2013; Leray et al., 2013; Wangensteen et al. 2018) and 650 bp (Ward et al., 2005) long *COI* gene mitochondrial fragments. For the latter a

combination of FishF1 and FishR1 primers was suitable for all 3 species. The PCR mixture consisted of 5x Taq Red buffer (5 μl), dNTPS (0.5 μl, 10 mM), a pair of primers (0.12 μl, 0.05 mM each), Taq polymerase (0.25 μl, 1.25 units per reaction), 1 μl of DNA solution (20-100 ng), and distilled water to a final reaction volume of 25 μl. The PCR thermal algorithm started with a pre-denaturation for 10 min at 95 0C. This was preceded by 35 cycles according to the following scheme: denaturation at 94 ºC for 1 min, annealing at 45 ºC for 1 min, and 1 min elongation at 72 ºC with a final elongation for 10 min. To check the results of amplification, we performed the electrophoresis of the fragments in 1% agarose gel followed by exposure in ethidium bromide solution and visualization under UV. The successfully amplified samples were purified by alcohol precipitation and the DNA pellet was dissolved in deionized water. A forward and reverse sequencing reactions were performed using purified amplicons together with corresponding primers used in PCR step according to the BrilliantDye Terminator Cycle Sequencing Kit manufacturer's instructions (NimaGen, Netherlands). The capillary electrophoresis of the fragments produced by cycle sequencing was performed on Applied Biosystems Genetic Analyzer 3500. Consensus sequences were assembled from the obtained chromatograms using Geneious program (Kearse et al. 2012). These sequences were used to search for the homologous ones in NCBI via the BLAST algorithm (Boratyn et al., 2012). The alignment together with the reference sequences selected from BLAST results and a read frame search using the translation tool was performed in MEGA 7.0 software (Kumar et al. 2016). Sequence matrices were then generated for each species separately and haplotypic variation was assessed in DnaSP v.6 (Rozas et al., 2017). Based on the results of hydrobiont genotyping (2 haplotypes for *P. dybowskii* and *P. latirostris*, as well as 3 haplotypes for *H. octogrammus*, see Table 1), individual aquatic DNA samples corresponding to the detected haplotypes of the 313 bp *COI* fragment were selected for the further work, one for each haplotype. After the detection of OTUs identified from the local database, it was necessary to clarify their relationships to the original haplotypes. In addition, the *COI* barcode sequences from *H. octogrammus* and *P. dybowskii* species were used to verify the species identity of the haplotypes obtained. Since no reference data are available for *P. latirostris*, we assembled *COI* fragments from reads of 4 transcriptomes of this species (Kawahara-Miki et al., 2011) using NOVOPlasty (Dierckxsens et al., 2016). Based on the combined matrix, a phylogenetic NJ tree was constructed in the MEGA 7.0 program (Figure 4). The robustness of the topology was assessed using 1,000 replicates of the nonparametric bootstrap test.

1.4 The DNA extraction from syringe filters, COI Leray fragment amplification, sequencing, and reads processing

The extraction of DNA from the syringe filters was performed using M-Sorb-OOM kit (Syntol, Russia) with modification of the manufacturer's protocol, according to which at the initial stage the lysis buffer was heated to 65oC and passed through the filter tip in the opposite direction to the filtration (backflushing method, after Kesberg and Schleheck, 2013), draining the entire volume of the resulting liquid into a clean test tube. Based on the isolated DNA, a 313-bp long *COI* fragment was amplified (Geller et al. 2013; Leray et al., 2013; Wangensteen et al. 2018) with three replicates per sample. For each sample, we used a pair of primers with an individual 7-nucleotide tag at the 5'-end (doubly-tagged approach) that were developed in ecotag (Boyer et al. 2016). The negative PCR control was also performed by using separate pair of tagged primers. The reaction mixture included 10 μl of AmpliTaq Gold 360 Master Mix, 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 the final volume of 20 μl. The PCR thermal cycling profile included preheating at 95ºC for 10 min with subsequent 35 cycles according to the following scheme: 1 min at 94ºC, 1 min at 45ºC and 1 min at 72ºC. The final elongation was at 72ºC for 5 min. The results of amplification were checked in the same way as described above. The amplicons were purified using Cleanup S-Cap (Evrogen, Russia) and normalized (see (Elbrecht and Steinke 2019)) before pooling. The volume of control reaction was taken as an average of the obtained volume of normalized samples. The normalized amplicons were then combined with the control and sequenced at Novogene (Tianjin, China). The library was created using a PCR-free NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, England) and sequenced on an Illumina high-throughput sequencer by a 250 bp paired-end sequencing technology.

The resulting reads were processed according to the Begum metabarcoding pipeline (Yang et al., 2021;

Zepeda-Mendoza et al., 2016). After removing adapters and pre-assessing the quality of the reads with fastqc, we used Spades (Bankevich et al., 2012) to correct possible errors. Paired-end reads were merged into consensus reads using PandaSeq (Masella et al., 2012). The reads were then demultiplexed and filtered using Begum. Clustering was performed in Sumaclust (Mercier et al., 2013) with parameters -t 0.98 and -R 0.85. Determination of the taxonomic identity of the generated operational taxonomic units (OTUs) was performed using BLAST+ 2.12.0 (Camacho et al., 2009). In addition, a local reference base was generated to identify the target OTUs and ZOTUs (ASVs), which consisted of 7 nucleotide sequences of previously found unique haplotypes (Table 1). The taxonomic information was then summarized in two separate tables (derived from global and local reference databases, see Table S1 and Table 3, respectively) based on the output from the MEGAN community edition program (Huson et al. 2016). The results derived from the local reference were further condensed using the lulu package (Frøslev et al., 2017) to test the effect of preserving non-original OTUs. Since the most acceptable (Antich et al., 2021) existing method for identifying ZOTUs in vsearch (Edgar, 2016) was found to be extremely sensitive to coverage, after obtaining consensus sequences we had to use several sequential searches using the grep command in the shell of Ubuntu to sort the reads based on sample tags followed by the search for the haplotypes obtained during Sanger sequencing (see section 1.3) in the sample-sorted reads content.

2. Assessing the genetic variation of COI fragments in the population datasets retrieved from GenBank

The most common groups of multicellular organisms (phylum Mollusca, phylum Echinodermata, subphylum Crustacea of Arthropoda phylum, class Polychaeta of Annelida phylum, and class Actinopterygii of Chordata phylum). Taxonomic-based searches were performed in the popset database of the NCBI GenBank resource (Benson et al. 2018) among the sequence sets for population-genetic studies. An important reason for the selection of these groups was their good coverage in GenBank, as well as the presence of homologous Leray *COI* marker fragments in NCBI. A total of 83 datasets were selected. Nine to 20 sets for each group (Supplement A, Table S2). There were at least 17 sequences in each dataset. Reduction of the retrieved sequences to the Leray fragment length was performed through their alignment together with reference datasets, which included 313 bp *COI* fragments with a retrieved reading frame as well as several complete *COI* gene sequences from each group. The sequence set then joined in the MEGA program together with the reference sequence set and was translated given the mitochondrial code corresponding to the taxon and aligned using ClustalW (Larkin et al., 2007) based on the Protein Weight Matrix BLOSUM with default parameters. If the alignment was successful (the Leray region was located within a fragment of the examined *COI* sequences, between 130 and 236 amino acid sites of the complete translated gene sequence), matrix was truncated according to the Leray fragment length. The haplotype function of the pegas package (Paradis, 2010) was used to estimate the haplotypic variability of all the data sets obtained in this manner (both the original and trimmed ones). The number of population clusters in each dataset was estimated using the Geneland program (Gulliot et al., 2005; Gulliot, 2008) without prior information on the geographic or other subdivision of the samples. The calculation was based only on the sites with SNPs. They were extracted from the sequence matrices using the SNP-sites program (Page et al., 2016). Next, we used the console version of the PGDSpider program (Lischer and Excoffier, 2012) to convert the sets into the format recommended by the Geneland authors (Guillot et al., 2011). During the preliminary stage, 2 independent MCMC runs were performed with the total number of generations equal to 100000 and 500000 and the number of generations accounted for the burn-in step equal to 200 and 250 (discarded after the search), respectively. Sampling from the parameters space was performed every 100 generations. The maximum number of populations simulated during the search was set to 20 with the correlated frequency model accounted. For those datasets that showed differences in the number of determined populations between the first two runs, or did not form a clear peak in the distribution of the number of populations along the chain after burn-in step and showed density below 0.5, an additional run was conducted using 1500000 generations to exclude possible undersampling during the search. Scripts providing simultaneous formatting as well as the analysis of all data sets are given in the supplementary material (Supplement B). Statistical processing and data visualization were performed in R (R Core Team, 2021).

Results

1. The genotyping of hydrobionts from the mock communities

The assessment of hydrobionts haplotypic variation by Leray fragment (313 bp) showed that seven sequences of *H. octogrammus* species contained 3 haplotypes in three variable sites. One of the haplotypes was present in five fish collected in Vityaz Bay. The other two were found in fish collected in Vostok Bay. The sequence matrix of the species *P. dybowskii* contains one variable site with two haplotypes without a clear association to localities. In shrimps *P. latirostris* we found two haplotypes located at one variable site. One of the haplotypes was found in a shrimp from Vityaz Bay, the second one is common to eight shrimps collected in Vostok Bay. The genotyping by the Folmer fragment (˜650 bp) revealed similar results, with an exception that in *H. octogrammus* 5 haplotypes were found in 6 variable sites (Table 1).

Table 1. The results of genotyping of aquatic organisms by the *COI* mitochondrial fragment

Species	Sampling location	Number of individuals		Number of haplotypes	
		At the beginning of the experiment	At the end of the experiment	Leray (313 bp)	Folmer (~650 bp)
H. octogrammus	Vostok	3	2	2	2
	Vityaz	4	4	1	3
P. dybowskii	Vostok	8	8	2	2
	Vityaz	6	6		
P. latirostris	Vostok	16	9	1	1
	Vityaz	22	1	1	1

2. The genetic variation of COI fragments from GenBank population data sets

The number of sequences in each set ranged from 17 to 350 (84.3±5.5)[1]. The original length of the fragments ranged from 366 to 1153 sites (652±13.7). Up to 11 groups based on the original fragment (3.2±0.2) and up to 9 groups based on the reduced fragment (2.4±0.2) were identified by the analysis of intraspecific structure. About 27.7% of the original datasets (23 sets) showed no intraspecific structure, and 12 (14.5%) of the reduced datasets did so (Table S2). Seventy-nine datasets (95.2%) gave a reliable determination of the number of clusters (the density value of 80% or higher according to calculations on the non-reduced set), while for the remaining 4 (*J. singaporensis* (Crustaceans), *G. adustus*, *R. formosanus* and *S. formosus* (Fishes)) even a repeated run using more generations did not increase the reliability of results (the density value was from 24 to 60%). A significant decrease in the number of clusters was expected when moving to a reduced dataset according to the results of the Wilcoxon pairwise test (Fig. 2A). In contrast, one species (*C. sapidus* (Crustaceans)) showed the opposite trend, but it was associated with a decrease in the confidence of the intraspecific analysis (see Table S2). A decrease in validity was also recorded in the case of the species *L. armata* (Crustaceans) when moving to the reduced dataset. Except for these 2 cases, the length reduction did not cause a decrease in validity for the other sets (see Table S2). Overall, without the pairwise test, when

looking at the sparse data presented in the original fragment analysis, neither the number of sequences in the set and sites with SNPs (Fig. 2B) nor the length of sequences (Fig. 2C) explained the increased number of detectable intraspecific clusters (R2?0.3 at p=0.05). In the reduced dataset, the reliably detectable number of clusters which are consistent with estimates based on the original, and not consistent with the original set, also show no detectable trends that could be noticed for extrapolation on the global level.

1Hereinafter are the values of the mean +- standard deviation

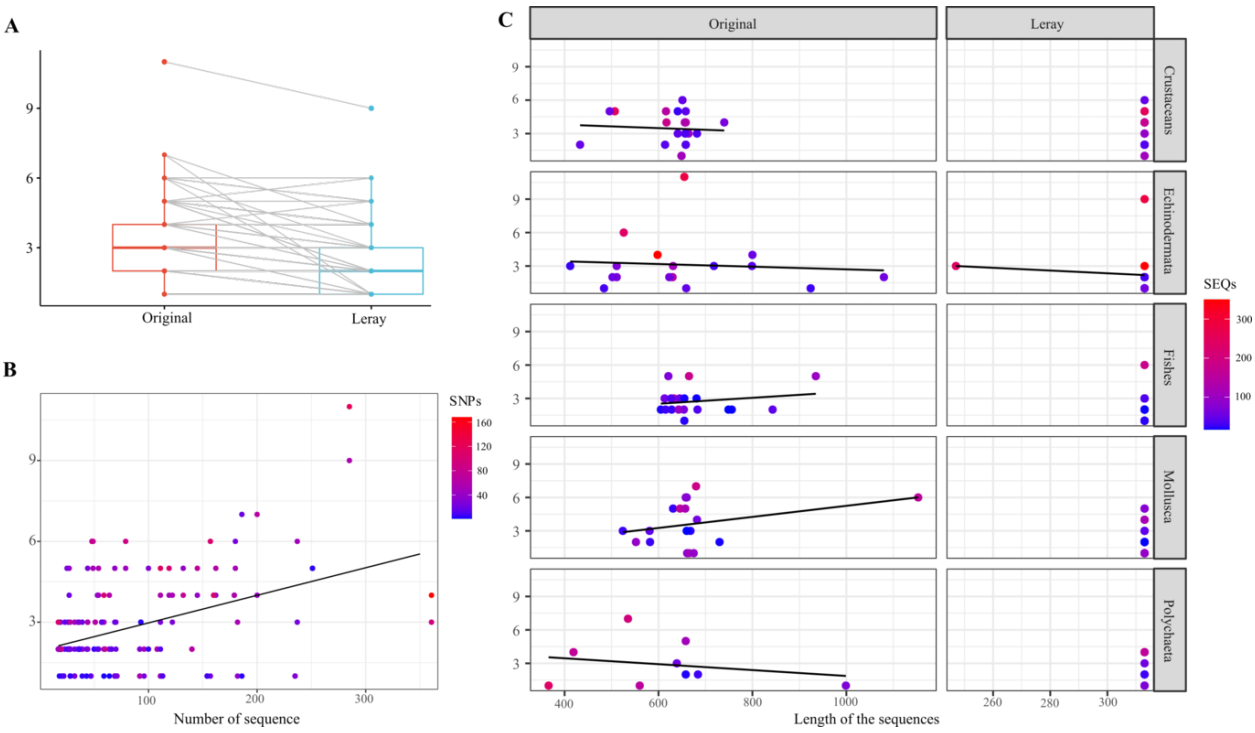


Figure 2. The number of intraspecific clusters estimated for *COI* datasets. A. The comparison of the average number of clusters of the original (Original) and reduced (Leray) datasets. The Wilcoxon signed rank test was used to assess the reliability of the decrease in the number of clusters when reducing the size of the fragment (p=0.000000005328). B. The dependence of the number of intraspecific clusters on the number of sequences in the dataset. Generated based on the original data. The line indicates a linear regression. The color scale, graded from blue to red, indicates the variation in the number of sites found with SNPs. C. The comparison of the number of intraspecific clusters across taxonomic groups based on sequence length and the number of sequences in the set (SEQs). The line indicates a linear regression. The color scale with gradation from blue to red indicates the variation in the number of sequences.

One of the species, *P. lividus* (Echinodermata), showed a noticeably shorter Leray region length among the reduced datasets based on the amino acid sequence alignment results (Figure 3). The Leray region in this species is 247 bp long, which is caused by a 66-base deletion located between 160 and 183 positions and affecting 22 amino acid residues. No other data on fragment length polymorphism were found in the analyzed data sets.

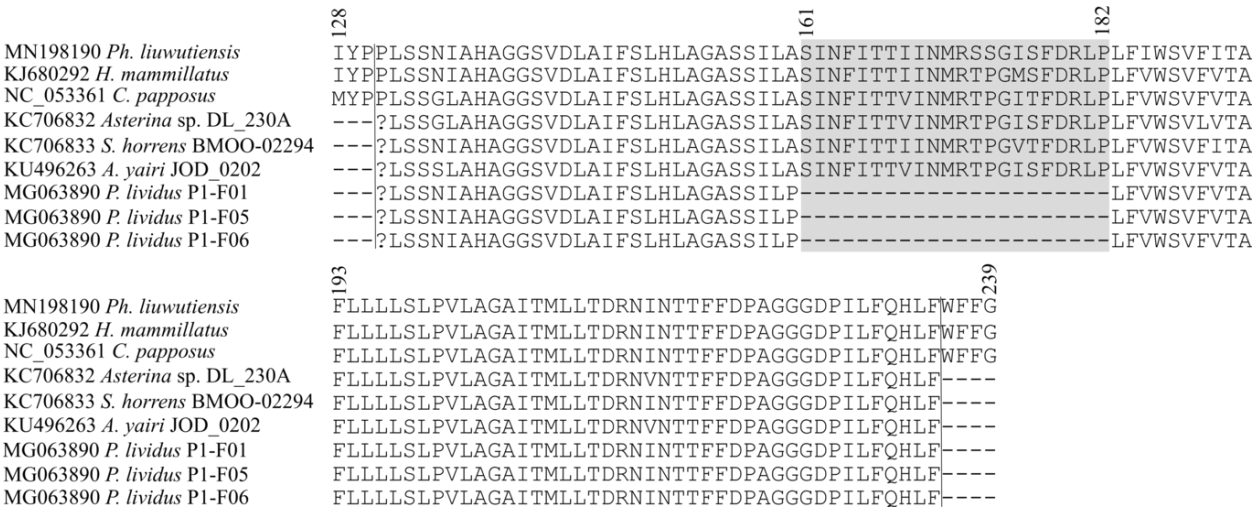


Figure 3. The fragment of the interleaved amino acid sequence matrix that contains the Leray site of *P. lividus* species. The horizontal lines indicate site boundaries. The numbers above the matrix indicate the positions of the sites in the complete amino acid sequence of the complete *COI* gene relative to *Phyllophorus liuwutiensis* species. For the purpose of comparison, five more representatives of Echinodermata were added to the matrix in addition to this species. The shaded area highlights the deletion sites in *P. lividus*.

3. The estimation of species and genetic diversity based on DNA metabarcoding data from the aquatic environments

3.1 On the basis of the global reference

After the filtering and demultiplexing procedure, the *COI* fragment had 2188718 reads, of which 513152 reads were unique (Table 2). The control detected 1308 reads (0.06% of the total). The clustering revealed approximately 3,000 OTUs. Most of them were not classified to any of the life domains. After the blasting procedure based on the global reference database, 326 OTUs belonging to Eukaryota (284) and Bacteria (42) were retained for subsequent work. After the reduction of the taxonomy, a total of 243431 (27.4%) reads remained in the table, with a minimum threshold of 5 reads per OTU (Table S1). The number of reads attributable to different OTUs was highly non-uniform. The average value was 477.3±159.3. The modal group, which was 5 reads, characterized more than 17% of all the OTUs in the different samples.

Table 2. Number of reads and OTUs that corresponded to each of the aquatic DNA samples based on the global reference library

Sample or haplotype name	Total number of reads/number of taxonomically referenced reads	Number of OTUs
PD20-1	72592/180	5
PD20-3	103077/422	19
PL20-1	100536/1722	44
PL20-2	174597/2275	38
HO20-4	120273/46674	30
HO20-6	117520/45469	37
HO20-7	91637/47615	23
RES[a]	103733/1091	38
RA[b]	198507/15072	47
LA[c]	217846/29230	50
V[d]	887092/53681	178
Control	1308	-
Total	2188718/243431	326

a - environmental DNA sample from the storage reservoir; b - mock community based on samples from Vostok Bay; c - mock community based on samples from Vityaz Bay; d - environmental DNA sample from natural conditions, Vostok Bay.

The highest number of unique OTUs (178) and reads (53681) were in the natural community of Vostok Bay. Of these, 156 OTUs with 93.7% of the reads belonged to Eukaryota. The Bacteria in this sample included representatives of the only family Flavobacteriaceae, which contained the genera *Aquimarina*, *Formosa*, and *Polaribacter*. The latter accounted for the majority of bacterial OTUs (83.3%) and reads (98.7%). The species-level identity of the bacterial OTUs was not possible to determine. Among the found eukaryotic OTUs there were 16 phyla, 30 classes, 51 orders, 65 families, 74 genera, and about one hundred unambiguously defined clusters of putatively species rank (Table S2). The most of the found OTUs (61) belonged to the plylum Bacillariophyta. This plylum also accounted for more than a half (51.8%) of all reads. The vast majority of reads per OTU were from the genus *Phoronopsis* (7511 or 14.9%). More than 1000 reads per OTU also came from the species *Ditylum brightwellii*, *Coscinodiscus wailesii*, *Thalassiosira nordenskioeldii*, *Minutocellus polymorphus*, *Asterionellopsis glacialis*, *Coscinodiscus* sp. and *Thalassiosira punctigera* of Bacillariophyta plylum, as well as genera *Hyaloperonospora* of Oomycota plylum and *Macrocystis pyrifera* of Ochrophyta plylum (Figure S1, Table S1). No OTUs belonging to *P. latirostris* or related decapods were detected, nor were the species of *Zostera* in whose belts the samples were collected.

Among the samples of artificial communities, as well as individual samples of environmental DNA and those from the reservoir tank in the aquarium, there were almost no common OTUs. With the exception of *Oncorhynchus keta* species OTU with a high number of reads (from 23 to 793 per sample). In the environmental DNA samples from hydrobionts genotyped by the first and second haplotype samples of *P.*

latirostris, 44 and 38 OTUs were detected, respectively; in the first, second and third haplotype samples of *H. octogrammus*, 30, 37 and 23 OTUs were found. The first and second haplotype samples of *P. dybowskii* accounted for the least number of OTUs – 5 and 19. Vostok and Vityaz Bays mock samples had 47 and 50 OTUs, respectively, and the storage tank had 39 OTUs (Table S1).

3.2 On the basis of the local reference

The search for taxonomic assignment based on the local reference, consisting of *COI* haplotypes of genotyped individuals from the mock communities, revealed 17 OTUs with 209499 reads (13966.6±5802.6) (Table 3). Of these, 10 OTUs corresponded to the species *H. octogrammus*, 6 OTUs belonged to *P. latirostris*, and one to *P. dybowskii*. The majority of the reads (68%) accounted for OTUs assigned to the species *H. octogrammus* (145653 or 70%). Regarding the original reference haplotypes, most of the reads (66.7%) also came from the 3 haplotypes of *H. octogrammus*. They form a common cluster together with OTUs 4, 5, and 6 (Figure 4) with intraspecific variability of no more than 0.013. Other haplotypes of this species form 2 separate phylogroups with divergence from 0.071 to 0.106. In the most divergent phylogroup, 3 OTUs had a deletion of 1 nucleotide, as well as 1 to 3 amino acid substitutions relative to the reference haplotypes. The *P. dybowskii* haplotypes accounted for 26% of all reads, which revealed similarities only to PD20-1 hap1. Accordingly, it completely matched OTE22. *P. latirostris* had the least number of reads (939 or 0.4%), but its homologous haplotypes were the most diverse, forming four phylogroups and containing from 1 to 3 amino acid substitutions relative to the reference sequences. The divergence from the cluster with the original haplotypes was between 0.023 and 0.126. The samples from the mock communities had 1308 (Vostok Bay) and 12703 (Vityaz Bay) reads.

The presence of additional OTUs with deep divergence in the species *H. octogrammus* and *P. latirostris* requires proof of their homology to these taxa. This cannot be done for shrimp, because there is no complete reference base for this genus of shrimps. The availability of a nucleotide sequence reference database for the species *H. octogrammus*, due to its completeness, allows us to find out whether additional OTUs belong to any of the known greenlings species (family Hexagrammidae). When comparing the identified OTUs with the specified database, it was found that the nearest OTU to the original cluster is determined as *H. octogrammus*, but occupies a basal position in relation to it (Figure S2). The remaining 4 OTUs form a separate cluster occupying an intermediate position between the genera *Hexagrammos* and *Pleurogrammus*.

The results of OTUs condensation using the lulu program left 6 OTUs (Table S3). The species *P. dybowskii* and *P. latirostris* retained one haplotype each as central OTUs (Figure 4, Table S3). At the same time, the species *H. octogrammus* with rather high intraspecific variability retains 3 OTUs in the cluster with the reference haplotype, as well as one of the haplotypes of the divergent cluster, carrying a deletion of 2 nucleotides.

As for the ASV detection approach implemented here, for two of the three species, the success of detecting of the exact genetic variants was inversely proportional to the success of detecting additional OTUs. Thus, only for the species *P. dybowskii* all (two) genotyped haplotypes were detected with high coverage. Both, in individual samples and in artificial communities (Table 4). Their reads in minor numbers were scattered throughout the samples, but showed absolutely no reciprocal cross-contamination. The original haplotypes of species *H. octogrammus* were almost not found, even in artificial communities. For *P. latirostris* species, we were able to find 106 reads of the PL_hap1 haplotype, linked to the locality of the Vityaz Bay, and 528 reads for haplotype PL_hap2 which is from the Vostok Bay community. At the same time, haplotype 2 turned out to be much more represented in the mock community of the Vityaz Bay. In the sample from natural environments there were only minor numbers of reads for ASV of the species *P. dybowskii* and *P. latirostris* presented.

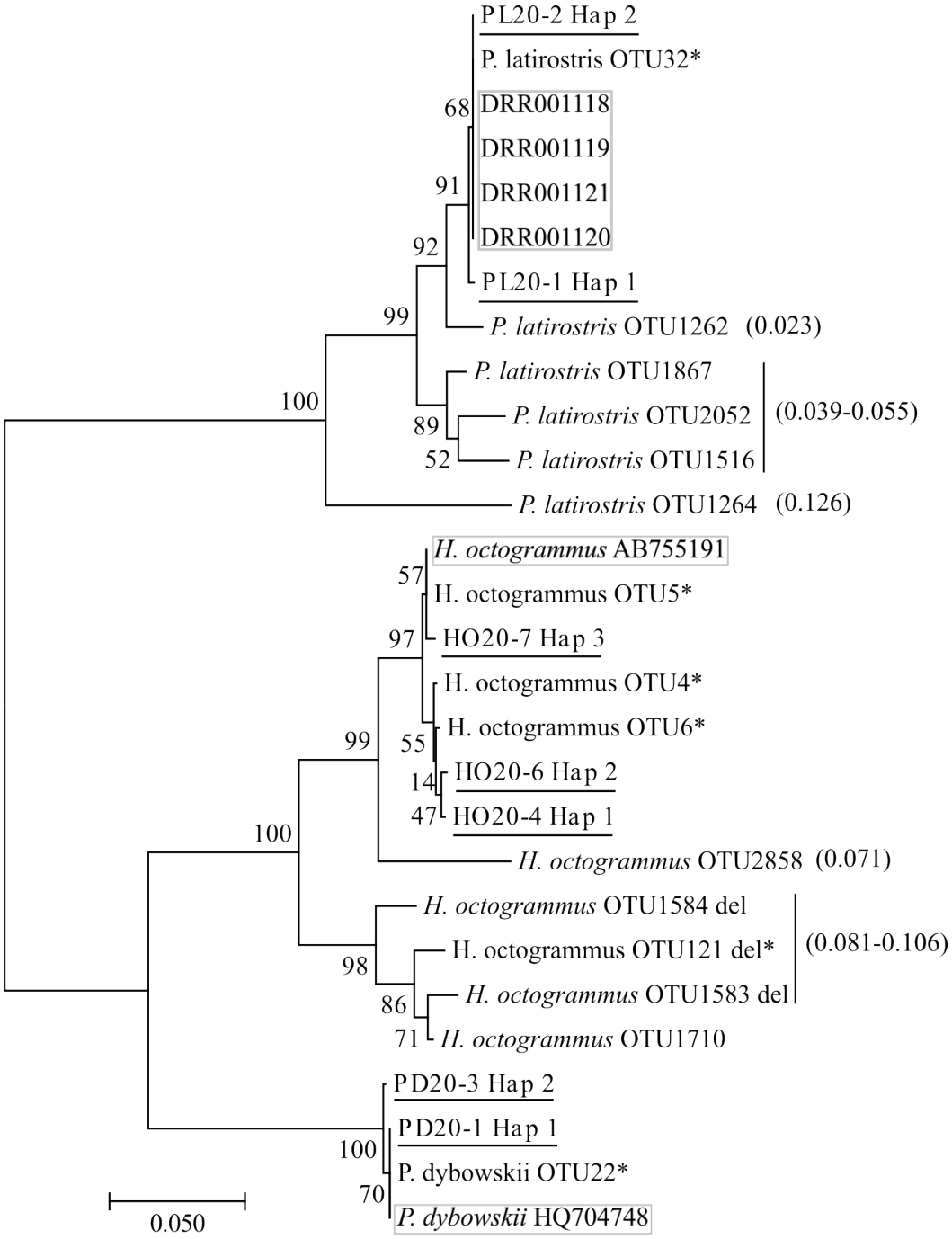


Figure 4. The phylogenetic NJ-tree showing the relationships between the original genotyped haplotypes (underlined), reference sequences (highlighted by a frame), and OTUs identified by eDNA sequencing (all others). The asterisk indicates OTUs retained based on the results of compression by the lulu package. The tree was constructed on the basis of uncorrected genetic p-distances. The nodes show the results of the topological robustness analysis, in %. The divergence values relative to the reference sequences of each species are shown in parentheses against the highlighted phylogroups (marked by vertical lines).

Discussion

This paper addresses, for the first time, the possibilities and limitations of rapid assessment of genetic variability among abundant marine species using a standardized *COI* metabarcode under the experimental conditions. In addition, using the data retrieved from GenBank we examined the extent to which the number of detectable populations (clusters) changes during the shift from the *COI* barcode to metabarcode.

Table 3. The OTUs found with the local reference database, indicating the number of reads

OTU	Species	V[a]	Vostok					Vityaz		LA[b]	RA[c]	Total
			PL_hap2	HO_hap2	HO_hap3	PD_hap1	PD_hap2	HO_hap1	PL_hap1			
4	*H. octogrammus*	5	0	24	51	0	0	46126	23	4936	0	51165
5		0	0	0	47064	0	0	0	0	0	499	47563
6		0	0	43719	0	0	0	19	0	0	323	44061
121 del[d]		0	0	1135	670	0	0	955	0	15	0	2775
1710 inc[e]		0	0	33	0	0	0	6	0	0	0	39
1584 del[d]		0	0	6	0	0	0	19	0	0	0	25
1583 del[d]		0	0	0	0	0	0	20	0	0	0	20
2858		0	0	0	0	0	0	0	0	5	0	5
32	*P. latirostris*	0	757	0	0	0	0	0	154	7004	106	8021
1516		0	0	0	0	0	0	0	0	41	0	41
1262		0	0	0	0	0	0	0	0	37	0	37
1264		0	0	0	0	0	0	0	0	22	0	22
1867		0	0	0	0	0	0	0	0	10	0	10
2052		0	0	0	0	0	0	0	0	8	0	8
22	*P. dybowskii*	0	7	5	5	25312	0	1	6	625	380	55707

a - environmental DNA sample from natural conditions; b - mock community based on samples from Vityaz Bay; c - mock community based on samples from Vostok Bay; d - deletion at the 237th position of the Leray site; e - stop codon.

We notice that it was not possible to achieve the detection of original *COI* haplotypes in reads originating from natural eDNA samples. The same was true for some individual eDNA samples. Combined with the small number of common OTUs between the artificial community samples, this may indicate rather low coverage, which can be noted as a major drawback of the study (see Table 2). It is possible that primer bias may also affect the results (Shu et al., 2021). In addition, all of the artificial samples were noticeably contaminated with the reads from the species *Oncorhynchus keta* contained in the tank located in the same room. But it appears that the water simply contains much more microbial DNA (in our case it is the plylum Bacillariophyta) than the commercial fish and invertebrate species of interest (Collins et al., 2021), and the reason lies in fundamental differences in primer design approaches for target species detection and metabarcoding surveys (Freeland, 2017). Typically, in such studies at least 30,000 reads per sample are required to reach a plateau at the number of OTUs (Dully et al., 2021). At the same time, there is an evidence that even for individually designed markers for target species, the number of detectable haplotypes using eDNA can be reduced relative to their actual number (Adams et al., 2022).

As recent studies show that the environmental DNA can provide detailed information on a species-specific haplotype diversity, although currently only a few studies have successfully used this approach to obtain population-genetic data (Sigsgaard et al., 2016; Elbrecht et al., 2018; Tsuji et al., 2020a,b; Andres et al.,

2021; Adams et al., 202). Factors such as sequencing errors that lead to false positives, as well as limitations in the understanding of the processes that affect the rate of environmental DNA production affect the validity of the results (Eble et al., 2020).

A study of haplotype diversity among *Plecoglossus altivelis altivelis* fishes using the mitochondrial *D-loop* marker showed high accuracy in determining variation by using ASV noise reduction algorithms, which eliminated false positives (Tsuji et al., 2019). Further studies in this area showed that the haplotype diversity estimated from invasive screening was lower compared to estimates obtained using environmental DNA at all screening sites. Non-invasive and invasive DNA sampling were found to be prone to overestimate and underestimate intraspecific genetic diversity, respectively. Similarly, the results showed variation depending on which of the noise reduction algorithms was used (Tsuji et al.,2020a).

Another work also indicates that the analysis of environmental DNA yields a result close to the estimated intraspecific diversity of the entire population. In addition, it has been observed that the difference in the number of haplotypes between estimates based on DNA extracted from tissues and environmental DNA is most likely due to the difference in the sample coverage (Tsuji et al.,2020b).

In the mentioned studies, the *D-loop* was used as a marker, which is a non-coding region with a level of variability higher than that of most other mitochondrial and some nuclear markers. At the same time, it is known that markers encoding a protein, which includes *COI*, have a lower level of mutations and, in addition, may tend to increase the detection accuracy due to noise suppression.

When comparing ASVs and OTUs isolated from individual specimens using specifically a local reference, we found an interesting pattern. Thus, the lack of detectability of ASVs is observed in those species which exhibit some additional OTUs. Accordingly, it makes sense to assume that they (the sequences forming the additional OTUs) can compete for primer hybridization during PCR in comparison to the original ASVs, which can ultimately result in a global underestimation of the genetic diversity of live organisms from eDNA based on the use of this marker (in particular). Previously, these two measures were considered in parallel, and the possibility of distorting one indicator at the expense of the other was not discussed, especially the reasons for such a bias.

To clarify the nature of the additional OTUs and their origin, the relevant assumptions have been made, considering them as: 1) sequencing errors (artifacts); 2) derivatives from duplicated regions of the mitochondrial genome; 3) cryptic (or overlooked) species diversity; and 4) NUMTs or pseudogenes.

Since the origin of identical sequencing errors with high coverage in different samples is unlikely, this version has to be rejected. Currently, there is no information on the complete sequences of the mitochondrial genome of *P. latirostris*, but a closely related species of this genus, *P. borealis*, does not exhibit such structural features (Viker et al., 2006). In fish, duplication of the mitochondrial genome fragments is generally quite rare phenomenon (however, see the flatfish Li et al., 2015), and species close to *H. octogrammus* also do not exhibit such a features (Ji et al., 2020). The presence of cryptic diversity cannot be excluded for *P. latirostris* not only because of the lack of a complete reference for the genus, but also because of the "eating" of several specimens in the artificial communities during the experiment (see materials and methods). Accordingly, at least a part of the OTUs of this species can be attributed to unidentified, but existing haplotypes in the artificial communities (Fig. 4, Table 3 – OTU 32). At the same time, comparison with the data from the sites of the original species description (Fig. 4, id. DRR) indicates that the intraspecific variation of *P. latirostris* on the basis of the marker used is homogeneous. Hence, we can conclude that OTUs diverging from the reference ones by at least 0.023 (Fig. 4) should be considered as originated from a different type of phenomenon. They, like such OTUs of *H. octogrammus* species, are highly likely to be nuclear copies of mitochondrial sequences or NUMTS, which is a type of pseudogenes (Hazkani-Covo et al., 2010; Marshall and Parson, 2021). The divergent OTUs (Fig. 3) are characterized either by nonsynonymous nucleotide substitutions (*P. latirostris*) or a shift in the reading frame (*H. octogrammus*), due to which they can be attributed to pseudogenes, having previously excluded other interpretations of their origin. In general, it is rather difficult to differentiate NUMTS from true sequences of the mitochondrial DNA (Hazkani-Covo et

al., 2010; Nugent et al., 2020; Porter, Hajibabaei, 2021), and it is much easier to work in this respect with model organisms, for which all kinds of references are available (Marshall, Parson, 2021), as well as try to prevent accumulation of NUMTs at the experimental stage (Wang et al., 2019). Organismal DNA samples, meanwhile, do not tend to detect many NUMTS during Sanger sequencing because of their small number in the amplicon pool to be generated (Hebert et al., 2004; Schultz, Hebert, 2021) as opposed to eDNA. In our case (with the exception of OTUs with 2-nucleotide deletions carrying a reading frame shift), we can be satisfied with the exclusion method only and only if a pre-collected complete barcode DNA reference library for the species in question is available. At this point, we can expect that almost all phylogeographic works based on environmental DNA metabarcoding is subject to a kind of "survivorship bias," where only what has passed selection by primers and sequencing coverage is analyzed. It cannot be excluded that our study also carries this error. The processing of the obtained OTUs using the lulu feature gave, basically, the expected results. For example, we showed that sufficiently divergent NUMTS are not eliminated, but remain as a separate OTUs (Figure 4, Table S3), especially in the presence of a large number of reads for them. The problem is that there is no proper database of pseudogenes, and we again return to the "survivorship bias". In this respect, we agree with our colleagues who point to the necessity of maintaining a pseudogenes database in addition to the well-curated barcode reference library (Schultz and Hebert, 2021).

Table 4. The ASVs found with local reference database indicating the number of reads

ASV	Species	V[a]	Vostok					Vityaz		LA[b]	RA[c]	Total
			PL_hap2	HO_hap2	HO_hap3	PD_hap1	PD_hap2	HO_hap1	PL_hap1			
HO20-4_Hap_1	*H. octogrammus*	0	0	0	0	0	0	0	0	0	0	0
HO20-6_Hap_2		0	0	7	0	0	0	0	0	0	0	7
HO20-7_Hap_3		0	0	0	1	0	0	0	0	0	0	1
PL20-1_Hap_1	*P. latirostris*	0	0	0	0	0	0	0	106	7	0	113
PL20-2_Hap_2		3	528	4	1	0	0	0	4	4384	119	5043
PD20-1_Hap_1	*P. dybowskii*	0	7	5	5	25312	0	1	6	135	47	25518
PD20-3_Hap_2		11	5	5	7	0	22929	0	7	395	302	23661

a - environmental DNA sample from natural conditions; b - mock community based on samples from Vityaz Bay; c - mock community based on samples from Vostok Bay.

A recent paper examining the expected frequency of NUMTS in various marine animals indicated that it is precisely the use of *COI* Leray that may pose the highest risk of detecting NUMTS in metabarcoding studies (Schultz, Hebert, 2021), as more than 58% of the pseudogenes identified in the study were of lengths up to 300 bp, although that research is more similar in its methodology to that based on PCR free approach or metagenomics (Singer et al., 2020) rather than metabarcoding. It should be emphasized that the results obtained in our work are most likely valid only for metabarcoding, where PCR causes a bias resulting in misrepresentation of the haplotypic diversity of the environmental samples (Tables 3, 4).

The calculations based on the data retrieved from the GeneBank do not allow us to formulate any recommendations for the correction of works to detect genetic diversity using environmental DNA metabarcoding (Figure 2). However, a natural variation in fragment length can quite rarely be expected, which can be used to correct for filtering fragments by length during the computation. The length reduction, at the same time, generally does not entail a decrease in the reliability of the results. The number of population-genetic clusters, which was calculated in this work, is a rather conservative measure, and is not customized for a particular data set with the choice of the exact model. However, it is clear that as one goes from *COI* barcode to metabarcode, the number of identifiable populations is lost by 1 cluster. For the sets that did not exhibit a decrease in them, one cannot detect any pattern other than the intuitive conclusion that length reduction did not affect them due to the strong divergence, and the random concentration of all information within the metabarcode region.

References

Adams, C. I., Hepburn, C., Jeunen, G. J., Cross, H., Taylor, H. R., Gemmell, N. J., ... & Knapp, M. (2022). Environmental DNA reflects common haplotypic variation. Environmental DNA.

Adams, C. I., Knapp, M., Gemmell, N. J., Jeunen, G. J., Bunce, M., Lamare, M. D., & Taylor, H. R. (2019). Beyond biodiversity: can environmental DNA (eDNA) cut it as a population genetics tool?. Genes, 10(3), 192.

Alemu, Y. (2021). Seal pose estimation using convolutional neural networks. Master Thesis. School of Engineering Science, Laskennallinen tekniikka.

Andres, K.J., Sethi, S.A., Lodge, D.M. and Andrés, J. (2021), Nuclear eDNA estimates population allele frequencies and abundance in experimental mesocosms and field samples. Mol Ecol, 30: 685-697. https://doi.org/10.1111/mec.15765

Antich, A., Palacin, C., Wangensteen, O. S., & Turon, X. (2021). To denoise or to cluster, that is not the question: optimizing pipelines for COI metabarcoding and metaphylogeography. BMC bioinformatics, 22(1), 1-24.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV., Sirotkin AV., Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19(5): 455–477.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K.D., and Sayers, E.W. 2018. GenBank. Nucleic Acids Res. 46(D1): D41–D47. Available from http://dx.doi.org/10.1093/nar/gkx1094.

Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., & Madden, T. L. (2012). Domain enhanced lookup time accelerated BLAST. Biology direct, 7, 12. https://doi.org/10.1186/1745-6150-7-12

Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E (2016) obitools: A unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources 16(1): 176-182. https://doi.org/10.1111/1755-0998.12428

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. BMC bioinformatics, 10(1), 1-9.

Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., ... & Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods in Ecology and Evolution, 10(11), 1985-2001.

Dierckxsens N., Mardulyn P. and Smits G. (2016) NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Research, doi: 10.1093/nar/gkw955

Dully, V., Wilding, T. A., Muhlhaus, T., & Stoeck, T. (2021). Identifying the minimum amplicon sequence depth to adequately predict classes in eDNA-based marine biomonitoring using supervised machine learning. Computational and structural biotechnology journal, 19, 2256-2268.

Edgar R.C. (2016), UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing, https://doi.org/10.1101/081257

Egerton, J. P., Johnson, A. F., Turner, J., LeVay, L., Mascarenas-Osorio, I., & Aburto-Oropeza, O. (2018). Hydroacoustics as a tool to examine the effects of Marine Protected Areas and habitat type on marine fish communities. Scientific reports, 8(1), 1-12.

Elbrecht V, Steinke D (2019) Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. Freshwater Biology 64(2): 380-387. https://doi.org/10.1111/fwb.13220

Elbrecht, V., Vamos, E. E., Steinke, D., & Leese, F. (2018). Estimating intraspecific genetic diversity from community DNA metabarcoding data. PeerJ, 6, e4644.

Field, K. A. et al. 2019. Publication reform to safeguard wildlife from researcher harm. – PLoS Biol. 17: e3000193.

Folmer, O., Hoeh, W. R., Black, M. B., & Vrijenhoek, R. C. (1994). Conserved primers for PCR amplification of mitochondrial DNA from different invertebrate phyla. Molecular Marine Biology and Biotechnology, 3(5), 294-299.

Freeland, J. R. (2017). The importance of molecular markers and primer design when characterizing biodiversity from environmental DNA. Genome, 60(4), 358–374. doi:10.1139/gen-2016-0100

Froslev, T. G., Kjoller, R., Bruun, H. H., Ejrnaes, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature communications, 8(1), 1-11.

G. Guillot, F. Mortier, and A. Estoup. Geneland: A computer package for landscape genetics. Molecular Ecology Notes, 5(3):708–711, 2005b.

G. Guillot. Inference of structure in subdivided populations at low levels of genetic differentiation.The correlated allele frequencies model revisited. Bionformatics, 24:2222–2228, 2008.

Geller J, Meyer C, Parker M, Hawk H (2013) Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. Molecular Ecology Resources 13(5): 851-861. https://doi.org/10.1111/1755-0998.12138

Guillot, G., Santos, F., & Estoup, A. (2011). Population genetics analysis using R and the Geneland program. Lyngby, Denmark: Technical University of Denmark

Hazkani-Covo, E., Zeller, R. M., & Martin, W. (2010). Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS genetics, 6(2), e1000834.

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator . Proceedings of the National Academy of Sciences of the United States of America, 101 (41), 14812–14817.

Hering D. et al. Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive // Water Research. 2018. Vol. 138. P. 192-205.

Hilborn, R., Quinn, T. P., Schindler, D. E., & Rogers, D. E. (2003). Biocomplexity and fisheries sustainability. Proceedings of the National Academy of Sciences, USA, 100(11), 6564–6568. https://doi.org/10.1073/pnas.1037274100.

Huson DH, Beier S, Flade I, Gorska A, El-Hadidi M, Mitra S, Ruscheweyh HJ, Tappu R (2016) MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Computational Biology 12(6): e1004957. https://doi.org/10.1371/journal.pcbi.1004957

Jeff A. Eble, Toby S. Daly-Engel, Joseph D. DiBattista, Adam Koziol, Michelle R. Gaither. (2020). Marine environmental DNA: Approaches, applications, and opportunities. Advances in Marine Biology, 86 (1), 141-169. https://doi.org/10.1016/bs.amb.2020.01.001.

Jerde C. L. Can we manage fisheries with the inherent uncertainty from eDNA? // Journal of fish biology. 2019. P. 1-44. doi:10.1111/jfb.14218.

Jerde C. L. et al. “Sight-unseen” detection of rare aquatic species using environmental DNA // Conservation Letters. 2011. Vol. 4. . 2. P. 150-157.

Ji, D., Liang, J., Li, P., Gao, T., & Xu, S. (2020). The complete mitochondrial genome of Hexagrammos agrammus (Scorpaeniformes: Hexagrammidae) by next-generation sequencing. Mitochondrial DNA Part B, 5(3), 2509-2511.

Kalchugin P.V. Long-term dynamics of biomass and dominant species of the bottom fish complex in Peter the Great Bay. Izvestiya TINRO. 2021;201(1):44-61. (In Russian) https://doi.org/10.26428/1606-9919-2021-201-44-61

Kawahara-Miki, R., Wada, K., Azuma, N., & Chiba, S. (2011). Expression profiling without genome sequence information in a non-model species, Pandalid shrimp (Pandalus latirostris), by next-generation sequencing. PloS one, 6(10), e26043.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics (Oxford, England), 28(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kesberg AI, Schleheck D (2013) Improved protocol for recovery of bacterial DNA from water filters: Sonication and backflushing of commercial syringe filters. Journal of Microbiological Methods 93(1): 55-57. https://doi.org/10.1016/j.mimet.2013.02.001

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular biology and evolution, 33(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., ... & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. bioinformatics, 23(21), 2947-2948.

Leitwein, M., Duranton, M., Rougemont, Q., Gagnaire, P. A., & Bernatchez, L. (2020). Using haplotype information for conservation genomics. Trends in ecology & evolution, 35(3), 245-258.

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. Frontiers in Zoology 10(1): 34. https://doi.org/10.1186/1742-9994-10-34

Li D., Hao Y., Duan Y. Nonintrusive methods for biomass estimation in aquaculture with emphasis on fish: a review // Reviews in Aquaculture. 2019. P. 1-22. doi: 10.1111/raq.12388.

Li, D. H., Shi, W., Munroe, T. A., Gong, L., & Kong, X. Y. (2015). Concerted evolution of duplicate control regions in the mitochondria of species of the flatfish family Bothidae (Teleostei: Pleuronectiformes). PLoS One, 10(8), e0134580.

Lischer, H. E., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics, 28(2), 298-299.

Marshall, C., & Parson, W. (2021). Interpreting NUMTs in forensic genetics: Seeing the forest for the trees. Forensic Science International: Genetics, 53, 102497.

Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD (2012) PANDAseq: Paired-end assembler for illumina sequences. BMC Bioinformatics 13(1): 31. https://doi.org/10.1186/1471-2105-13-31

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

Minteer, B. A. et al. 2014. Avoiding (re)extinction. – Science 344: 260–261.

Murakami, H., Yoon, S., Kasai, A. et al. Dispersion and degradation of environmental DNA from caged fish in a marine environment. Fish Sci 85, 327–337 (2019). https://doi.org/10.1007/s12562-018-1282-6

Nester, G. M., Heydenrych, M. J., Berry, T. E., Richards, Z., Wasserman, J., White, N. E., ... & Claassens, L. (2022). Characterizing the distribution of the critically endangered estuarine pipefish (Syngnathus watermeyeri) across its range using environmental DNA. Environmental DNA.

Nugent, C. M., Elliott, T. A., Ratnasingham, S., & Adamowicz, S. J. (2020). Coil: An R package for cytochrome c oxidase I (COI) DNA barcode data cleaning, translation, and error evaluation. Genome, 63 , 291–305. https://doi.org/10.1139/gen-2019-0206

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., & Harris, S. R. (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microbial genomics, 2(4).

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics, 26(3), 419-420.

Porter, T. M., & Hajibabaei, M. (2021). Profile hidden Markov model sequence analysis can help remove putative pseudogenes from DNA barcoding and metabarcoding datasets. BMC Bioinformatics, 22 (256), 1–20. https://doi.org/10.1186/s12859-021-04180-x

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rees H. C. et al. The detection of aquatic animal species using environmental DNA–a review of eDNA as a survey tool in ecology // Journal of Applied Ecology. 2014. Vol. 51. . 5. P. 1450-1459.

Rozas, Julio & Ferrer-Mata, Albert & Sanchez-DelBarrio, Juan & Guirao-Rico, Sara & Librado, Pablo & Ramos-Onsins, Sebastian & Sanchez-Gracia, Alejandro. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. Molecular biology and evolution. 34. https://doi.org/10.1093/molbev/msx248

Schindler, D. E., Hilborn, R., Chasco, B., Boatright, C. P., Quinn, T. P., Rogers, L. A., & Webster, M. S. (2010). Population diversity and the portfolio effect in an exploited species. Nature, 465(7298), 609–612.

Schultz, J., & Hebert, P. (2021). Do pseudogenes pose a problem for metabarcoding marine animal communities?. Authorea Preprints.

Shu, L., Ludwig, A., & Peng, Z. (2021). Environmental DNA metabarcoding primers for freshwater fish detection and quantification: In silico and in tanks. Ecology and Evolution, 11(12), 8281-8294.

Siddiqui S. A. et al. Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data // ICES Journal of Marine Science. 2018. Vol. 75. . 1. P. 374-389.

Sigsgaard, E. E., Jensen, M. R., Winkelmann, I. E., Moller, P. R., Hansen, M. M., & Thomsen, P. F. (2020). Population-level inferences from environmental DNA — Current status and future perspectives. Evolutionary Applications, 13(2), 245– 262. https://doi.org/10.1111/eva.12882

Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., Pedersen, M. W., Jaidah, M. A., Orlando, L., Willerslev, E., Moller, P. R., & Thomsen, P. F. (2016). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. Nature Ecology & Evolution, 1(1), 4– 4. https://doi.org/10.1038/s4155 9- 016- 0004

Singer, G. A., Shekarriz, S., McCarthy, A., Fahner, N., & Hajibabaei, M. (2020). The utility of a metagenomics approach for marine biomonitoring. BioRxiv.

Tsuji, S, Maruyama, A, Miya, M, et al. Environmental DNA analysis shows high potential as a tool for estimating intraspecific genetic diversity in a wild fish population. Mol Ecol Resour. 2020a; 20: 1248– 1258. https://doi.org/10.1111/1755-0998.13165

Tsuji, S., Miya, M., Ushio, M., Sato, H., Minamoto, T., Yamanaka, H. (2019). Evaluating intraspecific genetic diversity of a fish population using environmental DNA: An approach to distinguish true haplotypes from erroneous sequences. bioRxiv 429993. https://doi.org/10.1101/429993

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. 2020b, 20(5), 1323–1332. https://doi.org/10.1111/1755-0998.13200

Turon, X., Antich, A., Palacin, C., Praebel, K., & Wangensteen, O. S. (2020). From metabarcoding to metaphylogeography: separating the wheat from the chaff. Ecological Applications, 30(2), e02036.

Veilleux, H. D., Misutka, M. D., & Glover, C. N. (2021). Environmental DNA and environmental RNA: Current and prospective applications for biological monitoring. Science of the Total Environment, 782, 146891.

Viker, S. M., Klingberg, A. N., & Sundberg, P. (2006). The complete mitochondrial DNA sequence of the northern shrimp, Pandalus borealis. Journal of Crustacean Biology, 26(3), 433-435.

Vucetich, J. A. and Nelson, M. P. 2007. What are 60 warblers worth? Killing in the name of conservation. – Oikos 116: 1267–1278.

Wang et al., Assessment of fish composition and spatio-temporal distribution using acoustic and conventional netting methods in Xiangxi River, a large tributary of the Three Gorges Reservoir, China. Water. 2022. . . ..

Wang, D., Xiang, H., Ning, C., Liu, H., Liu, J. F., & Zhao, X. (2019). Mitochondrial DNA enrichment reduced NUMT contamination in porcine NGS analyses. Briefings in Bioinformatics, 21 (4), 1368–1377. https://doi.org/10.1093/bib/bbz060

Wangensteen OS, Palacin C, Guardiola M, Turon X (2018) DNA metabarcoding of littoral hardbottom communities: High diversity and database gaps revealed by two molecular markers. PeerJ 6: e4705. https://doi.org/10.7717/peerj.4705

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. (2005). DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1462), 1847-1857.

Yang, C., Bohmann, K., Wang, X., Cai, W., Wales, N., Ding, Z., . . . & Yu, D. W. (2021). Biodiversity Soup II: A bulk-sample metabarcoding pipeline emphasizing error reduction. Methods in Ecology and Evolution, 12(7), 1252-1264.

Zemanova, M. A. (2020). Towards more compassionate wildlife research through the 3Rs principles: Moving from invasive to non-invasive methods. Wildlife Biology, 2020.

Zepeda-Mendoza ML, Bohmann K, Carmona Baez A, Gilbert MTP (2016) DAMe: A toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses. BMC Research Notes 9(1): 1-13. https://doi.org/10.1186/s13104-016-2064-9

Additional files, tables and figures

Hosted file

`Supplement A.zip` available at https://authorea.com/users/519712/articles/593475-experimental-evaluation-of-species-and-genetic-variability-based-on-dna-metabarcoding-from-the-aquatic-environment-extra-otus-formed-by-numts-may-reduce-the-diversity-of-asvs

Hosted file

Supplement B.zip available at https://authorea.com/users/519712/articles/593475-experimental-evaluation-of-species-and-genetic-variability-based-on-dna-metabarcoding-from-the-aquatic-environment-extra-otus-formed-by-numts-may-reduce-the-diversity-of-asvs

Hosted file

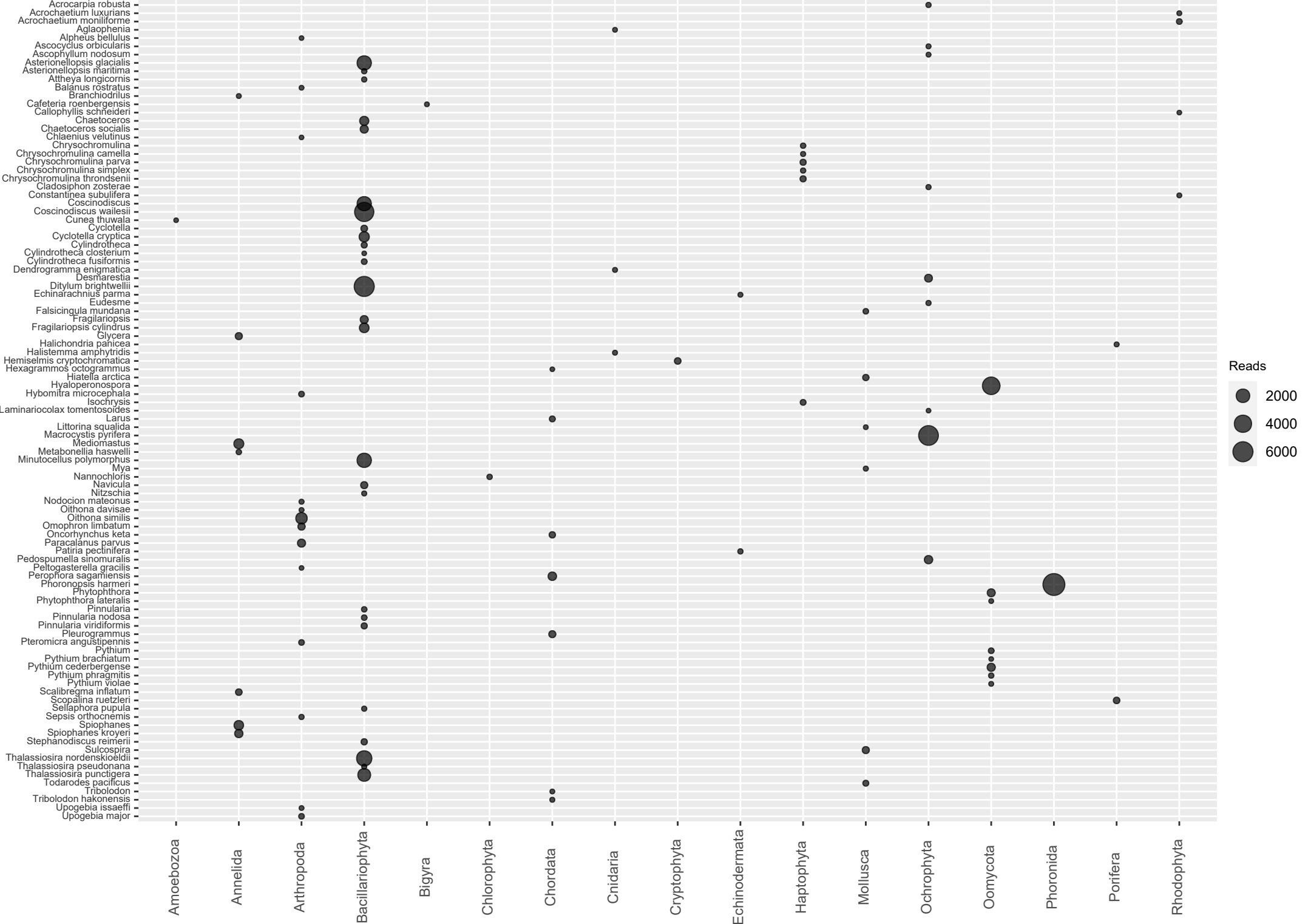
Table S1.xls available at https://authorea.com/users/519712/articles/593475-experimental-evaluation-of-species-and-genetic-variability-based-on-dna-metabarcoding-from-the-aquatic-environment-extra-otus-formed-by-numts-may-reduce-the-diversity-of-asvs

Hosted file

Table S2.csv available at https://authorea.com/users/519712/articles/593475-experimental-evaluation-of-species-and-genetic-variability-based-on-dna-metabarcoding-from-the-aquatic-environment-extra-otus-formed-by-numts-may-reduce-the-diversity-of-asvs

Hosted file

Table S3.xls available at https://authorea.com/users/519712/articles/593475-experimental-evaluation-of-species-and-genetic-variability-based-on-dna-metabarcoding-from-the-aquatic-environment-extra-otus-formed-by-numts-may-reduce-the-diversity-of-asvs


Acrocarpia robusta
Acrochaetium luxurians
Acrochaetium moniliforme
Aglaophenia
Alpheus bellulus
Ascocyclus orbicularis
Ascophyllum nodosum
Asterionellopsis glacialis
Asterionellopsis maritima
Attheya longicornis
Balanus rostratus
Branchiodrilus
Cafeteria roenbergensis
Callophyllis schneideri
Chaetoceros
Chaetoceros socialis
Chlaenius velutinus
Chrysochromulina
Chrysochromulina camella
Chrysochromulina parva
Chrysochromulina simplex
Chrysochromulina throndsenii
Cladosiphon zosterae
Constantinea subulifera
Coscinodiscus
Coscinodiscus wailesii
Cunea thuwala
Cyclotella
Cyclotella cryptica
Cylindrotheca
Cylindrotheca closterium
Cylindrotheca fusiformis
Dendrogramma enigmatica
Desmarestia
Ditylum brightwellii
Echinarachnius parma
Eudesme
Falsicingula mundana
Fragilariopsis
Fragilariopsis cylindrus
Glycera
Halichondria panicea
Halistemma amphytridis
Hemiselmis cryptochromatica
Hexagrammos octogrammus
Hiatella arctica
Hyaloperonospora
Hybomitra microcephala
Isochrysis
Laminariocolax tomentosoides
Larus
Littorina squalida
Macrocystis pyrifera
Mediomastus
Metabonellia haswelli
Minutocellus polymorphus
Mya
Nannochloris
Navicula
Nitzschia
Nodocion mateonus
Oithona davisae
Oithona similis
Omophron limbatum
Oncorhynchus keta
Paracalanus parvus
Patiria pectinifera
Pedospumella sinomuralis
Peltogasterella gracilis
Peronora sagamiensis
Phoronopsis harmeri
Phytophthora
Phytophthora lateralis
Pinnularia
Pinnularia nodosa
Pinnularia viridiformis
Pleurogrammus
Pteromicra angustipennis
Pythium
Pythium brachiatum
Pythium cederbergense
Pythium phragmitis
Pythium violae
Scalibregma inflatum
Scopalina ruetzleri
Sellaphora pupula
Sepsis orthocnemis
Spiophanes
Spiophanes kroyeri
Stephanodiscus reimerii
Sulcospira
Thalassiosira nordenskioeldii
Thalassiosira pseudonana
Thalassiosira punctigera
Todarodes pacificus
Tribolodon
Tribolodon hakonensis
Upogebia issaeffi
Upogebia major
Amoebozoa
Annelida
Arthropoda
Bacillariophyta
Bigyra
Chlorophyta
Chordata
Cnidaria
Cryptophyta
Echinodermata
Haptophyta
Mollusca
Ochrophyta
Oomycota
Phoronida
Porifera
Rhodophyta
Reads
2000
4000
6000


ANGBF4507-12 Hexagrammos otakii JQ738554
ANGBF56879-19 Hexagrammos otakii MK560700
ANGBF2952-12 Hexagrammos otakii HM180619
ANGBF2951-12 Hexagrammos otakii HM180621
ANGBF2286-12 Hexagrammos otakii JQ738555
GBMIN130163-17 Hexagrammos otakii KU236820
GBCCA12772-15 Hexagrammos otakii KP330537
ANGBF7277-12 Hexagrammos otakii JF511630
ANGBF7271-12 Hexagrammos otakii JF511642
ANGBF7263-12 Hexagrammos otakii JF511631
ANGBF7262-12 Hexagrammos otakii JF511633
ANGBF56880-19 Hexagrammos otakii MK560701
ANGBF2953-12 Hexagrammos otakii HM180617
ANGBF2950-12 Hexagrammos otakii HM180624
ANGBF2948-12 Hexagrammos otakii HM180618
ANGBF2947-12 Hexagrammos otakii HM180620
ANGBF2946-12 Hexagrammos otakii HM180622
ANGBF2945-12 Hexagrammos otakii HM180625
ANGBF24228-19 Hexagrammos otakii LC128741
ANGBF2287-12 Hexagrammos otakii JQ738557
ANGBF7258-12 Hexagrammos otakii JF511641
ANGBF7253-12 Hexagrammos otakii JF511651
ANGBF7255-12 Hexagrammos otakii JF511647
ANGBF8483-12 Hexagrammos otakii GU479055
GBMIN130164-17 Hexagrammos otakii KU236823
ANGBF2296-12 Hexagrammos otakii JQ738575
ANGBF4506-12 Hexagrammos otakii JQ738556
ANGBF7265-12 Hexagrammos otakii JF511654
ANGBF7274-12 Hexagrammos otakii JF511636
GBGCA12769-15 Hexagrammos otakii KP330534
GBMIN119977-17 Hexagrammos otakii KU236821
GBMIN119978-17 Hexagrammos otakii KU236822
ANGBF7261-12 Hexagrammos otakii JF511635
ANGBF7267-12 Hexagrammos otakii JF511650
ANGBF7252-12 Hexagrammos otakii JF511653
ANGBF7268-12 Hexagrammos otakii JF511648
FFES079-18 Hexagrammos otakii MW128554
ANGBF24229-19 Hexagrammos otakii MG835058
GBGCA12767-15 Hexagrammos otakii AB770486
GBGCA12774-15 Hexagrammos otakii NC 027195
GBGCA12768-15 Hexagrammos otakii KP330533
ANGBF7264-12 Hexagrammos otakii JF511629
GBGCA12771-15 Hexagrammos otakii KP330536
ANGBF7260-12 Hexagrammos otakii JF511637
ANGBF7275-12 Hexagrammos otakii JF511634
ANGBF4484-12 Hexagrammos otakii JQ738600
ANGBF7270-12 Hexagrammos otakii JF511644
ANGBF7273-12 Hexagrammos otakii JF511638
ANGBF7272-12 Hexagrammos otakii JF511640
ANGBF7278-12 Hexagrammos otakii JF511628
GBMNA14889-19 Hexagrammos otakii KR362879
ANGBF7256-12 Hexagrammos otakii JF511645
ANGBF7259-12 Hexagrammos otakii JF511639
GBMNA17715-19 Hexagrammos otakii AP017443
GBMTG005-16 Hexagrammos otakii NC 028630
ANGBF7257-12 Hexagrammos otakii JF511643
ANGBF7269-12 Hexagrammos otakii JF511646
ANGBF7276-12 Hexagrammos otakii JF511632
ANGBF7266-12 Hexagrammos otakii JF511652
ANGBF7254-12 Hexagrammos otakii JF511649
GBGCA12770-15 Hexagrammos otakii KP330535
ANGBF1797-12 Hexagrammos octogrammus HQ711871
ANGBF2309-12 Hexagrammos otakii JQ738601
ANGBF24220-19 Hexagrammos agrammus LC128742
ANGBF24216-19 Hexagrammos agrammus HM180616
ANGBF875-12 Hexagrammos agrammus HM180614
ANGBF876-12 Hexagrammos sp. JL-2010 HM180627
GBMNA14887-19 Hexagrammos agrammus AB763992
ANGBF24217-19 Hexagrammos agrammus HM180615
ANGBF2949-12 Hexagrammos agrammus HM180626
ANGBF56750-19 Hexagrammos agrammus MK560570
GBMTG3273-16 Hexagrammos agrammus NC 021459
ANGBF873-12 Hexagrammos agrammus HM180612
ANGBF874-12 Hexagrammos agrammus HM180613
Hexagrammos octogrammus OTU2858
GBGCA2840-13 Hexagrammos octogrammus ABT55158
Hexagrammos octogrammus OTU6
GBGCA2807-13 Hexagrammos octogrammus ABT55191
Hexagrammos octogrammus OTU5
GBGCA2809-13 Hexagrammos octogrammus ABT55189
GBGCA2812-13 Hexagrammos octogrammus X Hexagrammos agrammus AB755186
GBGCA2804-13 Hexagrammos octogrammus ABT55194
GBGCA2827-13 Hexagrammos octogrammus X Hexagrammos agrammus AB755171
GBGCA2830-13 Hexagrammos octogrammus ABT55168
GBGCA2805-13 Hexagrammos octogrammus ABT55193
GBGCA2817-13 Hexagrammos octogrammus ABT55181
ANGBF24224-19 Hexagrammos agrammus LC128737
GBGCA2808-13 Hexagrammos octogrammus ABT55190
GBGCA2814-13 Hexagrammos octogrammus X Hexagrammos agrammus AB755184
GBGCA2815-13 Hexagrammos octogrammus ABT55183
GBGCA2820-13 Hexagrammos octogrammus X Hexagrammos agrammus AB755178
GBGCA2824-13 Hexagrammos octogrammus ABT55174
GBGCA2825-13 Hexagrammos octogrammus ABT55173
GBGCA2831-13 Hexagrammos octogrammus ABT55167
GBGCA2832-13 Hexagrammos octogrammus ABT55166
GBGCA2834-13 Hexagrammos octogrammus ABT55164
GBGCA2835-13 Hexagrammos octogrammus ABT55163
GBGCA2837-13 Hexagrammos octogrammus ABT55161
GBGCA2839-13 Hexagrammos octogrammus ABT55159
GBGCA2844-13 Hexagrammos octogrammus X Hexagrammos otakii AB755154
GBGCA2845-13 Hexagrammos octogrammus X Hexagrammos otakii AB755153
GBGCA2848-13 Hexagrammos octogrammus X Hexagrammos otakii AB755150
GBGCA2803-13 Hexagrammos octogrammus ABT55195
GBGCA2806-13 Hexagrammos octogrammus ABT55192
GBGCA2810-13 Hexagrammos octogrammus X Hexagrammos agrammus AB755188
GBGCA2821-13 Hexagrammos octogrammus ABT55177
GBGCA2823-13 Hexagrammos octogrammus ABT55175
GBGCA2828-13 Hexagrammos octogrammus ABT55170
GBGCA2829-13 Hexagrammos octogrammus ABT55169
GBGCA2833-13 Hexagrammos octogrammus ABT55165
GBGCA2838-13 Hexagrammos octogrammus ABT55160
GBGCA2846-13 Hexagrammos octogrammus X Hexagrammos otakii AB755152
GBGCA2811-13 Hexagrammos octogrammus X Hexagrammos agrammus AB755187
GBGCA2813-13 Hexagrammos octogrammus X Hexagrammos agrammus AB755185
GBGCA2816-13 Hexagrammos octogrammus ABT55182
GBGCA2819-13 Hexagrammos octogrammus ABT55179
GBGCA2826-13 Hexagrammos octogrammus X Hexagrammos agrammus AB755172
GBGCA2841-13 Hexagrammos octogrammus ABT55157
GBGCA2842-13 Hexagrammos octogrammus X Hexagrammos agrammus AB755156
GBGCA2843-13 Hexagrammos octogrammus X Hexagrammos agrammus AB755155
GBGCA2847-13 Hexagrammos octogrammus X Hexagrammos otakii AB755151
GBGCA2818-13 Hexagrammos octogrammus ABT55180
GBGCA2822-13 Hexagrammos octogrammus X Hexagrammos agrammus AB755176
GBGCA2836-13 Hexagrammos octogrammus ABT55162
Hexagrammos octogrammus OTU4
DSFIB143-12 Hexagrammos stelleri
FMV080-08 Hexagrammos stelleri JQ354123
DBFOR015-12 Hexagrammos stelleri MW128558
DSFIB144-12 Hexagrammos stelleri
DBFOR016-12 Hexagrammos stelleri MW128556
FHAK117-19 Hexagrammos stelleri
DSFIB048-11 Hexagrammos stelleri
DSFIB107-11 Hexagrammos stelleri
FHAK238-21 Hexagrammos stelleri
ANGBF24230-19 Hexagrammos stelleri LC128740
DBFOR007-12 Hexagrammos stelleri MW128557
FHAK232-21 Hexagrammos stelleri
FMV105-08 Hexagrammos stelleri JQ354122
DSFAL485-09 Hexagrammos stelleri HQ712458
DSFIB142-12 Hexagrammos stelleri
DSFIB338-13 Hexagrammos stelleri
DSFIB046-11 Hexagrammos stelleri
MFCZ59-08 Hexagrammos stelleri GU440341
DSFIB336-13 Hexagrammos stelleri
DSFIB337-13 Hexagrammos stelleri
DSFIB235-12 Hexagrammos stelleri
DSFIB153-12 Hexagrammos stelleri
UKFBI318-08 Hexagrammos lagocephalus KF929970
FHAK213-21 Hexagrammos lagocephalus
GBMNA14888-19 Hexagrammos lagocephalus KP082334
FHAK215-21 Hexagrammos lagocephalus
GBMTG5555-16 Hexagrammos lagocephalus NC 026888
GBGCA12766-15 Hexagrammos lagocephalus KP827344
ANGBF52563-19 Hexagrammos lagocephalus LC125635
ANGBF24222-19 Hexagrammos lagocephalus LC128739
MFCS43-08 Hexagrammos lagocephalus GU440340
FHAK216-21 Hexagrammos lagocephalus
FHAK090-17 Hexagrammos
TZFPB908-08 Hexagrammos lagocephalus FJ164650
DSFAL072-07 Hexagrammos decagrammus HQ712454
FMV119-08 Hexagrammos decagrammus JQ354120
TZFPB179-05 Hexagrammos decagrammus FJ164646
TZFPB910-08 Hexagrammos decagrammus FJ164641
DSFAL053-07 Hexagrammos decagrammus HQ712451
FHAK038-17 Hexagrammos decagrammus
FHAK244-21 Hexagrammos decagrammus
TZFPB180-05 Hexagrammos decagrammus FJ164645
TZFPB914-08 Hexagrammos decagrammus FJ164642
TZFPB915-08 Hexagrammos decagrammus FJ164643
DSFAL070-07 Hexagrammos decagrammus HQ712452
DSFAL073-07 Hexagrammos decagrammus HQ712455
DSFAL074-07 Hexagrammos decagrammus HQ712456
DSFAL680-10 Hexagrammos decagrammus HQ712457
DSFIB157-12 Hexagrammos decagrammus
FHAK041-17 Hexagrammos decagrammus
FHAK239-21 Hexagrammos stelleri
FMV129-08 Hexagrammos decagrammus JQ354119
FMV143-08 Hexagrammos decagrammus JQ354121
MFCO77-08 Hexagrammos decagrammus GU440339
TZFPA155-07 Hexagrammos decagrammus FJ164644
TZFPB176-05 Hexagrammos decagrammus FJ164647
TZFPB177-05 Hexagrammos decagrammus FJ164649
TZFPB178-05 Hexagrammos decagrammus FJ164648
TZFPB904-08 Hexagrammos decagrammus FJ164640
UKFBI275-08 Hexagrammos decagrammus KF929969
ANGBF24221-19 Hexagrammos decagrammus LC128738
DSFAL071-07 Hexagrammos decagrammus HQ712453
UKFBK179-08 Hexagrammos decagrammus KF929968
Hexagrammos octogrammus OTU1584 del
Hexagrammos octogrammus OTU1710
Hexagrammos octogrammus OTU121 del
Hexagrammos octogrammus OTU1583 del
GBMNA14891-19 Pleurogrammus azonus AB744047
GBMTG5169-16 Pleurogrammus azonus NC 023129
GBGCA7721-09 Pleurogrammus azonus EU856711
DBFOR006-12 Pleurogrammus monopterygius MW128593
ANGBF24231-19 Pleurogrammus azonus LC128735
ABFJ112-06 Pleurogrammus azonus JF952816
ANGBF24232-19 Pleurogrammus monopterygius LC128736
DSFAL065-07 Pleurogrammus monopterygius HQ712718
DSFAL066-07 Pleurogrammus monopterygius HQ712719
DSFAL067-07 Pleurogrammus monopterygius HQ712720
DSFAL068-07 Pleurogrammus monopterygius HQ712721
DSFAL069-07 Pleurogrammus monopterygius HQ712722
DSFAL250-07 Pleurogrammus monopterygius HQ712723
DSFAL475-08 Pleurogrammus monopterygius HQ712714
DSFAL476-08 Pleurogrammus monopterygius HQ712715
DSFAL477-08 Pleurogrammus monopterygius HQ712716
DSFAL478-08 Pleurogrammus monopterygius HQ712717
FMV266-08 Pleurogrammus monopterygius JQ354276
FMV267-08 Pleurogrammus monopterygius JQ354278
FMV461-09 Pleurogrammus monopterygius JQ354277
GBGCT720-09 Pleurogrammus monopterygius EU856714
GBMNA14892-19 Pleurogrammus monopterygius AB753026
GBMTG5179-16 Pleurogrammus monopterygius NC 023475
MFC412-08 Pleurogrammus monopterygius GU440465
SDF60001-13 Pleurogrammus monopterygius KF415240
SDF60002-13 Pleurogrammus monopterygius KF415238
SDF60003-13 Pleurogrammus monopterygius KF415242
SDF60009-13 Pleurogrammus monopterygius KF415241
UKFBI355-08 Pleurogrammus monopterygius KF930278
FHAK192-19 Hexagrammos decagrammus
0.020