High-frequency eDNA metabarcoding survey data of coastal fish across the Japanese Archipelago

Hideyuki Doi[1], Masaki Miya[2,3], Hitoshi Araki[4], Akihide Kasai[5], Reiji Masuda[6], Toshifumi Minamoto[7], Satoquo Seino[8], Hiroki Yamanaka[9], Tetsuya Sado[10], Hisashi Yamakawa[11], Hyojin Ahn[5], Mitsuhiro Aizu-Hirano[8], Seiji Arakaki[12], Ryutaro Goto[13], Keisuke Koizumi[13], Nobuchika Hara[8], Yasuhisa Henmi[14], Masakazu Hori[15], Midori Iida[16,17], Tadashi Kajita[18], Yumi Kobayashi[4], Manabu Kume[19], Shin-ichiro Oka[20], Shunsuke Matsuoka[21], Yuki Minegishi[22], Mariko Nagano[23], Teruhiko Takahara[24,25], Hiroshi Takahashi[26], Sayaka Takahashi[6], Koji Takayama[27], Kunihiro Yamauchi[8], and Michio Kondoh[28,29]

1. Graduate School of Informatics, Kyoto University, Kyoto, Kyoto 606-8501, Japan

2. Natural History Museum and Institute, Chiba, Chiba 260-8681, Japan

3. Research Organization for Nano & Life Innovation, Waseda University, Tokyo 162-0041, Japan

4. Research Faculty of Agriculture, Hokkaido University, Sapporo, Hokkaido 060-8589, Japan

5. Faculty of Fisheries Sciences, Hokkaido University, Hakodate, Hokkaido 041-8611, Japan

6. Maizuru Fisheries Research Station, Field Science Education and Research Center, Kyoto University, Maizuru, Kyoto 625-0086, Japan

7. Graduate School of Human Development and Environment, Kobe University, Kobe, Hyogo 657-8501, Japan

8. Department of Urban and Environmental Engineering, Graduate School of Engineering, Kyushu University, Fukuoka, Fukuoka 819-0395, Japan

9. Department of Environmental Solution Technology, Faculty of Science and Technology, Ryukoku University, Otsu, Shiga 520-2194, Japan

10. National Institute for Environmental Studies, Lake Biwa Branch Office, Otsu, Shiga 520-0022, Japan

11. Department of Applied Genomics, Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan

12. Amakusa Marine Biological Laboratory, Kyushu University, Reihoku-Amakusa, Kumamoto 863-2507, Japan

13. Seto Marine Biological Laboratory, Field Science Education and Research Center, Kyoto University, Nishimuro, Wakayama 649-2195, Japan

14. Aitsu Marine Station, Center for Water Cycle, Marine Environment and Disaster Management, Kumamoto University, Kumamoto, Kumamoto 861-6102, Japan

15. Japan Fisheries Research and Education Agency, Yokohama, Kanagawa 236-8648, Japan

16. Marine Biological Station, Sado Island Center for Ecological Sustainability, Niigata University, Sado, Niigata 952-2135, Japan.

17. Usujiri Fisheries Station, Field Science Center for Northern Biosphere, Hokkaido University, Usujiri, Hakodate, Hokkaido 041-1613, Japan.

18. Iriomote Station, Tropical Biosphere Research Center, University of the Ryukyus, 870 Uehara, Taketomi, Yeyama, Okinawa, 907-1541, Japan

19. Faculty of Science and Engineering, Ishinomaki Senshu University, Ishinomaki, Miyagi 986-8580, Japan

20. Okinawa Churashima Foundation, Motobu, Okinawa 905-0206, Japan

21. Ashiu Forest Research Station, Field Science Education and Research Center, Kyoto University, Nantan, Kyoto 601-0703, Japan

22. Atmosphere and Ocean Research Institute, The University of Tokyo, Iwate 028-1102, Japan

23. Faculty of Bioenvironmental Sciences, Kyoto University of Advanced Science, Kameoka, Kyoto 621-8555, Japan

24. Faculty of Life and Environmental Sciences, Shimane University, Matsue, Shimane 690-8504, Japan

25. Estuary Research Center, Shimane University, Matsue, Shimane 690-8504, Japan

26. Department of Applied Aquabiology, National Fisheries University, Shimonoseki, Yamaguchi 759-6595, Japan

27. Department of Applied Aquabiology, National Fisheries University, Shimonoseki, Yamaguchi 759-6595, Japan.

28. Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi 980-8578, Japan

29. WPI-AIMEC, Tohoku University, Sendai, Miyagi 980-8578, Japan

Corresponding Authors
Hideyuki Doi (hideyuki.doi@icloud.com),
Masaki Miya (masaki_miya@me.com), or
Michio Kondoh (michio.kondo.b8@tohoku.ac.jp).

Abstract

The Japanese archipelago, spanning over 3,000 km from subarctic to subtropical zones (approximately 24°N to 45°N), hosts a diverse array of coastal fish communities shaped by pronounced latitudinal gradients. Here, we present a high-frequency environmental DNA (eDNA) metabarcoding dataset of coastal fishes collected across the archipelago. Using the MiFish primer set targeting a hypervariable region of the mitochondrial 12S rRNA gene, we conducted surveys at 195 sites, yielding 690 samples collected through one-time (annual), monthly, and biweekly sampling schemes. The inclusion of high-frequency sampling enabled the detection of fine-scale temporal variations in fish community composition. From January 2018 to March 2019, eDNA samples were collected along with metadata on geographic and environmental variables, including coastal structure, wave height, tide, currents, and water temperature. The resulting dataset comprises fish eDNA records for 1,477 species across 171 families. This dataset offers a valuable resource for investigating fish community structure, temporal dynamics, and biodiversity in the coastal ecosystems of the Japanese archipelago, supporting a range of applications in ecology, conservation, and eDNA-based research.

Background & Summary

The Japanese archipelago, stretching over 3,000 km along the northeastern edge of the Eurasia continent, encompasses a wide range of marine ecosystems, ranging from the subarctic waters of the Sea of Okhotsk to the subtropical regions of the East China Sea and the Pacific Ocean. This extensive latitudinal range, spanning from 45.5°N to 24.2°N, positions Japan as a critical region for investigating the ecological dynamics of coastal fish communities under varying climatic conditions. Japan is recognized as a global hotspot for marine biodiversity, particularly for fish species, with over 4,300 marine fish species recorded, accounting for approximately 25% of the world's total marine fish diversity [1][2]. This exceptional richness is attributed to the convergence of major ocean currents, such as the warm Kuroshio and cold Oyashio currents, which create diverse habitats ranging from coral reefs in the south to kelp forests in the north [3]. In addition, the archipelago also harbors unique assemblages of endemic and migratory species, further enhancing its significance as a biodiversity hotspot [1] [2].

Understanding the temporal and spatial dynamics of coastal fish communities is essential for effective marine conservation, particularly in the face of climate change and anthropogenic pressures such as overfishing and habitat degradation [4]. Traditional survey methods, such as net sampling and visual censuses, are often labor-intensive and limited in their ability to capture the full spectrum of biodiversity, especially for cryptic or rare species [5]. Environmental DNA (eDNA) metabarcoding has emerged as a transformative tool for biodiversity monitoring, enabling non-invasive detection of species from water samples, which is particularly effective for assessing fish communities in marine environments (e.g., [6][7]). eDNA-based methods allow for the simultaneous detection of multiple species, including those that are elusive or present in low abundance, thereby providing a more comprehensive view of community composition compared to traditional methods [8].

Our dataset is the result of an eDNA metabarcoding survey conducted across 195 sites from January 2018 to March 2019 (Fig. 1). The dataset includes 690 samples and

represents 1,496 species in 171 families of fishes, identified through eDNA metabarcoding using the MiFish primer set targeting the mitochondrial 12S rRNA gene [9]. All samples were collected using bucket sampling followed by on-site filtration, with environmental metadata recorded alongside the eDNA data (see Methods). Species identifications were verified through taxonomic, genotypic, and expert validation by ichthyologists. The families Gobiidae (180 species), Labridae (80 species), Blenniidae (58 species), Serranidae (48 species), and Muraenidae (46 species) exhibit the greatest species diversity (Fig. 2). This dataset captures approximately 30% of the estimated coastal fish diversity in the Japanese archipelago [1] and serves as foundational community data to enhance knowledge of species diversity, distributions, and community dynamics, supporting conservation efforts and ecological studies in this global biodiversity hotspot. The dataset also includes detailed metadata for each sample, such as the location, sampling date and time, coastal structure, wave height, tide, current, and water temperature, alongside sequence and metabarcoding data for coastal fish species.

The survey generated 690 samples through one-time (annual), monthly, and biweekly sampling schemes. In particular, the monthly and biweekly samplings captured fine-scale temporal variations in fish community composition. High-frequency sampling offers significant advantages for community analysis, especially in a region as geographically and climatically diverse as Japan. Monthly and biweekly datasets allow researchers to detect short-term ecological changes, such as seasonal migrations, spawning events, and responses to environmental fluctuations [10]. This temporal resolution is critical for understanding community turnover and stability, as it captures dynamic processes that annual or sporadic surveys may overlook [11]. Furthermore, the broad spatial coverage of 195 sites ensures that the dataset can be used to investigate biodiversity patterns and ecological drivers across subarctic, temperate and subtropical zones.

Methods

Sampling sites

Sampling was conducted at 195 sites across the Japanese archipelago (Fig.1). Sites were

selected based on four criteria: (1) accessibility from inland areas, (2) suitability for bucket sampling, (3) minimal anthropogenic impact (e.g., avoiding sites near sewage discharge from fish markets), and (4) a minimum straight-line distance of 10 km between sites to reduce spatial autocorrelation in fish community data.

Water sampling and on-site filtration

Water sampling was conducted between January 2018 and March 2019 using a 7.8-L polypropylene bucket (Soft Bucket 8, ISETO, Osaka, Japan) attached to a 15-m nylon rope. At each site, two seawater samples were collected using ten bucket casts per sample to minimize bias. On-site filtration was performed using Sterivex filter cartridges (0.45 μm pore size; Merck Millipore, MA, USA) and 50-mL syringes (TERUMO, Tokyo, Japan), targeting a filtration volume of 1,000 mL per sample (range: 200–900 mL in turbid waters, in the metadata). Filters were preserved with 1.6 mL of RNAlater (Thermo Fisher Scientific, DE, USA), sealed, and stored at –20°C. A daily filtration blank consisting of 500 mL of purified water was processed to monitor contamination.

Environmental data

Environmental conditions were recorded at the time of sampling to provide contextual information for the eDNA data. Weather conditions (sunny, cloudy, or rainy) were visually observed and noted during each sampling event. Wave height was visually estimated in meters (e.g., 0.5 m, 1 m). The coastal environment at each site was classified into one of four types based on visual inspection: rock, sand, concrete revetment, or rock revetment. Tidal conditions were recorded as lowest, low, high, or highest, based on tide tables and on-site observations. Current speed was estimated by observing surface water movement and categorized as either <1 m/s or >1 m/s. Surface water temperature was measured using a water temperature meter, such as the YSI Pro30 (YSI Inc., Yellow Springs, OH, USA).

DNA Extraction

eDNA was extracted from Sterivex filter cartridges using a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) following Miya et al. [12]. After enclosed RNAlater was removed, the filter cartridge was subjected to lysis using Proteinase K. Prior to lysis, a mixture of phosphate-buffered saline (PBS; 220 μL), Proteinase K (20 μL), and buffer AL (200 μL) was prepared and gently pipetted into the cartridge through the inlet port. The port was then sealed, and the cartridge was placed in a preheated incubator at 56˚C for 20 minutes with an oscillation mode using a rotator (RotoFlex Plus, Argos Technologies, Chicago, USA). After incubation, the seal was removed from the inlet port, which was then connected to a 2.0 mL DNA LoBind tube (Eppendorf Corporate, Hamburg, Germany) for DNA collection. The cartridge-tube assembly was placed inside a 50-mL conical tube and centrifuged at 6,000x g for 1 minute to recover the DNA extract. The resulting DNA extract (approximately 400 μL) was then purified using the DNeasy Blood & Tissue Kit according to the manufacturer's protocol, with a final elution volume set at 200 μL.

Library preparation and sequencing

A two-step PCR approach was employed for library preparation on the MiSeq platform (Illumina, CA, USA). For the first-round PCR (1st PCR), a mixture of four MiFish primers was used to co-amplify a hypervariable region of the fish mitochondrial 12S rRNA gene [9]. The primers were: MiFish-U-forward (5′–ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT NNN 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 ACG CTC TTC CGA TCT NNN 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′). Each primer included sequencing primer-binding sites at the 5′ end (before the six Ns) and six random bases (Ns) to improve cluster separation on the MiSeq flow cell during initial base call calibrations.

The 1st PCR for the samples and blanks was conducted with 35 cycles in a 12-μL reaction

volume containing 6.0 μL of 2× KAPA HiFi HotStart ReadyMix (KAPA Biosystems, MA, USA), 2.8 μL of an equimolar mixture of the four MiFish primers in equal volumes (U/E forward and reverse primers at 5 μM), 1.2 μL of sterile distilled water, and 2.0 μL of eDNA template (a mixture of duplicated eDNA extracts in equal volumes). To minimize PCR dropout, eight replicate reactions were performed for each eDNA template using an eight-tube strip (0.2 μL per tube). The thermal cycling conditions consisted of an initial denaturation at 95°C for 2 minutes, followed by 35 cycles of denaturation at 98°C for 20 seconds, annealing at 65°C for 15 seconds, and extension at 72°C for 15 seconds, with a final extension at 72°C for 5 minutes. A 1st PCR blank (1B) was included alongside the filtration blank (FB) and extraction blank (EB) to monitor contamination. However, to reduce experimental costs, only a single tube was used for each of the three blanks (FB, EB, 1B) rather than performing eight replicates. After the 1st PCR, equal volumes of PCR products from the eight replicates were pooled into a single 1.5-mL tube and purified using a GeneRead Size Selection Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The purified target products (~300 bp) were quantified using a TapeStation 2200 (Agilent Technologies, Tokyo, Japan), diluted to 0.1 ng/μL with Milli-Q water, and used as templates for the second-round PCR (2nd PCR).

The 2nd PCR was performed to append dual-index sequences (8 nucleotides, represented as Xs) and flow cell-binding sites for the MiSeq platform. The primers used were 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 reaction was carried out in a 12-μL volume containing 6 μL of 2× KAPA HiFi HotStart ReadyMix, 0.35 μL of each primer (10 μM), 3.8 μL of sterile distilled water, and 1.5 μL of the diluted 1st PCR product (0.1 ng/μL, except for the three blanks). The thermal cycling conditions included an initial denaturation at 95°C for 2 minutes, followed by 10 cycles of denaturation at 98°C for 20 seconds and a combined annealing and extension step at 72°C for 15 seconds, with a final extension at 72°C for 5 minutes. Prepared libraries were pooled, size-selected, and sequenced on the MiSeq platform using a MiSeq v2 Reagent Kit for 2 × 150 bp paired-end sequencing, with a 5% PhiX control spike-in to improve sequencing quality.

Data Preprocessing and Taxonomic Assignment

MiSeq raw reads were processed using USEARCH v10.0.240 [13]. The processing pipeline included the following steps: (1) merging paired-end reads, (2) trimming primer sequences, (3) quality filtering (removal of reads with >1% error rate or length <120 bp), (4) dereplication (excluding singletons, doubletons, tripletons), (5) denoising to generate amplicon sequence variants (ASVs) [14], and (6) taxonomic assignment using a custom reference database of 5,691 fish species. ASVs with ≥98.5% sequence identity to references were assigned species names (MOTUs); ASVs with identity between 80% and 98.5% were clustered as U98.5 MOTUs. Non-fish sequences (<80% identity) and deep-sea, oceanic, or freshwater species, were excluded from the dataset.

Data Records

The dataset comprises 690 eDNA samples collected from 195 coastal sites across the Japanese archipelago. It includes two main components: (1) metadata on sampling conditions and environmental variables (file name: CREST2018_data_site.csv), and (2) eDNA metabarcoding data (file name: CREST2018_data_eDNA.csv).

Metadata

Metadata for each sample is provided in the file CREST2018_data_site.csv, detailing sampling conditions and environmental variables. Two eDNA samples were collected per sampling event, each assigned a unique sample code (Sample_Code_1 and Sample_Code_2) representing the replicate samples for that event. Filtration blanks were also processed to monitor potential contamination, with associated blank codes (Blank_Code) and volumes (Blank_volume_mL). As summarized in Table 1, the metadata includes the following fields; Sample_Num, Sample_ID_Japanese, Sample_Type, Sampling_Date, Sampling_Time, Sampling_Weather, Site_Num, Site_Name_Japanese, Latitude, Longitude, Environment, Wave_Height_(m), Tide, Current, Water_Temperature, Water_Temperature_meter, BAC_addition,

Sample_1_volume_mL, Sample_2_volume_mL, Blank_volume_mL, Sample_Code_1, Sample_Code_2, and Blank_Code.

Metabarcoding Data

The eDNA sequence data derived from metabarcoding of the 12S rRNA gene are provided in a separate file, CREST2018_eDNA_data.csv, which contains taxonomic assignments and sequence read counts for fish species detected across samples. As summarized in Table 2, the file includes fields such as Family, Scientific_Name, Common_Name_Japanese, Ave_Identity, Ave_LOD, and the columns containing read counts corresponding to each sample code (e.g., C2018-HUA-0039). Raw sequence data are available in FASTA format and have been deposited in the DDBJ/EMBL/GenBank databases under accession number DRA****., linked to the corresponding taxonomic assignments.

Technical Validation

To ensure the accuracy of taxonomic assignments in this eDNA metabarcoding dataset, sequences were validated using molecular characteristics regarding the MitoFish database [17], a comprehensive repository of fish mitochondrial sequences. MitoFish was selected for validation due to its extensive collection of fish mitochondrial genomes, including 12S rRNA gene sequences, which provides a robust reference for marine fish species inhabiting the Japanese archipelago. Each taxon-specific alignment was manually inspected to confirm that it represented a single taxon and to identify any divergent sequences. Sequences were considered to belong to the same taxon if their sequence identity was ≥97.5%. Sequences with identity ≤97.4% were realigned with additional reference sequences from MitoFish, and taxonomic assignments were re-evaluated to ensure accuracy. Low-quality reads (expected error rate >1% or read length <120 bp), as well as sequences with read counts <10, were excluded to minimize false positives and ensure high-quality sequence data for downstream analyses.

Once each multiple-sequence alignment was confirmed to represent a single taxon based

on our criteria, a representative sequence from each taxon was queried against the MitoFish database to verify its identity. Phylogenetic validation was performed by constructing Neighbor-Joining (NJ) trees in MEGA7, based on aligned sequences and MitoFish references, ensuring that sequences formed monophyletic groups with their assigned taxa. To minimize amplification and sequencing bias, duplicate samples and eight PCR replicates per sample were used throughout the workflow.

Usage Notes

This dataset can be used to support eDNA-based biodiversity monitoring and ecological analyses. The eDNA data (CREST2018_data_eDNA.csv) can be merged with metadata (CREST2018_data_site.csv) using the Sample_Code_1 and Sample_Code_2 fields to correlate species presence with environmental variables. Users may also re-analyze sequence data using the raw FASTQ files available from the DRA repository, particularly as reference databases and bioinformatic tools are updated.

Code Availability

No custom code was used for data set construction.

Acknowledgments

We acknowledge the CREST2018 project team, Kyoka Kido, Ryoji Asada, and the others for their efforts on the eDNA sampling. This study was supported financially by JSPS CREST program (JPMJCR13A2) and JST COI-NEXT program (JPMJPF2206).

Author Contributions

H. Doi, M. Miya, and M. Kondoh conceived the study and drafted the manuscript. H. Doi, M. Miya, H. Araki, A. Kasai, R. Masuda, T. Minamoto, S. Seino, H. Yamanaka designed the sampling strategy and conducted field sampling. H. Ahn, M. Aizu-Hirano, S. Arakaki, R. Goto, K. Koizumi, N. Hara, Y. Henmi, M. Hori, M. Iida, T. Kajita, Y. Kobayashi, M. Kume, S. Oka, S. Matsuoka, Y. Minegishi, M. Nagano, T. Takahara, T.

Sado, H. Takahashi, S. Takahashi, K. Takayama, and K. Yamauchi conducted fieldwork. H. Yamakawa performed laboratory analyses. All authors contributed to data curation, provided critical feedback, and approved the final manuscript.

Competing Interests

The authors declare no competing interests.

References

[1] Kai, Y., Motomura, H., & Matsuura, K. (eds). Fish Diversity of Japan: Evolution, Zoogeography, and Conservation. (Springer, 2022).

[2] Fujikura, K., Lindsay, D., Kitazato, H., Nishida, S. & Shirayama, Y. Marine biodiversity in Japanese waters. PLoS ONE 5, e11836 (2010).

[3] Yamano, H., Sugihara, K. & Nomura, K. Rapid poleward range expansion of tropical reef corals in response to rising sea surface temperatures. Geophys. Res. Lett. 38, L04601 (2011).

[4] Lekve, K., et al. Spatio-temporal dynamics of species richness in coastal fish communities. Pro. Royal Soc. B 2691781-1789 (2002).

[5] Thomsen, P. F. et al. Monitoring endangered freshwater biodiversity using environmental DNA. Mol. Ecol. 21, 2565–2573 (2012).

[6] Ushio, M., et al. Temperature sensitivity of the interspecific interaction strength of coastal marine fish communities. eLife 12, RP85795. (2023).

[7] Oka, S., et al. Environmental DNA metabarcoding for biodiversity monitoring of a highly diverse tropical fish community in a coral reef lagoon: Estimation of species richness and detection of habitat segregation. Environmental DNA, 3, 55-69 (2021).

[8] Deiner, K. et al. Environmental DNA metabarcoding: transforming how we survey animal and plant communities. Mol. Ecol. 26, 5872–5895 (2017).

[9] Miya, M. et al. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 200 species from a single water sample. R. Soc. Open Sci. 2, 150088 (2015).

[10] Magurran, A. E. & Henderson, P. A. Temporal turnover and the maintenance of diversity in ecological assemblages. Phil. Trans. R. Soc. B 365, 3611–3620 (2010).

[11] Dornelas, M. et al. Quantifying temporal change in biodiversity: challenges and opportunities. Proc. R. Soc. B 280, 20121931 (2013).
[12] Miya, M. et al. Use of a filter cartridge for filtration of water samples and extraction of environmental DNA. J. Vis. Exp. 115, e54741 (2016).
[13] Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461 (2010).
[14] Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 11, 2639–2643 (2017).
[15] Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9, 286–298 (2008).
[16] Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874 (2016).
[17] Iwasaki, W. et al. MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. Mol. Biol. Evol. 30, 2531–2540 (2013).

Table 1 Specimen metadata associated with each sample in the dataset

Field	Description
Sample_Num	Unique identifier for the sampling event
Sample_ID_Japanese	Japanese identifier for the sampling event
Sample_Type	Type of sample (Sample or Blank)
Sampling_Date	Date of sample collection (YYYY/MM/DD)
Sampling_Time	Time of sample collection (HH:MM)
Sampling_Weather	Weather condition during sampling (Sunny, Cloudy, or Rainy)
Site_Num	Unique identifier for the sampling site
Site_Name_Japanese	Name of the sampling site in Japanese
Latitude	Latitude of the sampling site (degrees)
Longitude	Longitude of the sampling site (degrees)
Environment	Visually estimated coastal environment type (Rock, Sand, Concrete revetment, or Rock revetment)
Wave_Height_(m)	Visually estimated wave height at the time of sampling (meters)
Tide	Visually estimated tidal condition (Lowest, Low, High, Highest)
Current	Visually estimated water current speed (<1 m/s or > 1 m/s)
Water_Temperature	Surface water temperature at the sampling site (°C)
Water_Temperature_meter	Instrument used for temperature measurement (e.g., YSI Pro30)
BAC_addition	benzalkonium chloride (BAC) was added for the samples or not
Sample_1_volume_mL	Volume of the first sample filtered (mL)
Sample_2_volume_mL	Volume of the second sample filtered (mL)
Blank_volume_mL	Volume of the blank sample filtered (mL)
Sample_Code_1	Unique code for the first eDNA sample (e.g., C2018-HUA-0039)
Sample_Code_2	Unique code for the second eDNA sample (e.g. C2018-HUA-0040)
Blank_Code	Unique code for the blank sample (e.g. C2018-HUA-0041)

Table 2 Structure of the eDNA sequence data

Field	Description
Family	Taxonomic family of the detected fish species
Scientific_Name	Scientific name of the detected fish species
Common_Name_Japanese	Common name of the species in Japanese
Ave_Identity	Average sequence identity (%) to the reference database
C2018-HUA-0039	Read count of the species in sample C2018-HUA-0039

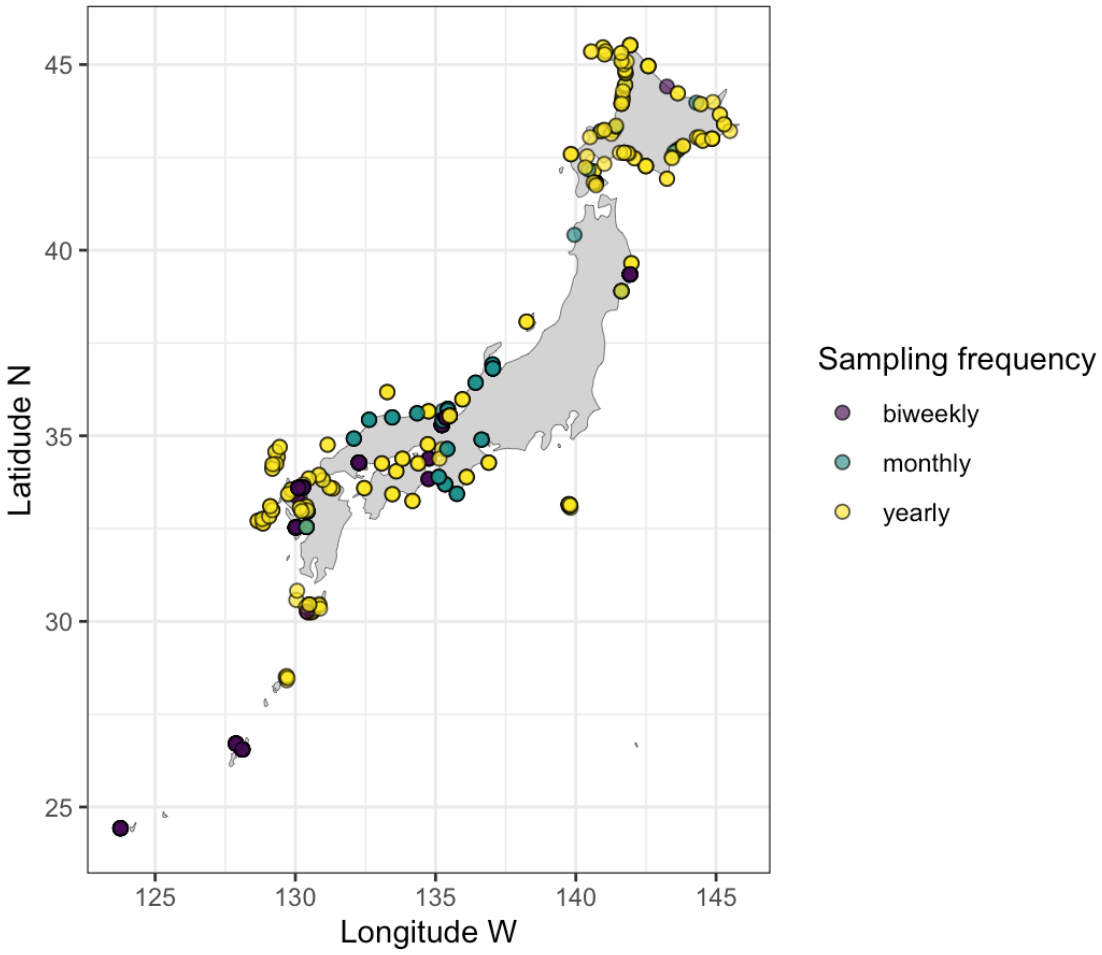

Figure 1. Map of the sampling sites across the Japanese archipelago.

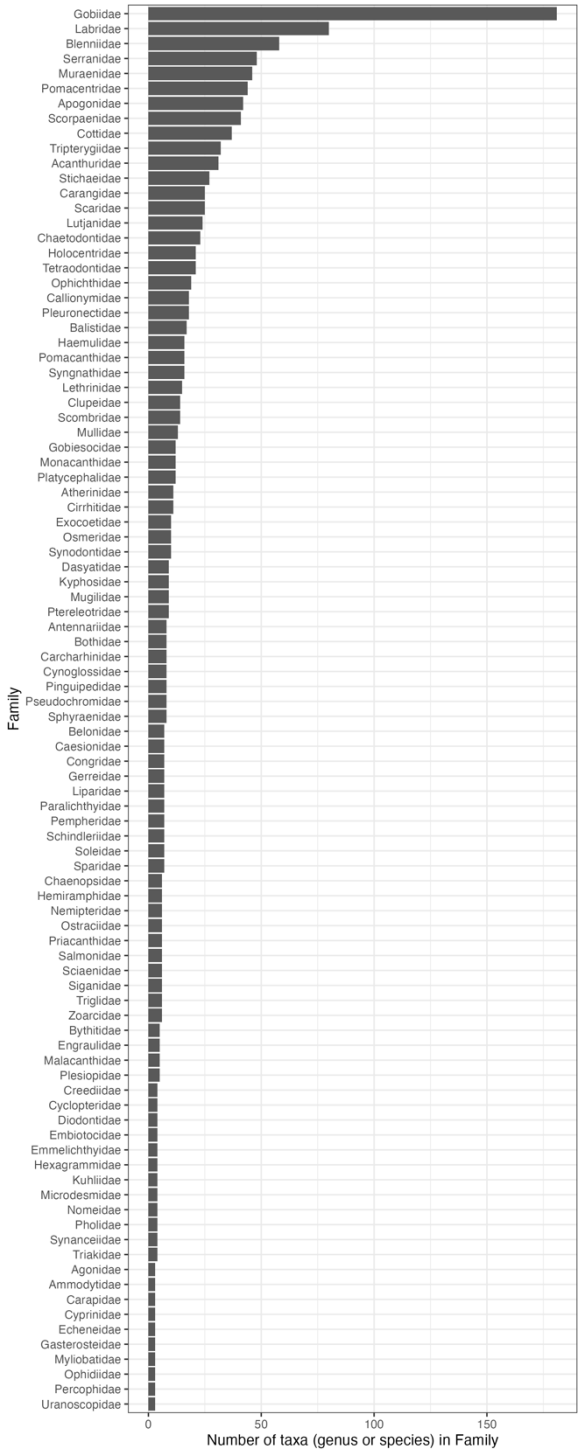

Figure 2. Species diversity (number of taxa, genus or species) across fish families in the dataset, showing the number of species per family for the 84 families with more than 3 species in 147 families.