

Observing Phylum-level Metazoan Diversity by Environmental DNA Analysis at the Ushimado Area in the Seto Inland Sea

Takeshi Kawashima^{1,2*}, Masa-aki Yoshida³, Hideyuki Miyazawa¹, Hiroaki Nakano⁴, Natumi Nakano⁵, Tatsuya Sakamoto⁶, and Mayuko Hamada^{6*}

¹National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan

²Department of Genetics, SOKENDAI University (National Institute of Genetics), Yata 1111, Mishima, Shizuoka 411-8540, Japan

³Marine Biological Science Section, Education and Research Center Biological Resources, Faculty of Life and Environmental Science, Shimane University, Shimane 685-0024, Japan

⁴Shimoda Marine Research Center, University of Tsukuba, 5-10-1, Shimoda, Shizuoka 415-0025, Japan

⁵Department of Biology, Nara Medical University, Nara 634-8521, Japan

⁶Ushimado Marine Institute, Okayama University, Okayama 701-4303, Japan

The dynamics of microscopic marine plankton in coastal areas is a fundamental theme in marine biodiversity research, but studies have been limited because the only available methodology was collection of plankton using plankton-nets and microscopic observation. In recent years, environmental DNA (eDNA) analysis has exhibited potential for conducting comprehensive surveys of marine plankton diversity in water at fixed points and depths in the ocean. However, few studies have examined how eDNA analysis reflects the actual distribution and dynamics of organisms in the field, and further investigation is needed to determine whether it can detect distinct differences in plankton density in the field. To address this, we analyzed eDNA in seawater samples collected at 1 km intervals at three depths over a linear distance of approximately 3.0 km in the Seto Inland Sea. The survey area included a location with a high density of Acoela (*Praesagittifera naikaiensis*). However, the eDNA signal for this was little to none, and its presence would not have been noticed if we did not have this information beforehand. Meanwhile, eDNA analysis enabled us to confirm the presence of a species of Placozoa that was previously undiscovered in the area. In summary, our results suggest that the number of sequence reads generated from eDNA samples in our project was not sufficient to predict the density of a particular species. However, eDNA can be useful for detecting organisms that have been overlooked using other methods.

Key words: eDNA, marine invertebrate, Xenacoelomorpha, Acoela, *Praesagittifera naikaiensis*, Placozoa, *Trichoplax adhaerens*

INTRODUCTION

Environmental DNA (eDNA) is a useful research tool for investigating a variety of organisms in diverse environments (reviewed by Thomsen and Willerslev, 2015). eDNA analysis was first applied to reveal bacterial diversity, and has subsequently been widely used on viruses, microorganisms, and macroorganisms. The application of eDNA analyses in aquatic environments is increasing. Marine eDNA analysis targeted for macroorganisms was originally performed on sediment samples, but recently it has been applied directly to water samples (Tsuji et al., 2017). Surveys for fishes, tetrapods, and some invertebrate species have been reported

in freshwater environments, such as lakes and rivers (Mächler et al., 2014; Hirai et al., 2015; Minamoto et al., 2017). However, zooplankton have not been included in marine eDNA analyses, except for some pioneering research.

We anticipate that an established eDNA analysis protocol for zooplankton surveys would contribute to various areas of zoology research. Investigating the dynamics of zooplankton communities moving in the ocean has long been challenging because visual observation by eye is almost impossible. Plankton nets are a popular tool for observing plankton, but recording the captured plankton is labor intensive. In the future, the development of automated image-recognition technology using artificial intelligence (AI) may reduce the labor required to record captured plankton, but it will still take some time for that technology to become available. Thus, eDNA can be a simple and powerful method of studying plankton at fixed points and fixed

* Corresponding author. E-mail: hamadam@okayama-u.ac.jp (MH); takeshik@nig.ac.jp (TK)

depths in the ocean, and for successive observations.

The international *Tara* Oceans expedition is probably the largest research program in this field (Bork et al., 2015). It included marine eDNA (12S rRNA) metagenomic analysis covering vast areas of ocean worldwide (de Vargas et al., 2015). The target species ranged from microorganisms, including zooplankton, to macroorganisms (Hingamp et al., 2013; Pesant et al., 2015; Gimmler et al., 2016; Alberti et al., 2017).

Nevertheless, marine metagenomics for zooplankton is still under development. Available reference sequences, which include mitochondrial COI and 18S sequences, are not yet sufficient to provide a comprehensive overview of zooplankton diversity. Moreover, there is little integrated knowledge of eDNA bioinformatics targeting zooplankton. The proper treatment for various types of plankton with different sizes and ecology, which are typically captured in a single water sample, is still unknown. Hence, there is a need to establish standardized protocols for analysis and interpretation of plankton-derived eDNA data.

As an initial trial model to address this problem, we looked for a marine invertebrate species that exhibits very high population density within a narrow area. This is because we expect such organisms to show a relatively simple dispersion of zooplankton in seawater. If eDNA analysis is performed at multiple evenly spaced points around a particular species habitat, the sequence read count of the species is expected to decrease gradually with increasing distance from the habitat.

For this purpose, we focused on an acoela species, *Praesagittifera naikaiensis* (formerly named *Convoluta naikaiensis*) (Yamasu, 1982; Jondelius et al., 2011) that is found in the Seto Inland Sea of Japan. In this area, six habitats with high population densities of *P. naikaiensis* were previously identified (Hikosaka-Katayama et al., 2015). Recently, the *P. naikaiensis* habitat survey was expanded and now covers 33 locations (Hikosaka-Katayama et al., 2020). We chose the easternmost of these locations, the Nishiwaki Coast of Ushimado in Okayama Prefecture, Japan, as our observation site (hereafter referred to as “Nishiwaki beach” or “Nishiwaki site”). We then verified whether a series of eDNA sequences obtained from seawater in this area reflected changes in the density of the animals.

While the published DNA sequence data for *P. naikaiensis* are still scarce, they are gradually being augmented. The genome of one species of acoel has been reported (Arimoto et al., 2019), and seven barcode sequences of Acoela are available in the JBIF (Japan Initiative for Biodiversity Information) database (Osawa, 2019). Therefore, insufficiency of the reference data was not a major obstacle for metabarcoding analysis.

This study aimed to investigate whether eDNA can detect distinct differences in plankton density in the field. Seawater samples were collected at about 1-km intervals up to approximately 3.0 km away from Nishiwaki site toward the Kuroshima Island area (Fig. 1). Mitochondrial COI barcode sequences were amplified from eDNA obtained from each seawater sample. We report here on some of the organisms with sequence read counts that showed a characteristic distribution.

MATERIALS AND METHODS

Around the same time as this project, the eDNA Society has proposed standards to be followed in eDNA analysis and compiled them into a protocol (Current version 2.2; The eDNA Society 2020; https://ednasociety.org/wp/wp-content/uploads/2020/09/eDNA_manual_ver2_2_1.pdf). Although we did not follow all of them in this study, we made an effort to follow them as much as possible at that time.

Seawater sampling and environmental DNA extraction

On 4 February 2019, we visited the Nishiwaki beach area (Seto Inland Sea, Okayama, Japan) to confirm the habitat. We collected sand from the shore and observed the conditions of mature individuals living there. In advance, we had observed the morphology of the Acoela species that is usually collected at this site (see Supplementary Figure S1A) and, confirmed that the morphology was identical to that reported by Yamasu (1982). Seawater sampling was conducted from a research boat on 5 February 2019, offshore from Nishiwaki beach. The sampling sites 1, 2, 3, and 4 are located approximately 0.1 km, 1.0 km, 2.0 km, and 2.9 km from the shore, respectively. At each site, water samples were collected at depths of 0, 3, and 5 m, with the exception of site 1 where only 0 and 3 m depth samples were collected because the water depth was less than 5 m. One liter of seawater was collected from the surface of the sea (0 m) using a bucket and from bottom waters (3 and 5 m depth) using a Van Dorn sampler. A total of 11 water samples were collected (Fig. 1, also see Supplementary Table S1).

(Note: The GPS latitude and longitude information for sampling sites in our note described on the ship differed greatly from the actual locations on the map. This was probably due to GPS errors. Therefore, the sampling sites described here are approximate locations based on the topography, as seen on the day of observation.)

Water samples were filtered and subsequent DNA extraction was performed following the methods of Miya et al. (2015), Yamamoto et al. (2016), and an environmental DNA experiment manual published by the eDNA Society. Immediately after collection, water samples were filtered through a Sterivex-HV filter (pore size 0.45 μm ; Merck Millipore) using a 50 mL syringe. The filter cartridges were filled with RNeasy Lysis Buffer (Thermo Fisher Scientific, Waltham, MA) to avoid DNA degradation prior to eDNA extraction. eDNA was extracted from each filter using a DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany), and eluted with 100 μL Buffer AE (QIAGEN), following the manufacturer's instructions.

Amplification of target sequences using polymerase chain reaction (PCR)

Metagenomic libraries were prepared using two-step PCR. In the first PCR, the mitochondrial COI region of the extracted eDNA was amplified using primers mICOLintF/HCO2198 (Folmer et al., 1994; Leray et al., 2013), with a linker sequence and random hexamer nucleotides (see Supplementary Table S2). Using gDNA from *P. naikaiensis* captured at the shore, we confirmed that the primers can amplify the desired length of sequence from the *P. naikaiensis* genome (see Supplementary Figure S1B).

Cycling conditions were as follows: 95°C for 3 min; followed by 16 cycles of 98°C for 20 s, 62°C (-1°C per cycle) for 15 s, and 72°C for 15 s; then 25 cycles of 98°C for 20 s, 46°C for 15 s, and 72°C for 15 s, and 72°C for 5 min. PCR was carried out with a 24 μL reaction volume containing 12 μL of 2 \times KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, WA), 0.72 μL of each primer (10 μM), 5.56 μL of sterile distilled H₂O, and 5 μL of template eDNA (5 μL distilled water for negative controls), with four replicates for each eDNA sample. The combined PCR products of four replicates for eDNA, and the negative control were cleaned up using Agencourt AMPure XP (Beckman Coulter, Brea, CA), according to a standard purification protocol. Purified PCR products were quantified using a

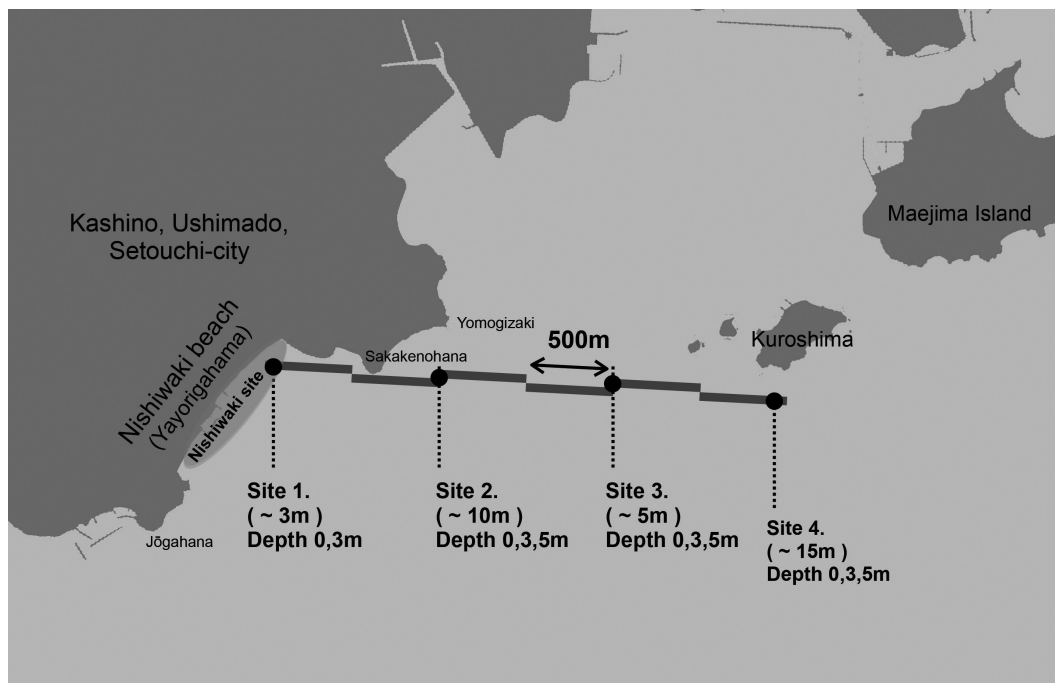


Fig. 1. The collection points of eDNA seawater. Sampling sites (1–4) and depth in the Ushimado area are indicated by black circles and text. The numbers in parentheses indicate the depth (m) from the sea surface to the seafloor at that point. Water samples were taken at depths of 0, 3, and 5 m at each site (no 5 m sample at sampling site 1). An electronic topographic Map 25000 (Geospatial Information Authority of Japan) was processed and created.

Qubit 2.0 fluorometer and the dsDNA HS Assay Kit (Thermo Fisher Scientific), and subsequently diluted to 1 ng/ μ L for use in the second PCR.

In the second PCR, primers corresponding to index 1 (i7) Adapters 701-706 and index 2 (i5) adapters 501 and 502 were used to add MiSeq adaptor sequences and eight-bp index sequences (see Supplementary Table S2). PCR was carried out with a 24 μ L reaction volume containing 12 μ L of 2 \times KAPA HiFi HotStart ReadyMix (KAPA Biosystems), 4 μ L of each primer (1.8 μ M), 2 μ L of sterile distilled water, and 2 μ L of the first PCR product. The PCR conditions were as follows: 95°C for 3 min; eight cycles of 98°C for 20 s, and 72°C for 20 s; and 72°C for 5 min. The amplification and length of the indexed PCR products were confirmed by agarose gel electrophoresis, and the PCR products were purified using Agencourt AMPure XP (Beckman Coulter). The obtained sequencing libraries were quantified using a Qubit 2.0 fluorometer and dsDNA HS Assay Kit (Thermo Fisher Scientific) and pooled at equal concentrations.

The libraries were sequenced using an Illumina MiSeq system with 600-cycle chemistry (2 \times 300 bp paired-end sequencing using MiSeq Reagent Kit v3) (Illumina, San Diego, CA). The total number of reads obtained was approximately 19 million pairs (18,939,064 pairs of reads; see Supplementary Table S3).

To confirm amplification of *P. naikaiensis* COI region using primers mlCOLintF/HCO2198, PCR was performed using *P. naikaiensis* gDNA. We extracted the gDNA from adult *P. naikaiensis* with a DNeasy Blood & Tissue Kit (QIAGEN) according to the manufacturer's protocol. PCR was performed using a T100 Thermal Cycler (Bio-Rad Laboratories, Hercules, CA) using the same conditions as those of the first PCR for preparation of metagenomic libraries. The resultant PCR amplicons were electrophoresed on 1% agarose gels.

Bioinformatic analyses for expression profiling

Seqprep (ver 1.1; John 2011; <https://github.com/jstjohn/SeqPrep>)

was employed to remove adapters, and to merge the forward and reverse reads. Finally, approximately 13 mega pairs of reads were merged into single amplicon sequences (12,968,412 merged reads in total). The length of the merged sequences ranged from 192 bp to 497 bp in length; 81% (10,548,638 reads) were 377 bp and more than 90% (11,700,871 reads) were between 360 bp and 380 bp in length. The size distribution was consistent with that of the target region of the mitochondrial genome.

To estimate the difference in the frequency of occurrence of organisms between samples, the RPKM (reads per kilobase per million mapped reads) was calculated using the following method.

Assignment to reference database

We used the COI-5P sequences provided by JBIF and the NCBI (National Center for Biotechnology Information) organelle genome database (mitochondrial complete genome) as a reference database. We used BLAST + (NCBI), with the option `-max_target_seqs = 1, -outfmt 6` for BLASTN and `-max_target_seqs = 1, -outfmt 6, -query_gencode 5, -db_gencode 5, -evalue 1e-40` for TBLASTX. The obtained BLASTN results were used to assign the reads to the expected taxonomy data based on the species information in the JBIF database and NCBI organelle genome database. In the process of assigning the reads to taxonomy data, we used the sequence match with a percent identity value > 80% and a score value > 100 for both BLASTN and TBLASTX. Of the two reference databases used, we preferentially used the search results from the JBIF database, and only used the search results from the NCBI organelle genome database when there were no hits in the JBIF database.

In obtaining taxonomic data from JBIF and GenBank data, we used the `names.dmp` and `nodes.dmp` files of the NCBI Taxonomy Database. For treating those data, we used an original data parser, Tax.rb (ver 1.0; Kawashima 2016; <https://github.com/tkwsml/Tax>).

Removal of contaminants

Sequence data detected from the controls were treated as contaminants that occurred during the experimental processing. If a sequence detected in the experimental sample group was identical to the contaminants, we excluded it from the subsequent steps of our analyses. We confirmed that the reads eliminated as contaminants in this process did not include the sequences of Acoela or its relatives.

Survey of Placozoa and haplotype check of collected animals

A survey of the Placozoa was performed using ethanol-treated substrate sampling (Miyazawa and Nakano, 2018) in tide pools of the Ushimado coast, Kuroshima Island, and Chiburi Island (Shodoshima Island). To check the haplotype of the collected animals, DNA was extracted using a DNeasy Blood and Tissue Kit (QIAGEN), and the mitochondrial 16S rRNA gene region was amplified via PCR using forward primers reported by Voigt et al. (2004) and the reverse primer reported by Signorovitch et al. (2006). PCR was carried out in a 20 μ L reaction volume containing 10 μ L of AmpliTaq Gold 360 DNA Polymerase (Thermo Fisher Scientific), 1 μ L of each primer (10 μ M), 2–8 μ L of extracted product, and sterile distilled H₂O. PCR conditions were as follows: 95°C for 10 min; 36 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 1 min; and 72°C for 7 min. The amplification and length of the PCR products were confirmed by agarose gel electrophoresis, and the PCR products were purified using Agencourt AMPure XP (Beckman Coulter). DNA sequencing of the purified PCR products was outsourced to MacroGen Japan (Kyoto, Japan). The obtained sequences were compared to the 16S rRNA sequences deposited in GenBank.

Raw data repository

The original raw sequence data used in this paper and the related metadata used in this study were registered in PRJDB9059.

RESULTS

Sequence read statistics

Sequences obtained from the 11 seawater samples were named S1-D0-C, S1-D3-C, S2-D0-C, S2-D3-C, S2-D5-C, S3-D0-C, S3-D3-C, S3-D5-C, S4-D0-C, S4-D3-C, and S4-D5-C. Here, S# indicates the water sampling site (e.g., S1), D# indicates the sampling depth in meters (e.g., D0), and the last letter indicates the primers used (e.g., C). These samples yielded a total of approximately 19 mega reads for both the forward and reverse directions (see Supplementary Table S3). After merging all reads using SeqPrep, a total of approximately 13 mega reads (83%) remained. The lowest number of reads per sample ranged from 769 kilo reads to about double that number (1.5 mega reads). The total number of merged reads was 12,968,412 of which 11,722,707 were from experimental samples and 1,245,705 were from control samples.

Contamination processing

The sequences were then searched by BLASTN against the COI sequences of the JBIF database and NCBI Organellar Genome Resources. Unfortunately, in the control sample, we also found some contamination-derived sequences that appeared to originate mainly from human and laboratory dust sources. The contamination-derived sequences found in the control samples were removed from the other samples.

Species estimation in sequence data

A total of 10,239,395 out of the 11,722,707 merged-reads resulted in some kind of BLASTN hit (see Supplementary Table S3). A total of 1,098,430 reads were assigned to a total of 21,887 species of metazoans under the specified threshold (a percent identity value > 80% and a score value > 100 for both of BLASTN and TBLASTX). To visualize the overall trend, we categorized the results for each of the phyla (Table 1). The BLASTN results were classified according to the NCBI Taxonomy classification, and 23 phyla were suggested. We could not find any clear biological differences or similarity between different depths of the sampling sites (see Supplementary Figure S2).

There is a risk of being assigned to the incorrect taxon if the DNA sequences of closely related species are not registered in the reference data. To avoid this problem, all eDNA sequence reads were matched to each database by TBLASTX as well. Gnathostomulida did not hit any reference by TBLASTX (Table 1). This suggested that the read-counts of this phylum's organisms by BLASTN are not correct. Onychophora can be given as another problematic example. Although multiple eDNAs have been assigned to some species of Onychophora, it is unlikely that these organisms inhabit the Seto Inland Sea (see Supplementary Table S4). When we performed BLASTN search for these eDNAs against NCBI-refseq, they were all assigned to one of the species of arthropods. This is because neither the JBIF nor the NCBI-Organella Database has sufficient coverage of arthropod COI sequences. The large difference in results between BLASTN and TBLASTX for Nemertea is also expected to be a problem with the reference database, but the cause is unknown.

The inadequate state of the reference database was also revealed. Although the JBIF database is a COI sequence database that covers relatively overall metazoan species, Placozoa, Hemichordata, Kinorhyncha, and Entoprocta are found only in another database: NCBI Organellar Database (dark gray cells in Table 1). The difference between the two databases is presumably related to the fact that the former is a partial sequence of the COI gene and the latter is a whole genome sequence.

The most abundant reads detected were from Arthropoda, followed by Cnidaria, Chordata, Mollusca, and Annelida. The reads from Arthropoda accounted for 74% of the total reads. The read count from these five major phyla accounted for more than 90%.

At the species level, many marine species were detected in each phylum, but this is probably also due to the lack of reference data. As an example, the read counts from Tardigrada show 110 counts from four species in BLASTN, and 22 counts from seven species in TBLASTX. However, the tardigrades registered in the reference database we used are all terrestrial tardigrades. Probably, sequences from marine tardigrades were obtained as eDNA, but they hit the sequences of terrestrial tardigrades as the most similar reference sequences. Similar trends were observed for other phyla. The species-level results included many species not found in the Seto Inland Sea, but in the higher taxa, animals often observed around the sampling site were actually included in the results. In addition, the results of some major species were quite reasonable. Therefore, the

Table 1. Summary of eDNA sequence hits to the reference database obtained by using BLASTN and TBLASTX. Summary of read counts of eDNA classified into the metazoan phyla. The results obtained by BLASTN are summarized on the left side, and the results obtained by TBLASTX are summarized on the right side of the table. Note that there are reads classified into four phyla marked with an asterisk (*), but in TBLASTX those read counts became zero. To avoid duplicate hits to the two databases, only the reads with no hit to JBIF were then searched in the NCBI Organella database.

Super Kingdom	Kingdom	Phylum	BLASTN						TBLASTX					
			Reads			Species			Reads			Species		
			JBIF	NCBI Organella	Total	JBIF	NCBI Organella	Total	JBIF	NCBI Organella	Total	JBIF	NCBI Organella	Total
Eukaryota	Metazoa	Arthropoda	547943	305763	853706	13245	3670	16915	339877	147355	487232	3113	2857	5970
		Nemertea	36314	3724	40038	103	56	159	20	5108	5128	7	66	73
		Cnidaria	33874	1489	35363	561	151	712	144484	195227	339711	541	596	1137
		Chordata	21079	11731	32810	244	467	711	80363	7925	88288	88	642	730
		Mollusca	21536	8150	29686	1308	448	1756	23540	79235	102775	593	924	1517
		Annelida	26786	2229	29015	339	60	399	25920	15066	40986	161	173	334
		Rotifera	16694	2	16696	72	2	74	14419	16795	31214	45	19	64
		Porifera	10537	2381	12918	220	163	383	14262	47406	61668	218	378	596
		Echinodermata	2081	9429	11510	165	84	249	321	40821	41142	27	152	179
		Platyhelminthes	2027	271	2298	69	36	105	2405	852	3257	18	15	33
		Nematoda	850	2127	2977	64	149	213	84	548	632	14	9	23
		Bryozoa	398	15	413	32	5	37	6515	4698	11213	24	46	70
		Chaetognatha	281	2	283	11	2	13	109	51	160	8	5	13
		Brachiopoda	198	2	200	9	2	11	191	190	381	8	11	19
		Xenacoelomorpha	13	90	103	4	4	8	8	330	338	3	13	16
		Tardigrada	110	0	110	4	0	4	6	16	22	2	5	7
		Gastrotricha	9	0	9	6	0	6	0	1398	1398	0	11	11
		Placozoa	0	85	85	0	11	11	0	2104	2104	0	13	13
		Hemichordata	0	1	1	0	1	1	0	304	304	0	38	38
		Kinorhyncha	0	4	4	0	3	3	0	6	6	0	1	1
		Entoprocta	0	1	1	0	1	1	0	2	2	0	1	1
		Gnathostomulida	0	1	1	0	1	1	0	0	0	0	0	0
		Acanthocephala	0	0	0	0	0	0	0	0	0	0	0	0
		Onychophora	26474	3729	30203	100	15	115	86	28	114	30	10	40
		Sub Total	747204	351226	1098430	16556	5331	21887	652610	565465	1218075	4900	5985	10885

BLASTN results are suitable for species-level identification, using the high percent identity (> 98%) as a criterion. On the other hand, the TBLASTX results are more suitable for surveying minor organisms at the phylum or class level, using a threshold for low percent identity.

We found much agreement between eDNA and the known biology of the areas we observed, although there are still many technical issues to be addressed, as described above. In the results of TBLASTX in Chordata, the following animals are frequently observed in the vicinity of the sampling sites: urochordates such as *Ascidia* and *Botrylloides*, cephalochordate *Branchiostoma belcheri*, cartilaginous fish *Dasyatis akajei*, and teleosts *Plotosus lineatus*, *Mugil cephalus*, *Epinephelus akaara*, and *Girella punctata* (see Supplementary Tables S5, S6). The finless porpoise *Neophocaena phocaenoides* is known to inhabit the inshore of Ushimado, too. On the other hand, the sequences of terrestrial mammals such as rodents and wild boars appeared in the results. It is possible that water from fields flowed into the sea. In addition, sequences with similarities to some primates (chimpanzees and gorillas) frequently appeared, which could be due to DNA contamination from humans.

As for echinoderms, the sequence of *Asterias amurensis*, one of the most common starfish in coastal areas of Japan, including the Seto Inland Sea, was found most frequently (see Supplementary Table S6). Other echinoderms, such as brittle stars, sea urchins, and sea cucumbers, are also observed around the sampling sites, and the presence of various species was supported by eDNA. It was reported that hemichordates *Glandiceps hacksii* and *Balanoglossus misakiensis* are found in the Seto Inland sea (Urata et al., 2012), although it is difficult to find them by field survey. In this analysis, the *Balanoglossus* sequence was detected, suggesting its habitation in the Ushimado inshore (see Supplementary Table S6, *B. carnosus*, *B. clavigerus*).

Arthropods accounted for the largest part of the results in terms of both species and read numbers (see Supplementary Tables S5, S6). Of these, copepods, barnacles, gammarids, and caprellids (skeleton shrimps) are commonly observed in the sea, seaweeds, and rocky shores around the sampling sites. On the other hand, a significant number of sequences similar to those of terrestrial chelicerates and insects appeared in the results. It is unclear whether they were derived from water from the land or whether many of

Table 2. Examples of a Read Counts profile, for Acoela and Placozoa. The numbers in the profile table are shown using FPKM. Actual read counts are shown in parentheses. The entire profile is described in Supplementary Tables S5 and S6.

BLASTN					TBLASTX				
a <i>Xenacoelomorpha</i>									
a.1 <i>Archaphanostoma ylvae</i>					<i>Archaphanostoma ylvae</i>				
	Site 1	Site 2	Site 3	Site 4		Site 1	Site 2	Site 3	Site 4
Depth 0 m	16 (5)	8 (5)	0 (0)	0 (0)	Depth 0 m	103 (31)	184 (108)	6 (2)	30 (10)
Depth 3 m	151 (76)	9 (4)	0 (0)	0 (0)	Depth 3 m	183 (92)	16 (7)	2 (1)	4 (2)
Depth 5 m	–	0 (0)	0 (0)	0 (0)	Depth 5 m	–	13 (6)	0 (0)	155 (67)
a.3 <i>Haplogonaria</i>					<i>Haplogonaria</i>				
	Site 1	Site 2	Site 3	Site 4		Site 1	Site 2	Site 3	Site 4
Depth 0 m	0 (0)	10 (6)	0 (0)	0 (0)	Depth 0 m	0 (0)	0 (0)	0 (0)	0 (0)
Depth 3 m	0 (0)	0 (0)	0 (0)	0 (0)	Depth 3 m	0 (0)	0 (0)	0 (0)	0 (0)
Depth 5 m	–	0 (0)	0 (0)	0 (0)	Depth 5 m	–	0 (0)	0 (0)	0 (0)
a.2 <i>Symsagittifera roscoffensis</i>					<i>Symsagittifera roscoffensis</i>				
	Site 1	Site 2	Site 3	Site 4		Site 1	Site 2	Site 3	Site 4
Depth 0 m	0 (0)	0 (0)	0 (0)	0 (0)	Depth 0 m	0 (0)	3 (2)	0 (0)	0 (0)
Depth 3 m	0 (0)	0 (0)	0 (0)	0 (0)	Depth 3 m	0 (0)	0 (0)	0 (0)	0 (0)
Depth 5 m	–	0 (0)	0 (0)	0 (0)	Depth 5 m	–	0 (0)	0 (0)	0 (0)
a.4 <i>Daku woorimensis</i>					<i>Daku woorimensis</i>				
	Site 1	Site 2	Site 3	Site 4		Site 1	Site 2	Site 3	Site 4
Depth 0 m	0 (0)	1 (1)	0 (0)	0 (0)	Depth 0 m	0 (0)	0 (0)	0 (0)	0 (0)
Depth 3 m	0 (0)	0 (0)	0 (0)	0 (0)	Depth 3 m	0 (0)	0 (0)	0 (0)	0 (0)
Depth 5 m	–	0 (0)	0 (0)	0 (0)	Depth 5 m	–	0 (0)	0 (0)	0 (0)
a.5 <i>Childia curinii</i>					<i>Childia curinii</i>				
	Site 1	Site 2	Site 3	Site 4		Site 1	Site 2	Site 3	Site 4
Depth 0 m	0 (0)	0 (0)	0 (0)	0 (0)	Depth 0 m	0 (0)	0 (0)	0 (0)	0 (0)
Depth 3 m	0 (0)	0 (0)	2 (1)	0 (0)	Depth 3 m	0 (0)	0 (0)	0 (0)	0 (0)
Depth 5 m	–	0 (0)	0 (0)	0 (0)	Depth 5 m	–	0 (0)	0 (0)	0 (0)
a.6 <i>Isodiametra bajaensis</i>					<i>Isodiametra bajaensis</i>				
	Site 1	Site 2	Site 3	Site 4		Site 1	Site 2	Site 3	Site 4
Depth 0 m	0 (0)	0 (0)	0 (0)	0 (0)	Depth 0 m	10 (3)	6 (4)	0 (0)	0 (0)
Depth 3 m	9 (5)	0 (0)	0 (0)	0 (0)	Depth 3 m	0 (0)	0 (0)	0 (0)	2 (1)
Depth 5 m	–	0 (0)	0 (0)	0 (0)	Depth 5 m	–	2 (1)	0 (0)	0 (0)
a.7 <i>Paratomella rubra</i>					<i>Paratomella rubra</i>				
	Site 1	Site 2	Site 3	Site 4		Site 1	Site 2	Site 3	Site 4
Depth 0 m	0 (0)	0 (0)	0 (0)	0 (0)	Depth 0 m	0 (0)	1 (1)	0 (0)	0 (0)
Depth 3 m	0 (0)	0 (0)	0 (0)	0 (0)	Depth 3 m	0 (0)	0 (0)	0 (0)	0 (0)
Depth 5 m	–	0 (0)	0 (0)	0 (0)	Depth 5 m	–	0 (0)	0 (0)	2 (1)
b.1 <i>Placozoan sp. BZ49</i>					<i>Placozoan sp. BZ49</i>				
	Site 1	Site 2	Site 3	Site 4		Site 1	Site 2	Site 3	Site 4
Depth 0 m	33 (10)	10 (6)	34 (10)	3 (1)	Depth 0 m	0 (0)	0 (0)	0 (0)	0 (0)
Depth 3 m	13 (7)	21 (9)	18 (7)	2 (1)	Depth 3 m	0 (0)	0 (0)	0 (0)	2 (1)
Depth 5 m	–	50 (23)	19 (6)	11 (5)	Depth 5 m	–	0 (0)	0 (0)	0 (0)
b.2 <i>Placozoan sp. 'Shirahama'</i>					<i>Placozoan sp. 'Shirahama'</i>				
	Site 1	Site 2	Site 3	Site 4		Site 1	Site 2	Site 3	Site 4
Depth 0 m	0 (0)	0 (0)	0 (0)	0 (0)	Depth 0 m	0 (0)	0 (0)	0 (0)	0 (0)
Depth 3 m	0 (0)	0 (0)	0 (0)	0 (0)	Depth 3 m	0 (0)	0 (0)	0 (0)	2 (1)
Depth 5 m	–	0 (0)	0 (0)	0 (0)	Depth 5 m	–	0 (0)	0 (0)	0 (0)
b.2 <i>Placozoan sp. BZ10101</i>					<i>Placozoan sp. BZ10101</i>				
	Site 1	Site 2	Site 3	Site 4		Site 1	Site 2	Site 3	Site 4
Depth 0 m	0 (0)	0 (0)	0 (0)	0 (0)	Depth 0 m	328 (98)	389 (228)	393 (114)	886 (291)
Depth 3 m	0 (0)	0 (0)	0 (0)	0 (0)	Depth 3 m	225 (113)	437 (187)	754 (291)	517 (212)
Depth 5 m	–	0 (0)	0 (0)	0 (0)	Depth 5 m	–	471 (214)	437 (136)	507 (218)

these arthropods had dropped into the sea.

The species belonging to Mollusca in our results showed the second greatest diversity. Bivalves *Modiolus* (mussel), and *Crassostrea gigas*, which are actively cultured in the Seto Inland Sea, and some cephalopods were detected by eDNA analysis, while gastropods predominated overall (see Supplementary Table S6). As for Annelida, lugworms are often found in the sand around the sampling site. In addition, *Urechis unicinctus* can be found in the muddy sand in Seto Inland Sea. In fact, these animals were included in the results. Of other spiralian detected by eDNA analysis, species belonging to Nemertea, Platyhelminthes, Bryozoa, and Brachiopoda can be collected in the inshore of Ushimado. Chaetognatha (Sagittoidea) and Rotifera are common plankton found around the sampling sites. In Gastrotricha, sequences similar to three species of Turbanella were included in the BLASTN results (see Supplementary Table S5).

In the basal metazoans, sequences from cnidarians and sponges were detected at high abundance. In the TBLASTX results of cnidarians, a large number of sequences from Actiniidae *Actinia equina* were included. Actiniidae *Anthopleura fuscoviridis* and *Anthopleura uchidai* are common cnidarians found in the intertidal zone of rocky shores near the sampling sites (see Supplementary Table S6). Probably eDNA analysis detected these species. In addition, stony corals, hydroids (*Metridium senile* and *Obelia*) which usually attached to seaweed, and jellyfish belonging to Cubozoa and Scyphozoa were also detected. Interestingly, several minor phyla were detected, including the phylum Placozoa and Acoela.

As mentioned above, we were able to confirm several examples of organisms that have been found in this area at the species or genus level in the eDNA analysis. However, for most of the eDNA reads, it was difficult to confirm whether they corresponded to past records. This indicates that bioinformatic methods that make it possible to easily compare the eDNA analysis with past records in published literature and databases are still required.

Detection and distribution of Acoela sequences

High population density of *P. naikaiensis* is found in the sand of the intertidal zone at Nishiwaki beach, and many of the animals with eggs can be observed in winter (see Supplementary Figure 1A). Thus, we expected that *P. naikaiensis*, including its eggs and larvae and/or their eDNA, should spread from the shore to the sea. However, none of the eDNA sequences obtained in this study were perfectly matched with the previously reported COI sequence of *P. naikaiensis* at the base level. In addition to the whole mitochondrial genome sequence of *P. naikaiensis* reconstructed using whole-genome sequencing (Arimoto et al., 2019), there are 42 partial COI sequences registered in GenBank (as of 16 June 2021). There was a high sequence similarity among these sequences, with even the most different sequences having 97.962% similarity (625 bp out of 638 bp matches for LC515738.1 and LC515764.1). We conducted a series of blast searches for the eDNA to the two reference databases; one is a mitochondrial genome database provided by GenBank and the other is a COI sequence database provided by JBIF.

As a result of BLASTN analysis for the eDNA against GenBank and JBIF, three reference sequences of acoela species other than *P. naikaiensis* were found to have a high similarity score. Ninety reads hit with 80–82% similarity to NC_034947.1 of *Archaphanostoma ylva* and one read hit with 77% similarity to NC_014578 of *Symsagittifera roscoffensis* in GenBank. One read hit with 81% similarity to GBSP4368-12 for *Haplogonaria* sp. in GBIF. The profile for the read-count and the fpkm of these reads that hit the acoela sequences in the experimental samples is shown in Table 2. a (note: all the profiles by the read counts and the fpkm are summarized in Supplementary Table S5 (by BLASTN) and S6 (by TBLASTX)). Although the number of detected reads was very small, the trend for *A. ylva* was similar to what we had initially expected for *P. naikaiensis*, in which the largest number of reads were detected at the depth 3 m of Site 1, and the amount of detection decreased away from there.

Detection of placozoan signals

Placozoans are small free-living animals with a simple body structure, and form one of the most basal lineages of metazoans. They are found in the temperate and tropical seas around the world. On the coast of Japan, surveys of placozoan habitats have demonstrated their wide distribution (Miyazawa and Nakano, 2018), but no placozoans have been reported in the Seto Inland Sea to date.

Our eDNA analysis detected Placozoa-like sequences at all sampling sites (Table 2. b). This suggests that placozoans also inhabit the Seto Inland Sea area. Therefore, we attempted to capture animals by sampling along the coasts of Ushimado, Kuroshima Island, and Chiburi Island (Shodoshima). We successfully captured seven placozoan individuals on Chiburi Island (approximately 6 km south of Nishiwaki coast) (Fig. 2). The mitochondrial 16S rRNA gene

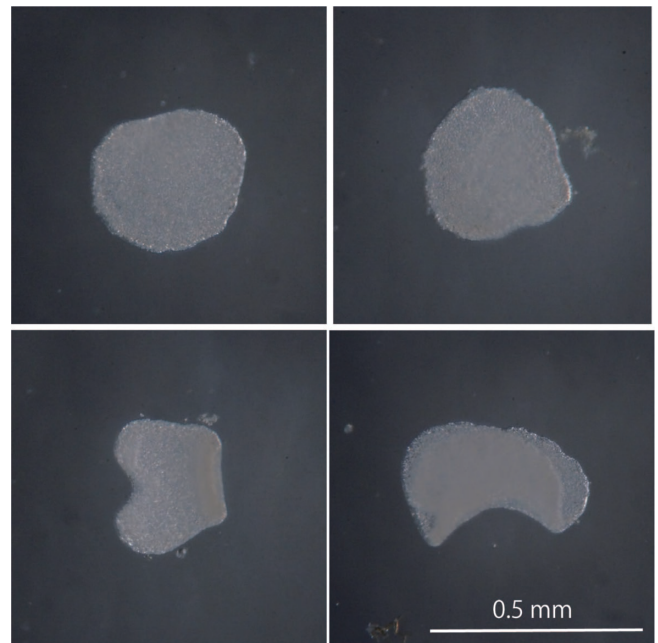


Fig. 2. Collected Placozoa. Four placozoans found on the beach of Chiburi Island.

haplotypes detected in DNA extracted from these individuals were all of the H2 type, which is commonly observed throughout Japan. This is the first record of placozoans in the Seto Inland Sea.

DISCUSSION

Current status of reference databases for eDNA analysis to understand phylum-level diversity

The two reference databases used in this study, the COI-5P dataset provided by the JBIF database and the mitochondrial complete genome dataset from the Organelle Genome Database provided by GenBank, included sequences from 1,814,715 and 9255 species, respectively. Thus, the JBIF database provided more comprehensive coverage of mitochondrial COI sequences in terms of the number of species. However, as evidenced by the fact that the presence of Placozoa could not be detected in the JBIF data set, the Organelle database of GenBank was also necessary for taxon-assignment of eDNA sequences. So, we used the Organelle database of GenBank as a supplement to the JBIF database. Currently, no single database is sufficient for eDNA analysis, and it is necessary to use a combination of multiple reference datasets. Due to this data bias, it became clear that qualitative analysis of eDNA is possible to some extent, but untargeted quantitative analysis is still difficult.

We developed a Ruby library (Tax.rb) to parse the NCBI Taxonomy database on the command line. Tax.rb allows easy conversion of phylogenetic names, species names, and genus names to Taxonomy IDs, and vice versa. The eDNA sequences can easily be mapped to the database-IDs of the database by using BLAST+. However, a convenient application tool has been required to obtain the species names from the blast result or to map the eDNA sequences to higher taxonomic groups. For merging results of blast search to multiple databases, Tax.rb is useful to perform conversion from species names to higher taxa.

Was the target organism captured in this study?

In our eDNA analysis, none of the sequence reads matched with the previously reported mitochondrial COI sequence of *P. naikaiensis*. However, some of our reads showed high sequence similarity to several other Acoela species in the reference databases. In particular, the distribution of reads that matched the sequence of *A. ylva* corresponded to our expectation of the distribution of *P. naikaiensis* eDNA. We therefore suggest that the read counts assigned to *A. ylva* reflect the distribution of the Acoela that we found at the seashore. However, according to Hikosaka et al. (2020), sequence differences between haplotypes within species of Acoela are at most 2%. In contrast, the Acoela sequences in the present eDNA study showed approximately 20% difference from *A. ylva* (65 bp / 319 bp), and 32.2% difference from *P. naikaiensis* (103 bp / 319 bp). However, these sequences showed greater similarity when compared at the amino acid level. For *A. ylva*, the alignment showed 87% (92/106) match and 93% (99/106) positive identity with no gaps. For *P. naikaiensis*, the alignment showed 80% (85/106) match and 90% (95/106) positive identity with no gaps. In addition, the top hits for these sequences in the GenBank database were all Acoela

sequences.

We conclude that our eDNA analysis most likely detected the presence of Acoela, but it remains unclear why the sequences differed significantly from that previously reported for *P. naikaiensis*. One possibility is that the distribution of *P. naikaiensis* eDNA was more localized in the intertidal zone than we expected. For photosynthesis of the symbiotic algae, *P. naikaiensis* shows positive phototaxis, and prefers to inhabit the surface of sand at the shore. Since Site 1 is ~ 100 m from the shore, it is possible that *P. naikaiensis*, including the eggs and larvae, and/or its eDNA, have not spread to that extent. On the other hand, there may be other unknown species of Acoela breeding at the deeper sites of Nishiwaki beach. Further more detailed surveys are required to confirm this.

What is the source of the DNA in eDNA?

The results of our eDNA and our visual observations do not give the same impression. At Nishiwaki beach, we confirmed that *P. naikaiensis* was present in very large numbers, in spite of the fact that eDNA analysis did not provide any clear signals of its presence. On the other hand, the signal of Placozoa was detected in almost all samples of the eDNA (Table 2), although Placozoa has never been found in the Seto Inland Sea before. Not only that, but the discovery of our seven individuals was also made through a very thorough field investigation. The discovery of Placozoa in this study can be attributed to the detection of signals by eDNA.

It is not yet known how the DNA detected as eDNA is present in seawater. For large animals such as fish, it is possible that naked DNA leaks from cells such as those in detached scales or mucus. However, for smaller animals, the source of the DNA being detected is not known. If the form of DNA in the seawater samples is clarified in the future, we may be able to understand the reason why *P. naikaiensis* was not clearly detected in this study.

ACKNOWLEDGMENTS

We would like to thank Drs. Satoshi Yamamoto (Kyoto University), Toshifumi Minamoto (Kobe University), and Yusuke Uchiyama (Kobe University) for their important advice, which was essential for conducting the analyses in this paper. We would like to express our appreciation to Mr. Godo, Mr. Saito, and other staff members of the Ushimado Marine Institute, Okayama University for their assistance with seawater sampling. We would like to thank Mr. Yuta Matsukawa, Mr. Akito Ohtsubo, Mr. Aoshi Kobayashi, and Mr. Shogo Inada who helped to collect the seawater samples from the research boat. Computations were partially performed on the supercomputer at DDBJ, National Institute of Genetics. This work was supported by a Challenging Exploratory Research Project for the Future grant from the Research Organization of Information and Systems (ROIS). We thank Mr. Hisanori Kohtsuka of Misaki Marine Biological Station, the University of Tokyo, and Dr. Tadashi Akiyama of Ushimado Marine Institute, Okayama University, for providing relevant information about our eDNA analysis. This research was conducted as part of the activities of the research group, RinkaiHack. We would like to thank all the people who supported us.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

TK, MY and MH designed the analysis, collected samples and

wrote the manuscript. HM and HN designed the Placozoa sampling and haplotype check. NN planned the molecular experiments and performed the DNA extraction work. TS planned for the seawater collection operation. All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIALS

Supplementary materials for this article are available online. (URL: <https://doi.org/10.2108/zs210073>)

Supplementary Figure S1. Image of *Praesagittifera naikaiensis* and PCR products using primers mICOLintF/HCO2198 and *P. naikaiensis* gDNA.

Supplementary Figure S2. Heatmap of fpkm for the eDNAs.

Supplementary Table S1. Sites for the seawater sampling.

Supplementary Table S2. Primer sequences.

Supplementary Table S3. Summary of sequence reads.

Supplementary Table S4. Phylum level summary of eDNA hits to the reference database by using BLASTN and TBLASTX.

Supplementary Table S5. Profile of read counts using BLASTN.

Supplementary Table S6. Profile of read counts using TBLASTX.

REFERENCES

- Alberti A, Poulain J, Engelen S, Labadie K, Romac S, Ferrera I, et al. (2017) Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Sci Data* 4: 170093
- Arimoto A, Hikosaka-Katayama T, Hikosaka A, Tagawa K, Inoue T, Ueki T, et al. (2019) A draft nuclear-genome assembly of the acoel flatworm *Praesagittifera naikaiensis*. *GigaScience* 8: giz023
- Bork P, Bowler C, de Vargas C, Gorsky G, Karsenti E, Wincker P (2015) Tara Oceans studies plankton at planetary scale. *Science* 348: 873
- de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* 348: 1261605
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3: 294–299
- Gimmler A, Korn R, de Vargas C, Audic S, Stoeck T (2016) The Tara Oceans voyage reveals global diversity and distribution patterns of marine planktonic ciliates. *Sci Rep* 6: 33555
- Hikosaka-Katayama T, Hikosaka A (2015) Artificial rearing system for *Praesagittifera naikaiensis* (Acoela, Acoelomorpha). *Ningen-Kagaku-Kenkyu* 10: 17–23
- Hikosaka-Katayama T, Watanuki N, Niiho S, Hikosaka A (2020) Geographical distribution and genetic diversity of *Praesagittifera naikaiensis* (Acoelomorpha) in the Seto Inland Sea, Japan. *Zool Sci* 37: 314–322
- Hingamp P, Grimsley N, Acinas SG, Clerissi C, Subirana L, Poulain J, et al. (2013) Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J* 7: 1678–1695
- Hirai J, Kuriyama M, Ichikawa T, Hidaka K, Tsuda A (2015) A meta-genetic approach for revealing community structure of marine planktonic copepods. *Mol Ecol Res* 15: 68–80
- John JS (2011) SeqPrep: tool for stripping adaptors and/or merging paired reads with overlap into single reads. URL: <https://github.com/jstjohn/SeqPrep> Accessed 10 April 2019
- Jondelius U, Wallberg A, Hooge M, Raikova OI (2011) How the worm got its pharynx: phylogeny, classification and Bayesian assessment of character evolution in Acoela. *Syst Biol* 60: 845–871
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, et al. (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. *Front Zool* 10: 34
- Mächler E, Deiner K, Steinmann P, Altermatt F (2014) Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. *Freshw Sci* 33: 1174–1183
- Minamoto T, Fukuda M, Katsuhara KR, Fujiwara A, Hidaka S, Yamamoto S, et al. (2017) Environmental DNA reflects spatial and temporal jellyfish distribution. *PLOS ONE* 12: e0173073
- Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, et al. (2015) MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *R Soc Open Sci* 2: 150088
- Miyazawa T, Nakano H (2018) Multiple surveys employing a new sample-processing protocol reveal the genetic diversity of placozoans in Japan. *Ecol Evol* 8: 2407–2417
- Osawa T (2019) Perspectives on biodiversity informatics for ecology. *Ecol Res* 34: 446–456
- Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, Gorsky G, et al. (2015) Open science resources for the discovery and analysis of Tara Oceans data. *Sci Data* 2: 150023
- Philip FT, Willerslev E (2015) Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. *Biol Conserv* 183: 4–18
- Signorovitch AY, Dellaporta SL, Buss LW (2006) Caribbean placozoan phylogeography. *Biol Bull* 211: 149–156
- Thomsen PF, Willerslev E (2015) Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. *Biol Conserv* 183: 4–18
- Tsuji S, Yamanaka H, Minamoto T (2017) Effects of water pH and proteinase K treatment on the yield of environmental DNA from water samples. *Limnology* 18: 1–7
- Urata M, Iwasaki S, Ohtsuka S (2012) Biology of the swimming acorn worm *Glandiceps hacksii* from the Seto Inland Sea of Japan. *Zool Sci* 29: 305–310
- Voigt O, Collins AG, Pearse VB, Pearse JS, Ender A, Hadrys H, et al. (2004) Placozoa - no longer a phylum of one. *Curr Biol* 14: R944–R945
- Yamamoto S, Minami K, Fukaya K, Takahashi K, Sawada H, Murakami H, et al. (2016) Environmental DNA as a 'snapshot' of fish distribution: A case study of Japanese jack mackerel in Maizuru Bay, Sea of Japan. *PLOS ONE* 11: e0149786
- Yamasu T (1982) Five new species of acoel flat worms from Japan. *Galaxea* 1: 29–43

(Received June 25, 2021 / Accepted December 6, 2021 /

Published online January 28, 2022)