Supporting Information – McCartin *et al.* 2024

Nuclear eDNA Meta-barcoding Primers for Anthozoan Coral Biodiversity Assessment

Luke J. McCartin[1,2], Emma Saso[3], Samuel Vohsen[1,2], Nicole C. Pittoors[1,2], Penny Demetriades[1], Catherine S. McFadden[4], Andrea M. Quattrini[3], and Santiago Herrera[1,2,3*]

1. Department of Biological Sciences, Lehigh University, Bethlehem, PA, USA
2. Lehigh Oceans Research Center, Lehigh University, Bethlehem, PA, USA
3. Department of Invertebrate Zoology, Smithsonian National Museum of Natural History, Washington D.C., USA
4. Department of Biology, Harvey Mudd College, Claremont, CA, USA
* santiago.herrera@lehigh.edu

https://doi.org/10.1101/2023.10.26.564240

1. Supporting Methods

1.1 Sample collection and conventional DNA barcoding of coral samples from the northwestern Gulf of Mexico

ROVs were equipped with a 4K or high-definition camera pan and tilt video camera, an Ultra-Short BaseLine (USBL) positioning beacon, and conductivity, temperature, and depth sensors. Branches or entire coral colonies were collected using "coral cutters" on the manipulator arm of the ROV. Samples were stowed in seawater in either an insulated "biobox" or a PVC cylinder with a rubber stopper until ROV recovery. Immediately after the ROV was recovered, the corals were transferred to cooled seawater, and subsamples of coral branches were either flash-frozen in liquid N_2 and stored at -80°C or immersed in 95% ethanol. Genomic DNA was purified from these samples by digestion with Proteinase-K, impurity removal with Ammonium Acetate, precipitation with isopropanol, washing with ethanol, and rehydration in 50 µL of Tris-HCl and EDTA (TE) buffer (dx.doi.org/10.17504/protocols.io.bypypvpw). DNA concentration and quality were assessed using a Nanodrop Spectrophotometer (Thermo Fisher Scientific).

PCR Primers (McFadden and Ofwegen 2012) were synthesized by Eurofins Genomics, purified by standard desalting, and normalized to 100 µM in TE buffer (pH 8.0). PCR reactions were performed in 25 µL final volume using the Promega GoTaq G2 HotStart Colorless Mastermix in 8-strip tubes. Primer stocks in TE buffer were diluted to 10 µM in molecular-grade water, and 0.25 µL of each forward and reverse primers were added to each reaction for a final concentration of 100 nM. 1 µL of template DNA was amplified in each reaction. The total reaction volume was completed to 25 µL using molecular-grade water. Cycling conditions were as follows: initial denaturation at 95°C for 2 minutes; then 30 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 45 seconds, and elongation at 72°C for 25 seconds; and a final elongation at 72°C for 5 minutes. PCR products were held at 4°C short-term (hours) and frozen at 20°C. Amplification success was confirmed by visualization of the target product size on a 1% agarose gel stained with GelRed (Invitrogen) in TBE buffer run at 110 volts for nearly the entire gel length. Samples that did not amplify were diluted 1:10 in TE buffer, and the PCR was repeated. If the reaction failed again, the DNA extracts were further purified using the DNEasy

PowerClean Cleanup Pro Kit (Qiagen) following the manufacturer's instructions, and the PCR was repeated. PCR products were purified before shipment using the QIAquick PCR Purification Kit following the manufacturer's protocol or purified at the sequencing facility as a paid service.

Sanger sequencing was conducted in both directions by Eurofins Genomics. Forward and reverse .ab1 chromatogram files were imported into Geneious Prime Version 2021.0.3 (https://www.geneious.com), and bases with an error probability limit of > 1% were trimmed from the ends of the sequences. Trimmed forward and reverse sequences were *de novo* assembled, and the consensus sequences were extracted. Base calls were made as the highest peak in the aligned chromatograms. Consensus sequences were aligned using MUSCLE (Edgar 2004) with the default parameters. The resulting alignment was visually inspected, and the aligned sequences were trimmed to an equal length by removing lower-quality bases at the alignment's 3' and 5' ends.

1.2 eDNA Field Sampling

Seawater was sampled during ROV dives using 1.7L General Oceanics Niskin bottles (Model 1010) mounted to the port side of the ROV *Global Explorer* (Oceaneering, Houston, TX). Two to four Niskin bottles were remotely triggered to collect replicate water samples at the seafloor near the corals seen in the video. Once the ROV was recovered, water from each Niskin was drained through rubber tubing into a sterile, 2L stand-up Whirl Pak bag (Nasco). The seawater was then filtered over a 0.22 µm pore size polyethersulfone Sterivex filter using a Masterflex L/S peristaltic pump with Easy Load II pump heads and Masterflex L/S 15 platinum-cured silicone tubing. The Sterivex filter was connected to the tubing via a luer-lock nylon barb, and the pump was set to 100 RPM. The effluent was collected in a bucket, and the volume filtered was measured using a graduated cylinder. The average filtered volume was 0.92 ± 0.15 (SD) liters. Once the entire volume was filtered, the Sterivex filter was placed in a Sterile Whirl-Pak and frozen at -80°C. The filters were transported back to the laboratory at Lehigh University on dry ice and stored at -80°C before DNA extraction.

In the evenings after ROV dives were completed, conductivity, temperature, and depth (CTD) casts were conducted at ROV dive sites at Bright Bank and VK826 using a Seabird 911plus CTD unit on a Niskin bottle rosette with twelve ~12 L General Oceanics and Ocean Test Equipment Niskin bottles. The ship was moved to the position on the seafloor of eDNA sampling during ROV dives, and the Niskin bottle rosette was lowered to the seafloor. Altitude from the seafloor was measured with an altimeter, and water samples were collected in duplicate or triplicate by remote triggering of the Niskin bottles as close to the seafloor as possible and at intervals in altitude from the bottom. After recovery of the rosette, samples were processed and preserved in the same way as the samples from Niskins mounted to the ROV, except that filtration was conducted directly from the Niskin bottle rather than after transferring the water to a Whirl-pak. To sample directly from Niskin bottles, a short segment of bleach-sterilized Masterflex L/S 24 C-flex tubing was connected to the bottle's petcock and stepped down to the L/S 15 tubing via a bleach-sterilized nylon barbed straight reducer. The average volume of the samples collected from these two casts was 3.3 ± 0.4 (SD) and 5.0 ± 0.7 (SD) liters, respectively.

1.3 Controlling for Contamination

During sample collection, library preparation, and data analysis, steps were taken to control for cross-contamination between field samples or contamination from other sources in the field and laboratory. eDNA filtration from ROV and CTD-mounted Niskin bottles were

conducted on the aft deck of the *R/V Point Sur* on a foldable table, separate from the wet lab where animal specimens were processed. Before filtration, the work surface was sterilized by wiping it with a 10% solution of household bleach and high-purity water, letting it air dry, and then rinsing it with high-purity water. The water was obtained via a Beckman B-pure system in the wet lab on the ship. For sampling at the Viosca Knoll dive site, two Thermofisher D0803 cartridges were run in parallel to purify the water. The downstream filter was exchanged for a Thermofisher D0809 ultrapure water cartridge for all other sampling. L/S 15 peristaltic pump tubing was sterilized by circulating 10% bleach solution through the tubing for at least five minutes. The tubing was rinsed by pumping fresh, high-purity water for five minutes (approximately 1 liter). Tubing segments and nylon connectors were sterilized by submersion in a 10% bleach solution for at least 15 minutes and rinsed by submersion in high-purity water for 15 minutes. Nylon connectors were air-dried on a clean paper towel afterward. Clean nitrile gloves were worn during sampling and the sterilization of sampling materials. After each sampling day, a sampling negative control was taken by pumping an average of 3.1 liters of high-purity water over a Sterivex filter using the same sampling equipment sterilized for that day and used to process the field samples.

eDNA purification was conducted with tools and supplies dedicated to eDNA extraction and in a separate part of the lab from where DNA was extracted from coral tissue. The work surface was bleach sterilized by wiping it with a 10% solution of household bleach and deionized water, letting it air dry, and then rinsing it with deionized water. Before extraction, pipettes were wiped with a 10% solution of household bleach and UV-irradiated in a UV pipette carousel (nUVaClean). Tube racks were immersed in a 10% solution of household bleach for at least 15 minutes, air-dried, and rinsed with deionized water. Additionally, extraction negative controls, consisting of the DNeasy kit reagents, were extracted alongside the eDNA samples and sampling negative controls.

PCR reactions were conducted in a My-PCR Prep Station (Mystaire, Creedmoor, NC, USA) hood with positive airflow, air filtration through a HEPA filter, and an overhead UV lamp. PCR preparation was also conducted with pipettes dedicated to PCR prep. The work surface and pipettes were wiped with a bleach solution and rinsed in the same manner as the extraction materials and UV-irradiated for 15 minutes before they were used. PCR products were visualized on gels and pooled using pipettes only used to handle PCR products at a separate laboratory bench dedicated to post-PCR work. PCR products were only handled at this bench and were not opened elsewhere in the lab. Sterile, filtered pipette tips were used at all stages of laboratory work. PCRs were performed in 96-well plates, and duplicate PCR negative controls (NTCs), consisting of the PCR mastermix and molecular-grade water in place of the DNA template, were conducted in each plate of PCR reactions.

Sampling negative controls were filtered, extracted, PCR amplified, and sequenced alongside the field samples to monitor for potential contamination sources in the field during sampling and in the lab during library preparation. Libraries of PCR negative controls were also prepared and sequenced with the field samples and sampling negative controls to monitor for contamination during library preparation. None of the sampling negative controls or PCR negative controls amplified, as visualized with agarose gel electrophoresis. Since the sampling negative controls, also extracted alongside the field samples, did not amplify, we did not prepare libraries for the separate extraction negative controls. During the initial sequencing run on an Illumina MiniSeq, equal volumes of each library were pooled and sequenced, including the sampling negative controls and PCR negative controls.

Only 1 µL of sampling negative controls and PCR no-template controls were pooled in the subsequent MiSeq run since these samples did not produce visible amplicons and yielded comparatively fewer reads than the eDNA samples in the MiniSeq run. In the MiniSeq run, sampling negative controls and PCR no-template controls yielded a maximum of 6 read pairs after initial filtering with *cutadapt* to identify amplicons in the libraries and trim primer sequences (**Table S4**). In comparison, the field samples yielded 8,215 read pairs on average after the same filtering step.

To determine if any sampling or laboratory contamination was present in the data, first, we assessed the prevalence of coral detections in the negative control samples. In the data generated using the *Scler-28S-eDNA* primers, five coral genera were detected in negative control samples: *Lateothela* (detected in 1 negative control with a maximum number of reads in any sample replicate of 18), *Incrustatus* (1 control, maximum reads = 4), *Parantipathes* (1 control, maximum reads = 2), *Stichopathes* (2 controls, maximum reads = 2), and *Thesea* (2 controls, maximum reads = 2). In the data generated using the *Scler-28S-eDNA* primers, two coral genera were detected in negative control samples: *Desmophyllum* (4 controls, maximum reads = 6) and *Madracis* (2 controls, maximum reads = 4). In the data generated using the *mtMutS* primers designed by Everett & Park (2018), 1 read from *Paramuricea* was detected in 1 negative control sample.

In addition to assessing the prevalence of contamination in the negative control samples, we also compared the depths at which coral genera were detected to their known depth ranges to assess the likelihood of each detection. In the data generated using the *28S-Anth-eDNA* primers, in samples collected at GC354 at 527 meters depth, we detected eDNA from two characteristically mesophotic octocorals: a maximum of 40 reads from *Thesea nivea* and a maximum of 3 reads from *Swiftia*. A substantial number of reads from *Swiftia* were also detected in other deep-sea samples collected VK826 (170 total). Similarly, eDNA from a few characteristically deep-sea genera was detected in mesophotic samples, including *Acanthogorgia* (27 reads total across two depths at Bright Bank), *Lateothela* (23 reads across three depths at Bright Bank), *Parantipathes* (2 reads at S. of Stetson Bank), *Paracalyptrophora* (1 read at one depth at Bright Bank), and *Sibopathes* (1 read at one depth at Bright Bank). In the data generated using the *28S-Scler-eDNA* primers, we detected eDNA from the genus *Desmophyllum* in mesophotic samples (28 reads across samples at EFGB and one depth at Bright Bank). In the *MutS* data, we detected reads from the mesophotic genera *Thesea* (11 reads across 2 samples), *Ctenocella* (8 reads in 1 sample), *Muricea* (6 reads in 1 sample), and *Bebryce* (2 reads in 1 sample) in deep-sea samples taken from GC354. Further, we detected *Paramuricea* (16 reads across 4 samples) and *Paracalyptrophora* (3 reads in 1 sample) in mesophotic samples taken at Bright Bank and S. of Stetson Bank.

Detections of mesophotic coral genera in deep-sea samples are unlikely. Although eDNA originated at shallower depths and could settle deeper, these detections are more likely to represent cross contamination during sample processing or our two-step library preparation. Therefore, we removed reads from *Swiftia* from the deep-sea samples and otherwise used a filter of 100 reads for any genera in deep-sea samples and 20 reads for mesophotic samples to remove implausible detections of taxa at low read abundances well outside of their depth ranges.

1.4 PCR Optimization

To successfully amplify our samples using the *28S rRNA* primer sets, we tested three master mixes of increasing cost and specificity (as advertised by the manufacturers); increasing

numbers of cycles; increasing primer concentrations; increasing concentrations of $MgCl_2$ and multiple annealing temperatures. Optimization reactions were performed on a subset of the eDNA samples collected from mesophotic and deep-sea sites. Specifically, these samples were collected at the seafloor using the ROV at Bright Bank and lease block VK826. Ultimately, we found that these samples were the most challenging to amplify among those tested herein, and thus the conditions presented represent the best strategy we found to encourage amplification across all the samples in our sample set. Purified genomic DNA from *Stichopathes* sp., *Tanacetipathes* sp., *Swiftia exserta*, *Callogorgia delta*, and *Lophelia pertusa* were also included in these reactions as positive controls for each order targeted: Antipatharia, Malacalcyonacea, Scleralcyonacea, and Scleractinia. In all cases, we used the manufacturer-suggested denaturation and elongation durations for the size amplicon we were targeting (approximately 400 bp) and used primers without the added CS1 and CS2 adapters. To assess the success of each reaction, 4 µL of each PCR product from the eDNA samples was visualized on a 1% agarose TBE gel stained with GelRed at 110V for nearly the entire length of the gel. Amplification success was assessed based on the band's intensity (strong, faint, or absent) and the presence of any products of a different size than the intended product (i.e., off-target bands).

We tested the Promega GoTaq G2 HotStart Mastermix (Promega), KAPA HiFi HotStart Mastermix (KAPA), and Platinum SuperFi II MasterMix (SuperFi II) with primer concentrations of 500 nM, an annealing temperature of 55°C, and 30 to 35 amplification cycles. We found that while the coral tissue extractions amplified in all cases, the eDNA samples did not amplify with either the Promega or KAPA mastermixes. Using the KAPA mastermix, we also tested 1:10 dilutions of our samples under the suspicion that the samples may be inhibited, but we did not see any improvement with the dilution. The SuperFi II mastermix has been shown to be successful in the amplification of fish eDNA from deep-sea environments, where low concentrations of template eDNA are suspected (Kawato et al. 2021). Concordantly, using the Superfi II mastermix, we found that our eDNA samples amplified, albeit weakly, using this mastermix and primer concentrations of 500 µM, an annealing temperature of 55°C and 35 amplification cycles. Increasing the primer concentrations to 1.0 µM and $MgCl_2$ concentrations to 3.25 mM and 4.75 mM, resulted in stronger amplification. However, with stronger amplification, we noted the increased prevalence of off-target bands in some samples. To potentially reduce the amplification of off-target templates while preserving amplification strength, we tested increasing annealing temperatures from 56°C to 59°C. Amplification was achieved at annealing temperatures of 58°C. However, the presence of off-target bands was not resolved. Thus, to encourage specificity to our target as much as possible, we settled on using a ‘touchdown’ protocol, with 15 cycles of annealing temperatures from 70°C to 56°C, prior to 25 cycles at 55°C (Korbie et al. 2008).

1.5 ROV Video Annotation

The video was recorded using the 4K resolution camera with pan, tilt, and zoom functions. The recorded video was converted to high definition 1080p and annotated using the Video Annotation and Reference System (VARS) version 8.3.4 (Monterey Bay Aquarium Research Institute, Monterey, CA). Annotations with geolocation and depth data were compiled with frame grab images (.png file extension) to generate a catalog of all coral morphospecies observed at each ROV dive site. Morphospecies were identified to the lowest possible taxonomic level from the video imagery and according to published species guides for the region (Shuler and Etnoyer 2020; Opresko et al. 2016). Identifications were only made to more specific taxonomic

levels than the video permitted if DNA sequencing data was generated for those morphospecies or if collections were identified by taxonomic experts from material deposited to the Smithsonian National Museum of Natural History.

Supplementary Figure Captions

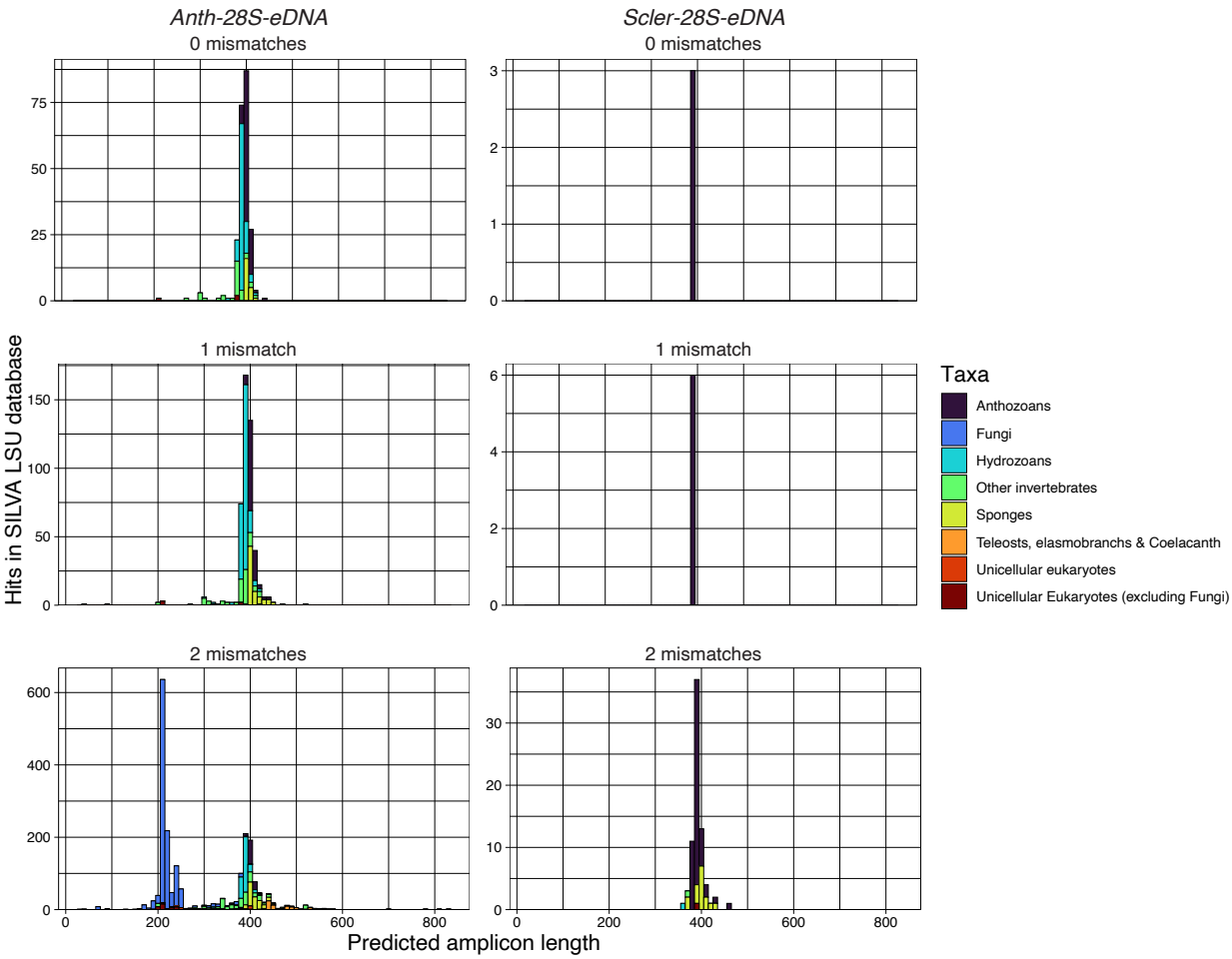

Figure S1: Results of *in silico* taxonomic specificity testing of the *Anth-28S-eDNA* and *Scler-28S-eDNA* primers against the SILVA large subunit ribosomal RNA database.

Amplicons were predicted by querying the database with the primers using *cutapapt* and zero, one, and two permitted mismatches to the primer sequences.

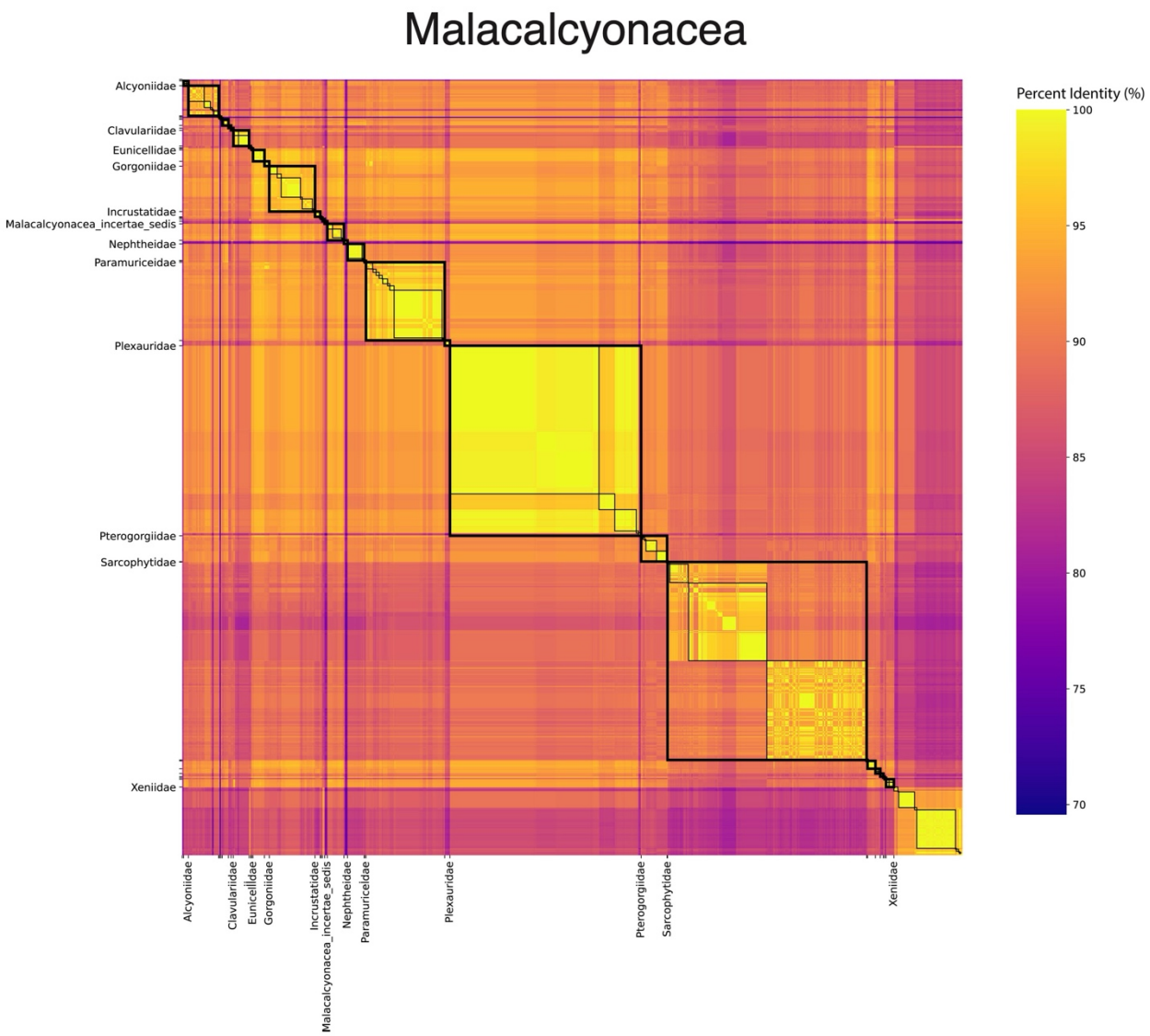

Figure S2: Percent Identities of *mtMutS* barcode sequences of Malacalcyonacea amplified *in silico* with the PCR primers described by Everett and Park (2018).

Boxes with thick black lines delineate comparisons within families, and boxes with thin black lines delineate comparisons within genera. Only families represented by the largest numbers of barcode sequences are labeled to highlight these comparisons and improve readability.

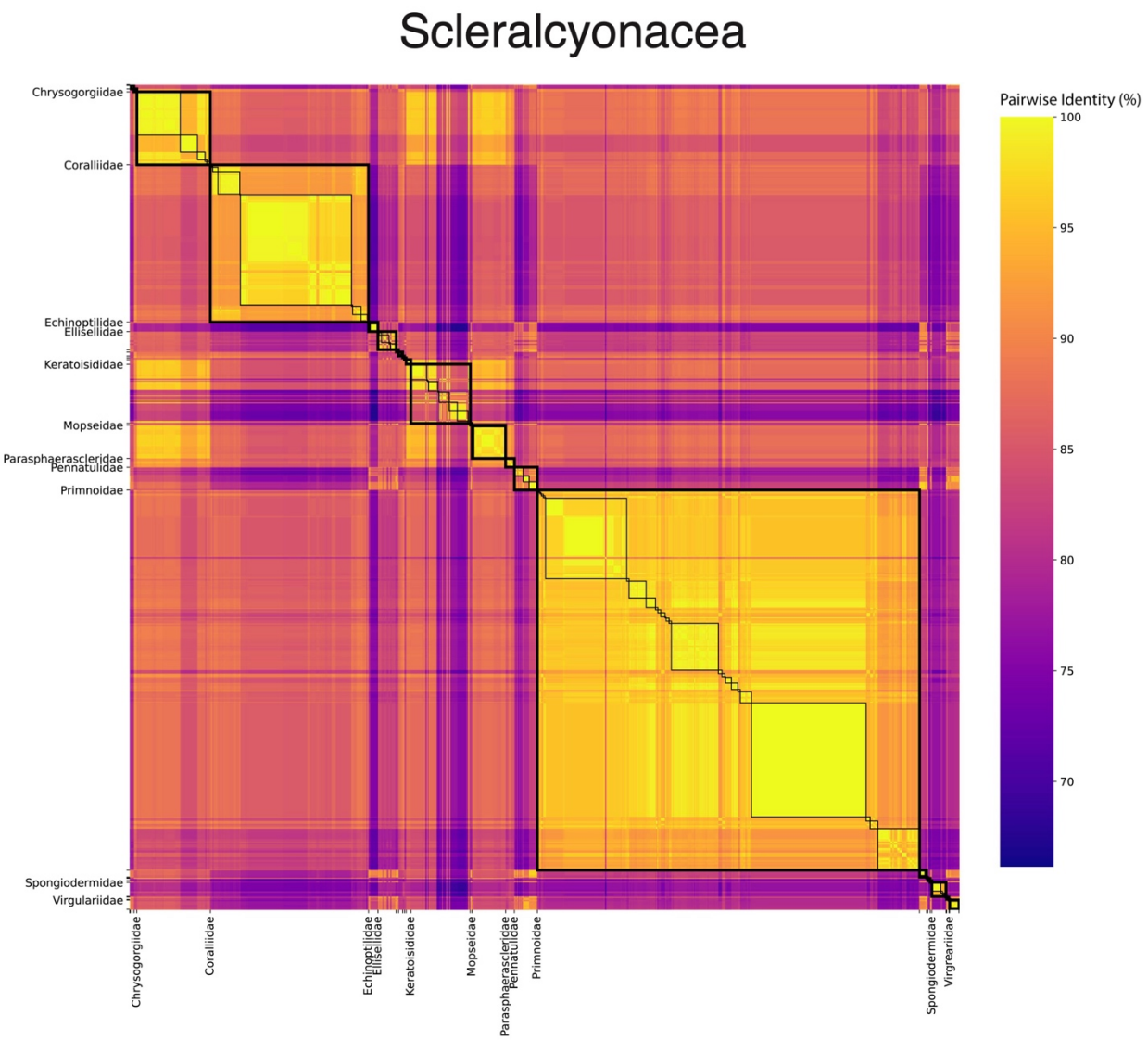

Figure S3: Percent Identities of *mtMutS* barcode sequences of Scleralcyonacea amplified *in silico* with the PCR primers described by Everett and Park (2018).

Boxes with thick black lines delineate comparisons within families, and boxes with thin black lines delineate comparisons within genera. Only families represented by the largest numbers of barcode sequences are labeled to highlight these comparisons and improve readability.

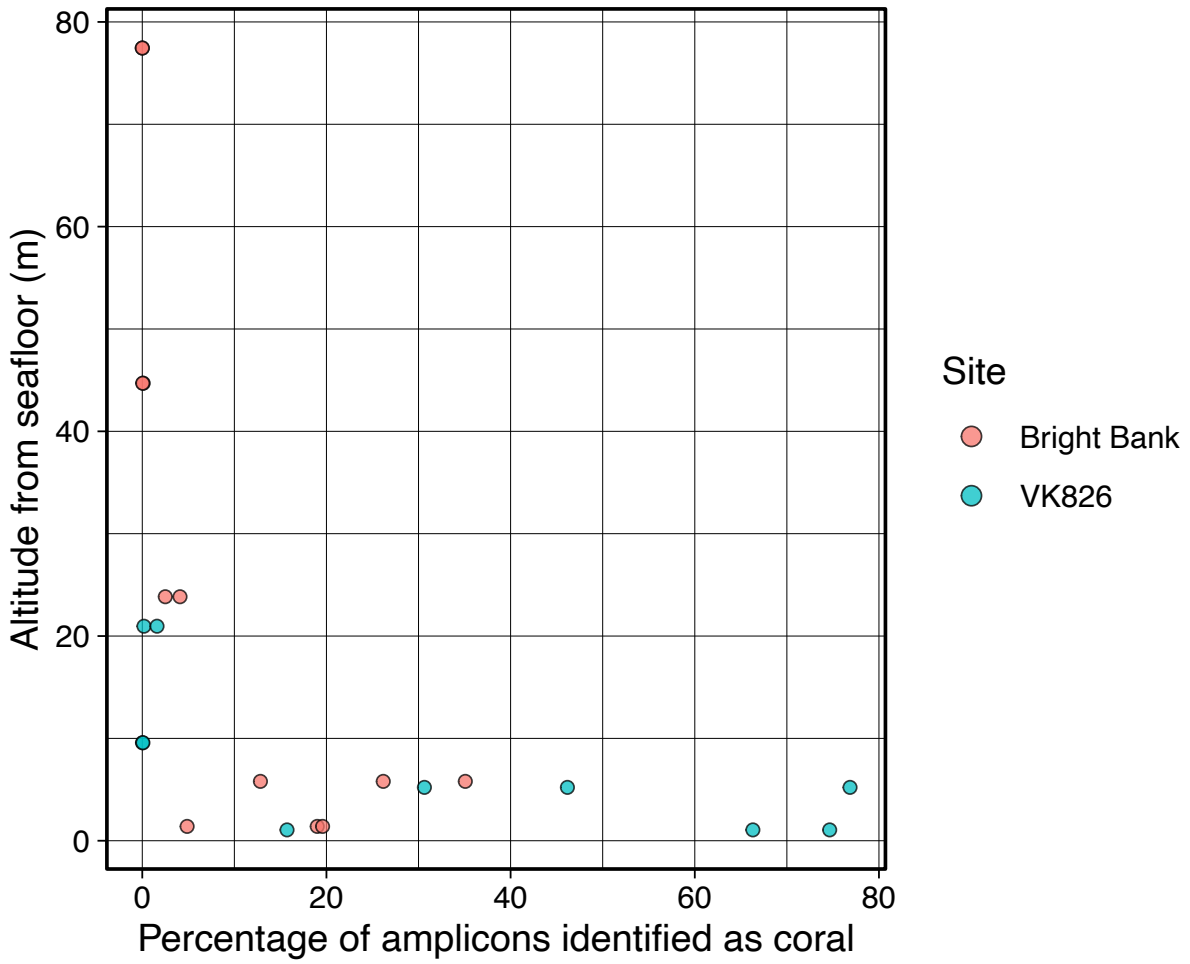

Figure S4: Percentage of sequencing reads identified as coral across libraries generated from eDNA water samples taken at different altitudes from the seafloor.

ASVs were identified as corals if the top BLASTN hit for that ASV was at least 90% identical to an anthozoan coral sequence in GenBank.

Figure S5: Maximum-likelihood phylogenetic tree of amplicon sequence variants (ASVs) recovered from eDNA samples collected in the Gulf of Mexico and amplified using the *28S* primers described herein.
ASVs were aligned with *MAFFT* and the phylogenetic tree was constructed using *IQTree* and *modelfinder* without bootstrapping. The model with the lowest BIC is a General Time Reversible model with empirical base frequencies (+F), and rate heterogeneity modeled using a gamma function (+G4). Tips are labeled with the *28S* primer set used,

ASV numerical identifier and their taxonomic classification. The tree is rooted at the note representing the most recent common ancestor to hexacorals and octocorals.

Supplementary Table Captions

Table S1: Sample metadata for collections of corals made from field sites in the northwestern Gulf of Mexico that were DNA barcoded with conventional PCR/Sanger sequencing and/or genome skimming.

Table S2: *28S* barcode sequences of anthozoan corals generated in this study and downloaded from GenBank for taxonomic classification, as well as analyses of primer complementary and the ability of the *28S* barcode amplified with the primers described herein to delineate taxonomic groups.

Table S3: *mtMutS* sequences of octocorals generated in this study of downloaded from GenBank for taxonomic classification and analyses of the ability of the *mtMutS* barcode amplified with the primers described in Everett and Park (2018) to delineate taxonomic groups.

Table S4: eDNA sample metadata and sequencing statistics for samples amplified using the *28S rRNA* and *mtMutS* primers and sequenced in this study.

Table S5: ASVs and their taxonomic classifications from eDNA metabarcoding of field samples collected in the northern Gulf of Mexico.

References

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Everett, M. V., and Park, L. K. (2018). Exploring deep-water coral communities using environmental DNA. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *150*(September 2017), 229–241. https://doi.org/10.1016/j.dsr2.2017.09.008

Kawato, M., Yoshida, T., Miya, M., Tsuchida, S., Nagano, Y., Nomura, M., Yabuki, A., Fujiwara, Y., & Fujikura, K. (2021). Optimization of environmental DNA extraction and amplification methods for metabarcoding of deep-sea fish. *MethodsX*, *8*, 101238. https://doi.org/10.1016/j.mex.2021.101238

McFadden, C. S., & Ofwegen, L. P. van. (2012). Stoloniferous octocorals (Anthozoa, Octocorallia) from South Africa, with descriptions of a new family of Alcyonacea, a new genus of Clavulariidae, and a new species of Cornularia (Cornulariidae). *Invertebrate Systematics*, *26*(4), 331–356. https://doi.org/10.1071/is12035

Opresko, D. M., Nuttall, M. F., & Hickerson, E. L. (2016). Black Corals of the Flower Garden Banks National Marine Sanctuary. *Gulf of Mexico Science*, *33*(1). https://doi.org/10.18785/goms.3301.05

Shuler, A. & Etnoyer, P. (2020). Alcyonacean octocorals of the Pinnacle Trend: A photo-identification guide. *NOAA technical memorandum NOS NCCOS, 282*. https://dx.doi.org/10.25923/xzd1-z382