fishes

Supplementary Materials

MDPI

Environmental DNA characterization of the fish species composition in the Mukawa River and adjacent habitats

Hyojin Ahn[1,2] and Michael J. Miller[3]

[1]*Research and Educational Unit for Studies on Connectivity of Hills, Humans and Oceans, Kyoto University, Kyoto, Japan*

[2]*Faculty of Fisheries Science, Hokkaido University, Hakodate, Hokkaido, Japan (present address)*

[3] *Department of Aquatic Bioscience, The University of Tokyo, Bunkyo, Tokyo 113-8657, Japan* (Retired)

Correspondence: mjm.ocean.sci@gmail.com; ahn@fish.hokudai.ac.jp

CONTENTS

Supplementary Methods

eDNA extraction

Total DNA was extracted from Sterivex filter units using a DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany), following procedures described by [1] and the manufacturer's protocol, with minor modifications. After removing RNAlater during centrifugation (4000 × *g* for 2 min), the filter units were rinsed with sterilized distilled water. For lysis of eDNA attached to membranes, proteinase K (20 μL) and lysis buffer AL (200 μL) were added to the filter units and incubated inside a 56°C preheated oven for about 20 min. The roller was turned on to enable uniform collection of DNA from the membrane. After incubation, the spin column was centrifuged at 4000 × *g* for 2 min to collect DNA, and then 200 μL of absolute ethanol was added and mixed well. The resulting solution was transferred into a spin column, centrifuged (6000 × *g* for 1 min), and then purified twice using wash buffer (AW1 and AW2). After the purification steps, DNA was eluted with buffer (110 μL) provided in the kit. Extracted DNA was then stored in a LoBind tube at -30°C.

Library preparation and sequencing

A two-step PCR for paired-end library preparation was used in the MiSeq platform (Illumina, San Diego, CA, USA). For the first-round PCR (1st PCR), a mixture of the following four primers was used: MiFish-U-forward (5′–ACA CTC TTT CCC TAC ACG CTC TTC CGA TCT NNN GTC GGT AAA ACT CGT GCC AGC–3′), MiFish-U-reverse (5′–GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TNN NNN NCA TAG TGG GGT ATC TAA TCC CAG TTT G–3′), MiFish-E-forward-v2 (5′–ACA CTC TTT CCC TAC ACG CTC TTC CGA TCT NNN RGT TGG TAA ATC TCG TGC CAG C–3′), and MiFish-E-reverse-v2 (5′–GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TNN NNN NGC ATA GTG GGG TAT CTA ATC CTA GTT TG–3′). These primer pairs amplified a hypervariable region of the mitochondrial 12S rRNA gene (*ca*. 172 bp; hereafter called "MiFish sequence") and appended primer-binding sites (5′ ends of the sequences before six Ns) for sequencing at both ends of the amplicon. The six random bases (Ns) were used in the middle of the primers to enhance cluster separation in the flow cells during initial base call calibrations of the MiSeq platform.

The 1st PCR was carried out with a 12-μL reaction volume containing 6.0 μL 2 × KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, MA, USA), 2.8 μL of a mixture of the four MiFish primers in equal volumes (U/E forward and reverse primers; 5 μM), 1.2 μL sterile distilled

water, and 2.0 μL eDNA template (a mixture of the duplicated eDNA extracts in equal volumes). To minimize PCR dropouts during the 1st PCR, eight replications were performed with the same eDNA template using a strip of eight tubes (0.2 μL). After an initial 3 min denaturation at 95°C, the thermal cycle profile (38 cycles) was as follows: denaturation at 98°C for 20 sec, annealing at 65°C for 15 sec, and extension at 72°C for 15 sec. There was a final extension at 72°C for 5 min. The 1st PCR blanks were prepared during this process along with a non-template control.

After completion of the 1st PCR, equal volumes of the PCR products from the eight replications were pooled and purified using an AMPure XP (BECKMAN COULTER, Brea, CA, USA) following the manufacturer's protocol. Subsequently, the purified target products (ca. 300 bp) were quantified using Agilent 2100 bioanalyzer (Agilent Technologies, Tokyo, Japan). The diluted products were employed as templates for the second-round PCR (2nd PCR). For the 2nd PCR, the following two primers were used to append dual-index sequences (eight nucleotides indicated by Xs) and flow cell-binding sites for the MiSeq platform (5′ ends of the sequences before eight Xs): 2nd-PCR-forward (5′–AAT GAT ACG GCG ACC ACC GAG ATC TAC ACX XXX XXX XAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T–3′) and 2nd-PCR-reverse (5′–CAA GCA GAA GAC GGC ATA CGA GAT XXX XXX XXG TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TCT–3′). The 2nd PCR was performed in a 15-μL reaction volume containing 7.5 μL 2 × KAPA HiFi HotStart ReadyMix, 0.88 μL of each primer (5 μM), 3.88 μL sterile distilled water, and 1.86 μL template (80 ng μL^{-1} except for the PCR blank). After an initial 3 min denaturation at 95°C, the thermal cycle profile (10 cycles) was as follows: denaturation at 98°C for 20 sec, combined annealing and extension at 72°C for 15 sec. There was a final extension at 72°C for 5 min.

All dual-indexed libraries were pooled in equal volumes into a 1.5-mL tube. Then, the pooled dual-indexed library was separated on a 2% E-Gel Size Select agarose gel (Life Technologies, Carlsbad, CA, USA) and the target amplicons (*ca*. 370 bp) were retrieved from the recovery wells using a micropipette. The concentration of the size selected libraries was measured using a Qubit dsDNA HS assay kit and a Qubit fluorometer (Life Technologies). The libraries were pooled and diluted to 4 nM with HT1 buffer (Illumina) and sequenced on the MiSeq platform using a MiSeq Reagent Kit v2 300 cycle (Illumina) following the manufacturer's protocol.

Data preprocessing and taxonomic assignment

Data preprocessing and analysis of MiSeq raw reads were performed with a pipeline (MiFish ver. 2.4) using USEARCH v11.0.667 [2]. The following steps were applied: (1) Forward (R1) and reverse (R2) reads were merged by aligning them with the *fastq_mergepairs* command. During this process, the following reads were discarded: low-quality tail reads with a cut-off threshold set at a quality (Phred) score of 2, reads that were too short (<100 bp) after tail trimming, and paired reads with multiple differences (>5 positions) in the aligned region (*ca*. 65 bp). (2) Primer sequences were removed from merged reads using the *fastx_truncate* command. (3) Reads without primer sequences underwent quality filtering using the *fastq_filter* command to remove low-quality reads with an expected error rate >1% and reads that were too short (<120 bp). (4) Preprocessed reads were dereplicated using the *fastx_uniques* command and all singletons, doubletons, and tripletons were removed from subsequent analysis as recommended [2]. (5) Dereplicated reads were denoised using the *unoise3* command to generate amplicon sequence variants (ASVs) without any putatively chimeric and erroneous sequences [3]. (6) Finally, ASVs were subjected to taxonomic assignments of species names (molecular operational taxonomic units; MOTUs) using the *usearch_global* command with sequence identity >98.5% to the reference sequences. ASVs with sequence identities of 80–98.5% were tentatively assigned "U98.5" labels before the corresponding species name with the highest identity (*e.g.*, U98.5_*Pagrus_major*) and they were subjected to clustering at the 0.985 level using the *cluster_smallmem* command. In an incomplete reference database, this clustering step enables the detection of multiple MOTUs under an identical species name. Those multiple MOTUs were annotated as "gotu1, 2, 3…" and tabulated all these outputs (MOTUs plus U98.5_MOTUs) with read abundances. ASVs with sequence identities <80% (saved as "no_hit") were excluded from the above taxonomic assignments and downstream analyses because all of them were found to be non-fish organisms.

As a reference database, MiFish sequences from Masaki Miya's laboratory were assembled [4]. In addition, whole mitochondrial genome and 12S rRNA gene sequences of all fish were downloaded from NCBI and extracted MiFish sequences using a custom Perl script. The MiFish sequences were combined from the two sources in a FASTA format and used the combined sequences as the custom reference database for taxonomic assignments. These automatic taxonomic assignments were refined with reference to family-level phylogenies based on MiFish sequences from MOTUs, U98.5_MOTUs, and the reference sequences from those families. For

each family, representative sequences (most abundant reads) were assembled from MOTUs and U98.5_MOTUs and added all reference sequences from that family and an outgroup (a single sequence from a closely-related family) in a FASTA format. The FASTA file was subjected to multiple alignment using MAFFT [5] with a default set of parameters. A neighbor-joining tree was constructed with the aligned sequences in MEGA7 [6] using Kimura two-parameter distances. The distances were calculated using pairwise deletion of gaps and among-site rate variations modeled with gamma distributions (shape parameter = 1). Bootstrap resamplings (n = 100) were performed to estimate statistical support for internal branches of the neighbor-joining tree and to root the tree with the outgroup, in order to reduce false sequence from the reference database. A total of 82 family-level trees were inspected and revised the taxonomic assignments. When U98.5_MOTUs placed within a monophyletic group consisting of a single genus, that genus was unidentified MOTUs with "sp" plus sequential numbers *(e.g.*, *Pagrus* sp1, sp2, sp3, ...). For the remaining MOTUs ambiguously placed in the family-level tree, the family name was assigned with "sp" plus sequential numbers (*e.g.*, Sparidae sp1, sp2, sp3, ...).

Species verification

The species obtained by pipeline still needed to be verified because sequencing results comprised only a short region (170 bp) of 12S rRNA [4], and similar sequences might correspond to different species. Also, multiple species could be incorporated into a single species, and vice versa. Therefore, all species on the list were checked with the original aligned sequences using the NCBI Basic Local Alignment Search Tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and applied MEGA7 [6] to construct a neighbor-joining tree for all stations characterized by occurrence of the same species. When several species shared the same or similar (>99%) aligned sequence, the species identity was confirmed by referring to species distribution reported by the IUCN (https://www.iucnredlist.org), FishBase (http://www.fishbase.de; [7]), illustrated books of Japanese fishes [8-10], and personal communications with local fishermen. Those species which were confirmed neither by sequence nor distribution, combined as genus spp. (*e.g.*, *Chelon* spp., *Pungitus* spp.) or family name (*e.g.*, Cyprinidae, Agonidae).

Species whose reads number amounted to <0.05% of total reads of library were deleted because they were potentially caused by contamination, as indicated by [11] with some modifications. If species that were obviously not expected in this area were detected, and

commonly consumed food items, they were regarded as contamination and removed as well. Negative control and PCR blanks were also analyzed by exactly same procedure. The number of reads corresponding to every fish detected in the negative control and PCR blanks were deleted and stone loach (*Noemacheilus barbatulus toni*) was removed from 500 m right side of adjacent shore after this process.

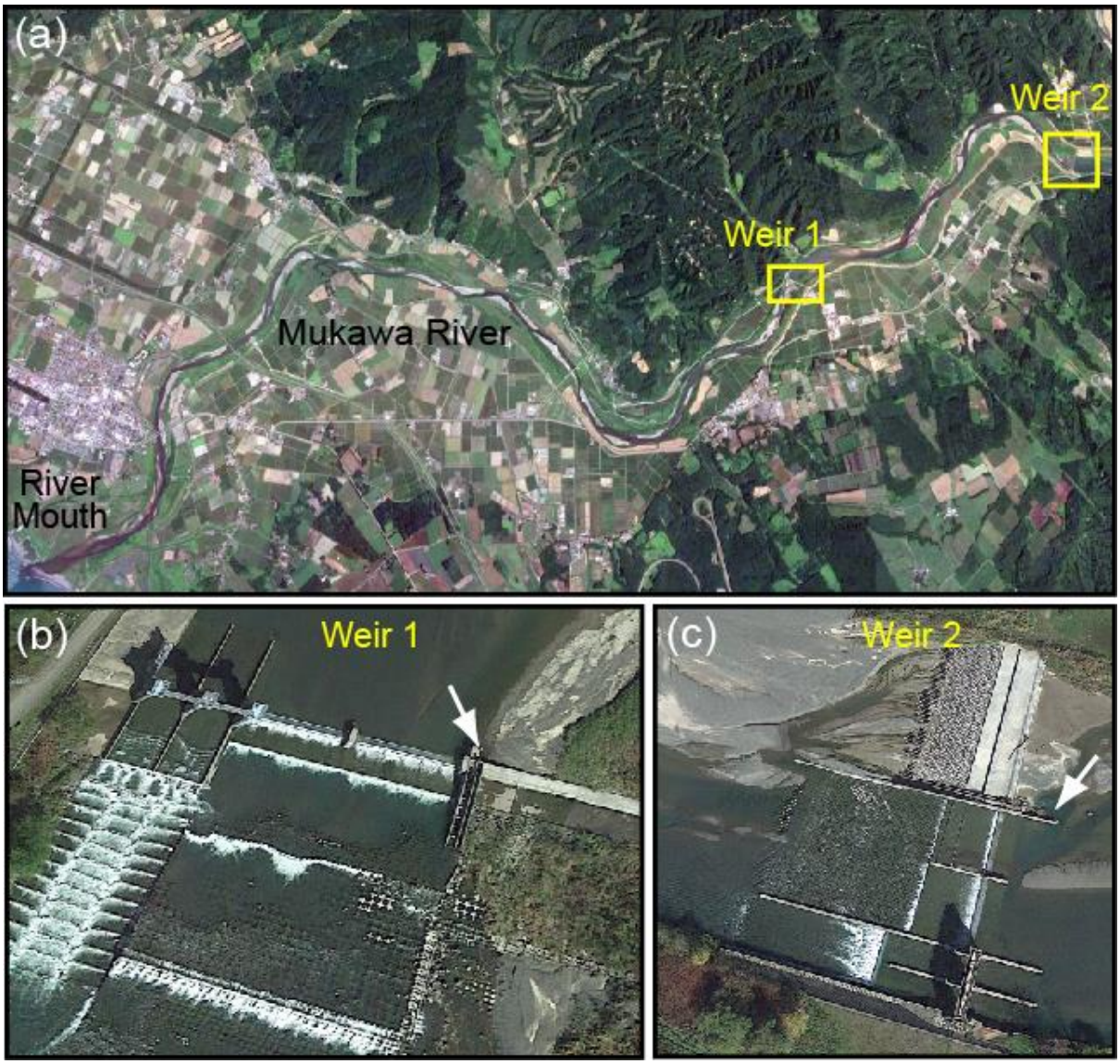

Figure S1. Satellite image of the lower Mukawa River (a) showing the locations of two water control weirs across the river (b, c), with arrows pointing out apparent fish ladder structures. Images were modified from Google Earth imagery.

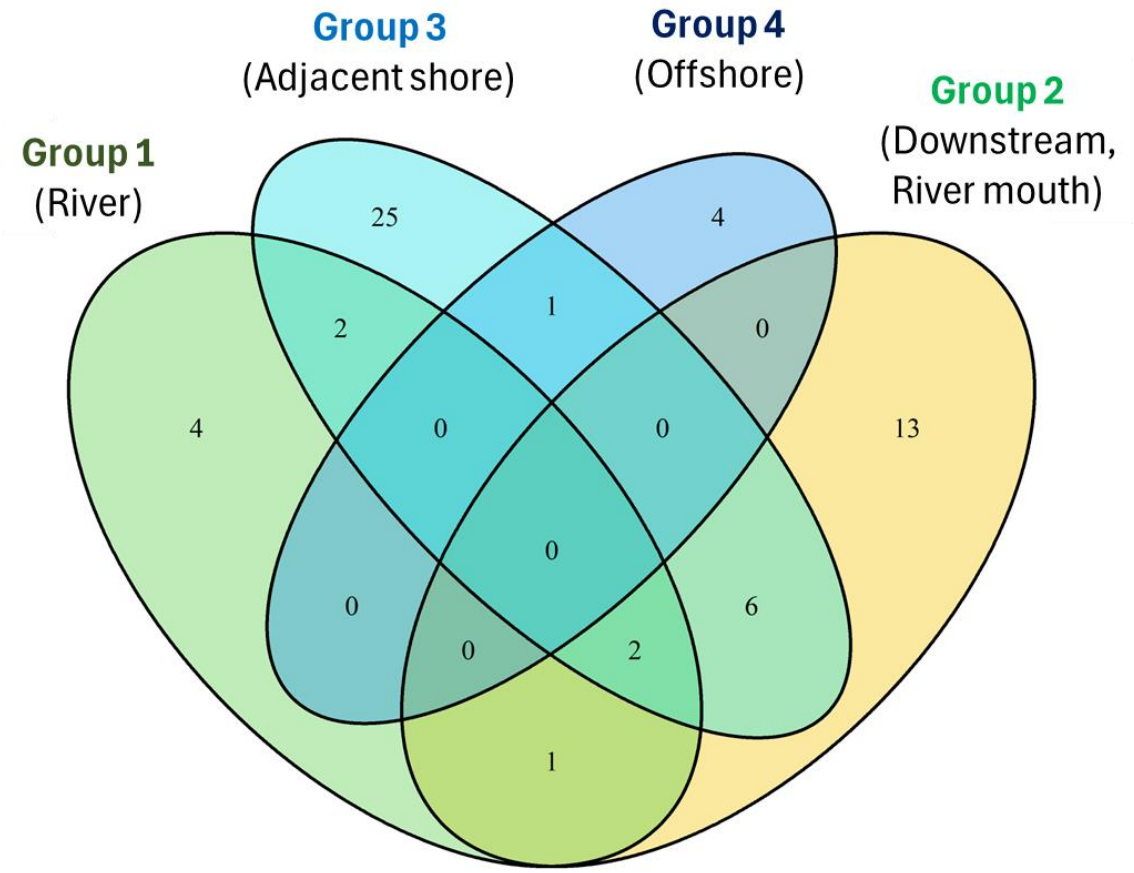

Figure S2. A Venn diagram showing the number of detected OTUs shared among the four groups of sampling sites of river (upstream, midstream), downstream and river mouth (high tide, low tide), adjacent shore (left and right sides), and offshore, which were determined by cluster analysis and ordination (see Fig. 7). Colors of the group labels follow those in Figure 7.

Table S1. List of OTUs detected and corresponding read numbers at each sampling station (US: upstream, MS: midstream, DS: downstream, LT: river mouth at low tide, HT: river mouth at high tide, RS1: right shore 500 m from river mouth, RS2: right shore 1 km from river mouth, LS1: left shore 500 m from river mouth, LS2: left shore 1 km from river mouth, OS: offshore). Different colors represent the original habitat of fish (dark green: freshwater, light-green: fresh-brackish water, yellow: fresh-brackish-seawater, light blue: brackish-seawater, dark blue: seawater) classified by Nelson [12]. †: endemic species, and ‡: endangered species by Ministry of the Environment of Japan (https://www.env.go.jp/).

Scientific Name	Common Name	Life history / habitat type	US	MS	DS	LT	HT	RS1	RS2	LS1	LS2	OS
Cyprinidae	Minnows and Carp	Freshwater	0	0	136	1591	224	0	0	86	0	0
Pseudorasbora parva	Topmouth gudgeon	Freshwater	0	0	0	61	45	0	0	0	0	0
Rhynchocypris percnurus sachalinensis †	Lake minnow	Freshwater	0	0	0	153	499	0	0	0	0	0
Misgurnus anguillicaudatus	Pond loach	Freshwater	0	33	157	876	643	0	0	0	0	0
Noemacheilus barbatulus toni	Brook loaches	Freshwater	13777	18124	5010	3141	4635	0	779	0	0	0
Lefua nikkonis †‡	Brook loaches	Freshwater	0	0	0	429	552	0	0	0	0	0
Pseudaspius hakonensis	Japanese dace	Freshwater-Brackish	0	1279	8506	4920	6720	4260	12073	2858	4018	0
Phoxinus spp.	Eurasian minnow	Freshwater-Brackish	0	0	0	0	0	109	0	98	0	0
Pungitius spp.	Stickleback	Freshwater-Brackish	0	0	0	804	1003	0	0	0	0	0
Rhinogobius sp.	Goby	Mostly-Freshwater	0	0	227	57	0	0	170	0	0	0
Tribolodon brandtii maruta	Pacific redfin	Anadromous	0	0	6908	5031	4586	17950	5193	1693	9079	0
Pseudaspius sachalinensis	Dace	Anadromous	0	8172	5381	7005	3480	483	553	100	0	0
Hypomesus japonicus	Smelt	Anadromous	0	0	0	0	0	0	0	131	23	0
Hypomesus olidus	Pond smelt	Anadromous	0	0	0	0	0	0	0	19	0	0
Osmerus dentex	Pacific rainbow smelt	Anadromous	0	0	0	0	0	0	0	0	63	0
Spirinchus lanceolatus †	Shishamo (Saltwater Smelt)	Anadromous	0	0	0	0	0	30	0	0	0	0
Oncorhynchus masou masou	Masou salmon	Anadromous	394	269	1808	1397	952	0	346	0	0	0
Oncorhynchus mykiss	Rainbow trout	Freshwater/Anadromous	208	71	36	0	0	0	0	0	0	0
Salvelinus leucomaenis leucomaenis	Whitespotted char	Freshwater/Anadromous	1725	0	0	0	0	851	0	0	0	0
Salvelinus malma kraschenninnikovi ‡	Dolly Varden trout	Freshwater/Anadromous	589	0	0	0	0	0	0	0	0	0
Eleginus gracilis	Saffron cod	Coastal-Brackish	0	0	0	0	0	0	0	0	0	115
Chelon spp.	Mullet	Coastal-Brackish	0	0	0	0	0	0	0	109	0	0
Mugil cephalus	Flathead grey mullet	Coastal-Brackish	0	0	121	1911	5523	0	261	0	0	0
Gasterosteus spp.	Stickleback	Freshwater-Brackish-Marine	0	0	0	119	0	0	0	0	0	0
Cottus amblystomopsis	Sakhalin sculpin	Amphidromous	12500	0	33	0	0	22	63	0	0	0
Acanthogobius lactipes	Goby	Amphidromous	0	0	0	377	251	0	0	0	0	0
Gymnogobius castaneus	Chestnut goby	Amphidromous	0	0	0	828	0	0	0	0	0	0
Gymnogobius opperiens	Goby	Amphidromous	0	0	53	215	150	0	0	0	207	0
Gymnogobius spp.	Goby	Amphidromous	0	0	95	454	386	0	0	0	0	0
Tridentiger brevispinis	Goby	Amphidromous	0	0	1232	1262	1302	634	0	92	0	0
Squalus suckleyi	Pacific spiny dogfish	Coastal-Brackish	0	0	0	0	0	0	0	22	0	0
Clupea pallasii	Pacific herring	Coastal-Brackish	0	0	0	0	0	0	527	379	822	0

Gadus chalcogrammus	Alaska pollock	Coastal-Brackish	0	0	0	0	0	0	0	130	0	0
Sardinops melanostictus	Japanese sardine	Marine	0	0	0	0	0	4557	8829	1013	3331	11818
Myctophum asperum	Prickly lanternfish	Marine-Oceanic	0	0	0	0	0	0	0	0	0	17497
Hexagrammos spp.	Greenling	Marine	0	0	0	0	0	0	0	0	2830	0
Pleurogrammus azonus	Okhotsk atka mackerel	Marine	0	0	0	0	0	0	0	0	0	161
Myoxocephalus spp.	Sculpin	Marine	0	0	0	0	0	0	1603	422	1028	0
Hemitripterus villosus	Sea raven	Marine	0	0	0	0	0	0	0	2584	125	0
Agonidae	Poacher	Marine	0	0	0	0	0	0	350	0	0	0
Podothecus spp.	Poacher	Marine	0	0	0	0	0	0	0	55	0	0
Liparis agassizii	Agassiz's snailfish	Coastal	0	0	0	0	0	483	0	1650	2028	0
Liparis punctulatus	Snailfish	Marine	0	0	0	0	0	0	0	133	0	0
Stichaeidae	Prickleback	Marine	0	0	0	0	0	0	0	0	492	0
Stichaeopsis nana	Eelpout	Marine	0	0	0	0	0	0	0	1893	5342	0
Pholis spp.	Gunnel	Intertidal	0	0	0	0	0	366	0	250	333	0
Scomber spp.	Mackerel	Marine-Oceanic	0	0	0	0	0	0	0	0	0	1769
Pleuronectidae	Righteye flounder	Coastal	0	0	0	81	51	0	160	8492	203	0
Pleuronectes punctatissimus	Speckled flounder	Marine	0	0	0	0	0	0	0	2905	0	0
Pseudopleuronectes spp.	Righteye flounder	Marine	0	0	0	0	0	1064	0	351	1341	0
Verasper moseri	Barfin flounder	Marine	0	0	0	0	0	604	361	0	0	0
Number of detected OTUs			**6**	**6**	**14**	**20**	**17**	**13**	**14**	**23**	**16**	**5**
Sampling stations			**US**	**MS**	**DS**	**LT**	**HT**	**RS1**	**RS2**	**LS1**	**LS2**	**OS**

Supplementary References

1. Miya, M., T. Minamoto, H. Yamanaka, S. I., Oka, K., Sato, S., Yamamoto, T., Sado, H. Doi, 2016. Use of a filter cartridge for filtration of water samples and extraction of environmental DNA. Journal of Visualized Experiments 117: e54741.
2. Edgar R. C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460–2461.
3. Callahan B. J., P. J. McMurdie & S. P. Holmes, 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME Journal 11: 2639–2643.
4. Miya, M., T. Sado, 2019. Multiple species detection using MiFish primers. In: eDNA Methods Standardization Committee editors. Environmental DNA sampling and experimental manual ver. 2.1. The eDNA Society, Otsu, Japan, pp. 55–92.
5. Katoh, K., H. Toh, 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinformatics 9: 286-298.
6. Kumar, S., G. Stecher, K. Tamura, 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33: 1870–1874.
7. Froese, R., D. Pauly, 2021. FishBase. World Wide Web electronic publication. www.fishbase.org
8. Hosoya, K., 2015. Freshwater fishes of Japan. Yama-Kei Publishers, Tokyo, Japan. (in Japanese)
9. Nakabo, T., 2013. Fishes of Japan with pictorial keys to the species. 3rd ed. Tokai University Press, Kanagawa, Japan.
10. Nakajima, J., 2017. Loaches of Japan. Yama-Kei Publishers, Tokyo, Japan. (in Japanese)
11. Andruszkiewicz, E. A., H. A. Starks, F. P. Chavez, L. M. Sassoubre, B. A. Block, A. B. Boehm, 2017. Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. PLOS ONE 12: e0176343.
12. Nelson, J. S., 2006. Fishes of the World. 4th ed. Wiley John & Sons, New Jersey, USA.