**SI Appendix for**

**The global distribution of environmental DNA sequences from coastal fishes in the Anthropocene**

**Supporting Methods**

**Method S1. Environmental DNA collection**

Between 1 and 4 replicates of seawater samples were sampled at each station, so the total number of samples is 584. Only samples filtered between 0 and 40m deep were considered in this study. For each sample collected with the two first sampling protocols (plastic bags and bottles), the seawater was filtered with sterile Sterivex filter capsules (Merck© Millipore; pore size 0.22µm) using disposable sterile syringes. The eDNA filtration device for the other two sampling protocols (transects) was composed of an Athena® peristaltic pump (Proactive Environmental Products LLC, Bradenton, Florida, USA; nominal flow of 1.0 L.min± 15%), a VigiDNA® 0.2µM cross flow filtration capsule (SPYGEN, le Bourget du Lac, France) and disposable sterile tubing for each filtration capsule. At the end of each filtration, the filter units were filled with CL1 Conservation buffer (SPYGEN, le Bourget du Lac, France) and stored in 50 mL screw-cap tubes at room temperature. The water inside the capsules were emptied, and the capsules were filled with 80 mL of CL1 Conservation buffer (SPYGEN, le Bourget du Lac, France) and stored at room temperature. The list of permits delivered for eDNA sampling can be found in Supporting Information Table S2.

**Method S2. eDNA extraction, amplification and sequencing**

eDNA extraction was performed in dedicated DNA laboratories. Decontamination procedures were conducted before and after all manipulations. Environmental DNA extractions were performed following the protocols in Pont et al. (2018) for SPYGEN capsules, and in Juhel et al. (2020) for the sterivex filters. A teleost-specific 12S mitochondrial rRNA primer pair (teleo, forward primer – ACACCGCCCGTCACTCT, reverse primer – CTTCCGGTACACTTACCATG (Valentini et al., 2016)) was used for the amplification of metabarcode sequences. The teleo barcode has been shown to be one of the most appropriate for fishes, even though it presents some amplification bias (Bylemans, Gleeson, Hardy, & Furlan, 2018). The purified PCR products were pooled in equal volumes, to achieve a theoretical sequencing depth of 1,000,000 reads per sample. Library preparation and sequencing were performed at Fasteris (Geneva, Switzerland). A total of 45 libraries were prepared using MetaFast protocol for Illumina sequencing platforms. A paired-end sequencing (2x125 bp) was carried out using an Illumina HiSeq 2500 sequencer with the HiSeq Rapid Flow Cell v2 using the HiSeq Rapid SBS Kit v2 (Illumina, San Diego, CA, USA) or a MiSeq (2x125 bp, Illumina, San Diego, CA, USA) using the MiSeq Flow Cell Kit v3 (Illumina, San Diego, CA, USA) or a NextSeq sequencer (2x125 bp, Illumina, San Diego, CA, USA) with the NextSeq Mid kit following the manufacturer’s instructions. This generated an average of 624,468 sequence reads (paired-end Illumina or Ion Torrent) per sample. Samples from the Cold\_Temperate\_Northwest\_Pacific region were filtered with 0.45µm pore size and then treated at the State Key Laboratory of Pollution Control & Resource Reuse in Nanjing University, China. DNA was extracted with the DNeasy Blood & Tissue Kit (Qiagen, Germany), with 3 negative controls, and the teleo fragment was amplified with one PCR replicate per sample. The library was prepared with a VAHTS® Universal DNA Library Prep Kit for Ion Torrent (Vazyme, China) and sequenced on Ion Torrent S5 sequencer (Life Technologies, USA).

**Method S3.** **Bioinformatic analysis**

The iterative process of SWARM leads to clusters composed of many sequences with more than d mismatches. Further, we used the -f (fastidious) option, which creates virtual sequences within clusters to link more dissimilar sequences together, hence limiting alpha-diversity inflation by joining low abundant MOTUs within larger ones. The minimum distance between clusters is 2 mismatches (d+1). Taxonomic assignment of MOTUs was carried out using the Lower Common Ancestor (LCA) algorithm *ecotag* implemented in the Obitools toolkit which assigns a taxonomy to sequences even when the sequence match is not perfect. The assignment was based on NCBI taxonomic tree of species to consider the current knowledge on molecular diversity per branch and assign a taxonomy at the lowest possible rank. If the sequence matches several identifications with equal percentages of similarity, *ecotag* assigns to the upper taxonomic level common between all possible matches. We then applied quality filters to be conservative in our estimates. The goal of LULU algorithm is to reduce the bias induced by intra-specific variability and the potential over-estimation of alpha-diversity due to d=1 in SWARM. Taxonomic assignments obtained from the LCA algorithm were further selected to ensure more conservative assignments, following results from previous studies (Juhel et al., 2020; Polanco Fernández et al., 2021). Assignments were accepted at the species level, as putative species, if the percentage of similarity with the reference sequence was 100%, at the genus level if the similarity was between 90 and 99%, and at the family level if the similarity was > 85%. If these criteria were not met, the MOTU was left unassigned. The post-LCA algorithm correction thresholds of 85% similarity for family and 90% for genus assignments were chosen to include a maximum of correct family and genus assignments while minimizing the risk of adding wrong assignments in the detections.

**Method S4.** **Explanatory factors**

We extracted environmental factors at their native spatial resolutions, over one year prior to the date of sampling and calculated the mean. In this way, we ensured that the highest spatial resolution of data was used in combination with temporally relevant metrics that describe the recent environmental conditions prior to sampling.

We used HDI values for 2019 from the Human Development Indicators and Indices (http://hdr.undp.org). We extracted the population size from the UN WPP-adjusted population count v4.11 for year 2020 at 30 arc-second resolution (<https://sedac.ciesin.columbia.edu>).

**Method S5. Statistical analyses**

**Sequence α-diversity**

Hill numbers provide a set of diversity indices, differing by a parameter “q”, which determines their sensitivity to relative abundances, and τ, their sensitivity to distance between sequences.

**Modeling MOTU and sequence α-diversity**

The model with Gaussian spatial correlation (function *corGaus*) had the lowest Akaike Information Criterion (AIC) compared to other correlation functions. The adjusted R² of the model was computed with the function *r2* from the package “performance”. Partial R² for each group of factors were obtained with the function *calc.relimp* from the package “relaimpo”.

**Modeling β-diversity**

To select dbMEMs to include in our dbRDA, we ran a dbRDA with the response variable and all dbMEMs as explanatory factors and selected the dbMEMs explaining most of the model’s variance (MEM1 to MEM5).

**Supplementary figures and tables**

Table S1. Information on sampling in each region

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Region** | **Date of sampling** | **Sampling method** | **number of stations** | **mean volume filtered per station (L)** | **Total volume filtered (L)** | **Sequencer** |
| Arctic | July-August 2020 | Transect | 19 | 30 | 570 | Miseq |
| Cold\_Temperate\_  Northwest\_Pacific | September 2020 | Bottle | 24 | 2.8 ±0.3 | 69 | IonTorrent |
| Lusitanian | October-November 2019  September 2020 | Transect | 13 | 60 | 780 | Miseq |
| Mediterranean\_Sea | March-June 2018  March & July 2019  July 2020 | Transect | 35 | 70.3 ±23 | 2460 | Miseq - Hiseq |
| Scotia\_Sea | January-February 2020 | Transect | 41 | 32.7 ±17.9 | 1344 | NextSeq - Miseq |
| Southeast\_Polynesia | June 2018 | Transect & Bag | 3 | 128 ±76 | 384 | Hiseq |
| Tropical\_East\_Pacific | March 2018 | Transect | 13 | 60 | 780 | Hiseq |
| Tropical Northwestern Atlantic | February, March & July 2018  January 2020 | Transect | 41 | 45 ±12.7 | 1844 | Hiseq - Miseq |
| Tropical Southwestern Pacific | October-December 2019  August 2020 | Benthic transect | 26 | 92.3 ±10.4 | 2400 | NextSeq - Miseq |
| Western\_Coral\_Triangle | October-November 2017 | Bag | 32 | 3.95±0.1 | 126.5 | Hiseq |
| Western\_Indian\_Ocean | April 2019 | Transect | 16 | 60 | 960 | Miseq |

Table S2. Permits issued by the relevant authorities in each sampling location requiring governmental authorization.

|  |  |
| --- | --- |
| Region | Permit detail |
| Scattered Islands | Data collected by a team of students and researchers from French institutions aboard a French ship belonging to the TAAF fleet, under permit 2019-45 from April 1st 2019, delivered by the Administration of the French Southern and Antarctic Lands. |
| New-Caledonia | Data collected by students and researchers from French institutions working in New-Caledonia, under permit N° 3066-2019/ARR/DENV delivered by the Southern Province of New-Caledonia, and permit N° 609011/2019/DEPART/JJC delivered by the Northern Province of New-Caledonia. |
| Indonesia | Fieldwork conducted by students and researchers from a French institution (IRD) and an Indonesian institution (BRIN) within the Lengguru project according to relevant guidelines by the government of the Republic of Indonesia and under research permit issued by RISTEK (Indonesia) (3179/FRP/E5/Dit.KI/IX/2017) and relevant Indonesian government collecting permit. |
| Arctic | Data collected by a team of student and researchers from French and Swiss institutions, under the research permit n° 20/01465-7 for research application RiS-ID 11544, delivered by the Governor of Svalbard. |
| Antarctic | Samples collected under a permit (RWS-2019/40813) provided by Rijkswaterstaat Zee en Delta, Government of the Netherlands, as per the Protection of Antarctica Act, to Stichting Greenpeace Council to carry out such research in the Antarctic area. |
| Colombia (Malpelo, Santa Marta, Providencia) | Samples collected by the INVEMAR, an entity attached to the Ministry of Environment and Sustainable Development. The INVEMAR does not require permit to sample as it is clarified in Paragraph 1, Article 2.2.2.2.8.1.2, Section 1 (Permits), Chapter 8 (Scientific Research) of Decree 1076 of 2015, Sole Regulatory Decree of the Environment and Sustainable Development Sector. "The Ministry of Environment and Sustainable Development, its affiliated entities, the National Natural Parks of Colombia, Regional Autonomous Corporations and/or Sustainable Development Corporations and the Large Urban Centers will not require the Specimen Collection Permit referred to in this decree (...)." |

Table S3. Read, MOTU and species counts after each cleaning and treatment step in the whole dataset. Numbers in parenthesis for LULU steps correspond to reads, MOTUs and species added with the China dataset.

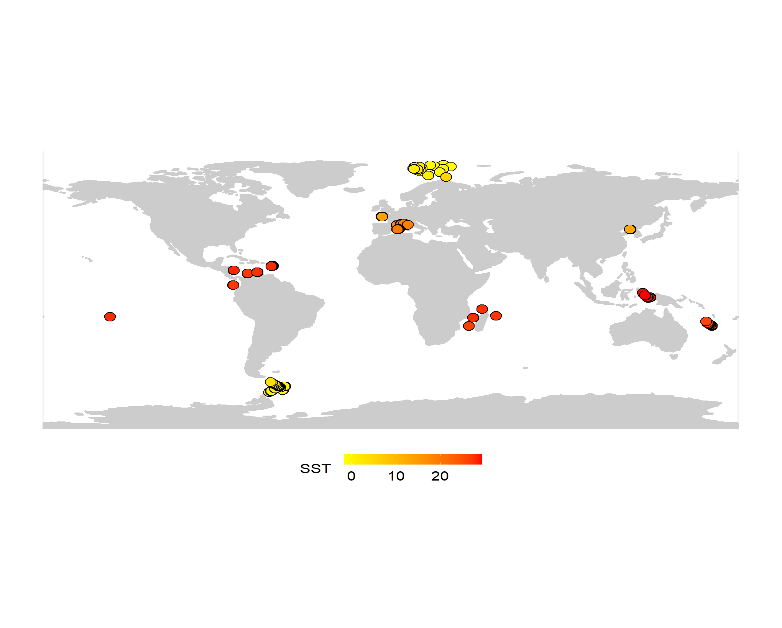
|  |  |  |  |
| --- | --- | --- | --- |
| **step** | **Reads** | **MOTUs** | **Species** |
| before | 368436373 | 124922 | 945 |
| tenreads | 368099042 | 9297 | 902 |
| blanks | 368073065 | 8792 | 902 |
| fishonly | 317176168 | 4844 | 874 |
| readlength | 317162303 | 4817 | 874 |
| PCR\_all | 308854365 | 2983 | 786 |
| LULU | 271694868 (+12014) | 2862 (+40) | 778 (+15) |
| LULU\_family | 245262610 (+11520) | 2239 (+37) | 778 (+15) |

Table S4. Read, MOTU and species counts after each cleaning and treatment step for each region.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **region** | **step** | **Reads** | **MOTUs** | **Species** |
| Arctic | before | 6346919 | 2557 | 21 |
| tenreads | 6340103 | 327 | 10 |
| PCR\_blanks\_chimeras | 6339817 | 314 | 10 |
| fishonly | 5059617 | 57 | 3 |
| readlength | 5059591 | 56 | 3 |
| PCR\_all | 5037842 | 33 | 2 |
| LULU | 5037842 | 33 | 2 |
| LULU\_family | 4945738 | 28 | 2 |
| Cold Temperate Northwest Pacific | LULU | 12014 | 42 | 18 |
| LULU\_family | 11520 | 39 | 18 |
| Lusitanian | before | 6320079 | 2079 | 60 |
| tenreads | 6314698 | 206 | 58 |
| PCR\_blanks\_chimeras | 6314653 | 204 | 58 |
| fishonly | 6160427 | 110 | 55 |
| readlength | 6160427 | 110 | 55 |
| PCR\_all | 6153448 | 96 | 52 |
| LULU | 6153448 | 96 | 52 |
| LULU\_family | 4362220 | 92 | 52 |
| Mediterranean Sea | before | 69838075 | 15412 | 134 |
| tenreads | 69778918 | 1079 | 129 |
| PCR\_blanks\_chimeras | 69776913 | 1032 | 129 |
| fishonly | 64681216 | 394 | 120 |
| readlength | 64671959 | 384 | 120 |
| PCR\_all | 64462922 | 249 | 100 |
| LULU | 64462922 | 249 | 100 |
| LULU\_family | 63876182 | 227 | 100 |
| Scotia Sea | before | 20412024 | 5326 | 19 |
| tenreads | 20396150 | 484 | 15 |
| PCR\_blanks\_chimeras | 20396065 | 480 | 15 |
| fishonly | 16405013 | 62 | 9 |
| readlength | 16404333 | 61 | 9 |
| PCR\_all | 16041636 | 40 | 8 |
| LULU | 16041636 | 40 | 8 |
| LULU\_family | 11005415 | 24 | 8 |
| Southeast Polynesia | before | 7442809 | 2408 | 118 |
| tenreads | 7438309 | 373 | 112 |
| PCR\_blanks\_chimeras | 7438275 | 370 | 112 |
| fishonly | 7179074 | 307 | 111 |
| readlength | 7179074 | 307 | 111 |
| PCR\_all | 6859662 | 197 | 83 |
| LULU | 6859662 | 197 | 83 |
| LULU\_family | 6293496 | 169 | 83 |
| Tropical East Pacific | before | 4957433 | 2055 | 66 |
| tenreads | 4950268 | 380 | 65 |
| PCR\_blanks\_chimeras | 4950242 | 379 | 65 |
| fishonly | 4739633 | 221 | 62 |
| readlength | 4739588 | 220 | 62 |
| PCR\_all | 4676658 | 153 | 48 |
| LULU | 4676658 | 153 | 48 |
| LULU\_family | 4401537 | 126 | 48 |
| Tropical Northwestern Atlantic | before | 32286077 | 12387 | 164 |
| tenreads | 32243468 | 975 | 135 |
| PCR\_blanks\_chimeras | 32237820 | 927 | 135 |
| fishonly | 29328813 | 734 | 132 |
| readlength | 29328137 | 731 | 132 |
| PCR\_all | 28831276 | 466 | 116 |
| LULU | 28825577 | 452 | 116 |
| LULU\_family | 23452276 | 314 | 116 |
| Tropical Southwestern Pacific | before | 52564185 | 20134 | 352 |
| tenreads | 52523440 | 2392 | 335 |
| PCR\_blanks\_chimeras | 52519753 | 2341 | 335 |
| fishonly | 47897730 | 1328 | 331 |
| readlength | 47896969 | 1327 | 331 |
| PCR\_all | 45546471 | 954 | 295 |
| LULU | 34292060 | 917 | 288 |
| LULU\_family | 30246965 | 742 | 288 |
| Western Coral Triangle | before | 149939309 | 57015 | 472 |
| tenreads | 149800467 | 3693 | 447 |
| PCR\_blanks\_chimeras | 149787194 | 3385 | 447 |
| fishonly | 119647080 | 2120 | 436 |
| readlength | 119644760 | 2112 | 436 |
| PCR\_all | 115574720 | 1228 | 397 |
| LULU | 89675333 | 1145 | 390 |
| LULU\_family | 82359410 | 941 | 390 |
| Western Indian Ocean | before | 18329463 | 8294 | 162 |
| tenreads | 18313221 | 713 | 159 |
| PCR\_blanks\_chimeras | 18312333 | 681 | 159 |
| fishonly | 16077565 | 550 | 156 |
| readlength | 16077465 | 548 | 156 |
| PCR\_all | 15669730 | 357 | 133 |
| LULU | 15669730 | 357 | 133 |
| LULU\_family | 14319371 | 301 | 133 |

Table S5. Information and sources of environmental variables

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Source** | **Unit** | **Original spatial resolution** | **Temporal resolution** | **Temporal duration** | **Web link** | **Reference** |
| Temperature | CoralReefWatch v3.1 | Celcius | 0.05° | Daily | 1985-2021 | <https://coralreefwatch.noaa.gov/product/5km/index.php> | Liu, G., et al. (2014) |
| Degree heating weeks | CoralReefWatch v3.1 | Daily temperature 1°C above the maximum monthly mean SST from 1985-1993, over a 12 week period | 0.05° | Daily | 1985-2021 | <https://coralreefwatch.noaa.gov/product/5km/index.php> | Liu, G.,et al. (2014) |
| pH | Norwegian Earth System Model forced ocean simulation (NorESM2) | pH | 1° | Monthly | 1980-2018 | https://www.noresm.org/resources/ | Norwegian Earth System Model forced ocean simulation (NorESM2) |
| Net primary productivity | Standard Vertically Generalized Production Model | mgC m-2 day-1 | 0.083° | Monthly | 2003-2021 | <http://orca.science.oregonstate.edu/2160.by.4320.monthly.hdf.vgpm.m.chl.m.sst.php> | Behrenfeld, M. J., & Falkowski, P. G. (1997). |
| Salinity | Global SSS/SSD L4 Reprocessed dataset | PSU | 0.25° | Monthly | 1994-2021 | [https://resources.marine.copernicus.eu/product-detail/MULTIOBS\_GLO\_PHY\_S\_SURFACE\_MYNRT\_015\_013](https://resources.marine.copernicus.eu/product-detail/MULTIOBS_GLO_PHY_S_SURFACE_MYNRT_015_013/INFORMATION) | Buongiorno Nardelli, B., 2012 |

Figure S1. Map of sea surface temperature (SST)

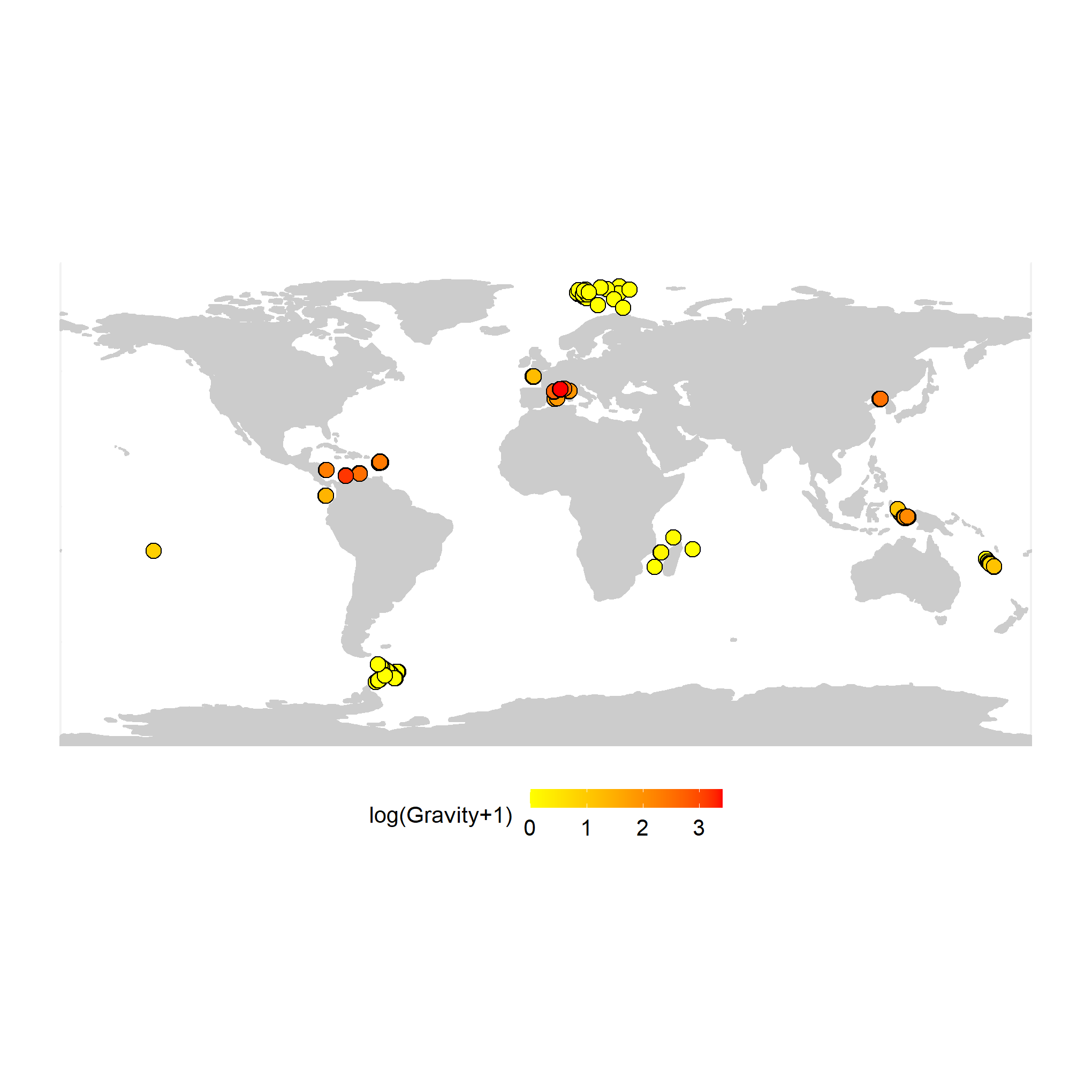


Figure S2. Map of gravity per station (log transformed)

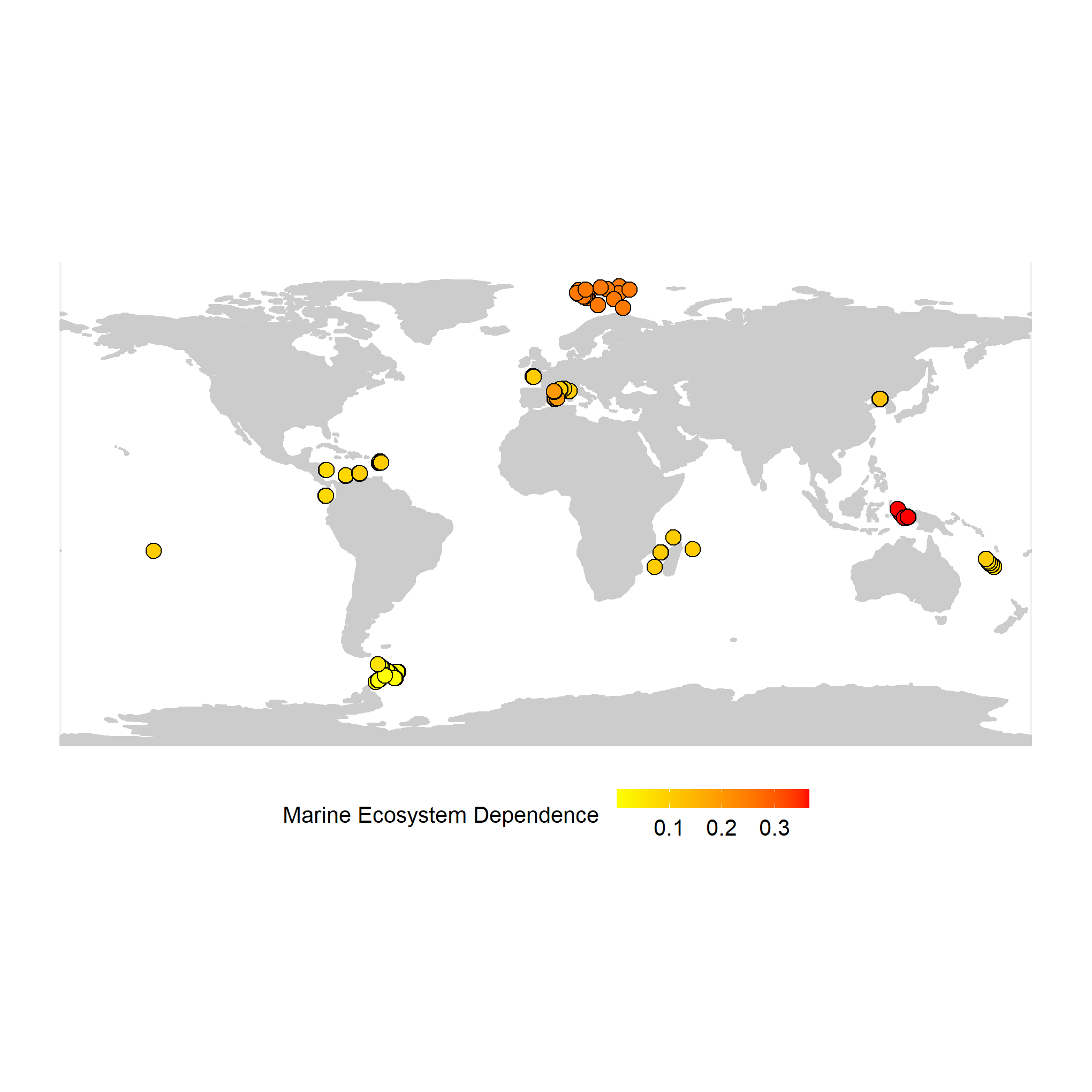
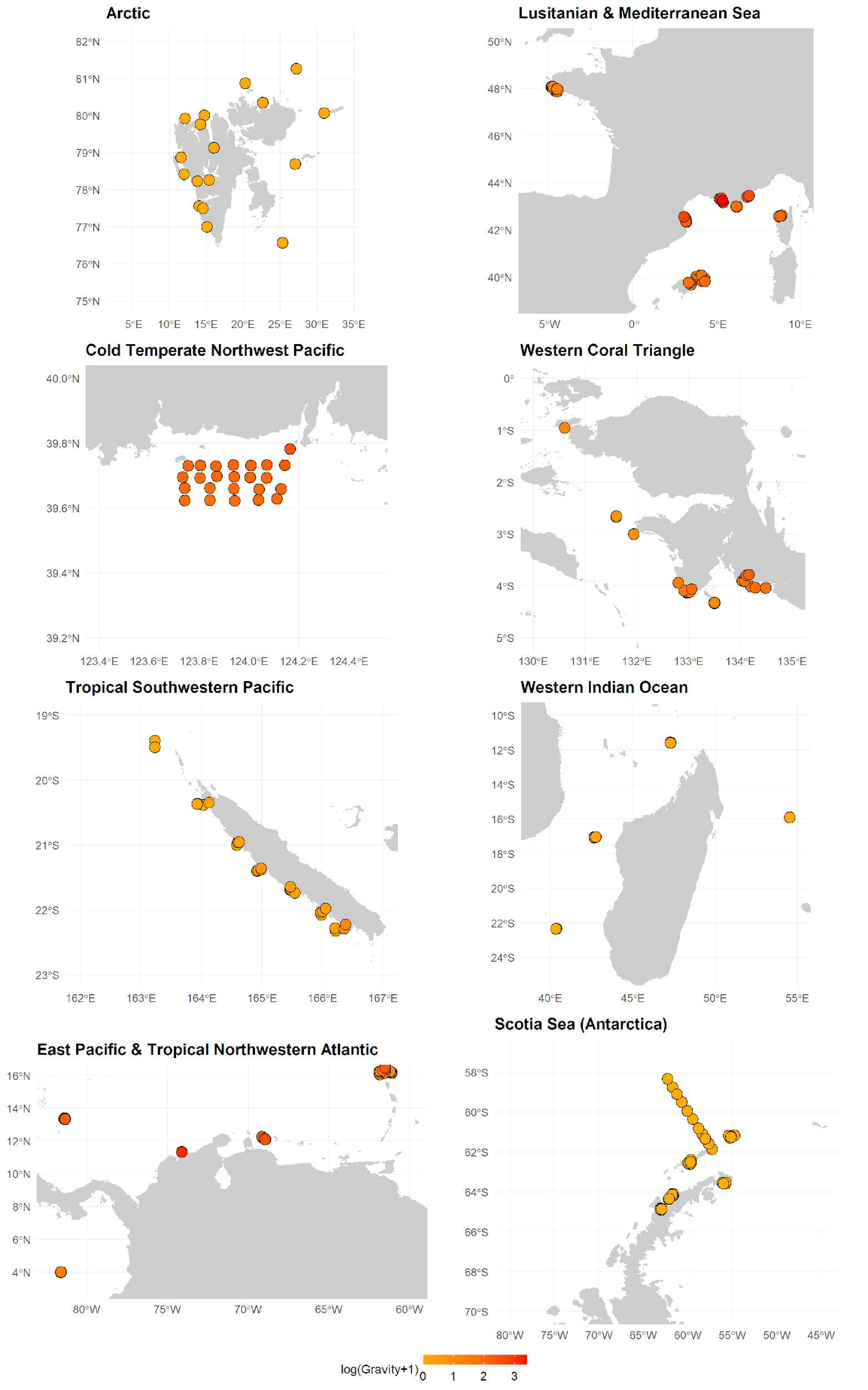


Figure S3. Map of marine ecosystem dependence per station

Figure S4. Maps of gravity (log transformed) by region

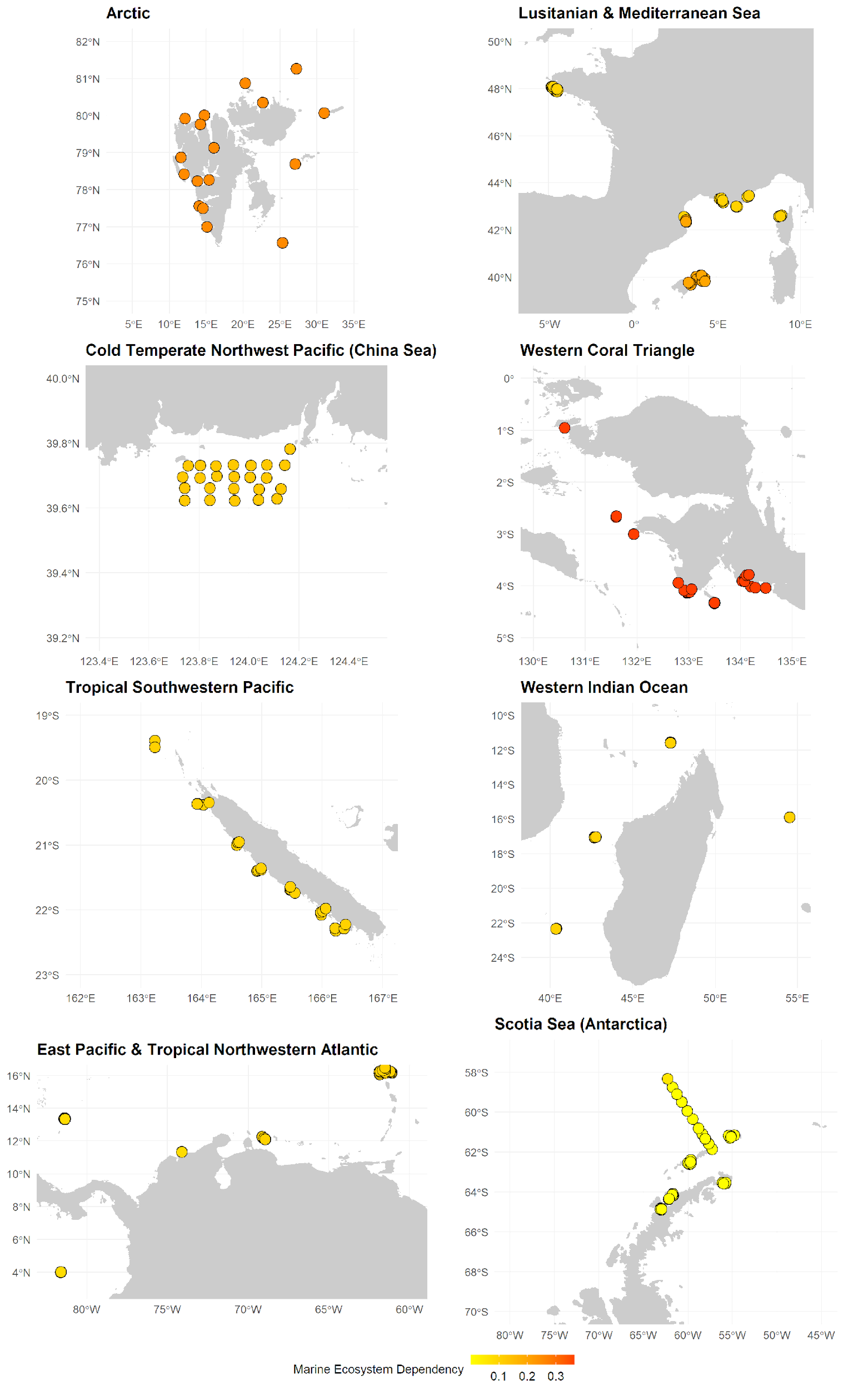


Figure S5. Maps of marine ecosystem dependence by region

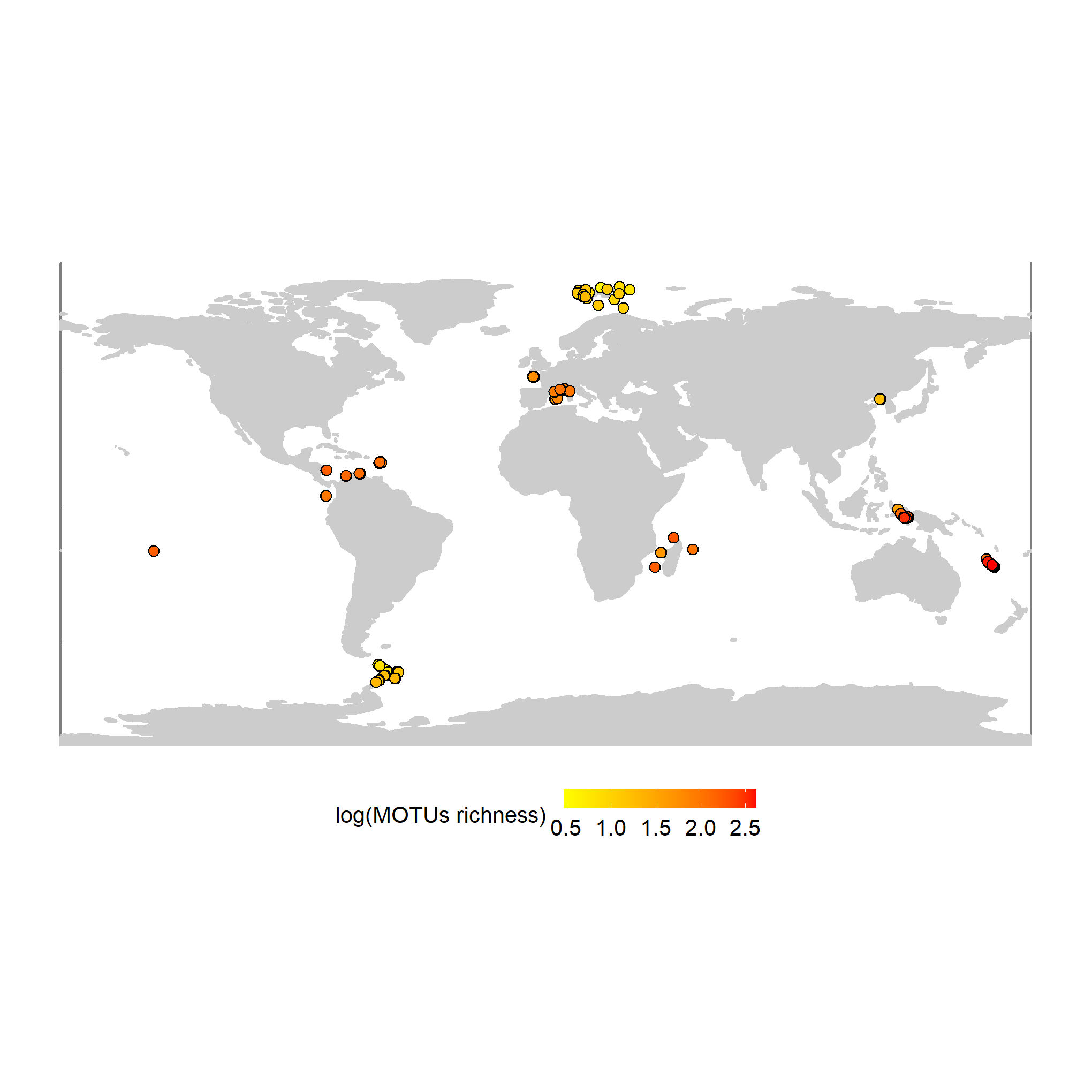


Figure S6. Map of all fish taxonomic α-diversity per station (log10 transformed). Raw values range from 2 to 414 MOTUs per station.

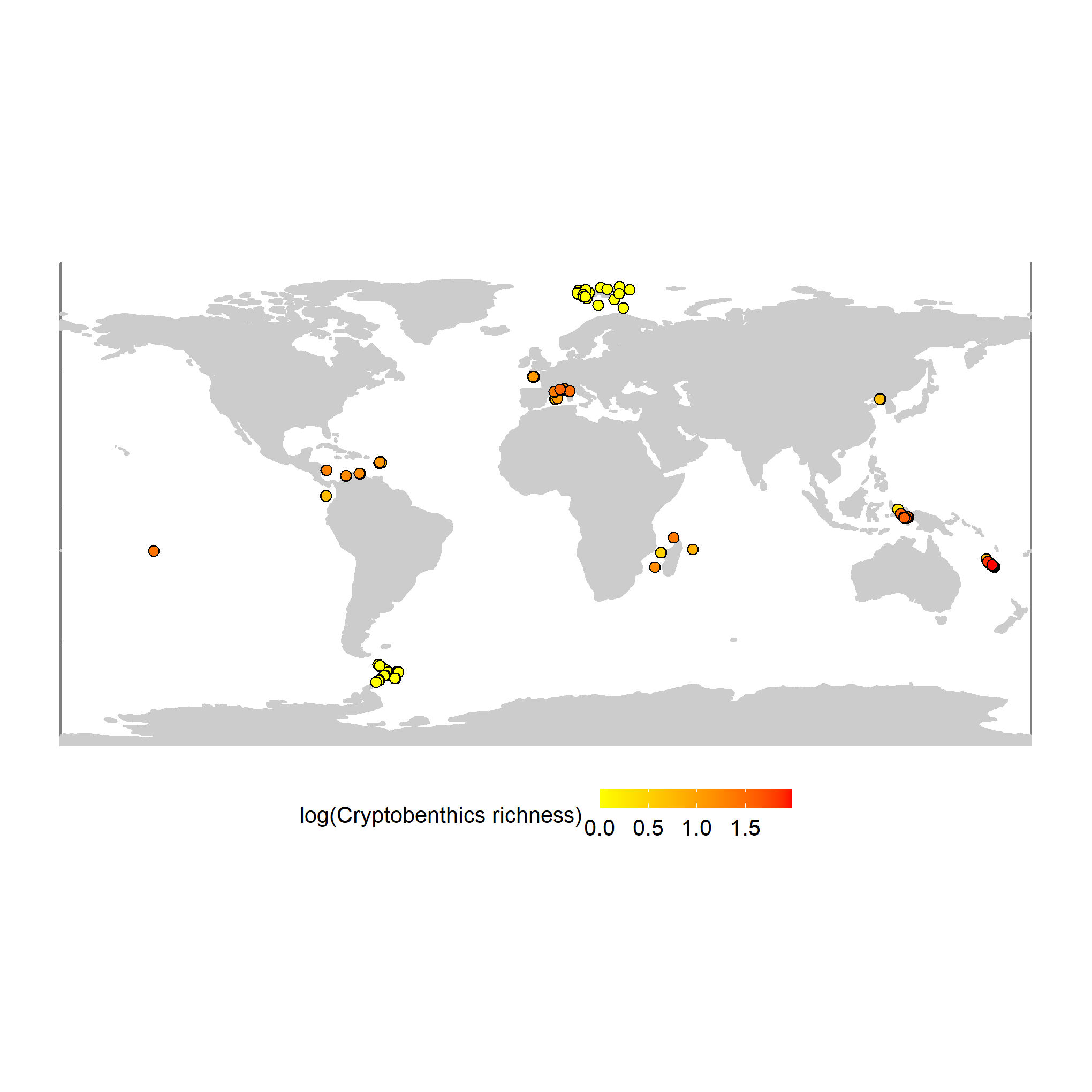
****

Figure S7. Map of Cryptobenthic fish taxonomic α-diversity per station (log10 transformed). Raw values range from 0 to 95 MOTUs per station.

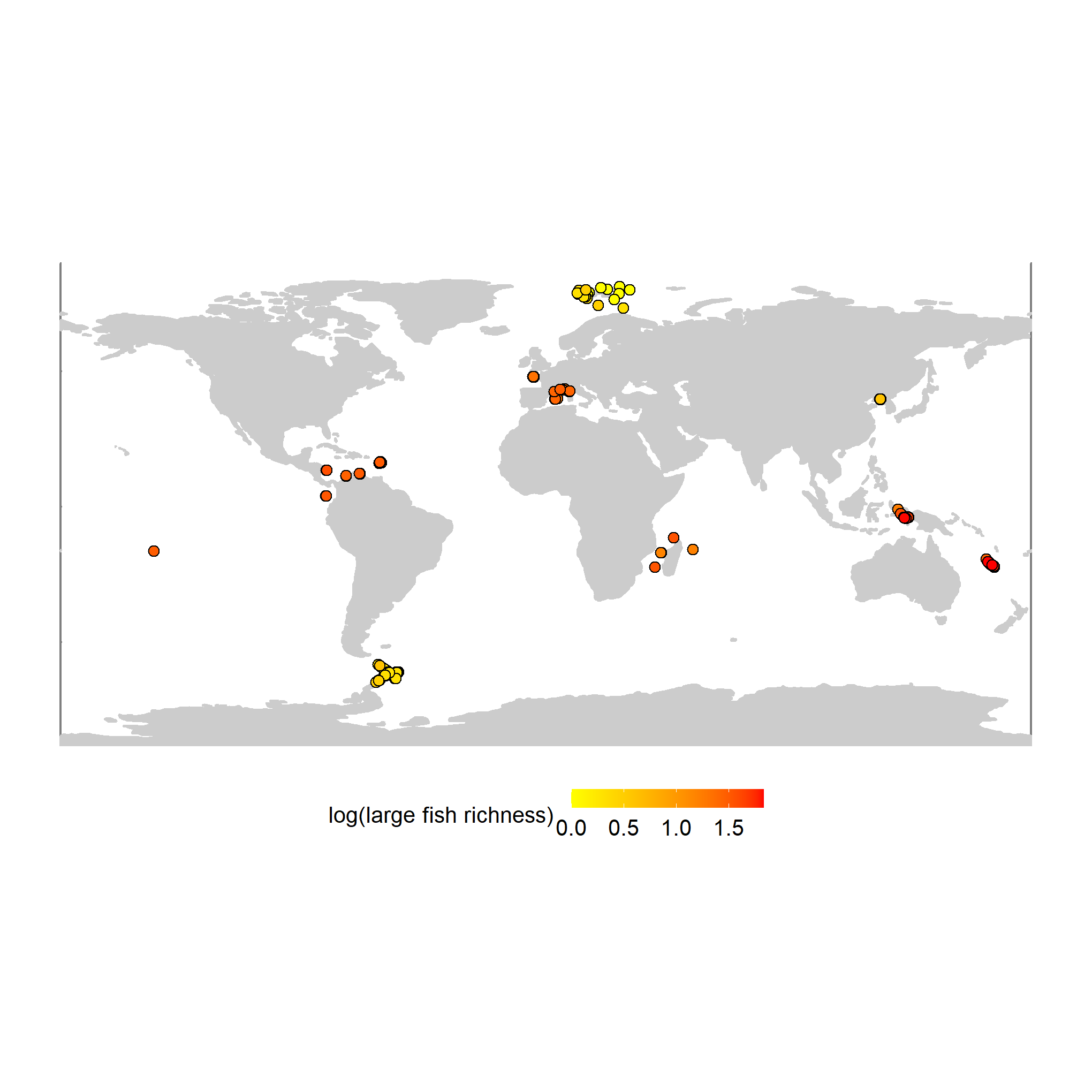
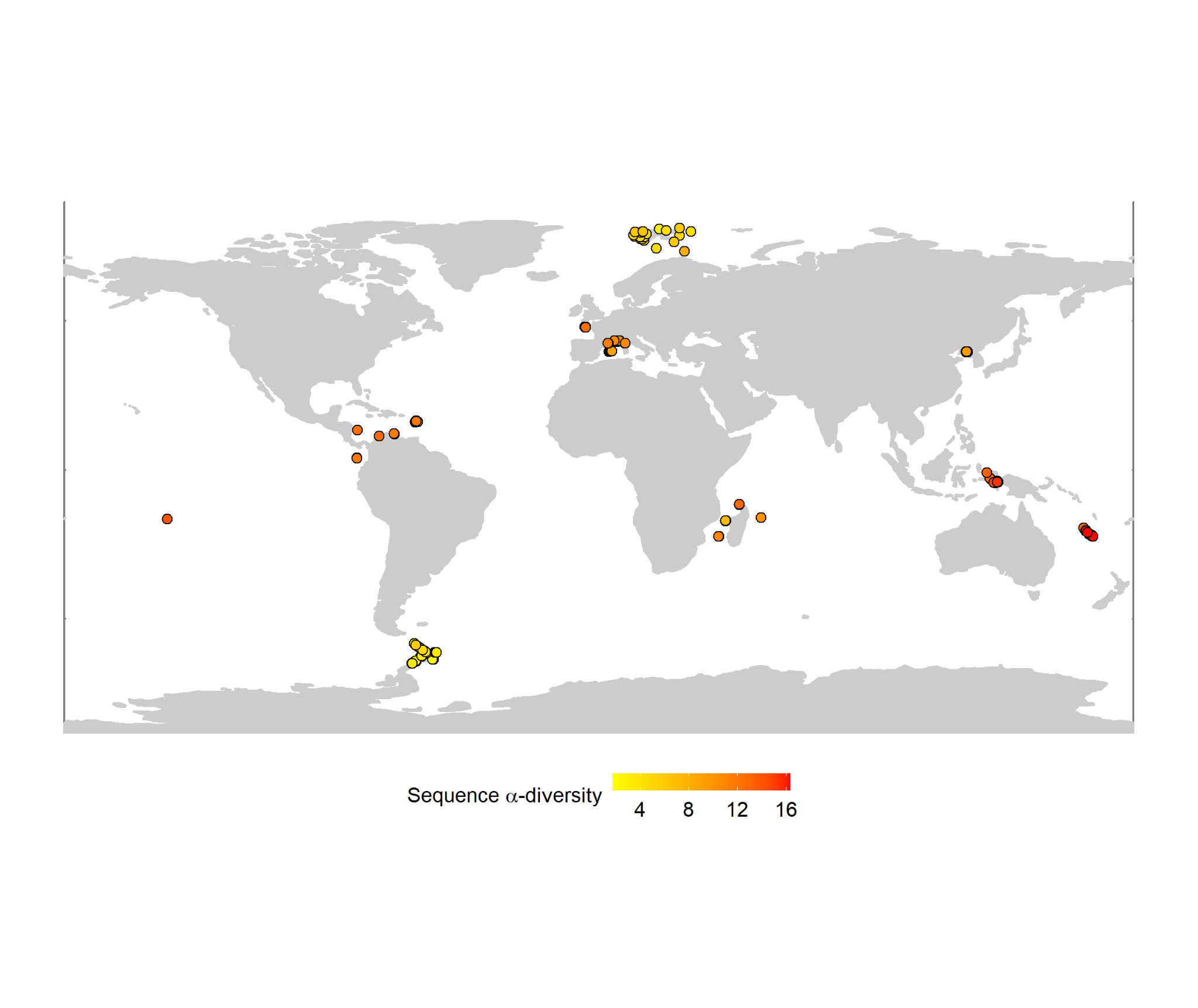


Figure S8. Map of large fish taxonomic α-diversity per station (log10 transformed). Raw values range from 0 to 67 MOTUs per station.

Figure S9. Map of sequence α-diversity per station.

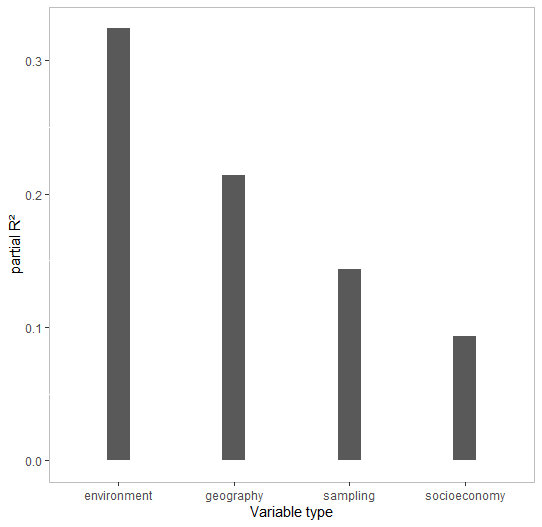


Figure S10. R² partitioned by variable groups in GLS predicting all fish taxonomic α-diversity.

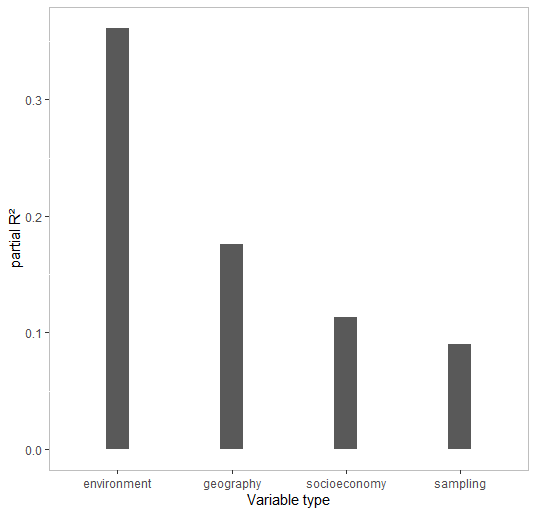


Figure S11. R² partitioned by variable groups in GLS predicting fish sequence α-diversity.

Table S6. Anova of GLS predicting fish taxonomic α-diversity within stations

numDF F-value p-value

(Intercept) 1 10.47266 0.0014

mean\_DHW\_1year 1 9.41162 0.0024

mean\_sss\_1year 1 4.72381 0.0307

mean\_SST\_1year 1 109.54371 <.0001

mean\_npp\_1year 1 3.65093 0.0572

HDI2019 1 0.64457 0.4228

Gravity 1 0.59028 0.4430

MarineEcosystemDependence 1 22.62229 <.0001

dist\_to\_CT 1 77.42021 <.0001

bathy 1 0.59233 0.4422

depth\_sampling 1 4.13457 0.0431

distCoast 1 17.38010 <.0001

volume 1 4.46858 0.0355

sample\_method 1 21.37818 <.0001

Table S7. Anova of GLS predicting sequence α-diversity within stations

numDF F-value p-value

(Intercept) 1 1.02278 0.3128

mean\_DHW\_1year 1 0.86087 0.3544

mean\_sss\_1year 1 2.73074 0.0997

mean\_SST\_1year 1 139.48278 <.0001

mean\_npp\_1year 1 31.30070 <.0001

HDI2019 1 1.00210 0.3178

Gravity 1 0.00607 0.9380

MarineEcosystemDependence 1 15.37866 0.0001

dist\_to\_CT 1 48.79961 <.0001

bathy 1 0.04562 0.8311

depth\_sampling 1 17.16315 <.0001

distCoast 1 6.49684 0.0114

volume 1 0.00431 0.9477

sample\_method2 1 27.39906 <.0001

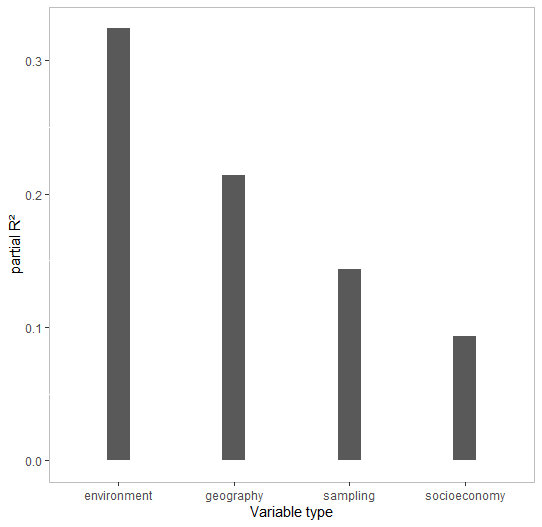


Figure S12. R² partitioned by variable groups in GLS predicting cryptobenthic taxonomic α-diversity (n=539 MOTUs).

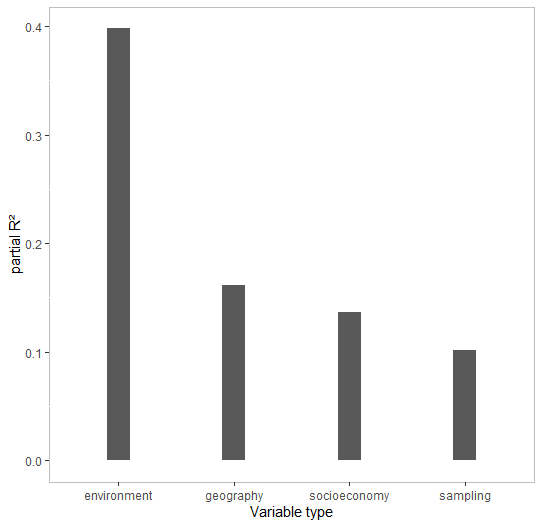


Figure S13. R² partitioned by variable groups in GLS predicting large fish taxonomic α-diversity (n=479 MOTUs).

Table S8. Anova of GLS predicting cryptobenthic taxonomic α-diversity within stations

numDF F-value p-value

(Intercept) 1 2.40720 0.1220

mean\_DHW\_1year 1 4.45024 0.0359

mean\_sss\_1year 1 15.95809 0.0001

mean\_SST\_1year 1 66.31302 <.0001

mean\_npp\_1year 1 25.78901 <.0001

HDI2019 1 0.01149 0.9147

Gravity 1 2.42389 0.1208

MarineEcosystemDependence 1 52.39797 <.0001

dist\_to\_CT 1 57.76497 <.0001

bathy 1 0.04206 0.8377

depth\_sampling 1 13.17270 0.0003

distCoast 1 2.70875 0.1011

volume 1 0.00693 0.9337

sample\_method 1 10.11247 0.0017

Table S9. Anova of GLS predicting large fish taxonomic diversity within stations

numDF F-value p-value

(Intercept) 1 2.56717 0.1104

mean\_DHW\_1year 1 0.05015 0.8230

mean\_sss\_1year 1 26.65398 <.0001

mean\_SST\_1year 1 82.81727 <.0001

mean\_npp\_1year 1 14.31105 0.0002

HDI2019 1 2.01311 0.1572

Gravity 1 0.16557 0.6844

MarineEcosystemDependence 1 14.82563 0.0001

dist\_to\_CT 1 45.73543 <.0001

bathy 1 0.78522 0.3764

depth\_sampling 1 3.06234 0.0814

distCoast 1 9.12800 0.0028

volume 1 0.00010 0.9920

sample\_method 1 14.00773 0.0002

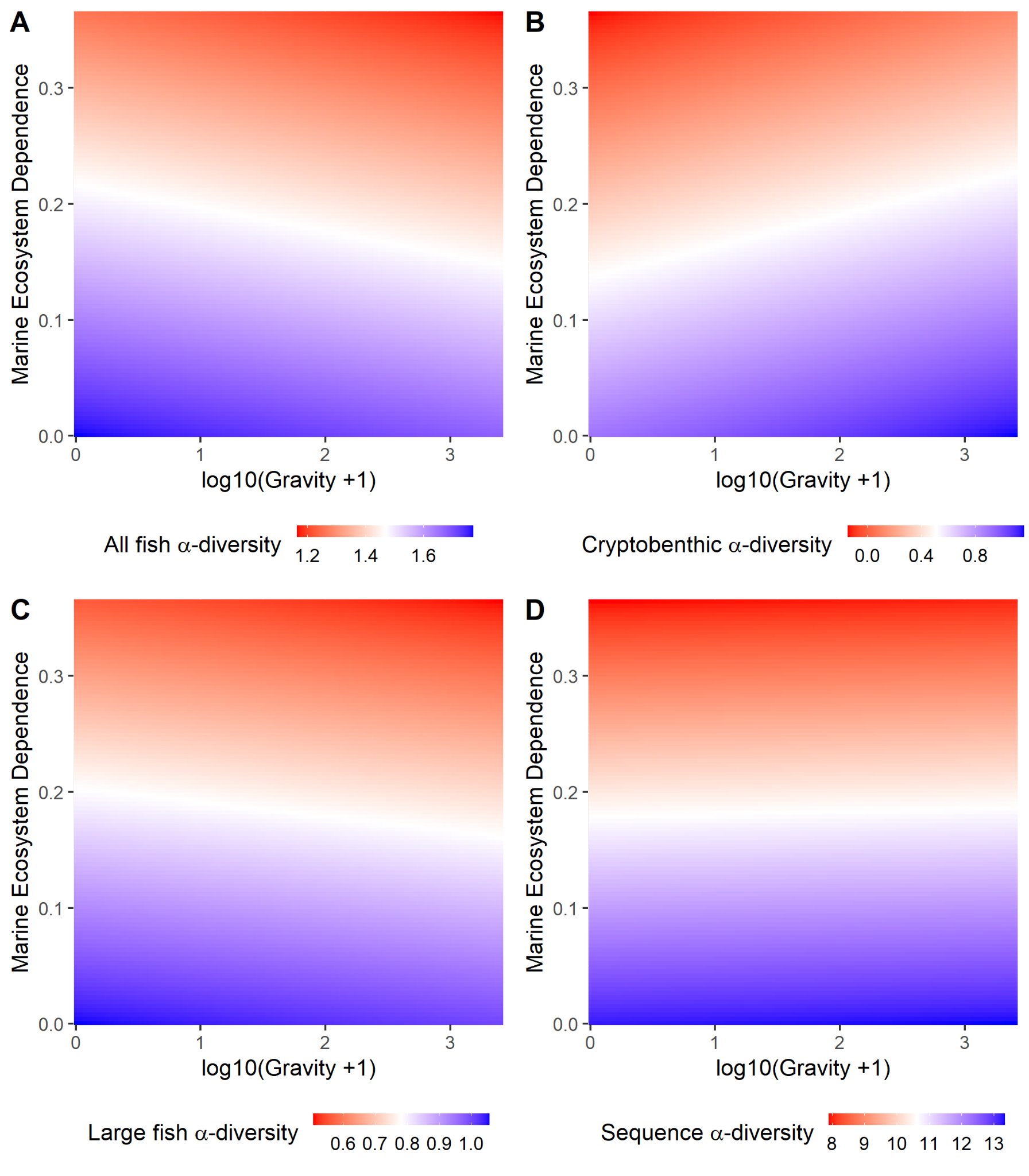


Figure S14. *Partial regression plots showing the relationships between the α-diversity of (A) all fish molecular operational taxonomic unit (MOTU), (B) large fish MOTUs (n = 479 MOTUs), (C) cryptobenthic MOTUs (n = 539 MOTUs), and (D) all fish sequences, and the combined influence of human gravity and marine ecosystem dependence. Blue indicates high diversity values and red indicates low diversity values.*

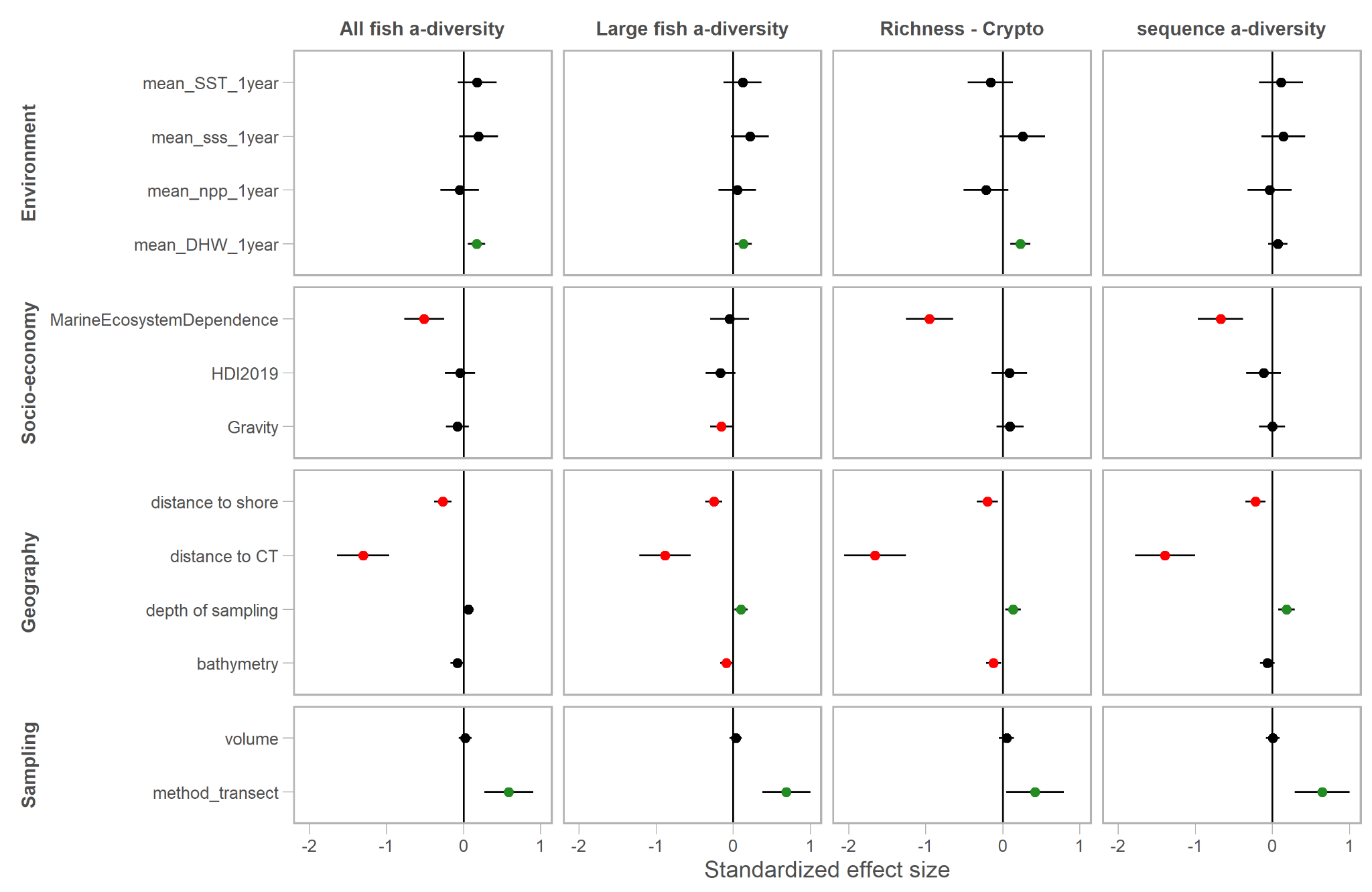
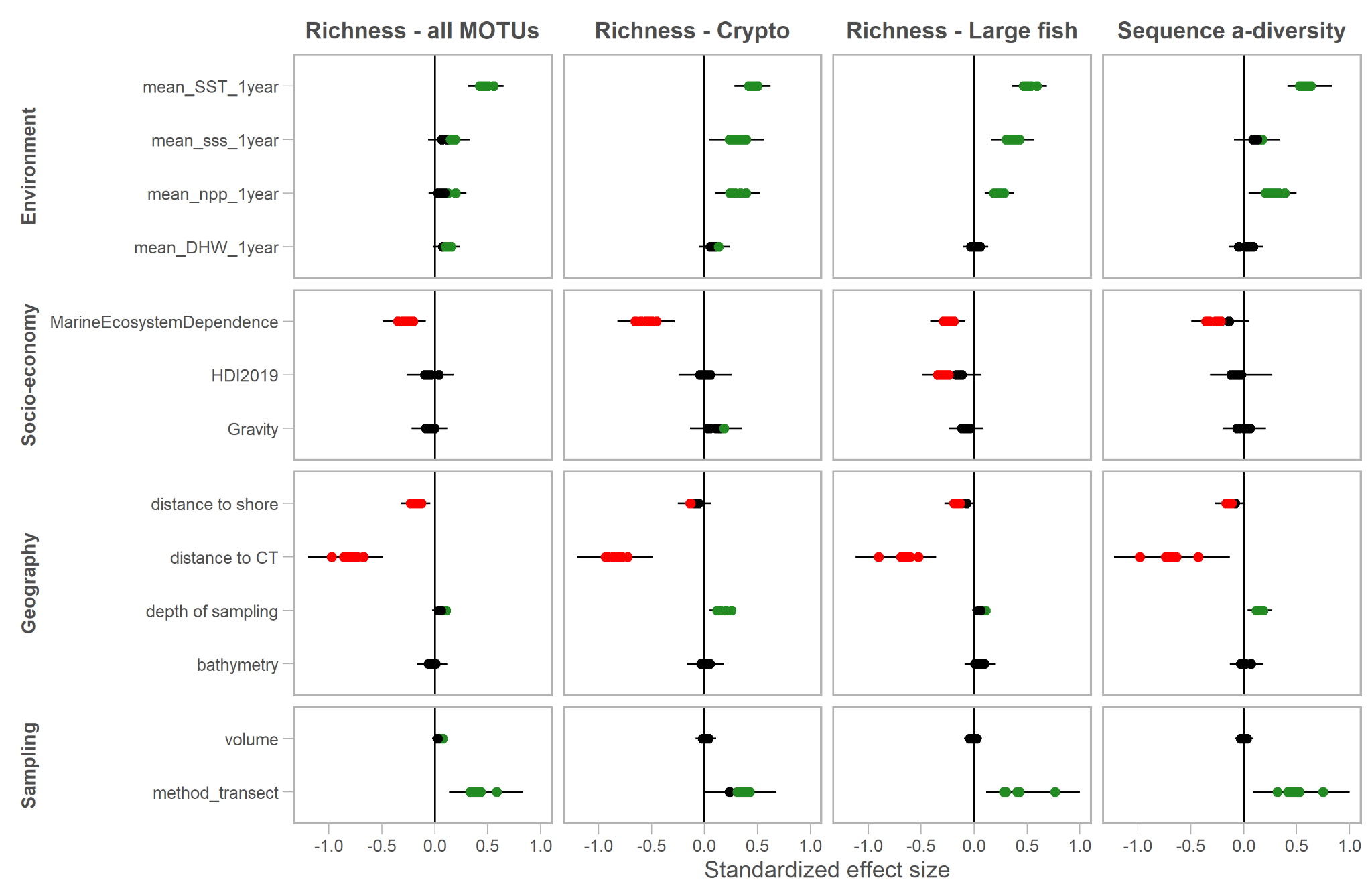
Figure S15. Sensitivity analysis: Effect size of 10 GLS with random subset of 80% of stations. (Cryptobenthic reduced dataset n=539 MOTUs, Large fish reduced dataset n=479 MOTUs).

Figure S16. Effect sizes of variables in GLS on all stations without polar regions (Scotia sea and Arctic). (Cryptobenthic reduced dataset n=539 MOTUs, Large fish reduced dataset n=479 MOTUs).

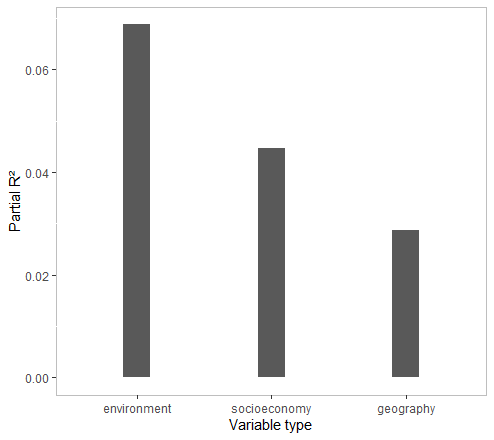


Figure S17. Partial R² by variable groups in dbRDA predicting taxonomic β-diversity between stations

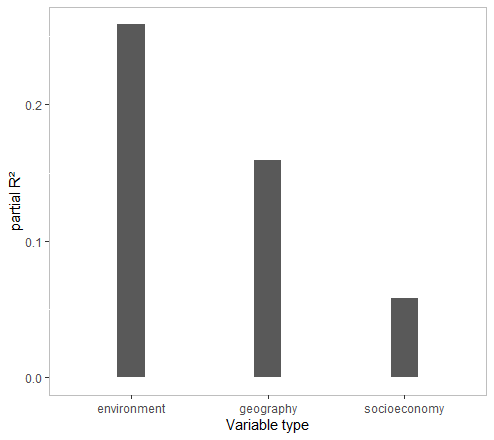


Figure S18. Partial R² by variable groups in dbRDA predicting sequence β-diversity between stations

Table S10. Anova of dbRDA predicting taxonomic β-diversity between stations

Df SumOfSqs F Pr(>F)

mean\_DHW\_1year 1 0.997 2.4709 0.01 \*\*

mean\_SST\_1year 1 2.754 6.8218 0.01 \*\*

mean\_sss\_1year 1 1.442 3.5709 0.01 \*\*

mean\_npp\_1year 1 1.908 4.7268 0.01 \*\*

HDI2019 1 1.494 3.7002 0.01 \*\*

Gravity 1 1.061 2.6271 0.01 \*\*

MarineEcosystemDependence 1 1.741 4.3135 0.01 \*\*

dist\_to\_CT 1 1.753 4.3430 0.01 \*\*

bathy 1 1.118 2.7697 0.01 \*\*

depth\_sampling 1 0.683 1.6924 0.01 \*\*

distCoast 1 0.760 1.8821 0.01 \*\*

Residual 249 100.521

Table S11. Anova of dbRDA predicting sequence β-diversity between stations

Df SumOfSqs F Pr(>F)

mean\_DHW\_1year 1 0.2400 4.0580 0.01 \*\*

mean\_SST\_1year 1 2.0332 34.3722 0.01 \*\*

mean\_sss\_1year 1 0.3325 5.6213 0.01 \*\*

mean\_npp\_1year 1 0.3097 5.2363 0.01 \*\*

HDI2019 1 0.1139 1.9253 0.09 .

Gravity 1 0.0807 1.3646 0.23

MarineEcosystemDependence 1 0.4128 6.9791 0.02 \*

dist\_to\_CT 1 0.2741 4.6330 0.01 \*\*

bathy 1 0.4450 7.5222 0.01 \*\*

depth\_sampling 1 0.0448 0.7578 0.50

distCoast 1 0.3732 6.3089 0.01 \*\*

Residual 248 14.6698

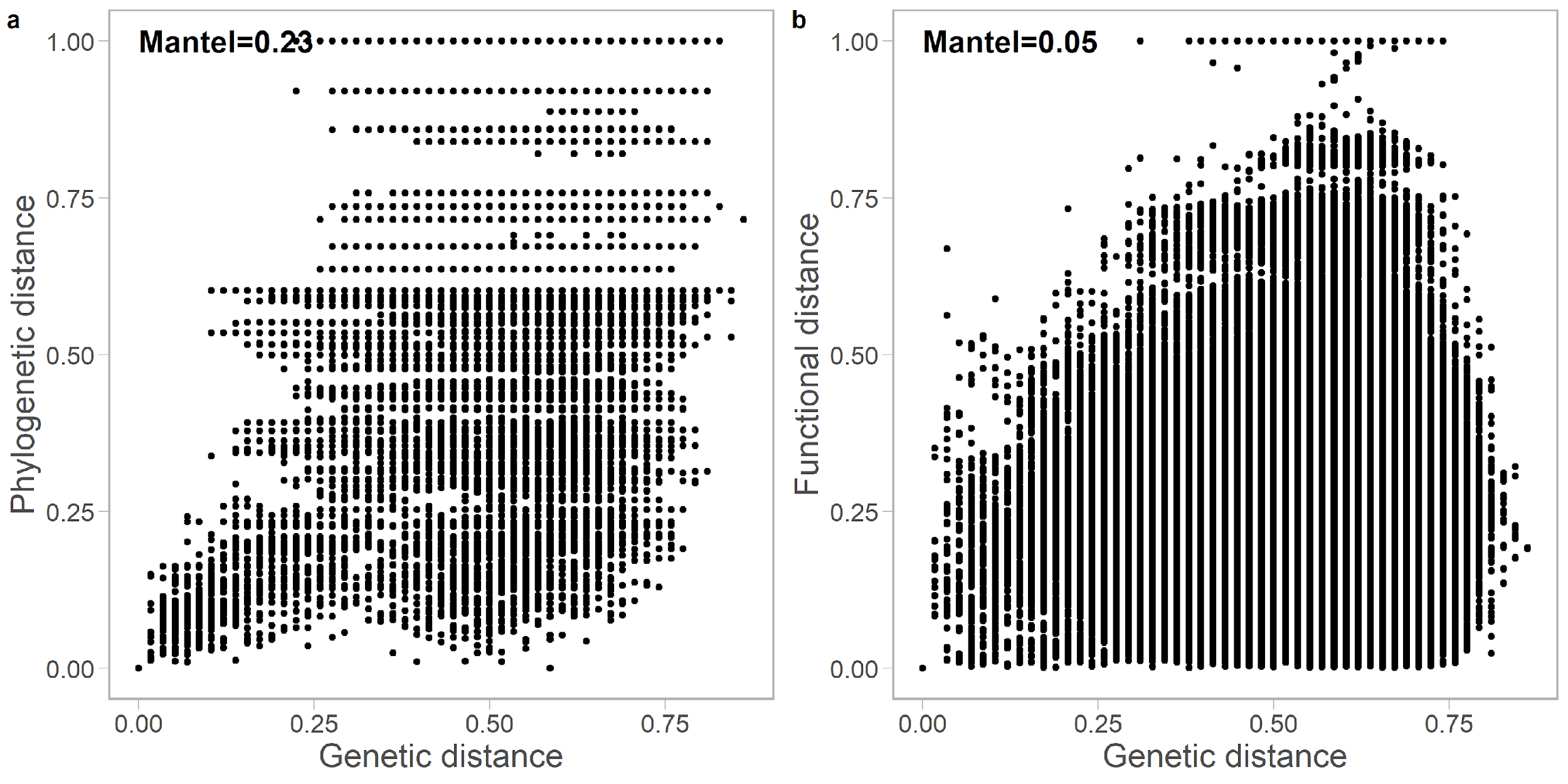


Figure S19. Correlation between **a,** phylogenetic pairwise distance and genetic pairwise distance, and **b**, between functional pairwise distance and genetic pairwise distance.

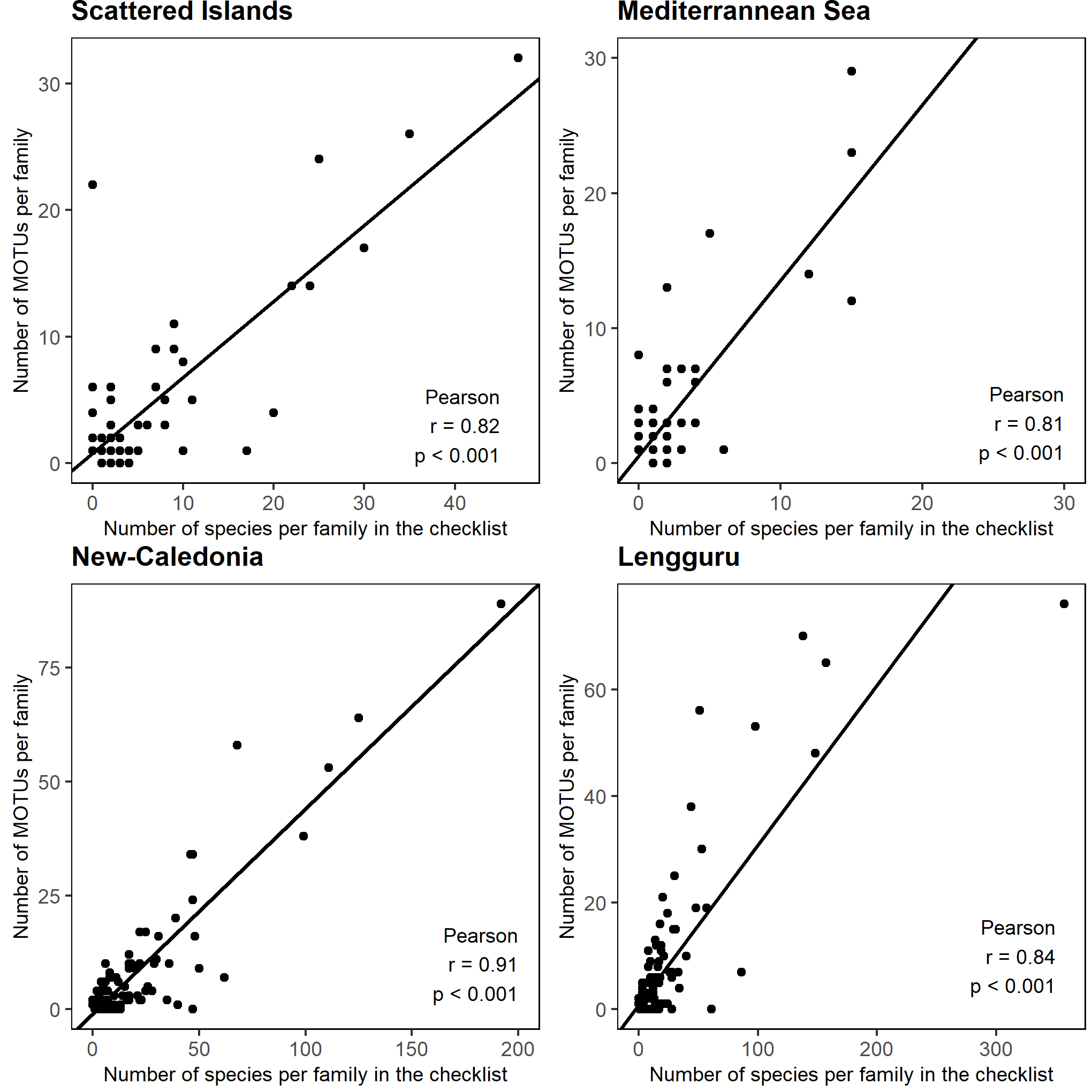


Figure S20. Correlation between number of MOTUs per family in our dataset and number of species per family in the regional checklists for the scattered islands, Mediterranean Sea, New-Caledonia and Lengguru.

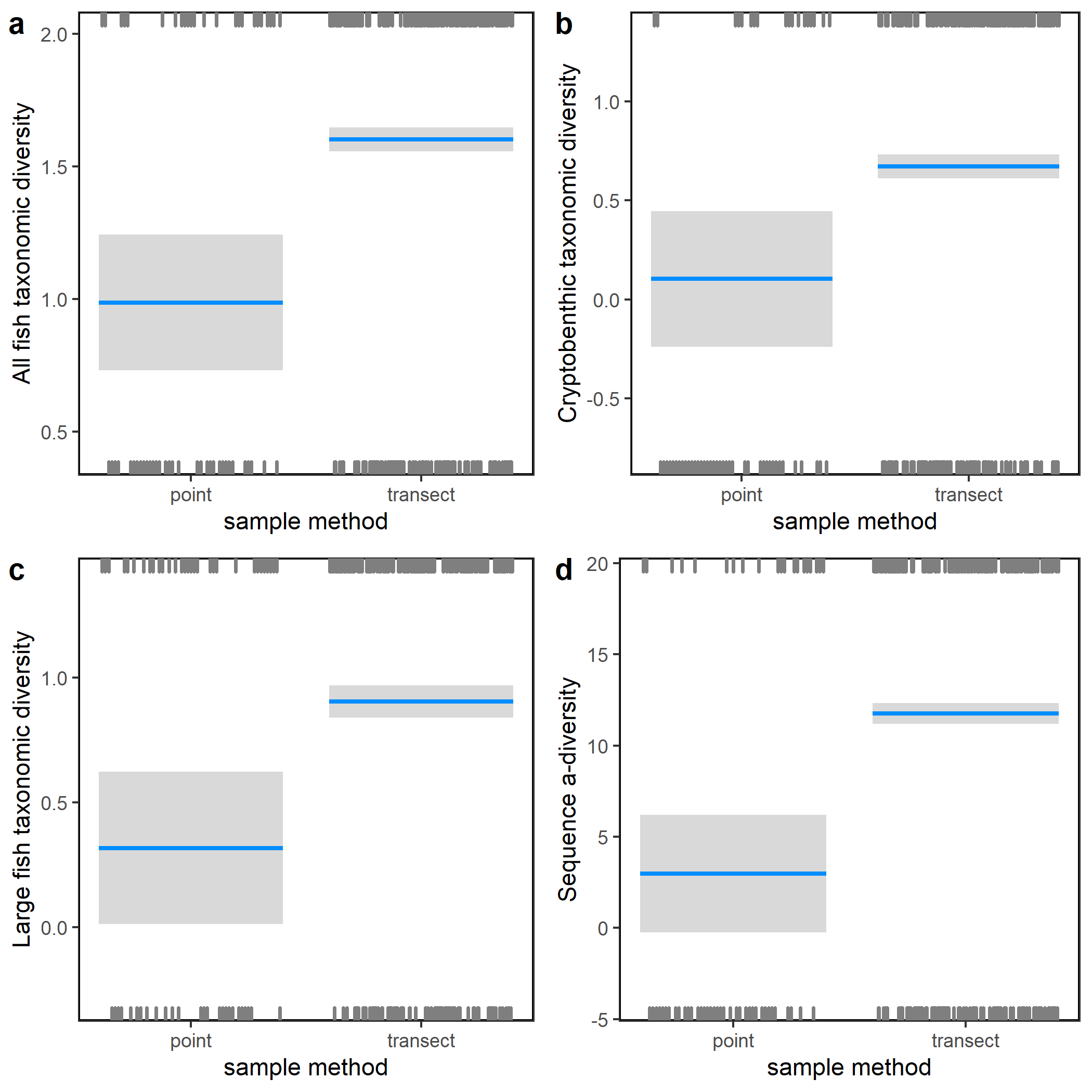


Figure S21. Partial regression plots showing the influence of the sampling method on **a,** all fish taxonomic α-diversity, **b,** Cryptobenthic taxonomic α-diversity (n=539 MOTUs), **c,** Large fish taxonomic α-diversity (n=479 MOTUs), and **d,** sequence α-diversity, conditioned on the median value of all other retained factors.

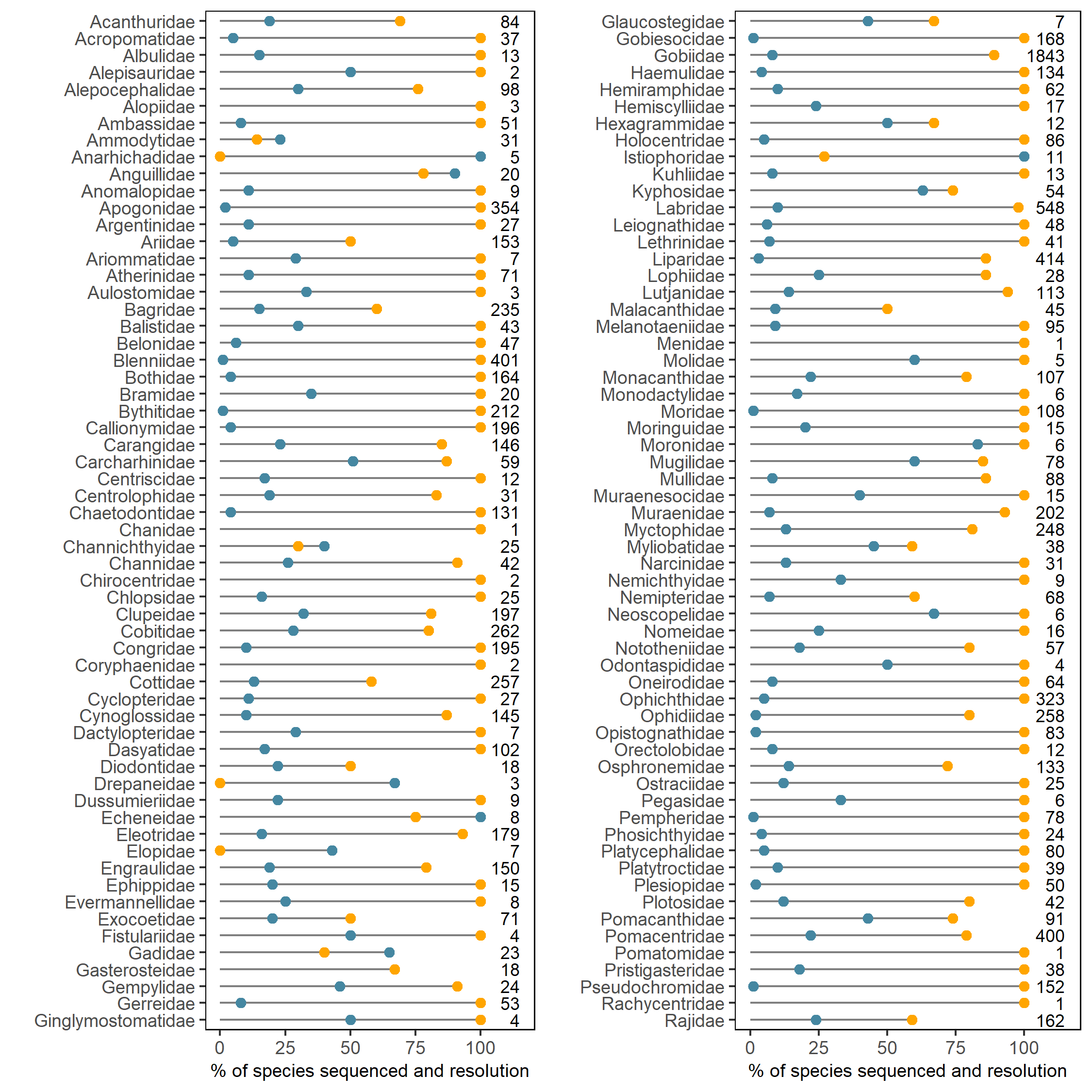


Figure S22. Percentage of species sequenced within each family (blue) and percentage of exclusive taxonomic resolution of the teleo barcode (BE = Proportion of species whose amplified sequences are unambiguously identified. Considering repeated species labels as ambiguity, Marquina et al (2018)) within each family (orange). Total number of species in the family is reported on the graph.

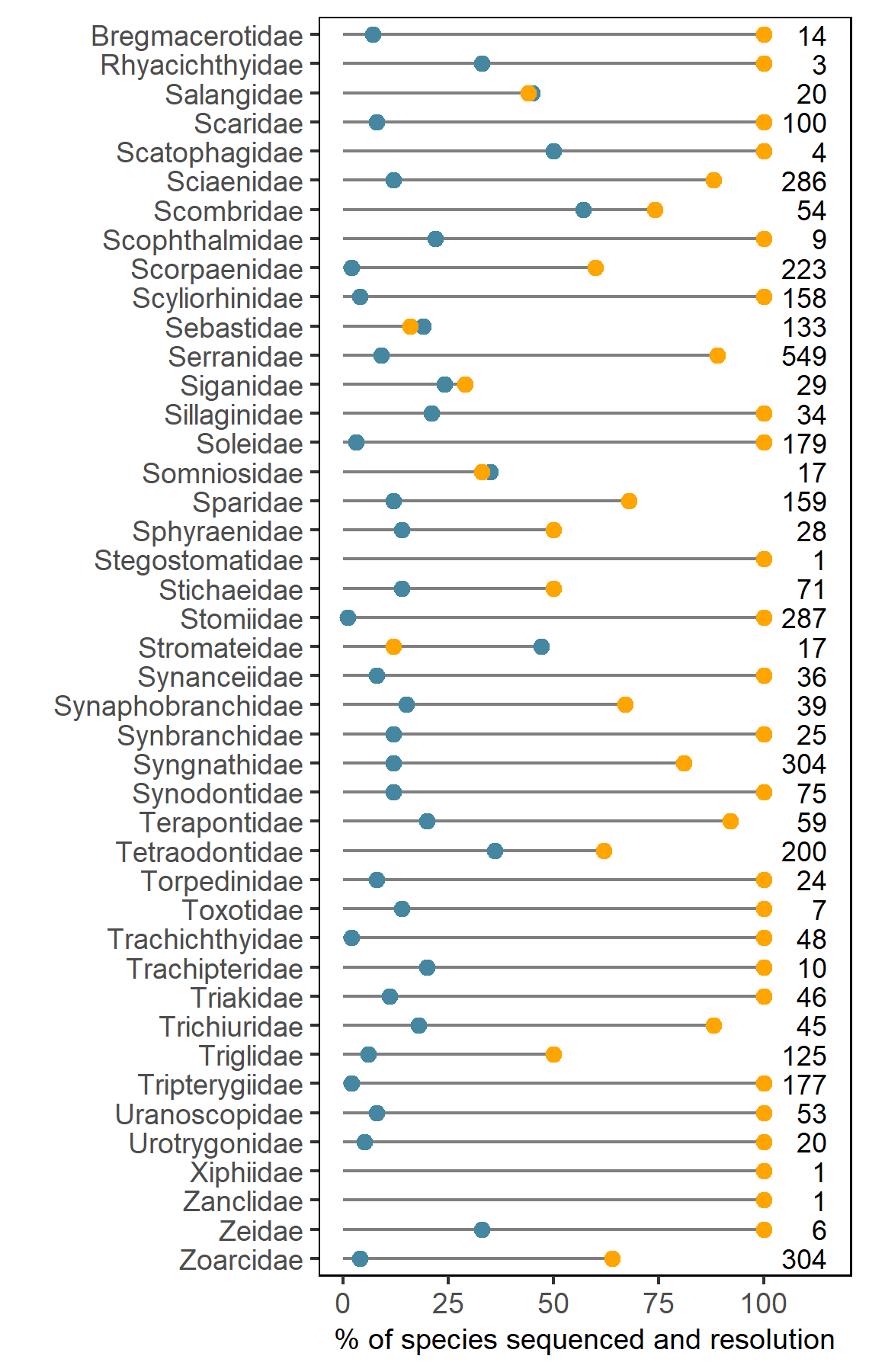


Figure S22 (suite). Percentage of species sequenced within each family (blue) and percentage of exclusive taxonomic resolution of the teleo barcode within each family (orange). Total number of species in the family is reported on the graph.