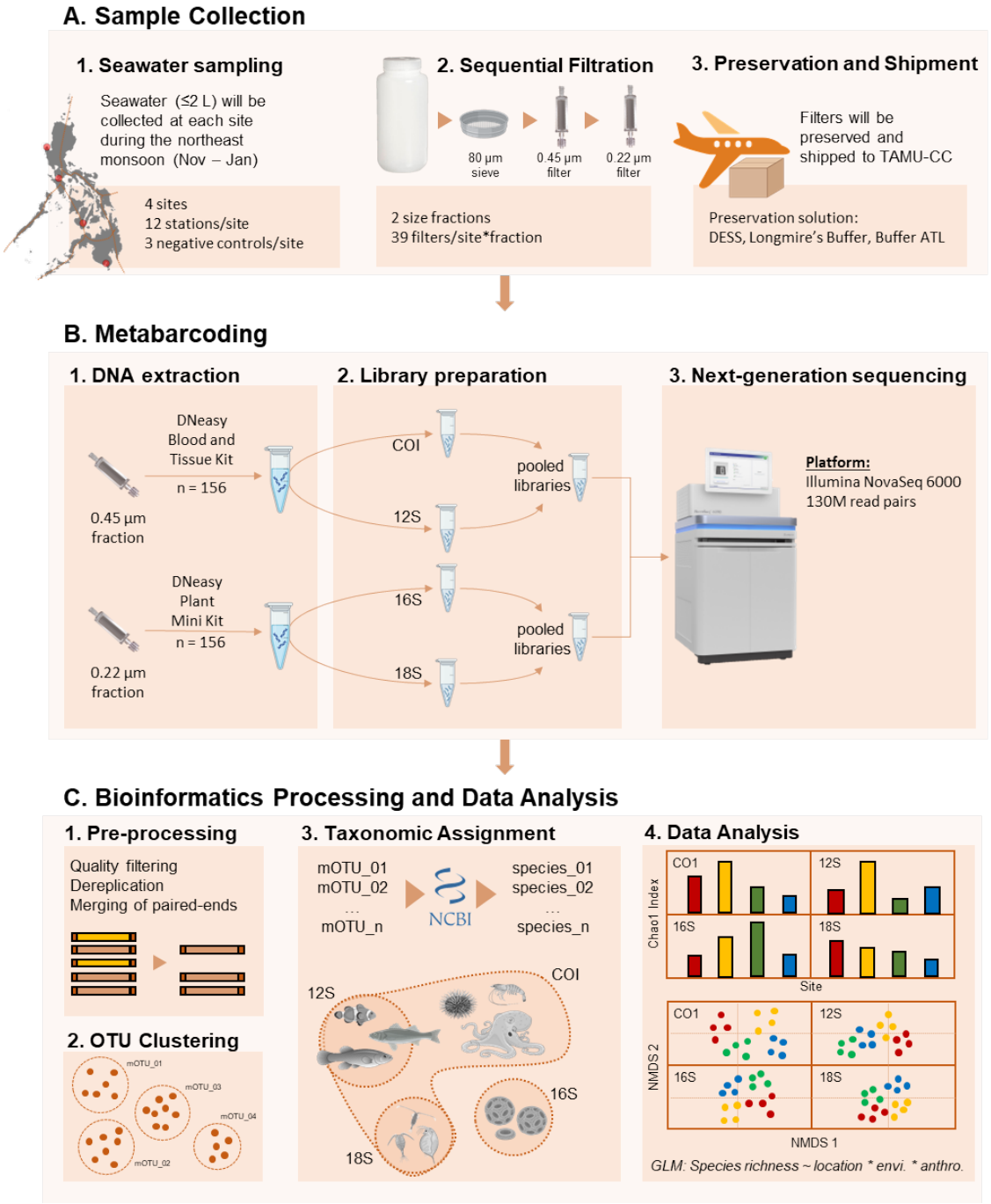
Appendix I. Materials and Methods



Appendix Figure 1. Overall schematics for the materials and methods. (A) Collected seawater samples will be sequentially filtered, preserved, and then shipped to Texas A&M University – Corpus Christi Genomics Core Laboratory (TAMU-CC GCL) for processing. (B) eDNA will be extracted and used as templates to amplify the target metabarcoding markers. The amplicon libraries will be pooled and then sequenced. (C) Pre-processed reads will be clustered based on a similarity threshold to obtain the molecular operational taxonomic units (mOTUs); these will be identified using a curated database, with each amplicon library targeting a specific taxonomic group. The read counts for each mOTU will then be used to test the hypotheses on patterns and drivers of marine biodiversity.

Invitation to Collaborate

Filipino researchers across four academic institutions were invited to collaborate on this project, each covering a sampling location: (1) University of the Philippines Diliman – The Marine Science Institute in Bolinao, Pangasinan; (2) University of the Philippines Mindanao in Sarangani Bay, (3) Batangas State University in the Verde Island Passage, and (4) Silliman University in Negros Oriental. In addition, the national agency tasked to conduct research and development for fisheries, the Department of Agriculture – National Fisheries Research and Development Institute (DA – NFRDI), was also invited to strengthen the partnership with the Philippine government. The research partnerships will be formalized with a Memorandum of Agreement for Educational and Scientific Cooperation which will stipulate the roles and responsibilities of each party, as well as the guidelines for data sharing and scientific output coming from this research grant.

Seawater Collection, Serial Filtration, and Preservation

There will be four sites in this study, each covering a different marine biogeographic region (Figure 1): (1) Bolino, Pangasinan in the West Philippine Sea (WPS), (2) Negros Oriental in the Visayas Region (VR), (3) and Sarangani in the Celebes Sea (CS); meanwhile, (4) the Verde Island Passage is a strait that connects WPS and VR. A total of 12 triplicate seawater samples (≤ 2 L; n = 36) will be collected from each site, and a negative field control (2 L sterile distilled water; n = 3) will be included to assess contamination[33,34]. To minimize seasonal variation, collection will only be done from November – January, which coincides with the northeast monsoon season. The water samples (n = 39/site) will be pre-filtered using 80 µm sieve into bleach-sterilized water containers to remove large particles (e.g., tissue debris, larvae), then serially filtered through 0.45 µm and 0.22 µm Sterivex cartridges using a sterile 300 mL syringe[34–36]. Filtration paraphernalia (e.g., silicone tubing, syringes) will be rinsed with 10% bleach and sterile distilled water before each use. Salt saturated DMSO buffer with EDTA (DESS), Longmire's Buffer, or Buffer ATL will be introduced to the cartridges for preservation[21,37]; the cartridges will then be sealed using Luer Lok cap, stored on ice, and then transported to a laboratory where it will be stored at -20 °C. The cartridges will be shipped to Texas A&M University – Corpus Christi Genomics Core Laboratory in dry ice for DNA extraction and sequencing.

DNA Extraction, Library Preparation, and Next-Generation Sequencing

eDNA will be extracted from 0.45 µm and 0.22 µm filter units using DNeasy Blood and Tissue Kit and DNeasy Plant Mini Kit, respectively[26,38], following the open Sterivex extraction method[35]. Briefly, DESS preservative will be flushed out from the Sterivex cartridge with a syringe, and the cartridge will be cut open by a PVC pipe cutter to allow removal of the filter. The filter will be placed in a Petri dish, cut into smaller pieces, and then placed in a fresh microtube for lysis. The remaining steps will be based following the manufacturer's protocol. All steps will be done aseptically under a laminar flow hood dedicated for eDNA extraction to minimize contamination. A negative laboratory control (ultrapure water) will also be processed alongside the samples.

Standard metabarcoding primers will be used to amplify target regions for each taxonomic group of interest (Appendix Table I). Library preparation will follow the Adapterama II protocol to allow pooling of libraries amplified using different primers[39], and sequencing of 2 x 250 paired-end reads will be done on Illumina NovaSeq 6000 platform[40].

Bioinformatics processing and data analysis

Bioinformatics pipeline will follow published protocols[35,41,42]. The raw sequences will be pre-processed by removing low quality reads and sequence pairs that do not contain the primer sequences. Forward and reverse reads will be merged in case they overlap, or will be concatenated otherwise; duplicate reads, singletons, chimeric sequences will then be removed. Pre-processed reads will be clustered based on a % similarity threshold (97 – 99%, depending on the marker) to identify the molecular operation taxonomic units (mOTUs). Taxonomic assignment will be done by querying a representative mOTU to a curated reference database; the identification will be done up to the species if possible. A

community matrix based on the read count of each mOTU for each sampling location will be generated for downstream analyses.

Data analysis will be done on R[43] following published analytical approaches[41], with the aid of the following packages: *vegan*[44], *ape*[45], and *tidyverse*[46]. The dataset will be partitioned based on the taxonomic groups recovered from each metabarcoding primer used. We will compare the alpha biodiversity indices (e.g., Chao1, Shannon, Simpson)[31] among the different biogeographic regions using multivariate mixed modeling[47]. We will use non-metric multidimensional scaling (NMDS) to visualize the differences in community composition among biogeographic regions, test for significant differences using Permutational Analysis of Variance (PERMANOVA), and then assess the effects of the explanatory variables of interest by fitting them in the ordination. The explanatory variables are: (1) geographical location, (2) sea surface temperature, (3) chlorophyll-a concentration, (4) population density, and (5) fishing effort. Where possible, ecological and anthropogenic variables will be retrieved from online databases such as the National Aeronautics and Space Administration's Earth Observing System Data and Information System (NASA - EOSDIS; https://oceancolor.gsfc.nasa.gov) and the CountryStat database of Philippine Statistics Authority (https://openstat.psa.gov.ph/Featured/CountrySTAT-Philippines).

Appendix Table I. Primers to amplify metabarcoding markers for various marine taxa.

Taxonomic Group	Target Marker	Primer Pair	Primer Sequence (5' – 3')	Reference
Fishes	12S (170 bp)	MiFish-U-F	GTCGGTAAAACTCGTGCCAGC	Miya et al., 2015
		MiFish-U-R	CATAGTGGGGTATCTAATCCCAGTTTG	
Eukaryotic Metazoa	COI (313 bp)	mlCOIintF-XT	GGWACWRGWTGRACWITITAYCCYCC	Wangensteen et al., 2018
		jgHCO2198	TAIACYTCIGGRTGICCRAARAAYCA	
Eukaryotic Plankton	18S rRNA (270 bp)	EukF	CCAGCASCYGCGGTAATTCC	Lin et al., 2019
		EukR	ACTTTCGTTCTTGAT	
Bacteria and Archaea	16S rRNA (290 bp)	515F	GTGYCAGCMGCCGCGGTAA	Lin et al., 2019
		805R	GACTACNVGGGTATCTAAT	