**Supplementary Data**

Contrasting regional cryptobenthic and conspicuous fish diversity by combining 12S eDNA metabarcoding, visual surveys and collections against historical records

Camila Mac Loughlin1, Tania Valdivia-Carrillo1,2, Salwa El Khattabi1,Fausto Valenzuela-Quiñonez1, Hector Reyes-Bonilla3, Adrian Munguia-Vega4,5\*

1 Centro de Investigaciones Biológicas del Noroeste, La Paz, Baja California Sur, Mexico.

2 Centro de Investigación Científica y de Educación Superior de Ensenada, Ensenada, Baja California, México

3 Universidad Autónoma de Baja California Sur, La Paz, Baja California Sur, Mexico.

4 Applied Genomics Lab, La Paz, Baja California Sur, Mexico.

5 Conservation Genetics Laboratory, The University of Arizona, Tucson, Arizona, USA.

\*Corresponding author e-mail: airdrian@arizona.edu

**Supplementary Data 1**. Methodology for the local 12S reference database

The PCR reactions consisted of 0.2 μM of each primer, 0.2 mM dNTP mix, 1X Mg+ free PCR Buffer, 1.5 mM MgCl2, 0.02% BSA, 1 U DNA Taq polymerase (Invitrogen), and 1 ng of DNA template in 25 μL reactions. The thermocycling protocol consisted of initial 5 min denaturation at 95°C, followed by 35 cycles of 95°C denaturation, 61°C annealing, and 72°C extension for 30 sec each; final extension was set for 5 min at 72°C. Amplifications were verified using 1.2% agarose gel electrophoresis, and successful PCR products were sent for paired-end Sanger sequencing at the University of Arizona Genetics Core, Tucson facilities in Arizona, USA.

Sequences were manually curated using Chromas Pro v2.1.10, and complete alignments were conducted using ClustalX v2.1 (Larkin et al., 2007), resulting in 58 unique high-quality sequences. Taxonomic annotations were manually retrieved from the NCBI taxonomy browser (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>; accessed January 15th, 2023), and curated sequences were registered in GenBank (Table S4). We obtained the ~64 bp teleo metabarcode reference for each species using the teleo primers to perform *in silico* PCR with Obitools-ecoPCR (Boyer et al., 2016). *In silico* PCR was unsuccessful for 14 species, for which the metabarcode region was extracted by performing complete alignment against the available references using ClustalX v2.1 (Larkin et al., 2007). To assess species discrimination of the teleo metabarcode within our new species, a complete alignment distance tree was plotted using FigTree v1.4.4 (Rambaut, 2010). To evaluate the availability of genetic references for the 12S teleo barcode, we created Venn diagrams using TaxonTableTools (Macher et al., 2021) to compare Gulf of California fish found in GenBank against our local reference and our method-combined fish species list for cryptobenthic and conspicuous fish.

**References**

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular ecology resources*, *16*, 176–182.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., … Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* , *23*, 2947–2948.

Macher, T. H., Beermann, A. J., & Leese, F. (2021). TaxonTableTools: A comprehensive, platform-independent graphical user interface software to explore and visualize DNA metabarcoding data. *Molecular ecology resources*, *21*, 1705–1714.

Rambaut, A. (2010). FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh.

**Supplementary Data 2.** Bioinformatic processing of 12S eDNA sequences

We employed a different bioinformatic approach compared to the annotated OTUs using fixed similarity thresholds outlined by Valdivia-Carrillo et al., (2021). We used the ANACAPA pipeline (Curd et al., 2019) to create a custom 12S reference database, obtain amplicon sequence variants (ASVs) from 12S sequences, and perform taxonomic annotation of ASVs within a Bayesian framework. First, we used the CRUX module with default parameters to create a genetic reference library of the 12S teleo metabarcode. We used the European annotated nucleotide (ENA) vertebrates repository (143rd version, downloaded June 7th, 2021; Kanz et al., 2005) as seed for ecoPCR *in silico* amplification (Boyer et al., 2016) and the NCBI annotated nucleotide repository (downloaded February 16th, 2023) for blastn search of all available 12S sequences. Our local fish references and associated taxonomy were manually incorporated into the CRUX reference and converted to Bowtie format.

Sequencing reads underwent quality control and ASV parsing using ANACAPA's second module. Primers and adapters were removed using cutadapt (Martin, 2011), allowing a 30% mismatch. Sequences with an average Phred33 score below 30 and shorter than 40 bp were discarded using FastX-toolkit (Gordon & Hannon, 2010). ASV parsing was performed using DADA2 (Callahan et al., 2016), allowing a minimum sequence length of 40 bp and 20 pb minimum overlap with a maximum of 2 mismatches for forward and reverse read alignment. The ANACAPA classifier module was then used to assign taxonomy to ASVs using a modified version of the BCLA algorithm in Bowtie 2 v2.3.5 (Langmead & Salzberg, 2012) under default parameters and 100 bootstrap replicates to assess the robustness of phylum to species-level annotations (see Table S4 for complete ANACAPA parametrization).

The resulting 12S raw ASV table contained read counts per sample, sequence identity, taxonomic annotations from species to phylum, bootstrap scores, and GenBank IDs from the matched references. Firstly, we removed all ASVs found in the negative control pool to avoid false positives due to contamination. We also removed global singletons, which pose a higher risk of including false positives, and poly G-tailed ASVs that matched as likely PCR or sequencing errors (see Table S7 for filtering steps). Taxonomic annotations were manually curated, eliminating uncertain annotations with a bootstrap score below 70 (i.e., any ASV annotation that didn’t align to the same reference in at least 70% of bootstrap replicates).

Subsequently, we manually reviewed all taxonomic annotations to retain fish from the Gulf of California. Taxa were retained if species or genus appeared in any of the following taxonomic lists: (i) our local fish genetic references, (ii) visual surveys, (iii) collections, or (iv) historical records. For genera not found in the previous step, we performed a second verification by consulting the Shorefishes of the Tropical Eastern Pacific (Robertson & Allen, 2015) and FishBase (Froese & Pauly, 2002) repositories to confirm the genus distribution in the Gulf of California. These filtering criteria excluded conspicuous genera *Etheostoma* (55 ASVs, ~800k reads)*, Genicanthus* (8 ASVs, 484 reads)*, Grammistes* (2 ASVs, 26 reads)*, Naso* (1 ASV, 8 reads)*, Oreochromis* (2 ASVs, 95 reads)*, Paracaesio* (2 ASVs, 7 reads), *Percina* (4 ASVs, 350 reads), *Pholidapus* (10 ASVs, 216 reads), *Ruvettus (1 ASVs, 4 reads), Sargocentron (7 ASVs, 1,966 reads)* and *Siganus* (1 ASV, 2 reads).

Finally, we excluded all reads and ASVs exclusive to three samples with low amplicon richness (fewer than two standard deviations below the mean of kept samples, i.e., lower than 2,340 ASVs). All filtering steps and resulting ASVs and sequence counts are available in Table S7. We kept a clean and curated ASV contingency table for the 25 high-quality samples, containing phylum-to-species annotations for subsequent analyses.

**References**

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular ecology resources*, *16*, 176–182.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods*, *13*, 581–583.

Curd, E. E., Gold, Z., Kandlikar, G. S., Gomer, J., Ogden, M., O’Connell, T., … Meyer, R. S. (2019). Anacapa Toolkit : An environmental DNA toolkit for processing multilocus metabarcode datasets. *Methods in ecology and evolution / British Ecological Society*, *10*, 1469–1475.

Froese, R., & Pauly, D. (2002). FishBase: A Global Information System on Fishes. *World Wide Web electronic publication: www.fishbase.org.*

Gordon, A., & Hannon, G. J. (2010). *Fastx-toolkit. FASTQ/A Short-Reads Preprocessing Tools*. http://hannonlab.cshl.edu/fastx\_toolkit/index.html.

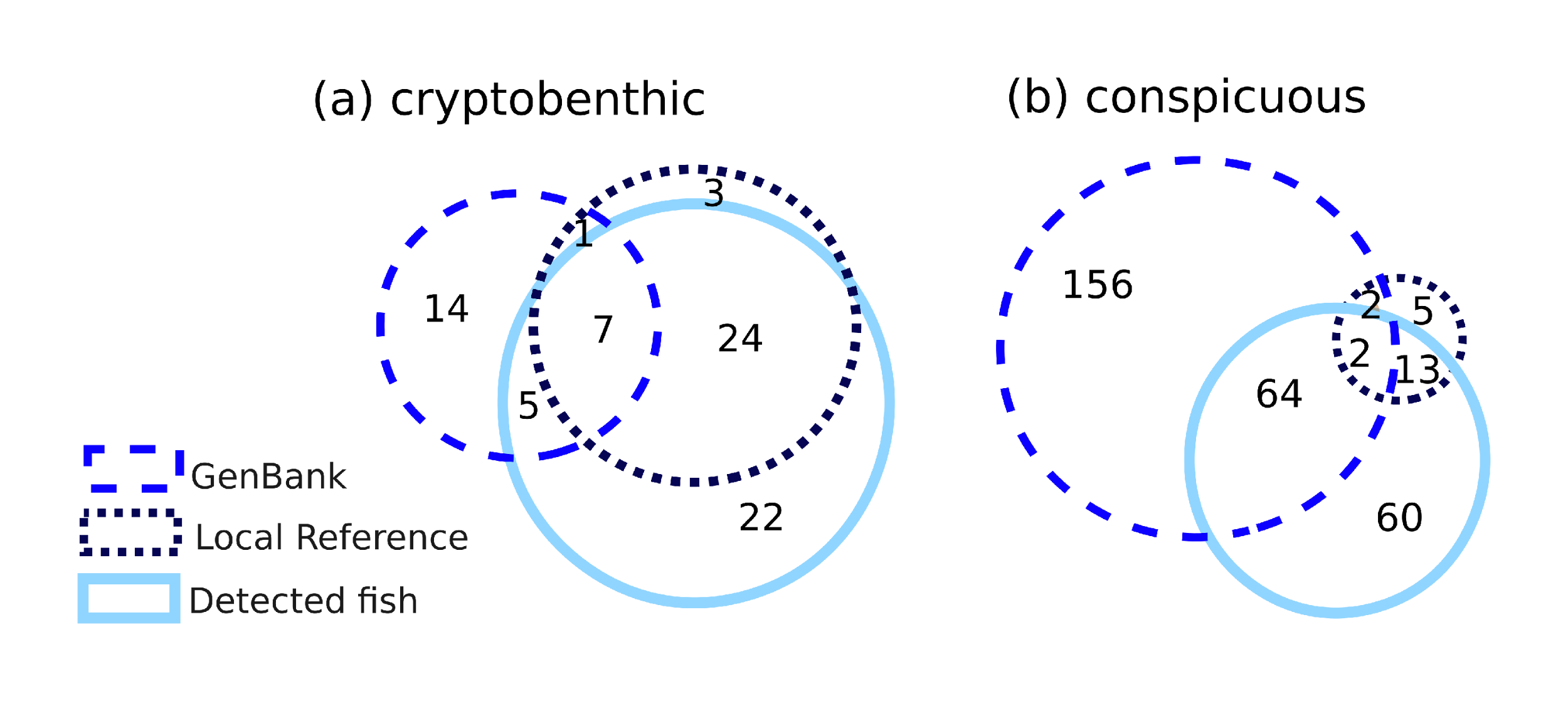
Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., … Apweiler, R. (2005). The EMBL nucleotide sequence database. *Nucleic acids research*, *33*, 21–26.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*, 357–359.

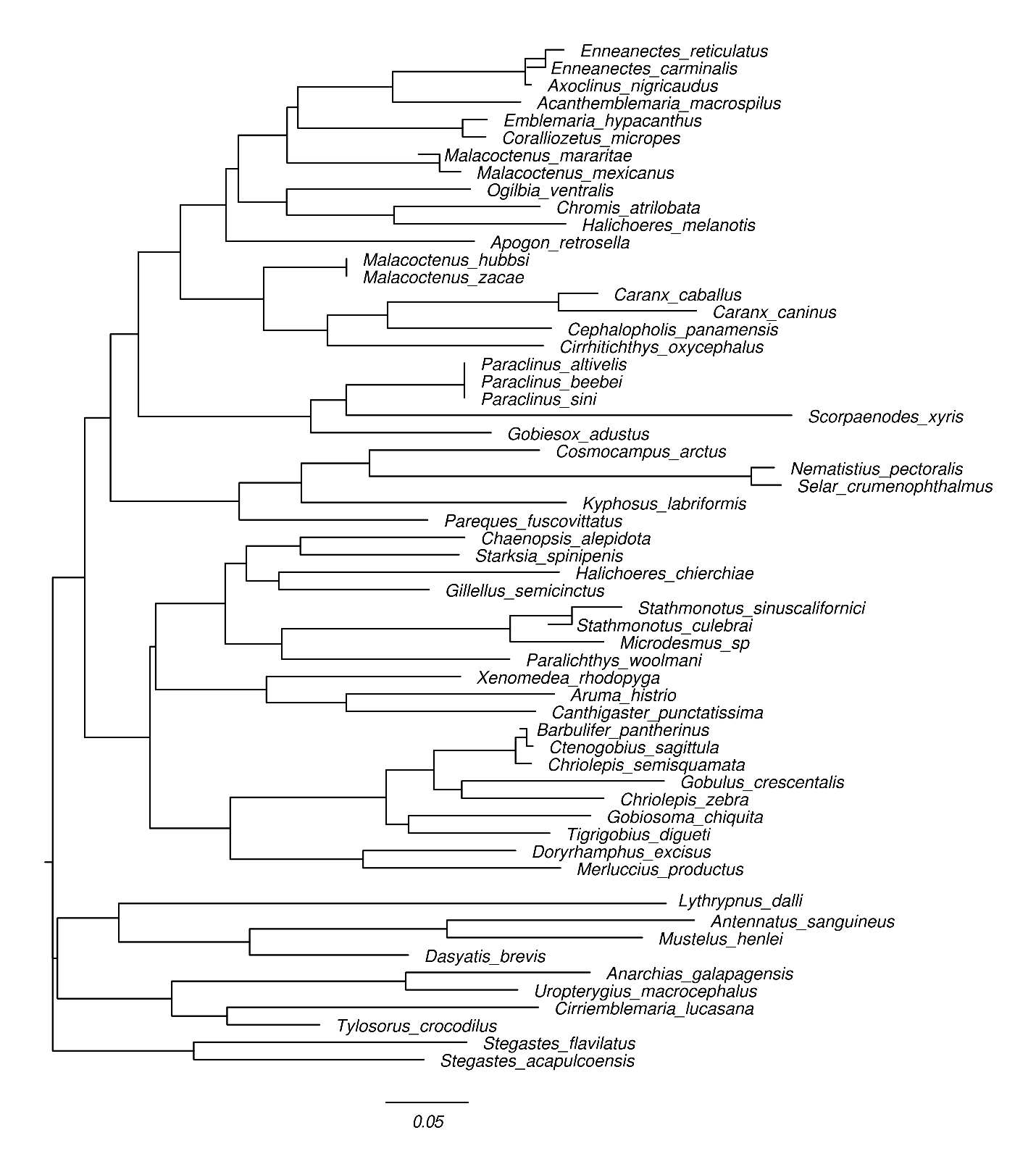
Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*, 10.

Robertson, D. R., & Allen, G. R. (2015). Shorefishes of the Tropical Eastern Pacific: online information system. Version 2.0. Balboa, Panama: Smithsonian Tropical Research Institute 2015. https://biogeodb.stri.si.edu/sftep/en/pages

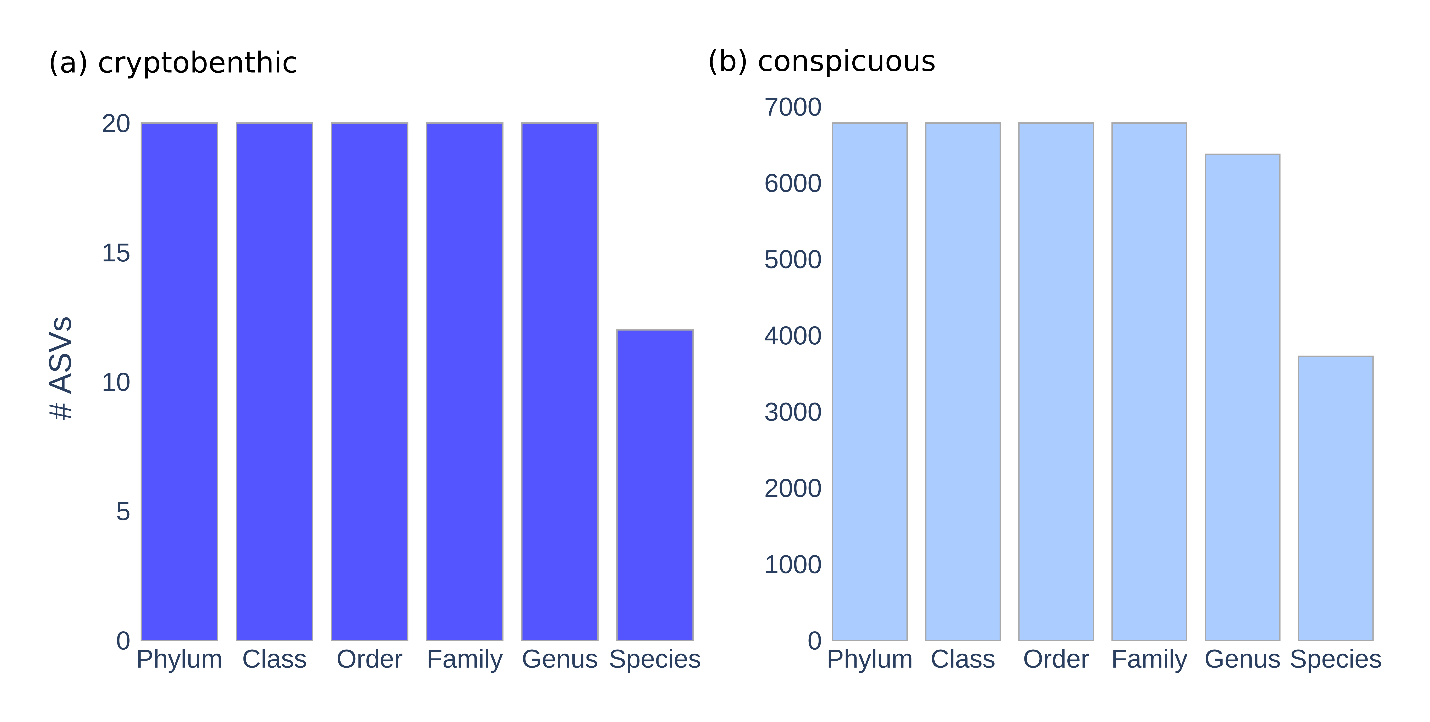
Valdivia-Carrillo, T., Rocha-Olivares, A., Reyes-Bonilla, H., Domínguez-Contreras, J. F., & Munguia-Vega, A. (2021). Integrating eDNA metabarcoding and simultaneous underwater visual surveys to describe complex fish communities in a marine biodiversity hotspot. *Molecular ecology resources*, 21, 1558-1574.



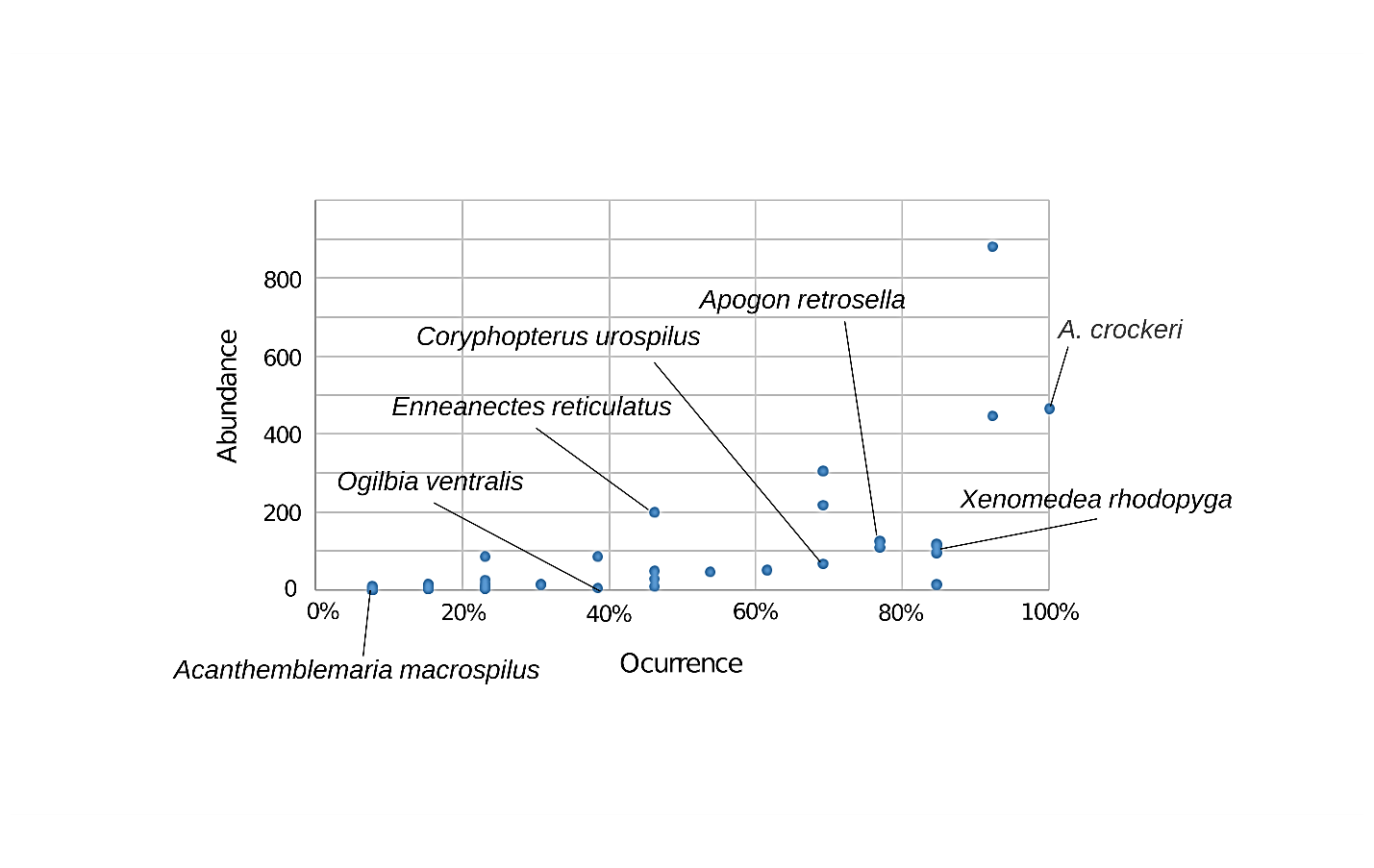
**Figure S1.** Comparison of genetic reference availability for cryptobenthic and conspicuous fish species from the Gulf of California. Venn diagrams showing 12S teleo metabarcodes found in GenBank, in our local reference, and all detected species in the field.



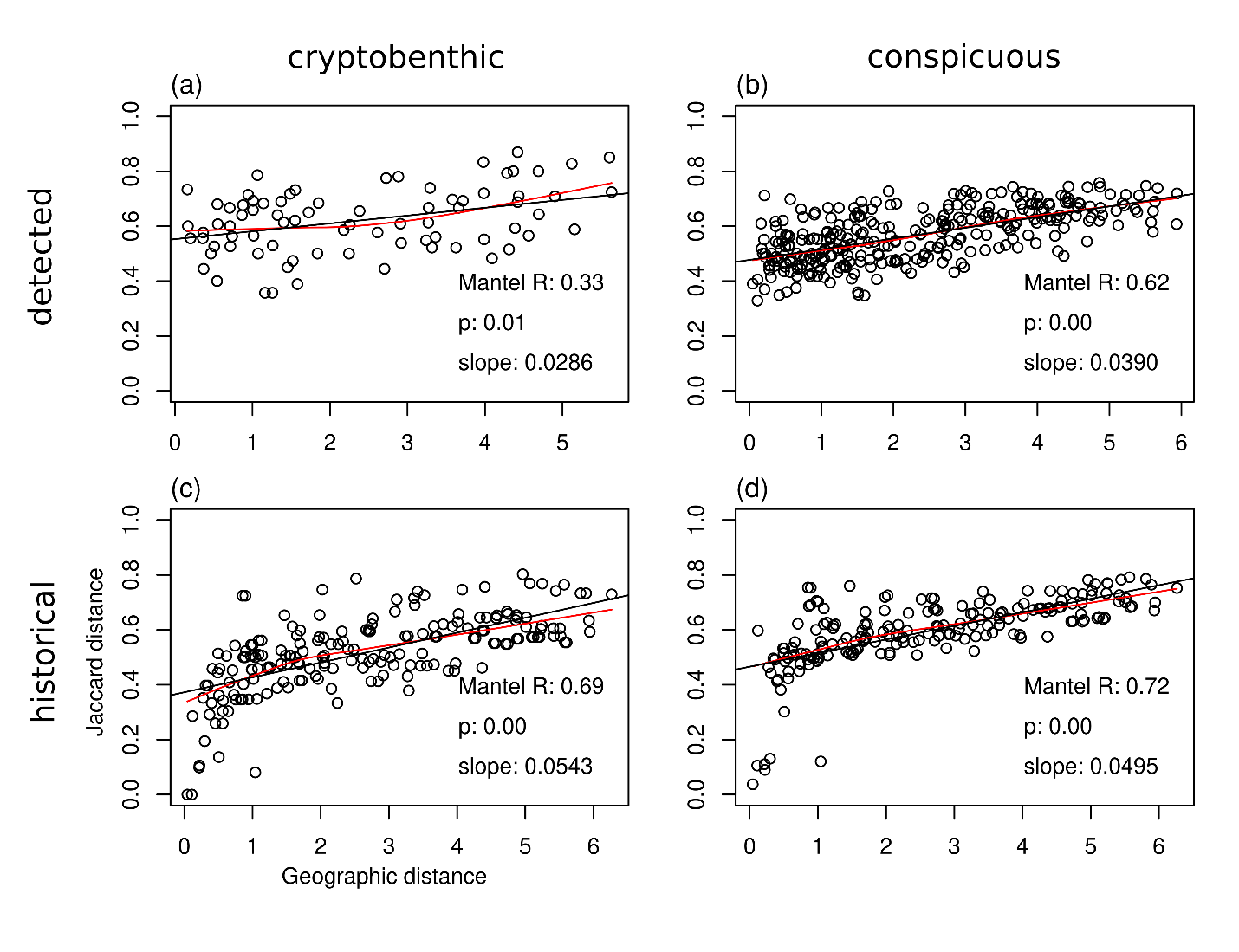
**Figure S2.** Genetic distance tree created from complete alignment of the teleo metabarcode obtained from 59 fish species added to a local reference database.



**Figure S3**. Taxonomic rank resolution for 12S eDNA metabarcoding. Number of ASVs identified per taxonomic rank for cryptobenthic and conspicuous fish.



**Figure S4.** Abundance and frequency distribution plot from collected cryptobenthic species. The species noted were collected by 12S eDNA metabarcoding. A noteworthy exception was *Acantemblemaria macrospilus*, a low abundance and infrequent sister species to *A. crockeri*. Unfortunately, despite several attempts with different individuals and varying PCR protocols, we were unable to produce a reference barcode for *A. crocker*i. Still, we suspect it might be *A. crockeri* that was identified as their closest referenced relative, *A. macrospilu*s.



**Figure S5.** Mantel test for distance dependence in community composition for the detections combining all methods and historical cryptobenthic and conspicuous fish in the Gulf of California. The straight line is a modeled regression among distance matrices, and the red smoothed line follows raw data dispersion. Mantel’s R statistic and p value are noted.