

1

2 Supporting Information for

3 Quantitative, Multispecies Monitoring at a Continental Scale

4 Gledis Guri, Owen Liu, Ryan P. Kelly, Megan Shaffer, Kim Parsons, Krista M. Nichols, Pedro F. P. Brandão-Dias and Andrew Olaf
5 Shelton

6 Gledis Guri.

7 E-mail: gguri@uw.edu

8 This PDF file includes:

- 9 Supporting text
- 10 Figs. S1 to S7
- 11 Table S1
- 12 SI References

13 **Supporting Information Text**

14 **Extended Methods**

15 **qPCR.** Environmental samples were analyzed in triplicate using the multiplexed assay targeting Pacific hake, Eulachon and
16 Pacific lamprey as described in (1) on a QuanStudio 6 (Applied Biosystems). Only Pacific hake qPCR data was used in this
17 study due to the remaining qPCR targets having little positive amplification. The samples underwent real-time thermocycler
18 protocol including an initial denaturation step at 95°C 10 min followed by 40 cycles of 15 s at 95°C and 1 min at 58°C. All
19 thermocycler reactions were run in 20 μL volume consisting of 10 μL of TaqMan Environmental Master Mix 2.0, 1 μL of
20 each primer (forward: AAATGTTAAACTAGAGCCGAATAGC and reverse: TCGTGGAGTCAAAGTGGGGTAGA, 10 nM
21 concentration each), 0.5 μL of probe (AAATGTTAAACTAGAGCCGAATAGC; 10nM concentration), 3 μL of dH2O and 2
22 μL of DNA template. To monitor quality control, all runs included an internal positive control (IPC) to detect PCR inhibition
23 and a negative template control (NTC) containing water instead of DNA template. Any IPC delay exceeding 0.5 cycles in the
24 NTC was considered inhibition and inhibited samples were diluted and re-analyzed.

25 Alongside environmental samples standard samples were analyzed constructed from a 120 bp synthetic DNA fragment
26 (gBlock; IDT) representing the 12S region of Pacific hake (*Merluccius productus*), which encompasses the 101 bp 12S qPCR
27 target (sequence information available in Ramón-Laca et al. 2021). The synthetic DNA fragment was diluted to create a series
28 of standards with final concentrations ranging from 10⁰ to 10⁵ copies/μL.

29 **Metabarcoding.** In total, 568 samples, including 554 environmental samples, 7 PCR blanks and 7 positive controls (holding only
30 kangaroo DNA) were amplified using MiFish-U universal primers (2) with Illumina tails (forward TCGTCGGCAGCGTCA-
31 GATGTGTATAAGAGACAGGCCGTAAACTCGTGCCAGC; reverse GTCTCGTGGCTCGGAGATGTGTATAAGA-
32 GACAGCATAGTGGGGTATCTAACCCAGTTG) by using a two-step PCR protocol. The DNA was amplified in the first
33 PCR reaction (PCR1) in a 20 μL reaction consisting of: 10 μL of Phusion Master Mix (2X), 0.4 μL of the forward primer (10
34 μM); 0.4 μL of the reverse primer (10 μM), 0.6 μL of 100% DMSO, 0.5 μL of rAlbumin (20 μg/μL), 4.4 μL of nuclease-free
35 water and 2 μL of DNA template. Reactions were run with the following cycling conditions: an initial denaturation of 98°C for
36 30 sec; followed by 35 cycles of 98°C for 10 sec, 60°C for 30 sec, and of 72°C for 3 sec; with a final extension of 72°C for 10 min
37 and hold at 4°C.

38 PCR product was cleaned using Ampure Beads (1.2x) and then indexed in a second PCR reaction (PCR2), with the
39 following recipe: 12.5 μL of KAPA HiFi HotStart ReadyMix (Roche Diagnostics), 1.25 μL of an index from IDT for Illumina
40 DNA/RNA UD Indexes (Sets A-D), 5 μL of PCR1 product, and 6.25 μL of nuclease free water. Cycling conditions included
41 an initial denaturation of 95°C for 5 min; 8 cycles of: 98°C for 20 sec, 56°C for 30 sec, 72°C for 1 min; and a final extension
42 of 72°C for 5 min. Resulting products were visualized on a 2% agarose gel and quantified using Quant-iT™ dsDNA Assay
43 Kit (Thermo Fisher Scientific, USA) with Fluoroskan™ Microplate Fluorometer (Thermo Fisher Scientific, USA). Indexed
44 products were normalized by concentration, pooled into libraries for sequencing, and then size- selected to extract only the
45 target fish band using the E-Gel™ SizeSelect™ II Agarose Gels (Thermo Fisher Scientific, USA). Subsequently the libraries
46 were sequenced on Illumina MiSeq platform using the v3 600 cycle kit.

47 **Mock community.** We constructed mock communities with even and skewed species total genomic DNA using Qubit HS
48 assay quantification method. We then quantified the concentration of the 12S rRNA gene using MarVer1 primers (forward:
49 CGTGCCAGCCACCGCG; reverse: GGGTATCTAACCCYAGTTG (3)), which perfectly match the template and give
50 unbiased estimates of concentration, using ddPCR (Bio-Rad, Inc., QX200 Droplet Digital PCR system). Each tissue of each
51 species was quantified in a 22 μL reaction consisted of 2 μL of DNA template from genomic DNA , 11 μL of ddPCR EvaGreen
52 (Bio-Rad), 0.22 μL of each forward and reverse primers (10 UM), and 0.56 μL of nuclease free water. The thermocycler reactions
53 were run in C1000 Touch Thermal Cycler with 96-Deep Well Reaction Module (Bio-Rad) using the PCR program as follows: 2
54 min at 50°C for enzyme activation, 2 min at 95°C for initial denaturation, and 40 cycles of denaturation for 1 sec at 95°C and
55 primer annealing and elongation for 30 sec min at 60°C, with a ramp rate of 2°C per s and heldat 4°C until droplets were read.
56 Droplets were determined to be positive after drawing a threshold based on NTCs.

57 **Bioinformatics.** Starting with Illumina-demultiplexed sequencing files, we first removed primer and adapter sequences using
58 Cutadapt v4.9, applying the default mismatch tolerance parameter (-e 0.1) (4). Next, we performed quality filtering using the
59 filterAndTrim function from the dada2 R package, retaining its default parameters while specifying the optional truncLen
60 argument to match the minimum MiFish amplicon length. After filtering, we denoised the sequences using dada2 with default
61 settings (5), which resulted in a list of Amplicon Sequence Variants (ASVs) across all samples. Each resulting ASV was then
62 assigned a unique 28-character hash code using the sha-1 algorithm with the R function map_chr. This allowed us to create a
63 common database of ASVs by hash, enabling us to track which sequences have already been annotated for consistency and to
64 avoid redundant computational efforts.

65 For taxonomic assignment of ASVs, we used standalone BLAST (6) to compare each unique ASV against the NCBI nr
66 eukaryote database as of January 2025. BLAST arguments were as follows: "-word_size 30 -eval 1e-30" -max_target_seqs
67 50". To streamline annotation, we also used the "-negative_taxids" argument, including a list of target organisms closely
68 related to those present in the sampled region, but known to be absent. The complete list of taxids that were excluded is
69 available alongside the pipeline code. Finally, we performed a least common ancestor (LCA) analysis on the top BLAST hits
70 using TaxonKit v0.17 (7). Sequences with a 100 percent identity match to any database sequence were prioritized for LCA. In

71 the absence of perfect matches, we conducted LCA analysis on the top hits with identity scores above 99.3, , 98, or 96 percent
72 respectively. In the absence of any hits above 96 identity, taxonomic assignment was truncated. Finally, for mock communities
73 were identify of species was known, we manually curated annotations before analysis.

74 **Bayesian model. OLE**

The above statistical model provides predictions of species-specific DNA concentrations (D) at each depth sampled ($d = 0$,
50, 150, 300 or 500 m). We used the sdmTMB package (8) to generate distribution models for the concentration of eDNA , as:

$$\ln D_{xyz} = \tau + \eta(s) + \varepsilon_d(s)$$

75 where x,y,z index longitude, latitude, and depth, respectively, τ is scalar intercept, $\eta(s)$ is a spatial smooth shared among
76 depths, and $\varepsilon_d(s)$ is a depth-specific spatial smooth. Both $\eta(s)$ and $\varepsilon_d(s)$ are a zero-centered Gaussian random field estimated
77 using the stochastic partial differntial equation (SPDE) approximation (see (8)). Each species was fitted individually while we
78 created and selected the best mesh for approximating spatial Gaussian random fields as part of the model.

79 **Model convergance. OLE + GLED**

80 **References**

- 81 1. A Ramón-Laca, A Wells, L Park, A workflow for the relative quantification of multiple fish species from oceanic water
samples using environmental DNA (eDNA) to support large-scale fishery surveys. *PLOS ONE* **16**, e0257773 (2021).
- 82 2. M Miya, et al., MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of
more than 230 subtropical marine species. *Royal Soc. Open Sci.* (2015).
- 83 3. E Valsecchi, et al., Novel universal primers for metabarcoding environmental DNA surveys of marine mammals and other
marine vertebrates. *Environ. DNA* **2**, 460–476 (2020).
- 84 4. M Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
- 85 5. BJ Callahan, et al., DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583
(2016).
- 86 6. SF Altschul, W Gish, W Miller, EW Myers, DJ Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410
(1990).
- 87 7. W Shen, H Ren, TaxonKit: A practical and efficient NCBI taxonomy toolkit. *J. Genet. Genomics* **48**, 844–850 (2021).
- 88 8. SC Anderson, EJ Ward, PA English, LAK Barnett, JT Thorson, sdmTMB: An R Package for Fast, Flexible, and
89 User-Friendly Generalized Linear Mixed Effects Models with Spatial and Spatiotemporal Random Fields (2022).

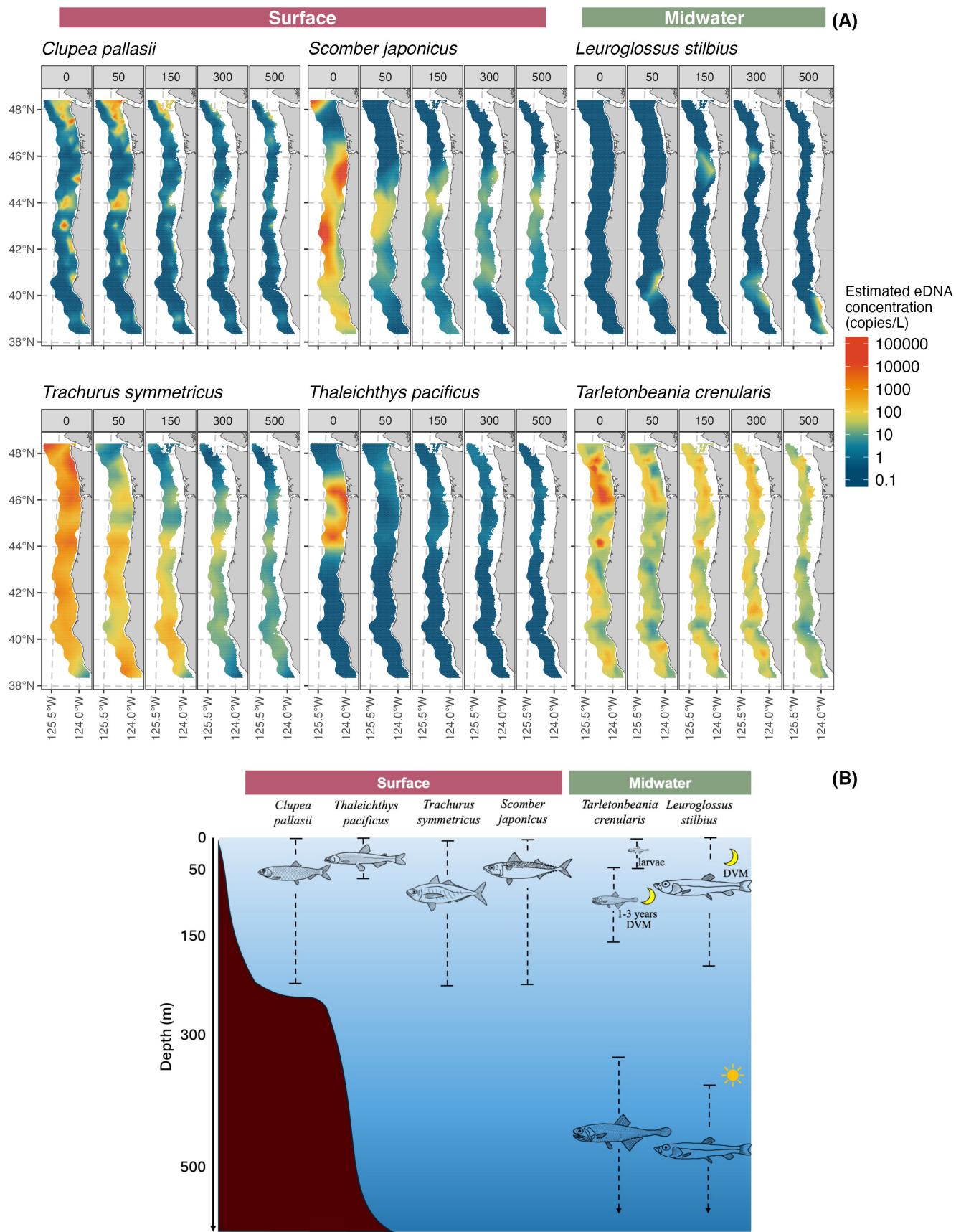


Fig. S1. First figure

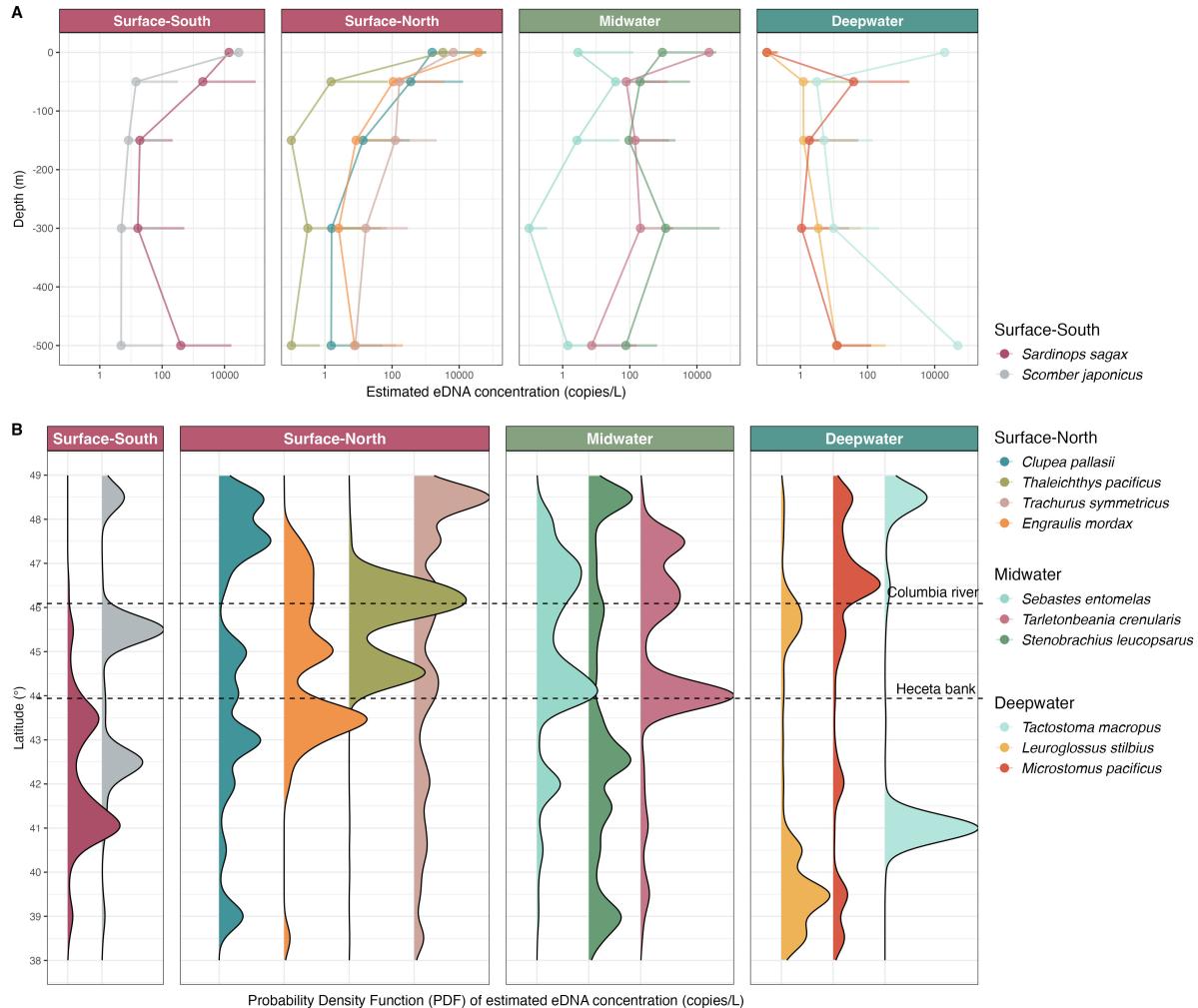


Fig. S2. Vertical (A) and latitudinal (B) distribution of eDNA concentrations grouped in ecological groups (surface south, surface north, midwater, and deepwater). Mean eDNA concentration of vertical distribution is shown with dots and 99% upper quantile with ticks (A). Probability density functions (PDF) of estimated eDNA concentrations across latitudinal gradients (B; square root transformed of x-axis to enhance low concentration peaks). Key geographic features, such as the Columbia River and Heceta Bank, are marked to indicate alignment of species distributions.

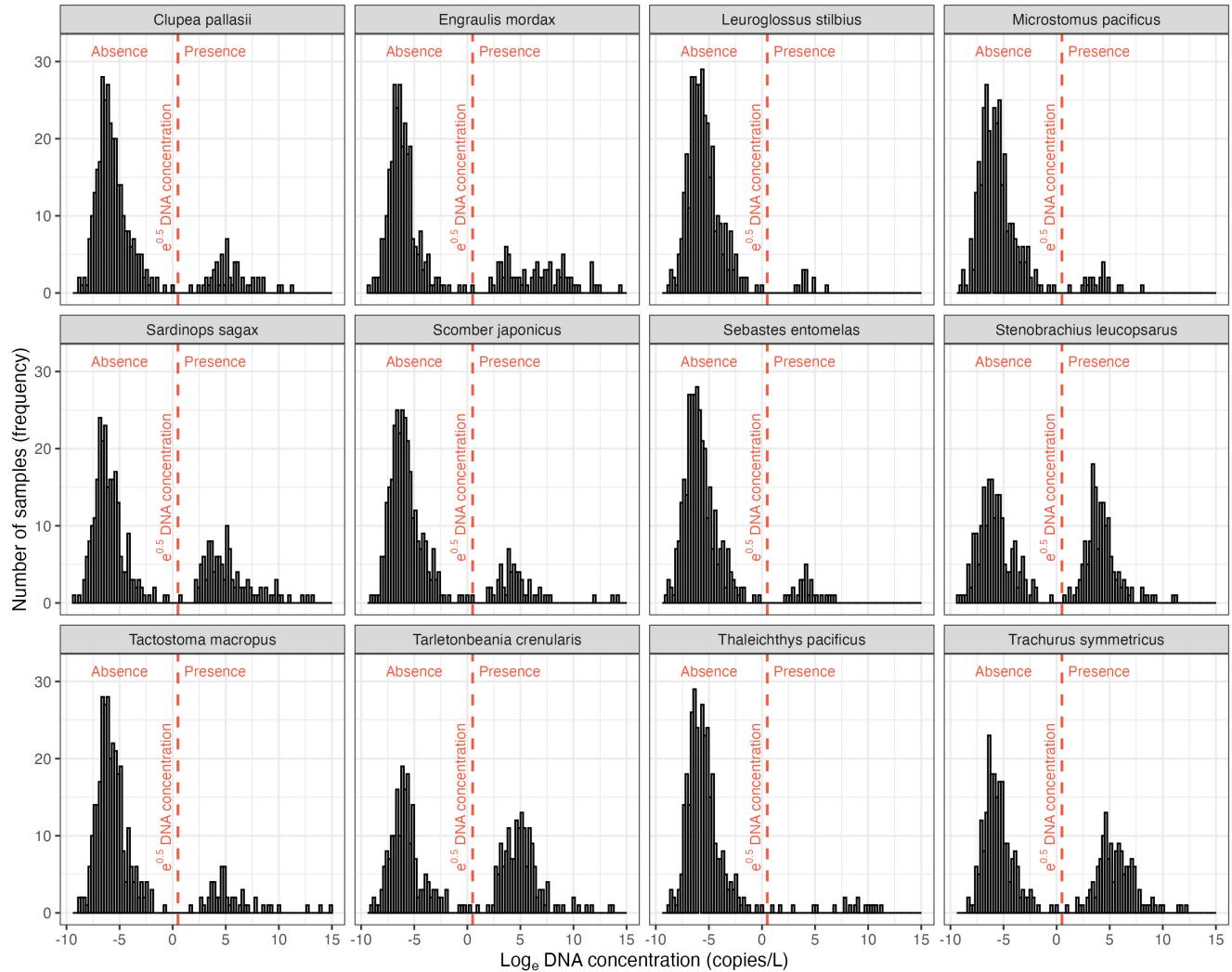


Fig. S3. First figure

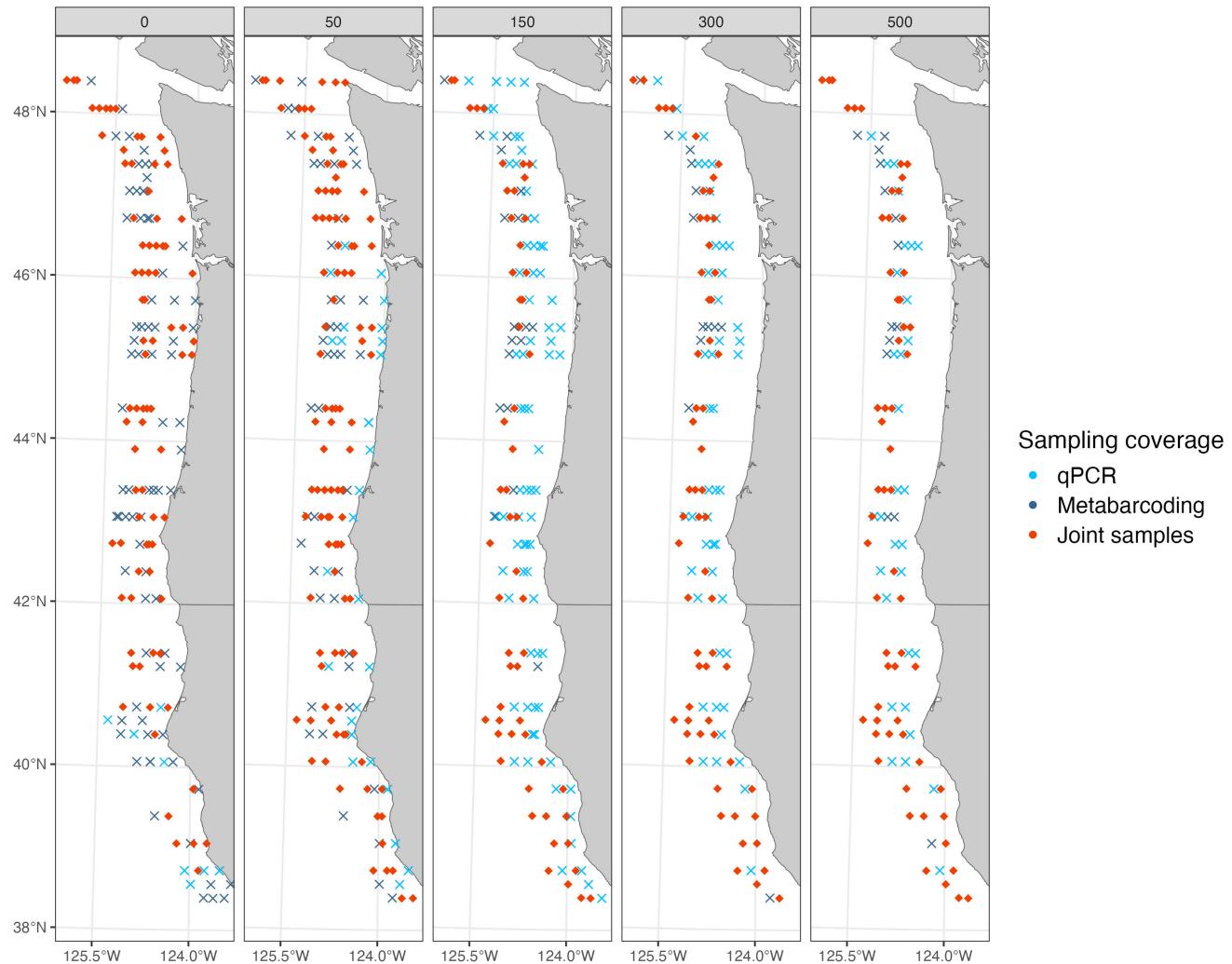


Fig. S4. First figure

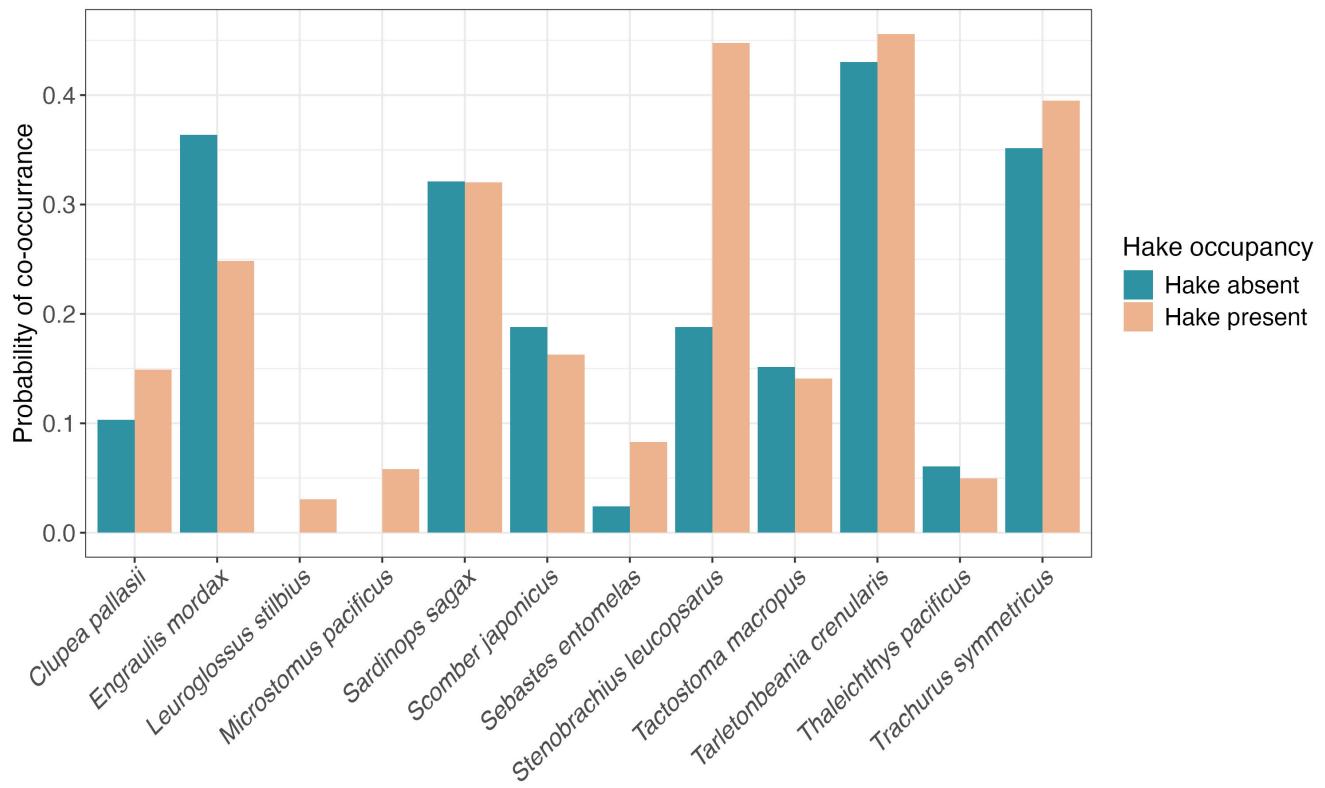


Fig. S5. First figure

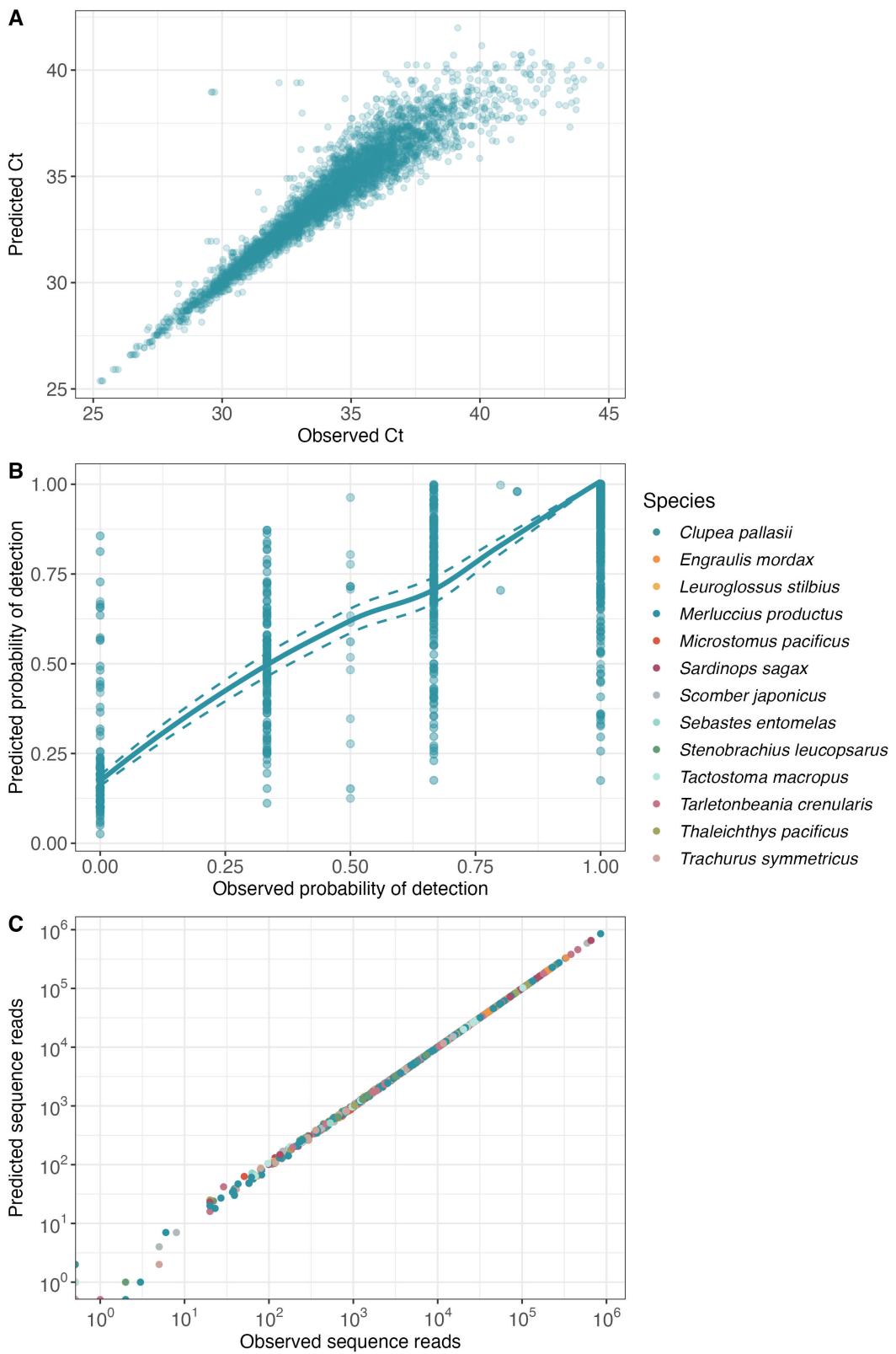


Fig. S6. First figure

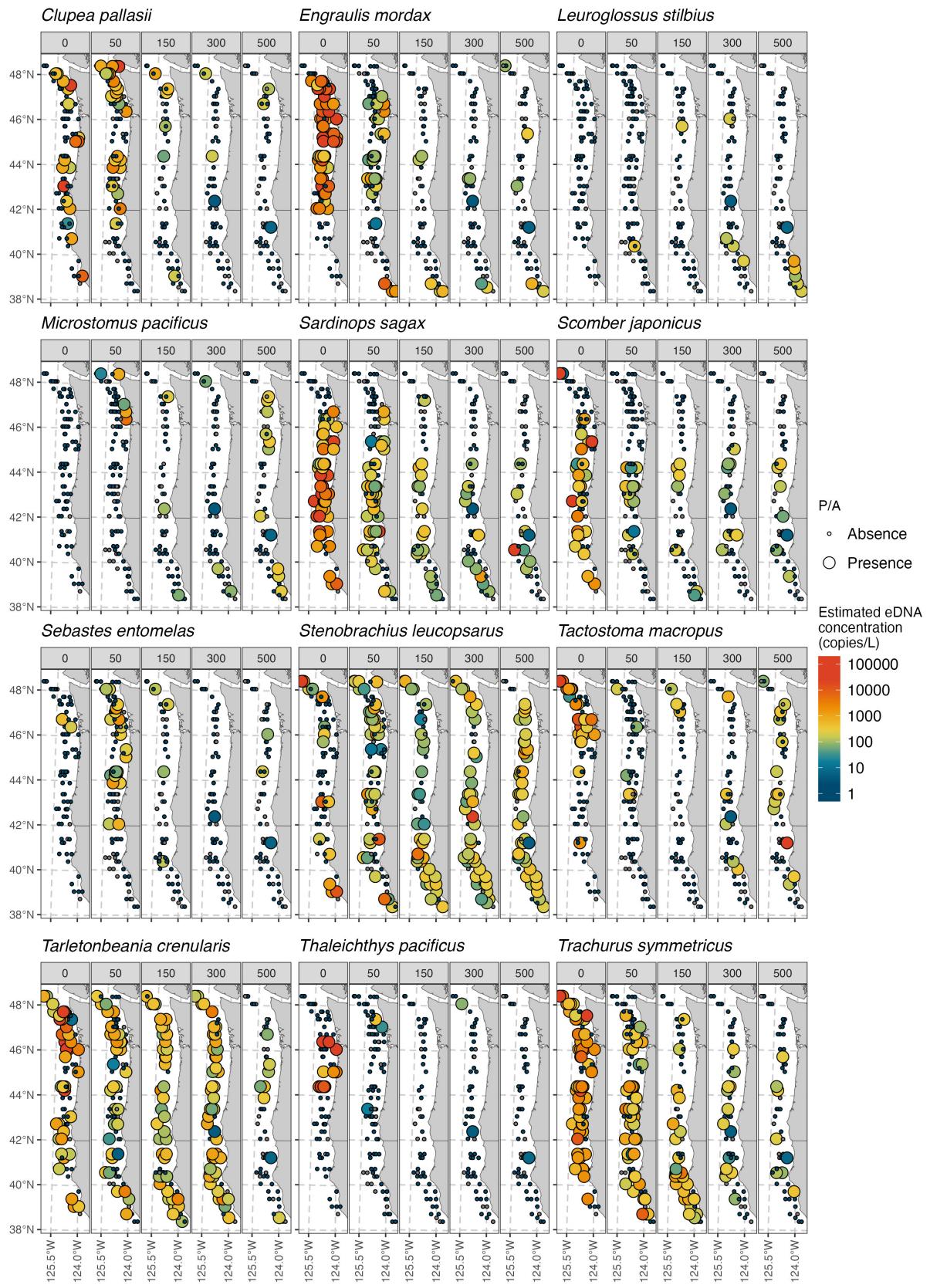


Fig. S7. First figure

Table S1. Initial relative abundances of the selected species across different mock communities.

Species	Mock1		Mock2		Mock3		Mock4	
	Even	Skew	Even	Skew	Even	Skew	Even	Skew
<i>Clupea pallasii</i>	0.437	0.325	0.042	0.009	NA	NA	0.025	0.038
<i>Engraulis mordax</i>	0.122	0.034	0.15	0.15	0.146	0.16	0.308	0.478
<i>Leuroglossus stilbius</i>	0.108	0.29	NA	NA	NA	NA	NA	NA
<i>Merluccius productus</i>	0.079	0.144	0.12	0.06	0.117	0.075	0.247	0.287
<i>Microstomus pacificus</i>	NA	NA	0.057	0.215	NA	NA	NA	NA
<i>Sardinops sagax</i>	0.028	0.033	0.14	0.035	0.137	0.025	0.289	0.131
<i>Scomber japonicus</i>	NA	NA	0.122	0.053	0.119	0.027	NA	NA
<i>Sebastes entomelas</i>	NA	NA	NA	NA	0.068	0.218	NA	NA
<i>Stenobrachius leucopsarus</i>	NA	NA	NA	NA	0.356	0.487	NA	NA
<i>Tactostoma macropus</i>	NA	NA	0.13	0.324	NA	NA	NA	NA
<i>Tarletonbeania crenularis</i>	NA	NA	0.083	0.124	NA	NA	NA	NA
<i>Thaleichthys pacificus</i>	0.123	0.113	0.099	0.012	NA	NA	0.013	0.021
<i>Trachurus symmetricus</i>	0.103	0.06	0.057	0.018	0.056	0.008	0.118	0.04