Title

Delineating Macro and Micro Marine Biodiversity in the Coral Triangle Using Autonomous Reef Monitoring Structures and DNA Metabarcoding

Author

Cahyani, Ni Kadek Dita

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Delineating Macro and Micro Marine Biodiversity in the Coral Triangle Using Autonomous Reef Monitoring Structures and DNA Metabarcoding

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biology

by

Ni Kadek Dita Cahyani

2021

ABSTRACT OF DISSERTATION

Delineating Macro and Micro Marine Biodiversity in the Coral Triangle Using Autonomous Reef Monitoring Structures and DNA Metabarcoding

by

Ni Kadek Dita Cahyani

Doctor of Philosophy in Biology

University of California, Los Angeles, 2021

Professor Paul Henry Barber, Chair

The exceptional concentration of marine biodiversity in the Coral Triangle is among the best-known biogeographic patterns in the ocean. Marine biodiversity peaks in the islands of Eastern part of Indonesia and the Philippines, the heart of the Coral Triangle, and significantly decreases moving away from this global biodiversity hotspot. However, data supporting this pattern largely come from fishes, corals and larger metazoans, and exclude smaller organisms that comprise the majority of marine biodiversity. This study utilized Autonomous Reef Monitoring Structure (ARMS) and DNA metabarcoding to examine biodiversity patterns of marine communities across Indonesia, the largest and most biologically diverse region of the Coral Triangle. In Chapter 1, I examine eukaryote biodiversity patterns of marine communities across Indonesia. Results demonstrate that smaller cryptofauna display similar biodiversity patterns to larger metazoans; the most diverse parts of Indonesia had more diversity per unit area, and greater heterogeneity and beta diversity across all spatial scales, individual ARMS, reefs, or regions. The results show that

processes shaping biodiversity hotspots appear consistent in marine and terrestrial ecosystems, and across size and spatial scales. In Chapter 2, I examine patterns marine bacterial diversity across Indonesia, comparing microbial diversity to eukaryotic and metazoan diversity from ARMS. Results showed strong regional differentiation in microbial communities. Microbial diversity tracked eukaryote and metazoan diversity, and displayed a significant pattern of isolation by distance, strongly indicating that associations with larger eukaryotes and physical limitations to dispersal differentiate microbial communities in the Coral Triangle. These results are counter to the hypothesis that "everything is everywhere, but the environment selects", and provide novel insights into the processes shaping marine microbial diversity in the world's most diverse marine ecosystem. In Chapter 3, I re-examine data from Chapter 1 to determine how strategies for marine ecosystem monitoring in Indonesia could be developed to yield the best results for the least cost, allowing resource managers to harness the power of metabarcoding to better monitor this region's biodiversity. Comparisons of cytochrome oxidase 1 (COI) and 18S rRNA metabarcoding data across three separate organismal size classes recovered from ARMS indicate that metabarcoding the 100 µm size fraction with COI captures the largest amount of diversity at the highest resolution. Results indicate that metabarcoding the 100 µm size fraction with COI provides the most accurate and economical approach to monitoring diversity in megadiverse regions where limited research investment precludes sequencing multiple size fractions with multiple metabarcoding markers. Combined, the results of this thesis demonstrate the power of ARMS and metabarcoding for the study and monitoring of marine biodiversity, providing new tools for the study and management of the exceptional marine biodiversity of the Coral Triangle.

The dissertation of Ni Kadek Dita Cahyani is approved.

Peggy Marie Fong

Thomas Bates Smith

Forest Rohwer

Paul Henry Barber, Committee Chair

University of California, Los Angeles,

2021

TABLE OF CONTENTS

Abstract …….. ii

List of Tables ……..……..…….. vii

List of Figures ……...……...…..…….. xii

Acknowledgments ……...……...……...…… xxii

Biographical Sketch ……...……...…….. xxvii

CHAPTER 1: DNA Metabarcoding Reveals Pronounced Biodiversity Gradients Benthic Marine Cryptofauna Across the Indonesian Archipelago ……..…..…….. 1

Abstract ……..……..….. 1

Introduction ……..…….. 2

Materials and Methods …….……..……..…….. 6

Results ……..……..…….. 14

Discussion ……..……..…….. 36

Supplemental Tables and Figures ……..…….. 46

References ……..…….. 77

CHAPTER 2: The Diversity of Indonesian Marine Bacteria: Assessing Baas Becking's Hypothesis "Everything Is Everywhere" Using Autonomous Reef Monitoring Structures and Metabarcoding …….……..... 89

Abstract ……..……..…….. 89

Introduction ……..……..……. 90

Materials and Methods …….……..…….. 95

Results ……..……..……..…….. 101

Discussion ……..……..…….. 114

Supplemental Tables and Figures ……..……..…….. 124

References ……..……..…….. 135

CHAPTER 3: Biodiversity Monitoring in High Diversity Marine Ecosystems ……..…….. 145

Abstract ……..……..….. 145

Introduction ……..…….. 146

Materials and Methods ……..……..……..…….. 150

Results ……..……..……..…….. 158

Discussion ……..……..……...…….. 170

Supplemental Tables and Figures ……..……..…….. 180

References ……..……..…….. 190

LIST OF TABLES

TABLES

Chapter 1.

Table 1-1. Operational Taxonomic Units (OTU) diversity from COI metabarcoding of autonomous reef monitoring structures spanning five sampled regions of Indonesia, including total diversity and diversity of three individual size fractions. Numbers in parentheses are diversity totals that include only OTUs unique to that size fraction (e.g. excluding any OTUs shared among size fractions).

Table 1-2. Operational Taxonomic Units (OTU) diversity based on 18S rRNA metabarcoding of autonomous reef monitoring structures spanning five sampled regions of Indonesia, including total diversity and diversity of three individual size fractions. Numbers in parentheses are diversity totals that include only OTUs unique to that size fraction (e.g. excluding any OTUs shared among size fractions).

Table 1-3. Beta diversity across five sampling locations across Indonesia obtained from analysis of COI and 18S metabarcoding data using the *Adespatial* package in R.

Chapter 2.

Table 2-1. Amplicon Sequence Variants (ASVs) diversity captured in a set of three ARMS based on 16S rRNA metabarcoding.

Table 2-2. Amplicon Sequence Variants (ASVs) diversity captured in a single ARMS based on 16S rRNA metabarcoding.

Table 2-3. Isolation by distance correlation (Mantel's *r* and correlation coefficient) between matrices of geographic distance and Jaccard dissimilarity distance among ARMS size fractions.

SUPPLEMENTAL TABLES

Chapter 1 Supplemental Tables.

Supplemental Table S1-1. Location and number of metabarcoding samples used on this study.

Supplemental Table S1-2. Total samples, sequence reads, and Operational Taxonomic Units (OTUs) for the COI datasets as revealed by DNA metabarcoding of Autonomous Reef Monitoring Structure (ARMS) from Indonesia.

Supplemental Table S1-3. Total samples, sequence reads, and Operational Taxonomic Units (OTUs) for the 18S rRNA datasets as revealed by DNA metabarcoding of Autonomous Reef Monitoring Structure (ARMS) from Indonesia.

Supplemental Table S1-4. Alpha diversity indices (ANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers, rarefied to a standardized number of OTUs per ARMS to account for variation in sequencing depth.

Supplemental Table S1-5. Tukey HSD post hoc tests pairwise comparisons on eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers.

Supplemental Table S1-6. Alpha diversity indices (ANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers, without rarefaction (e.g. raw OTU data).

Supplemental Table S1-7. Tukey HSD post hoc tests pairwise comparisons on eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers. The ANOVA were calculated without rarefication (e.g. raw data).

Supplemental Table S1-8. Alpha diversity indices (ANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers. Data were analysed per size fraction (100 μm, 500 μm, and sessile fraction) using the rarefied datasets to account for variation in sequencing depth.

Supplemental Table S1-9. Tukey HSD post hoc tests pairwise comparisons on eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers. Data were analyzed per size fraction (100 μm, 500 μm, and sessile fraction).

Supplemental Table S1-10. Alpha diversity indices (ANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers. Data were analysed per size fraction (100 μm, 500 μm, and sessile fraction). ANOVA were generated from rarefied data after excluding OTUs among size fraction to account for the possibility of carryover during ARMS sample processing.

Supplemental Table S1-11. Tukey HSD post hoc tests pairwise comparisons on eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers. Data were analysed per size fraction (100μm, 500μm, and sessile fraction). Post hoc tests were generated from rarefied data after excluding the shared OTUs between size fraction.

Supplemental Table S1-12. Beta diversity summary (PERMANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Cenderawasih Bay) in Indonesia for A) COI and B) 18S rRNA.

Supplemental Table S1-13. Beta diversity indices (PERMANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Cenderawasih Bay) in Indonesia. Analysis were calculated using total diversity per ARMS unit (e.g. summing diversity across all size fractions from a single ARMS).

Supplemental Table S1-14. Beta diversity indices (PERMANOVA) of eukaryote diversity based on site and ARMS within locations (Aceh, Pulau Seribu, Bali, Raja Ampat and Cenderawasih Bay) in Indonesia for A) COI and B) 18S rRNA.

Chapter 2 Supplemental Tables.

Supplemental Table S2-1. Location and number of metabarcoding samples used on this study.

Supplemental Table S2-2. Total samples, sequence reads, and Amplicon Sequence Variants (ASVs) for the 16S rRNA datasets as revealed by DNA metabarcoding of Autonomous Reef Monitoring Structure (ARMS) from Indonesia.

Supplemental Table S2-3. Alpha diversity indices (ANOVA) of microbial diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using 16S rRNA markers, rarefied to a standardized number of Amplicon Sequence Variants (ASVs) per ARMS to account for variation in sequencing depth.

Supplemental Table S2-4. Tukey HSD post hoc tests pairwise comparisons on microbial diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using 16S rRNA.

Supplemental Table S2-5. Alpha diversity indices (ANOVA) of microbial diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using 16S rRNA markers. Data were analysed per size fraction (100 μm, 500 μm, and sessile fraction) using a rarefied datasets to account for variation in sequencing depth.

Supplemental Table S2-6. Tukey HSD post hoc tests pairwise comparisons on microbial diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using 16S rRNA.

Supplemental Table S2-7. Tukey HSD post hoc tests pairwise comparisons on microbial diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using 16S rRNA markers. Post hoc test were generated from rarefied data after excluding Amplicon Sequence Variants (ASVs) shared among size fraction.

Supplemental Table S2-8. Beta diversity summary (PERMANOVA) of microbial diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Cenderawasih Bay) in Indonesia using 16S rRNA.

Chapter 3 Supplemental Tables.

Supplemental Table S3-1. Location and number of metabarcoding samples used on this study.

Supplemental Table S3-2. Total samples, sequence reads, and Operational Taxonomic Units (OTUs) for the COI datasets as revealed by DNA metabarcoding of Autonomous Reef Monitoring Structure (ARMS) from Indonesia.

Supplemental Table S3-3. Total samples, sequence reads, and Operational Taxonomic Units (OTUs) for the 18S rRNA datasets as revealed by DNA metabarcoding of Autonomous Reef Monitoring Structure (ARMS) from Indonesia.

Supplemental Table S3-4. Alpha diversity indices (ANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using A) COI and B) 18S rRNA.

Supplemental Table S3-5. Beta diversity summary (PERMANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia for A) COI and B) 18S rRNA.

Supplemental Table S3-6. Beta diversity indices (PERMANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia. Analysis were calculated from total diversity per ARMS (e.g. summing diversity of all 3-set of fraction per ARMS.

Supplemental Table S3-7. Operational Taxonomic Units (OTU) richness across from five regions of Indonesia from A) COI and B) 18S rRNA metabarcoding data, including total OTU diversity, diversity of the 100 µm fraction, 500 µm fraction, and sessile fractions.

Supplemental Table S3-8. Unique Operational Taxonomic Units (OTU) richness across from five regions of Indonesia from A) COI and B) 18S rRNA metabarcoding data, including total unique OTU diversity, unique diversity of the 100 µm fraction, 500 µm fraction, and sessile fractions.

Supplemental Table S3-9. Operational Taxonomic Units (OTU) diversity captured in a set of three ARMS based on A) COI and B) 18S rRNA metabarcoding.

Supplemental Table S3-10. Operational Taxonomic Units (OTU) diversity captured in a single ARMS based on A) COI and B) 18S rRNA metabarcoding.

LIST OF FIGURES

FIGURES

Chapter 1.

Figure 1-1. A) Autonomous Reef Monitoring Structure (ARMS) structure photographed underwater. B) ARMS plate colonized by benthic marine organisms. C) Map of the Coral Triangle with five sampling locations: (1) Pulau Weh, Aceh, (2) Kepulauan Seribu, Jakarta, (3) Pemuteran, Bali, (4) Raja Ampat, West Papua, and (5) Teluk Cenderawasih, West Papua. The Coral Triangle Scientific Boundary (red line) is based on Veron et al. (2009).

Figure 1-2. Rarefaction plots showing numbers of Operational Taxonomic Units (OTUs) as a function of sequencing depth for the five regions based on A) COI and B) 18S rRNA, as well for each individual ARMS unit for C) COI and D0 18S rRNA.

Figure 1-3. Taxonomic composition of eukaryote communities identified across 100 μm, 500 μm, and sessile ARMS sample fractions, spanning five regions of Indonesia based on A) relative abundance of sequence reads and C) numbers of OTUs (Operational Taxonomic Units) based on COI, and B) relative abundance of sequence reads and D) numbers of OTUs based on 18S rRNA, excluding all taxa with <2% relative abundance.

Figure 1-4. Average per site OTU diversity captured in a single ARMS (Autonomous Reef Monitoring Structure) unit across all sites for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Figure 1-5. Total Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Figure 1-6. Percentage of total regional diversity captured in an individual ARMS unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Figure 1-7. Total Operational Taxonomic Units (OTUs) diversity captured in a single sampling site across five sampling regions across Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Figure 1-8. Percentage of regional diversity captured in an individual sampling site across five sampled regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Figure 1-9. Total endemic Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Figure 1-10. Total endemic Operational Taxonomic Units (OTUs) diversity captured per sampling site across five sampling regions across Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Figure 1-11. Total Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit in the A) 100 μm B) 500 μm and C) Sessile size fractions across five sampled regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers. Scale on Y axes vary.

Figure 1-12. Total Operational Taxonomic Units (OTUs) diversity captured per sampling site in the A) 100 μm B) 500 μm and C) Sessile size fractions across five sampling regions across Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers. Scales on y axes vary.

Figure 1-13. Total endemic Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit in the A) 100 μm B) 500 μm and C) Sessile size fractions across five sampled regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers. Scale on y axes vary.

Figure 1-14. Total endemic Operational Taxonomic Units (OTUs) diversity captured per sampling site in the 100 μm, 500 μm and sessile size fractions across five regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers. Scales on y axes vary.

Figure 1-15. Principal Coordinates Analysis (PCoA) analysis illustrating dissimilarities in eukaryote community composition across 59 ARMS units deployed across the Indonesia archipelago based on COI (A and B) and 18S rRNA (C and D) using Jaccard and Bray-Curtis similarities.

Figure 1-16. Unweighted Pair Group Method with Arithmetic mean (UPGMA) clustering analysis with heatmap and dendrogram illustrating dissimilarities in marine benthic eukaryote community composition based on Jaccard similarity of OTU diversity from 59 ARMS from 5 regions of Indonesia for both A) COI and B) 18S rRNA.

Figure 1-17. Ternary plots of Jaccard similarity and the partitions of beta diversity (replacement, richness difference, and Jaccard similarity) for 59 ARMS (Autonomous Reef Monitoring Structure) deployed across Indonesia for A) COI and B) 18S rRNA metabarcoding.

Chapter 2.

Figure 2-1. Theoretical size threshold determining ubiquity (modified from Finlay, 2002), including where ARMS size fractions fall along the size threshold.

Figure 2-2. A) Autonomous Reef Monitoring Structure (ARMS) photographed underwater. B) ARMS plate colonized by organisms. C) Map of the Coral Triangle and sampling locations: (1) Pulau Weh, Aceh, (2) Kepulauan Seribu, Jakarta, (3) Pemuteran, Bali, (4) Raja Ampat, West Papua, and (5) Teluk Cenderawasih, West Papua. The Coral Triangle Scientific Boundary (red line) is based on Veron et al. (2009).

Figure 2-3. Alpha diversity rarefaction plot of full dataset generated with Ranacapa (Kandlikar et al., 2018) and iNEXT packages (Chao et al., 2014; Hsieh et al., 2016) in R (R development core team). Number of amplified sequence variants (ASVs) (left axis) plotted against sequencing depth (bottom axis) for A) each individual ARMS unit, and B) each of the five regions.

Figure 2-4. Taxonomic composition of eukaryote diversity at Phylum level for A) microbial communities for entire ARMS units, and B) for each of the three size fractions. Bar plot showing taxa relative abundance of the sample across five different location in Indonesia. The bar plot constructed based on phyla contribute more than 2% of the relative abundance of each sample.

Figure 2-5. Microbial diversity patterns from autonomous reef monitoring structures across the Indonesian archipelago, including A) the average diversity at a site captured in a single ARMS unit, B) the average regional diversity captured in a single sampling site, C) ASV richness per ARMS, D) endemic ASV richness per ARMS, E) ASV richness per site and F) endemic ASV richness per site based on 16S rRNA metabarcoding.

Figure 2-6. Boxplot showing the diversity indices (Chao1 and Shannon) of microbial community composition across Indonesia. Analyses include A) the full dataset, and the different fractions, B) 100 µm–500 µm, C) 500 µm-2 mm, and D) sessile fraction. Black diamonds represent mean alpha diversity from each location. Colored boxes represent the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The letters at the top of the box are the results of Tukey test of multiple comparisons.

Figure 2-7. Plot of microbial diversity as a function of eukaryote/metazoan diversity based on A) total COI ASVs and microbial ASVs, B) endemic COI ASVs and endemic microbial ASVs, D) total 18S rRNA ASVs and microbial ASVs, D) endemic 18S rRNA ASVs and endemic microbial ASVs, E) metazoan 18S rRNA ASVs and microbial ASVs, D) endemic metazoan 18S rRNA ASVs and endemic microbial ASVs, including best fit line and R^2 values, all of which are significant.

Figure 2-8. Principal Coordinates Analysies (PCoA) illustrating dissimilarities in microbial community composition across Indonesia caculated on total microbial diversity for individual ARMS unit (e.g. summed across all three size fractions). PCoA was performed using Bray-Curtis and Jaccard similarity on the A) full dataset, and B) the dataset without shared amplified sequence variants (ASVs) across different locations.

Figure 2-9. Principal Coordinates Analysis (PCoA) analysis illustrating dissimilarities in microbial community composition across Indonesia. Shared amplified sequence variants (ASVs) between size fraction were excluded from this dataset and were rarefied even depth to 1,038 reads per samples. Analyses using Bray-Curtis and Jaccard similarity were undertaken on the different fractions (106–500 µm, 500 µm - 2mm, and sessile) across the five sampling regions.

Figure 2-10. Isolation by distance correlation (Mantel's *r*) between matrices of the natural log of geographic distance and Jaccard community dissimilarity distance among ARMS size fractions: A) 100 µm, B) 500 µm, and C) sessile.

Figure 2-11. Number and distribution of microbial ASVs revealed from16S rRNA metabarcoding of 100 µm, 500 µm, and sessile size fractions obtained from autonomous reef monitoring structures from across the Indonesian archipelago. Plot represents data rarefied to an even depth of 36,719 reads per ARMS unit.

Chapter 3.

Figure 3-1. A) ARMS structure photographed underwater. B) Autonomous Reef Monitoring Structure (ARMS) plate colonized by organisms. C) Map of the Coral Triangle with five sampling locations: (1) Pulau Weh, Aceh, (2) Kepulauan Seribu, Jakarta, (3) Pemuteran, Bali, (4) Raja Ampat, West Papua, and (5) Teluk Cenderawasih, West Papua. The Coral Triangle Scientific Boundary (red line) was based on Veron et al. (2009).

Figure 3-2. Rarefaction plot showing numbers of Operational Taxonomic Units (OTUs) as a function of sequencing depth for the five regions based on A) COI and B) 18S rRNA, as well as for each individual ARMS unit for C) COI and D) 18S rRNA.

Figure 3-3. Venn diagram showing overlap in Operational Taxonomic Units (OTUs) among the 500µm, 100µm, and sessile fractions based on A) COI and B) 18S rRNA.

Figure 3-4. Boxplot of Chao1 and Shannon diversity indices of eukaryote communities based on A) COI and B) 18S rRNA summarized across each of three Autonomous Reef Monitoring Structure (ARMS) fractions. The diamond shapes represent the mean of alpha diversity from each fraction size. The box in the boxplot represent the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with the points outside considered outliers. The letters at the top of the box are the results of Tukey test of multiple comparisons.

Figure 3-5. Taxonomic composition of eukaryote communities identified across 500μm, 100μm, and sessile sample fractions across five regions of Indonesia based on A) relative abundance of sequence reads and C) numbers of OTUs (Operational Taxonomic Units) based on COI, and B) relative abundance of sequence reads and D) numbers of OTUs based on 18S rRNA, excluding all taxa with <2% relative abundance.

Figure 3-6. Numbers of unique and shared Operational Taxonomic Units (OTUs) among five sampled regions of Indonesia based on A) COI and B) 18S rRNA.

Figure 3-7. Boxplot showing the diversity indices (Chao1 and Shannon) of eukaryote community composition from 59 ARMS units across five regions of Indonesia. The diamond shapes represent the mean of alpha diversity from each location. The box in the boxplot represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with point outside consider outliers. The letters at the top of the box are the results of Tukey test of multiple comparisons.

Figure 3-8. Principal Coordinates Analysis (PCoA) analysis illustrating dissimilarities in eukaryote community composition from 59 ARMS representing 5 regions of Indonesia. Analysis was undertaken using Jaccard and Bray-Curtis similarities on the full dataset of COI (A and B) and 18S rRNA (C and D) across different all sampling locations.

Figure 3-9. Operational Taxonomic Units (OTU) richness across five regions of Indonesia for 18S rRNA (red) and COI (blue). Solid lines are total OTU diversity, while dashed lines are OTUs that are unique to a single region.

Figure 3-10. Operational Taxonomic Units (OTU) richness by fraction across five regions of Indonesia for 18S rRNA (solid line) and COI (dashed line).

SUPPLEMENTAL FIGURES

Chapter 1 Supplemental Figures.

Supplemental Figure S1-1. Average site diversity captured in a single ARMS (Autonomous Reef Monitoring Structure) unit across five sampling regions for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.

Supplemental Figure S1-2. Total Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.

Supplemental Figure S1-3. Percentage of regional diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.

Supplemental Figure S1-4. Total Operational Taxonomic Units (OTUs) diversity captured in a single sampling site across five sampling regions across Indonesia for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.

Supplemental Figure S1-5. Percentage of regional diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.

Supplemental Figure S1-6. Total endemic Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.

Supplemental Figure S1-7. Total endemic Operational Taxonomic Units (OTUs) diversity captured in a single sampling site across five sampling regions across Indonesia for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.

Supplemental Figure S1-8. Total Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit in the A) 100 μm B) 500 μm and C) Sessile size fractions across five sampled regions of Indonesia for A) COI and B) 18S rRNA. Box plots were generated from rarefied data after excluding OTUs shared among size fractions.

Supplemental Figure S1-9. Total Operational Taxonomic Units (OTUs) diversity captured in a single sampling site in the A) 100 μm B) 500 μm and C) Sessile size fractions across five sampling regions across Indonesia for A) COI and B) 18S rRNA. Box plots were generated from rarefied data after excluding OTUs shared among size fractions.

Supplemental Figure S1-10. Total endemic Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit in the A) 100 μm B) 500 μm and C) Sessile size fractions across five sampled regions of Indonesia for A) COI and B) 18S rRNA. Box Box plots were generated from rarefied data after excluding OTUs shared among size fractions.

Supplemental Figure S1-11. Total endemic Operational Taxonomic Units (OTUs) diversity captured in a single sampling site in the A) 100 μm B) 500 μm and C) Sessile size fractions across five sampling regions across Indonesia for A) COI and B) 18S rRNA. Box plots were generated from rarefied data after excluding OTUs shared among size fractions.

Supplemental Figure S1-12. Ternary plots of Jaccard similarity and the partitions of beta diversity (replacement, richness difference, and Jaccard similarity) for the 100 μm, 500 μm, and sessile fraction obtained from the COI metabarcoding. Ternary plots are shown for the total experiment as well as between sites and between locations. The mean values are represented by numbers in the bigger circle.

Supplemental Figure S1-13. Ternary plots of Jaccard similarity and the partitions of beta diversity (replacement, richness difference, and Jaccard similarity) for the 100 μm, 500 μm, and sessile fraction obtained from the 18S rRNA metabarcoding. Ternary plots are shown for the total experiment as well as between sites and between locations. The mean values are represented by numbers in the bigger circle.

Chapter 2 Supplemental Figures.

Supplemental Figure S2-1. Taxonomic composition of 7,082 shared microbial amplified sequence variants (ASVs) between 500 µm - 2 mm, 106-500 µm, and sessile size fraction. Bar plots showing taxa relative abundance of different size fraction from the sample across five different location in Indonesia. The bar plots were constructed based on phyla contribute more than 2% of the relative abundance of each sample.

Supplemental Figure S2-2. Boxplots showing the microbial diversity indices (Chao1 and Shannon) across Indonesia. Analysis was undertaken on A) full dataset after removing the ASVs shared among the different fractions B) 106-500 µm, C), 500 µm - 2 mm, and D) Sessile fraction. The diamonds represent mean alpha diversity from each location, the box represents the 1st and 3rd quartiles and the vertical line is the median of the dataset. The letters at the top of the box are the results of Tukey test of multiple comparisons.

Supplemental Figure S2-3. Principal Coordinates Analysis (PCoA) analysis illustrating dissimilarities in microbial community composition across Indonesia. Shared amplified sequence variants (ASVs) between size fraction were excluded from this dataset and were rarefied even depth to 1,038 reads per samples. Analysis was undertaken using Bray-Curtis and Jaccard similarity on the different fractions (106-500 µm, 500 µm-2 mm, and Sessile) across all sampling locations.

ACKNOWLEDGMENTS

This whole Ph.D. journey is a significant life experience that I will forever cherish and grateful for. I dedicated these pages to my mentors, friends, family, and those who supported and encouraged me throughout this amazing journey.

First, I would like to express my gratitude and appreciation for my advisor, Paul Barber. I do not have many things to say, except "*Terima Kasih*." Thank you for being an amazing mentor, not only in the academic setting but also in the research setting as a whole. Thank you for believing in me, and motivating me to be a better researcher. Thank you for all the experiences and knowledge and for teaching me that family is always number one. "*Terima kasih Pak Paul dan keluarga*".

Second, I would like to acknowledge my committee members, Peggy Fong and Thomas Smith from UCLA, and Forest Rohwer from San Diego State University. Thank you for the guidance you have provided throughout my research and manuscript writing. Your comments and feedback have enriched the manuscript. I hope this study will be a tremendous addition to biodiversity research in Indonesia.

Third, I want to say thank you to Christopher Meyer from the Smithsonian Institution. Thank you for hosting me while doing my lab work there. Washington DC has always offered a different vibe from Los Angeles. The fact that I was always there during winter is perfect. I had a great time working at the Smithsonian National History Museum's LAB (the Laboratories of Analytical Biology). Working there is a dream come true. Ever since my undergrad years, I always wondered how it would feel to be working inside the museum, and now I have a bit of that taste. Thank you for all the people there, the LAB members, researchers, post-docs, and staff

who were always there when I need some enlightening. I especially want to thank Jordan Casey, who patiently taught me how to do the Next Generation Sequencing lab work.

This whole Ph.D. journey will not happen without significant support from the Ecology and Evolutionary biology (EEB) Department of UCLA. Jocelyn Yamadera, Tessa Villaseñor, and Melissa Carrillo. Thank you for helping me settle in to campus life, working with all the paperwork, and assuring me that everything would be okay when I hit some deadlines. I also want to say thank you to the EEB community and my cohort for all the great experiences.

This achievement accumulates a significant research partnership, friendships, love, and support from so many people, even before I decided to go for a Ph.D. Many people have been my inspiration throughout the year, and I am thankful I still call most of them my friend. Thank you.

Thank you for the amazing team from Yayasan Biodiversitas Indonesia (BIONESIA), Bali, Indonesia. Andrianus Sembiring, Ni Putu Dian Pertiwi, Astria Yusmalinda, Yuliana Syamsuni, Eka Maya Kurniasih, Danie el Malik, and Enex Yuli Artiningsih. You are the best team to work with. We have been doing it for more than ten years, and I am looking forward to many fun years ahead. Do not give up and keep doing good science.

And, of course, my biggest thank you is for the Barber Lab members. Ten years ago, I was introduced to the Barber lab members in Indonesia, Rita Rachmawati, Samantha Cheng, Sara Simmons, Allison Fritts-Penniman, and Abril Iñiguez (and family). All of your journeys to Bali-Indonesia inspired me to take the same path as yours; I am so thankful for the friendships. For my partner in crime, Aji Wahyu Anggoro, I don't know what to say. Just thank you for being you. Often you frustrated me with all the crazy ideas in your head, but you were always be there for me, so thank you. Zack Gold, thank you for your constant assistance, especially for the

ARMS samples back in 2016. Kelcie Chiquillo, my seagrass queen, hope we can meet again in Bali and do more fun stuff together. Eric Caldera, you have been a great resource in improving my knowledge in the microbial world; thank you for all your advice, comments, and input for my microbial chapter; that's been a great help. Sam Degregori, Erick Zerecero, Onny Marwayana, Candice Cross, Satoshi Tomano, and all current lab members, thank you for the friendship. I hope we can meet in Bali.

Thank you to my fellow Indonesian students and family in LA; thank you for keeping me entertained with all those road trips and the great family pictures. Krama Bali Los Angeles and the extended family; these people are my Balinese/Indonesian family in LA. They invited me into their family, embraced me so warmly, helped me ease down my longing for home, and made me realize that family is not always blood-related.

Thank you to my roommates, Dian Tri Irawaty and Novia Kusumayani, and the ladies who keep us busy, Fika Nikendary, Indri Sukmaputri, and Farah Anissa. The six of us have different personalities, but Los Angeles has glued us together as good friends Thank you for the excellent food, movies, laughter, the girl trips, and all those unimportant yet invigorating talks.

Thank you to my husband for letting me spread my wings. I know this was not an easy decision to make. Still, your encouragement is the biggest reason I agreed to go on this journey. For my son, I Putu Padmabumi (Bumi), thank you for always keeping me grounded. I hope you are proud of me, as I am always proud of you. Thank you to my family, my parents, brothers, and sisters who keep supporting me all this time.

Furthermore I would like to thank the rest of the graduate and undergraduate research team and their supervisors for their collaborative effort during data collection across Indonesia. Lembaga Ilmu Pengetahuan Indonesia (LIPI): Irma Arliza, Ismiliana Wirawati and Dedy

Kurnianto. Universitas Udayana, Bali: Eloq Faiqoh, I Gede Budi Astrawan, Budi Santoso, Putu Satya Pratama Atmaja, I Gusti Ayu Ricca Mahatma Putri, I Gusti Ngurah Agung Dhananjaya, Febriyanto Arifin, Dika Madyawan, Ratih Permitha Syury, Luh Putu Puspita Dewanti, Annassita Gianie. Universitas Negeri Papua: Suparno and Jenly Haurissa. Institut Pertanian Bogor (IPB): Hawis Madduppa, Beginer Subhan, Samsul Bahri, Ahmad Taufik Ghozali, Gensten Hazery, Mutia Ramadhaniaty, La Ode Abdul Fajar Hasidu, M. Andre Nugraha, M. Ismatullah Jay, Muhammad Fatoni Ranchman, Destia Handayani, Nia Amanda, Susi Awaliyah, Fauzan Dzulfanazhir, Jordan Tito Lubis. Universitas Syah Kuala, Aceh: Muh. Fadli, Muhammad Tawakkal, Fadhlurahman, Afrita Ida Utami, Rizki Syahputra, Aprilina Saragih. Muhammad Dailami, Ayu indah Lestari, Dandi Saleky, Yunus Baab, Rimer H Biloro, Andri Wahyu Kuncoro, Masriana, and everybody that I may not have mentioned here.

This work was funded by the National Science Foundation with a Partnerships for International Research and Education (PIRE) grant (OISE-1243541), entitled "Assembly of Marine Biodiversity Along Geographic and Anthropogenic Stress Gradients". I would like to say thank you for the PI, co-PI and all the collaborators on this grant; Forest Rohwer (San Diego State University), Paul Barber (University of California, Los Angeles), Jonathan Geller (Moss Landing Marine Laboratories and San Jose State University), Nancy Knowlton, Chris Meyer, and Allen Collins (The Smithsonian Institution, National Museum of Natural History), Russell Brainard, and Molly Timmers (NOAA Pacific Islands Fisheries Science Center), The National Evolutionary Synthesis Center (NESCent), I Gusti Ngurah Kade Mahardika (Udayana University), Ambariyanto (Diponegoro University), Hamid Toha (State University of Papua), Mark Erdmann (Conservation International), and Dr. Ir. Zainal Arifin (Indonesian Institute of Sciences or LIPI). Thank you as well as the Laboratory of Marine Molecular Genetics, Research

Center for Oceanograpy, Indonesian Institute of Sciences, Jakarta, Indonesia for providing us with laboratory space and equipment to do DNA extractions for this study. I also want to thank Emma Ransome, and Aaron Hartman for giving me so much important input for my research and analysis. Thank you, Gerry Allen, from the Western Australia Museum for providing me with data on reef fish biodiversity.

I am forever grateful for the Fulbright Scholarship. I never thought I would be a part of the Fulbright family. This opportunity has given me such a fantastic journey, meeting many wonderful people and travel across states and countries. I have learned a lot and will keep learning new things. Thank you for the American Indonesian Exchange Foundation (AMINEF) Indonesia and the Institute of International Education (IIE). Thank you to all the mentors from the American Language Institute (ALI) of San Diego State University for the best three weeks before we head to our campus and start our classes.

Finally, I thank the following agencies for the financial support that made my dissertation research possible. Thank you for the Center for Southeast Asia Study (CSEAS), UCLA, for providing a travel grant that funded my sampling activities to Indonesia. I also thank the UCLA EEB Department for providing travel grants that funded various conferences and field works.

BIOGRAPHICAL SKETCH

Previous Degrees Awarded

Universitas Gadjah Mada, Yogyakarta, Indonesia – B.S. in Biology. Awarded 2003

Universitas Udayana, Denpasar, Bali, Indonesia – M.S. in Environmental Study. Awarded 2007

Awards and Grants

2018 - Indonesian Studies Travel Grant from the UCLA Center for Southeast Asian Studies

2017 - Women in Marine Science in Indo-Pacific travel grant. The 10th Indo-Pacific Fish.

Conference, Papeete, Tahiti, French Polynesia. October 2-6, 2017.

2016 - Indonesian Studies Travel Grant from the UCLA Center for Southeast Asian Studies

2015 - Fulbright Presidential Scholarship program (Ph.D) (2015-2018)

Publications

1. Winterbottom, R., Erdmann, M. V, & Cahyani, N.K.D. (2015). New species of Trimma (Actinopterygii, Gobioidei) from Indonesia, with comments on head papillae nomenclature. Zootaxa 3973 (2): 201-226.

2. Sembiring, A., Pertiwi, N.P.D., Mahardini, A., Wulandari, R., Kurniasih, E.M., Kuncoro, A.W., Cahyani, N.K.D., Anggoro, A.W., Ulfa, M., Madduppa, H., Carpenter, K., Barber, P.H. and Mahardika, G.N. (2014). DNA Barcoding reveals targeted fisheries for endangered sharks in Indonesia. Fisheries Research. 164 (2015). 130-134

3. Winterbottom, R., Erdmann, M. V, & Cahyani, N.K.D. (2014). Three new species of Trimma (Pisces; Gobioidei) from Indonesia. Zootaxa 3838 (3): 367-384.

4. Winterbottom, R., Erdmann, M. V, & Cahyani, N.K.D. (2014). Trimma helenae (Pisces; Gobioidei), a new species of gobiid fish from Indonesia. Zootaxa 3760 (3): 420–428.

5. Barber, P. H., Ablan-lagman, M. C. A., Cahyani, D., Crandall, E. D., Ravago-gotanco, R., Juinio-meñez, M. A., Mahardika, I.G.N., Shanker, K., Starger, C.J., Toha, A.H.A., Anggoro, A.W. and Willette, D.A. (2014). Advancing biodiversity research in developing countries : the need for changing paradigms. Buletin of Marine Science 90(1). 2014 Rosenstiel School of Marine & Atmospheric Science of the University of Miami

Conference and Presentations

1. September 11th, 2020. The 1st International Conference on Biotechnology and Food Sciences (INCOBIFS). Faculty of Fisheries ad Marine, Universitas Airlangga, Surabaya. (Invited Speaker). *Molecular Genetic for Indonesian Marine Biodiversity*.
2. July 14, 2019 - July 19, 2019. Gordon Research Conference. HKUST, HK (Poster session). *Employing the power of metabarcoding and autonomous reef monitoring structures to assess Indonesian marine macro and micro biodiversity*
3. August 3rd, 2018. Faculty of Fisheries and Marine Sciences, Udayana University, Bali, Indonesia (Speaker). *General Lecture on Indonesian Marine Diversity, Introducing ARMS (Autonomous Reef Monitoring Structures as a standardized method for monitoring Indonesian marine biodiversity in Indonesia*
4. July 18th, 2018. Faculty of Marine Sciences, Diponegoro University, Semarang, Indonesia (Speaker). *Autonomous Reef Monitoring Structures (ARMS): A standardized method for monitoring Indonesian marine biodiversity in Indonesia*
5. May 23rd, 2018. 21st Annual Biology Research Symposium. Dept. of Ecology and Evolutionary Biology, UCLA (Poster Session). *Autonomous Reef Monitoring Structures (ARMS): A standardized method for monitoring Indonesian marine biodiversity in Indonesia*
6. May 13th, 2018. The 4th World Conference on Marine Biodiversity. Montreal, Canada (Presenter). *Metabarcoding on Autonomous Reef Monitoring Structures (ARMS): A standardized method for monitoring Indonesian marine biodiversity in the 'omicx era*
7. October 2nd-6th, 2017. The 10th Indo-Pacific Fish Conference, Papeete, Tahiti, French Polynesia. Presenter for Women in Marine Science in Indo-Pacific Session

CHAPTER 1

DNA Metabarcoding Reveals Pronounced Biodiversity Gradients Benthic Marine Cryptofauna Across the Indonesian Archipelago

Abstract

The exceptional concentration of marine biodiversity inside the Coral Triangle is among the best-known biogeographic patterns in the ocean. However, data supporting this pattern largely come from fishes, corals and larger metazoans, and exclude smaller organisms that comprise the majority of marine biodiversity. This study utilized Autonomous Reef Monitoring Structure (ARMS) and DNA metabarcoding to examine biodiversity patterns of marine communities across Indonesia, the largest and most biologically diverse region of the Coral Triangle. Using metabarcoding data from COI and 18S rRNA in a geographically nested design across size-fractionated communities, we demonstrate that encrusting and smaller cryptofauna display similar biodiversity patterns to larger metazoans. The most diverse parts of the Coral Triangle had more diversity per unit area, and greater heterogeneity and beta diversity across all spatial scales, similar to patterns from terrestrial biodiversity hotspots. Additionally, patterns were consistent across size fractions, suggesting that the processes structuring diversity in this region act broadly across the diversity of life and organism sizes. Surprisingly, the smallest organisms (106-500 μm) always displayed the strongest patterns across all metrics of diversity examined and across all spatial scales, results counter to the "everything is everywhere" hypothesis. Given that biodiversity patterns of small cryptofauna parallel larger marine metazoans, and that the packing of this diversity is similar to terrestrial biodiversity hotspots, results suggest that processes shaping biodiversity hotspots are likely similar in marine and terrestrial ecosystems, and across size and spatial scales.

Introduction

Tropical marine biodiversity studies largely focus on macrofauna such as fishes, corals, and molluscs, because they are large, conspicuous, and relatively well-known taxonomically (Bellwood, 2001; Bellwood et al., 2005; Mustika et al., 2012; Mustika et al., 2013). However, it is unknown whether diversity patterns in these taxa are representative of all marine biodiversity, the majority of which are "cryptofauna", small, cryptic marine species that are largely undocumented (Knowlton et al., 2010; Plaisance et al., 2011). In limited studies that compare larger metazoans to smaller taxa like cowries, foraminifera, and euphasiids, results show varied degrees of concordance (e.g. Bellwood & Meyer, 2009; Tittensor et al., 2010), suggesting that biodiversity patterns derived from larger metazoans may not be representative of marine biodiversity, limiting our understanding of both the patterns and processes shaping marine biodiversity.

Often described as the "Amazon of the Oceans", the Coral Triangle is defined by the presence of >500 hard coral species (Veron et al., 2009), and is the global epicenter of marine biodiversity. Spanning only 6 million km^2 and containing less than 30% of global coral reef area, the Coral Triangle is home to 76% of all species of scleractinian corals and 37% of the world's reef fish species 8% of which are endemic or locally restricted species (Allen, 2008; Veron et al., 2009). Indonesia is both the largest geographic area of the Coral Triangle and the most diverse; it is the center of a biodiversity "bullseye" with sharply decreasing biodiversity gradients with increasing distance from this region (Roberts et al., 2002; Bellwood et al., 2005; Bellwood & Meyer 2009).

Although the biodiversity gradients that define the Coral Triangle biodiversity hotspot have been known for decades (Ekman, 1953; Ladd, 1960; Woodland, 1983; Woodland, 1986)

the mechanisms driving this pattern are still vigorously debated (Bowen et al., 2013; Barber and Meyer, 2015). The Coral Triangle is described as an evolutionary source of biodiversity (a "Center of Origin"; Ekman 1953), or an accumulation of diversity that evolved in peripheries of the Indian and Pacific Oceans ("Center of Accumulation"; Ladd 1960; Kay, 1984; Jokiel & Martinelli, 1992). Other studies support neither of these hypotheses (Barber & Bellwood, 2005; Halas & Winterbottom, 2009), suggesting instead that these biodiversity gradients might result from a "mid-domain effect" (Connoly et al, 2003; Bellwood et al., 2005) or have pluralistic origins (Barber, 2009; Gaither et al., 2013; Barber & Meyer, 2015).

Neglected in this debate is the fact that marine biodiversity patterns, and the inferences derived from them, are determined almost exclusively from a small set of highly visible reef-dwelling taxa: corals, fish, and conspicuous gastropods (Roberts et al., 2002; Bellwood et al., 2005; Carpenter & Springer, 2005; Bellwood & Meyer, 2009). These taxa represent a small portion of total coral reef biodiversity and may not be representative of other more speciose groups (e.g. crustaceans; Malay and Paulay, 2010). Furthermore, biodiversity patterns in these groups are commonly calculated from species range maps interpolated from individual point observations (Roberts et al., 2002) or the United Nations FAO (the Food and Agriculture Organization) fisheries data (Carpenter & Springer, 2005) that may greatly overestimate or underestimate actual species distributions. Even studies that directly measure biodiversity (e.g. Karlson et al., 2004) likely harbor inaccuracies due to the difficulty of taxonomic identifications, and/or the presence of cryptic species (Knowlton, 1986; Knowlton, 1993; Hellberg, 2009; Bucklin et al., 2011) in corals (Souter, 2010), fish (Rocha et al., 2008), and molluscs (Williams & Reid, 2004; Meyer & Paulay, 2005; Marko & Moran, 2009). As such, it is unclear whether our

knowledge of marine biodiversity patterns, and thus our understanding of the evolution and assembly of marine biodiversity, is correct.

There are numerous challenges to expanding the taxonomic scope of marine biodiversity studies in the sea. First, approximately 30-90% of marine species are undescribed (Mora et al., 2011; Appeltans et al., 2012). Second, the majority of marine taxa are small and difficult to identify morphologically (Bouchet et al., 2002). Third, even among taxa that can be distinguished morphologically, specialized taxonomic expertise is typically required to identify them to species (Bouchet et al., 2002; Knowlton et al., 2010). Such problems are compounded in regions like the Coral Triangle where the scale of biodiversity, known and unknown (Barber & Boyce, 2006), far exceed taxonomic expertise and resources required to study that diversity, making it difficult to understand biodiversity patterns more broadly in this region.

Another challenge in expanding taxonomic breadth of biodiversity studies is the belief that small species (<1mm) that often make up the largest proportion of biodiversity have cosmopolitan ranges leading to an absence of clear biodiversity patterns (Fenchel & Finlay, 2004). Often known as the "*everything is everywhere*" hypothesis, this theory predicts that the large population sizes of microscopic species make them less prone to geographic differentiation and less likely to display biogeographic patterns (Finlay, 2002). While global analyses support this hypothesis (Fenchel & Finlay, 2004), there is growing evidence for more regional structure in microscopic organisms (Faurby & Funch, 2011), particularly in taxa spanning strong biogeographic boundaries (Faurby & Barber, 2015).

Molecular tools, such as DNA metabarcoding (Carugati et al., 2015; Pavan-Kumar et al., 2015; Wangensteen & Turon, 2016) are revolutionizing our ability to detect and document marine biodiversity, including rare or cryptic species (Pawlowski et al., 2016). Metabarcoding is

the basis of a relatively new approach to marine biodiversity monitoring, Autonomous Reef Monitoring Structure (ARMS). ARMS are structures comprised of a stack of PVC plates, with spaces in between, that are designed to mimic the structure of the matrix of coral reefs, providing a substrate for marine biota to colonize (Knowlton et al., 2010; www.oceanARMS.org). Because ARMS are identical and can be deployed in a standardized habitat type for the same amount of time, ARMS combined with metabarcoding provide a highly standardized way to sample and document marine biodiversity (Knowlton et al., 2010; Plaisance et al., 2011a; Leray & Knowlton, 2015; Al-Rshaidat et al., 2016; Pearman et al., 2016; Ransome et al., 2017; Obst et al., 2020). A substantial advantage of ARMS in marine biodiversity studies is that this approach can provide information on thousands of taxa spanning the diversity of life over a standardized area, simultaneously, without any specific taxonomic expertise. Moreover, because ARMS processing protocols separate organisms based on size (500-106 µm, 2 mm-500 µm, >2mm, and encrusting sessile organisms (Leray & Knowlton, 2015), it is possible to explicitly test whether biodiversity patterns are consistent across different size classes of organisms, expanding our understanding of marine biodiversity patterns and how diversity is assembled into biodiversity hotspots.

This study examines biodiversity of benthic marine cryptofauna across Indonesia, the largest and most biologically diverse region of the Coral Triangle. Specifically, we use ARMS combined with high-throughput DNA metabarcoding to test 1) whether marine biodiversity patterns based on larger metazoans like fish and corals are broadly representative of marine biodiversity more broadly, and 2) whether patterns of biodiversity are consistent across size classes. We do this in an explicitly hierarchical framework, using an explicitly standardized

sampling design, that allows us to measure how biodiversity is packed onto reefs of the Coral Triangle, providing novel insights into the assembly of this global biodiversity hotspot.

Materials and Methods

ARMS Deployment, Collection, and Sampling

To examine scaling of marine biodiversity across the Coral Triangle biodiversity hotspot, we used Autonomous Reef Monitoring Structures (ARMS). ARMS consist of nine 23 cm x 23 cm PVC plates stacked vertically in an open and obstructed format attached to a 35 cm x 45 cm base plate (Figure 1-1B; https://www.oceanarms.org/protocols/arms-assembly), and facilitate sampling of marine biodiversity in a highlight standardized fashion.

We employed a hierarchical geographic sampling design capture marine biodiversity across well-established marine biodiversity gradients. These sites ranged from Aceh in Western Indonesia, a site outside of the Coral Triangle (Hoeksema, 2007; Bellwood & Meyer, 2009; Veron et al., 2009) to Raja Ampat, a region of Eastern Indonesia that is known for having the highest coral (Veron et al., 2009) and fish diversity (Allen, 2008) in the world (Figure 1-1). Additional deployments occurred in the Seribu Islands (Java), Bali, and Cenderawasih Bay (West Papua). In each of these five regions, we deployed ARMS at four different sites. To minimize variation in community composition due to depth and/or habitat differences, at each site, we deployed a set of three individual ARMS on the benthos spaced 3 to 5 meters apart, in similar forereef habitats, at a standardized depth of ~10m (9.7-13 m).

After a three-year deployment, we recovered the ARMS from the seafloor, and collected and processed all the animals that had colonized them following the methods of Ransome et al. (2017). To recover ARMS units, we encapsulated them within 100 µm mesh-lined containers to

prevent the escape of motile organisms while they are brought up from the sea floor. After returning the ARMS to the lab in aerated filtered seawater, ARMS were disassembled and every layer was photographed on both sides and assigned an ID number.

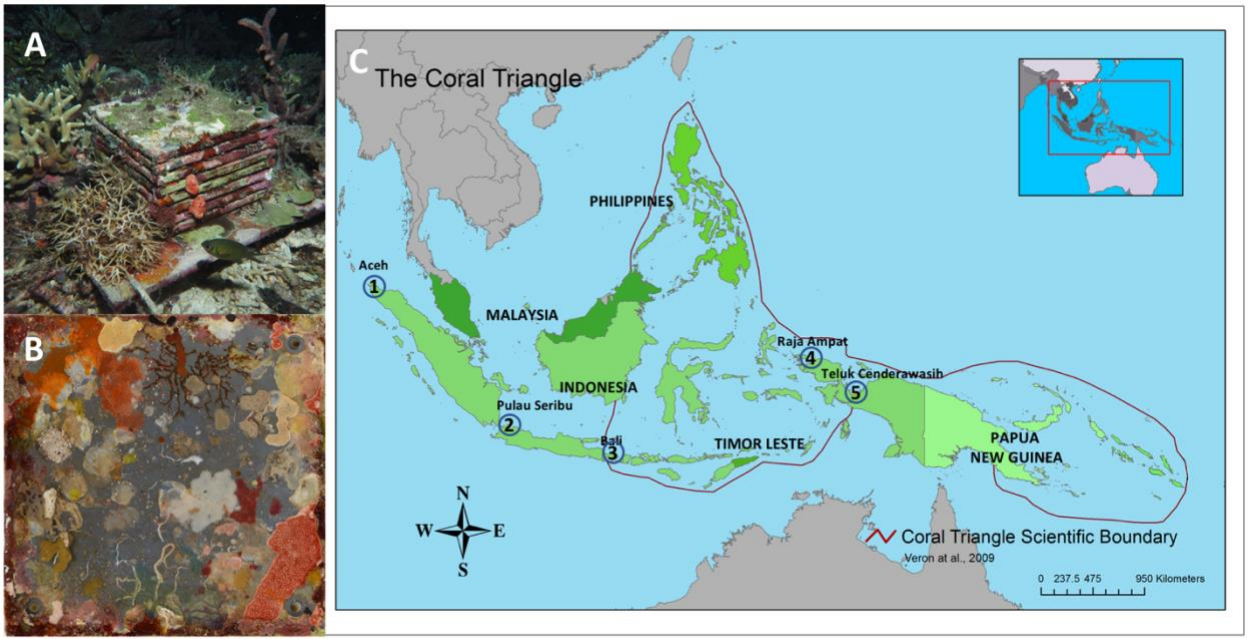


Figure 1-1. A) Autonomous Reef Monitoring Structure (ARMS) structure photographed underwater. B) ARMS plate colonized by benthic marine organisms. C) Map of the Coral Triangle with five sampling locations: (1) Pulau Weh, Aceh, (2) Kepulauan Seribu, Jakarta, (3) Pemuteran, Bali, (4) Raja Ampat, West Papua, and (5) Teluk Cenderawasih, West Papua. The Coral Triangle Scientific Boundary (red line) is based on Veron et al. (2009).

To test whether biodiversity patterns were consistent across size classes, we used a stainless sieves to separate motile organisms into two size fractions, a “100 µm fraction” (all motile organisms between 500-106 µm), and a “500 µm fraction” (all motile organisms from 2 mm-500 µm); all organisms >2 mm were vouchered for later study and are not included herein.

We washed these fractions with filtered seawater in a 45 μm Nitex net prior to preserving the samples with 95% ethanol and storing them at -20 °C until DNA extraction. Next, we scraped all encrusted or sessile biota (hereafter "sessile fraction") from the ARMS plates into filtered seawater, then homogenized it with a kitchen blender for 30 s at maximum speed. We rinsed the homogenate with filtered sea water taken from the ARMS recovery site into a 45 μm Nitex mesh collection net until the water ran clear, then placed approximately 10 g of the homogenate into a 50 ml falcon tube filled with DMSO, stored at -20 °C until DNA extraction.

DNA Preparation and Extraction

To remove inorganic material (e.g. sediment) that could inhibit DNA amplification, we performed a series of decantations (see Leray and Knowlton, 2015; https://www.oceanarms.org/protocols/molecular-analysis/bulk-dna-extractions/sample-preparation). Next, we extracted DNA from the 500 μm, 100 μm, and sessile samples using the MO-BIO Powermax® Soil DNA Isolation Kit according to the manufacturer's protocol with the addition of 400 μg/ml Proteinase K. To remove potential PCR inhibitors, we purified the DNA extractions using MO-BIO PowerClean® DNA Clean-Up Kits and quantified DNA concentrations using Qubit dsDNA HS Kit. The decantation and DNA extraction were performed at Yayasan Biodiversitas Indonesia (Bionesia), Denpasar, Bali, Indonesia and Laboratory of Marine Molecular Genetics, Research Center for Oceanograpy, Indonesian Institute of Sciences, Jakarta, Indonesia.

Because metabarcoding primers have known taxonomic biases (Giebner et al., 2020), we used two markers, *Cytochrome c Oxidase Subunit I* (COI), and the V4 region of 18S rRNA gene. We amplified COI using a dual-indexing approach with seven pairs of tagged COI PCR primers (mlCOIintF/jgHCO2198) (Geller et al., 2013; Leray et al., 2013). To account for potential PCR

bias (Ficetola et al., 2015; Nichols et al., 2018) and maximize probability of amplification of low copy templates, we performed PCR reactions in triplicate. Each PCR reaction was 20 µL in volume, consisting of 1 µL of 10 µM each forward or reverse primer, 1.4 µL of dNTPs, 0.4 µL of Taq Polymerase (CIontech), 2 µL of PCR buffer (CIontech), using 10 ng of extracted DNA. Thermocycling utilized a touchdown profile beginning with 16 initial cycles of denaturing at 95 °C for 10 s, annealing for 30 s at 62 °C (−1 °C per cycle), and extension for 60 s at 72 °C, followed by 20 cycles at an annealing temperature of 46 °C. We then pooled and visualized triplicate PCR products via electrophoreses in a 1.2% agarose gel.

To prepare the sequencing libraries, we pooled 1 ug of each tagged PCR sample, each comprised of a series of the seven tailed-primer pairs, into 12 samples and purified with Agencourt AMPure XP beads. We used a total of 1 µg of these pooled amplicons for end repair, A-tailing, and adaptor ligation using the TruSeq PCR-free kit (Illumina) following manufacturer protocols. We then validated the libraries via qPCR using the KAPA library quantification kit and diluted to a final concentration of 4nM before sequencing on an Illumina MiSeq using the 600 cycle reagent kit v3 (Illumina, San Diego, CA).

We amplified 18S rRNA using the V4_18SNext.For and V4_18SNext.Ref primers (Manzari et al., 2015) in 25uL reaction volumes consisting of 1.25 µL of 0.5 µM each forward or reverse primer, 2.5 µL of 0.2 mM dNTPs, 0.5 µL of Taq Polymerase (Phusion), 5 µL of PCR buffer (Phusion), and 2.5 ng DNA. Thermocycling parameters used an initial denaturing at 98 °C for 30 s, 10 cycles each at 98°C for 10 s, 44°C for 30 s, and 72 °C for 15 s, followed by 15 cycles each at 98°C for 10 s, 62°C for 30 s, and 72 °C for 15 s and finish with a final extension step at 72 °C for 7 min. We then confirmed and visualized successful PCRs through electrophoresis on a 1.2% agarose gel.

To create the 18S rRNA sequencing library, we employed a dual-indexing approach using the Nextera® index kit (Illumina), confirming successful indexing through electrophoresis on a 1.2% agarose gel. We then cleaned the libraries with Agencourt AMPure XP beads, pooled and diluted to a final concentration of 2nM. The sequencing was performed with MiSeq Illumina using a V2 500-cycle kit with 20% PhiX DNA added to each run as to improve data quality. All DNA sequencing was performed at the Laboratories of Analytical Biology, Smithsonian Institution National Museum of Natural History, Washington DC.

Data analyses

COI sequences were pre-processed, quality filtered and analyzed using *QIIME2* version 2017.8.0. (the Quantitative Insights Into Microbial Ecology 2 program, https://qiime2.org/). We used the Divisive Amplicon Denoising Algorithm 2 (*DADA2*) software, wrapped in *QIIME2*, to filter, trim, de-noise, and merge the data, removing chimeric sequences using the consensus method (Callahan et al., 2016).

For 18S rRNA, we merged the forward and reverse reads using *PEAR* (Zhang et al., 2014) allowing for a maximum of 10 differences in the overlap (default in *PEAR*) and only keeping aligned reads between 380 and 440 base pairs (bp) before quality filtering and analysis using *QIIME2* and *DADA2* to filter, trim, de-noise, and removing chimeric sequences using the consensus method (Callahan et al., 2016).

Next, we used the *LULU* algorithm (Frøslev et al., 2017) to filter out spurious sequences that may originate from PCR and/or sequencing errors, intra-individual variability (pseudogenes, heteroplasmy). *LULU* filters based on sequence similarity and co-occurrence rate with more abundant clusters, allowing us to curate the datasets while avoiding arbitrary abundance filters (Frøslev et al., 2017; Brandt et al., 2020). We ran LULU with a minimum relative co-occurrence

of 0.95 for both COI and 18S rRNA dataset, using a minimum similarity threshold (minimum match) at 84% (default) for COI and 18S rRNA.

Following quality filtering, trimming, de-noising, and chimera removal, we performed Operational Taxonomic Unit (OTU) clustering using the *Vsearch* (Rognes et al., 2016) plug-in in *QIIME2*. OTU clustering employed a 97% sequence identity threshold for COI and 99% sequence identity threshold for 18S rRNA using de novo clustering with "QIIME vsearch cluster-features-de-novo" command. This *de novo* OTU picking process clusters sequence reads by comparing sequences against one another without any external reference sequence collection. To test for saturation of OTU discovery, we created rarefaction curves with *iNEXT* function (Chao & Chiu, 2016; Hsieh et al., 2016).

To assign taxonomy to COI OTUs, we performed taxonomy assignment using *BlastN* (Altschul et al., 1990) against the *CRUX* (Curd et al., 2018) database, using a cutoff of 85% sequence identity. All assigned sequences were then aligned with *MAFFT* (Katoh & Standley, 2013), and used for phylogenetic reconstruction in *FastTree* (Price et al., 2010). We excluded all COI sequences that *BLAST* assigned to bacteria, keeping eukaryotes and unidentified taxa. To assign taxonomy for 18S rRNA sequences, we used a feature classifier in *QIIME2* trained against the PR2 database (Guillou et al., 2013), adopting a default confidence threshold of 0.7. To summarize the taxonomic composition of each sample we used *phyloseq* (McMurdie & Holmes, 2013) to generate stacked bar plots summarizing taxonomic composition and sequence abundance using *ggplot2* (Wickham, 2009) in R. Because of paucity of barcoding data from the Coral Triangle, we merged taxa at the phylum level and removed groups that represented less than 2% total abundance of the community.

Statistical Analyses

To test for saturation of OTU discovery, we created rarefaction curves with *iNEXT* (Chao et al., 2014; Hsieh et al., 2016) and *Ranacapa* package using *ggrare* command (Kandlikar et al., 2018). Then, we tested for differences in alpha and beta diversity across sampling sites and regions using multiple approaches. First, to determine whether marine biodiversity patterns based on large metazoans like fish and corals are representative of marine biodiversity more broadly, we calculated OTU diversity from ARMS for each of the five sampled regions. To test whether biodiversity patterns are consistent across size classes or whether smaller organisms are more widespread as predicted by the "everything is everywhere" hypothesis, we examined total diversity, as well as diversity of each size fraction for both COI and 18S rRNA. Because smaller life stages or pieces of larger organisms carried over during ARMS processing could influence patterns in smaller size fractions, we repeated the above examining all OTUs as well as OTUs remaining after excluding any shared among size fractions.

Next, to examine distribution of diversity across spatial scales, we calculated total OTU diversity for 1) individual ARMS, 2) individual sites (e.g. sets of 3 ARMS deployed on an individual reef), and 3) each of the 5 regions spanning the east-west biodiversity gradient in Indonesia.. At each spatial scale above we calculated total richness (e.g. total number of OTUs) across all size fractions and each individual size fraction for both barcoding markers. We then used ANOVA to test whether total OTU diversity per ARMS unit and per site differed among the five different sampling regions, following by a *post-hoc* Tukey test implemented in the package *FSA* (Ogle 2017) in R (R development core team) to examine for significant differences among specific sites. Because high diversity is often driven by large numbers of endemic

species, we examined the above patterns with total OTU diversity as well as examining OTUs endemic at the scale of a single ARMS unit, site, or region.

Next, we examined of taxonomic turnover (e.g. shared vs unique taxa) on multiple scales, within site (e.g. among a set of 3 ARMS), among sites within a region, and across the Indonesian archipelago, using multiple approaches. For each spatial scale, we used the the *eulerr* (Larsson et al., 2020) package in R and http://bioinformatics.psb.ugent.be/webtools/Venn/ to create a set of Venn diagrams to determine how many OTUs were shared at different spatial scales and total endemic richness (e.g. total number of OTUs unique to each of the three spatial scales).

To compare community composition across the five regions and reefs therein, we employed multiple approaches. First, we conducted multivariate analyses (PERMANOVA) based on Jaccard distances in the *vegan* package (Ogle, 2017) in R (R development core team) testing statistical significance using 9999 permutations and a significance level of $\alpha = 0.05$. We calculate the compositional dissimilarity using '*adonis*' command and the homogeneity of group dispersion using '*betadisper*' command in *vegan* package (Oksanen, 2017). Second, we conducted Principles Coordinates Analyses (PCoA) using the *Ampvis2* package (Andersen et al., 2018) with the ordination function of *phyloseq* for both Jaccard and Bray Curtis dissimilarity matrices and generate the ordination plot using *ggplot2* (Oksanen, 2017). Lastly, we generated clustered heatmaps and dendrograms illustrating dissimilarities in metazoan community composition using Python version 3.7.1 (2018; http://www.python.org) with the package *Pandas* version 0.23.4 (2018; http://pandas.pydata.org), *seaborn* (Wascom, et al., 2017), and *matplotlib* (Droettboom, et al., 2017).

Adespatial package: To examine the relative contributions to richness and replacement in driving patterns of beta diversity, we employed Podani's Jaccard-based indices (Wickham, 2009;

Podani & Schmera, 2011; Legendre 2014) calculated using *adespatial* package (Legendre & De Cáceres, 2013) within R (R development core team). This analysis calculates three indices that measure similarity, relative species replacement and relative richness differences among all pairs of sites (Dray et al. 2016), and then uses ternary plots to visualize these differences. We conducted this analysis comparing (1) all possible pairwise comparisons by combining OTUs diversity among all individual ARMS unit; (2) among sites (e.g., sets of 3 ARMS units); (3) among the five regions samples and (4) for each size fraction.

Results

We recovered a total of 59 of 60 ARMS; one ARMS unit from Sumur Tiga, Aceh could not be recovered. From the 500 µm, 100 µm, and sessile fractions, we amplified a total of 174 samples with both COI and 18S rRNA primers. Due to PCR amplification failure, we obtained data from 58 ARMS units using COI and 59 units using 18S rRNA. Sites with incomplete data (missing ARMS or failed PCRs) were included when considering absolute diversity, but were excluded from analyses such as percent regional diversity captured at a single site, where missing samples could impact results

DNA sequencing returned 19,052,584 reads and 46,633,073 reads from COI and 18S rRNA, respectively. OTU clustering yielded a total of 12,330 OTUs for COI and 14,350 OTUs for 18S rRNA (Supplemental Table S1-2, S1-3). While rarefaction curves for both COI and 18S approached the asymptote, curves for COI were flatter than 18S, indicating greater saturation in COI of OTU discovery. (Figure 1-2A, 3-2B), and both markers showed less saturation in OTU diversity in high diversity regions of Eastern Indonesia, compared to lower diversity regions to the west. Although sequencing depth did not saturate at the regional scale, sequencing depth was sufficient to saturate in most individual ARMS units (Figure 1-2C, 3-2D).

Taxonomic composition

Across Indonesia, the COI metabarcoding dataset was dominated by *Porifera*, making up 17% of total sequence reads but only 2% of total OTUs (Figure 1-3A, 3-3C). Other prominent taxa included *Arthropoda* (15% sequence reads and 18% taxa) and *Rhodophyta* (15% sequence reads and 6% taxa); 35% of the total reads and 58% OTUs were unidentified. Both the 500 µm and 100 µm (motile) fractions were dominated by *Arthropoda* (24% and 18%, respectively), and unidentified taxa (40% and 55%), while the sessile fraction was dominated by *Porifera* (40%), *Rhodophyta* (30%), unidentified species (11%), and *Cnidaria* (10%) (Figure 1-3A).

The18S rRNA metabarcoding dataset was dominated by *Arthropoda* (23% sequence reads and 15% taxa), *Annelida* (22% sequence reads and 4.8% taxa), *Porifera* (16% sequence reads and 2.5% taxa), and *Rhodophyta* (15% sequence reads and 3.6% taxa) and (Figure 1-3B, 1-3D), and only had 2% of the total reads and 28% OTUs were unidentified. Both the 500 µm and 100 µm fractions were dominated by *Annelida* (36% and 24%, respectively) and *Arthropoda* (27% and 40%), and the sessile fraction was dominated by *Rhodophyta* (35%) and *Porifera* (33%) (Figure 1-3B).

Differences in abundance (based on sequence reads) had minimal variation across locations (Figure 1-3). Pulau Seribu had more unidentified COI reads (43%) compared to the other locations (range from 32% to 35%). Aceh had the highest abundance of *Porifera* (27%) and Bali had the highest abundance of *Rhodophyta* (29%), with similar patterns in 18S rRNA.

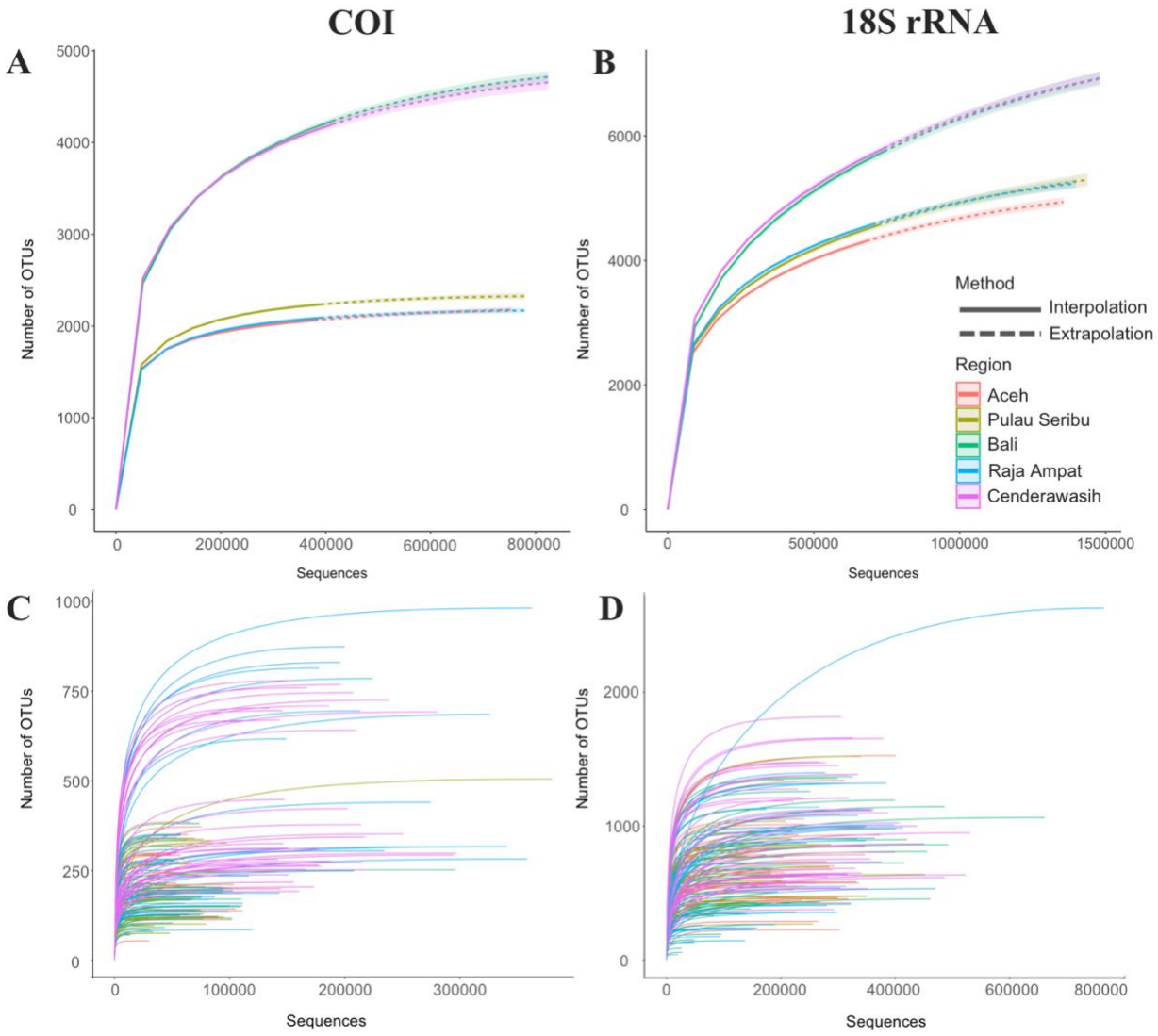


Figure 1-2. Rarefaction plots showing numbers of Operational Taxonomic Units (OTUs) as a function of sequencing depth for the five regions based on A) COI and B) 18S rRNA, as well for each individual ARMS unit for C) COI and D0 18S rRNA.

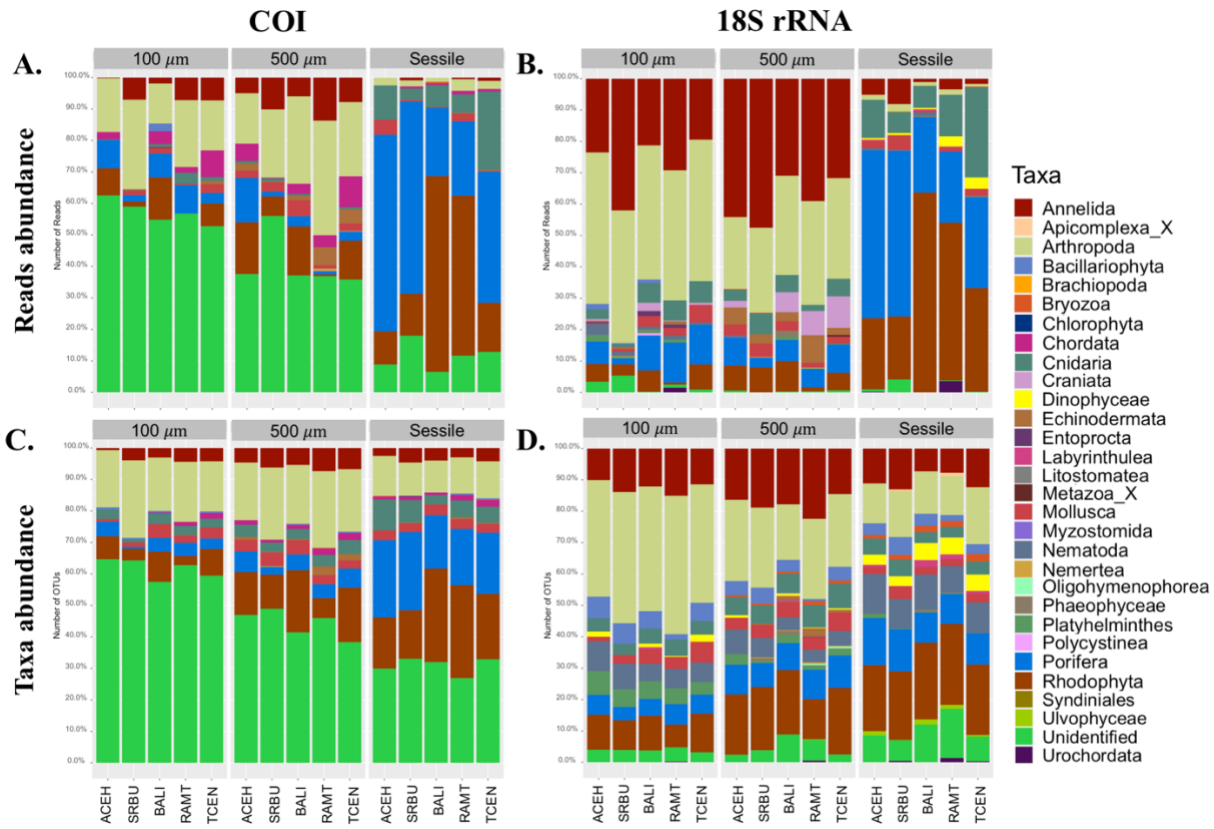


Figure 1-3. Taxonomic composition of eukaryote communities identified across 100 μm, 500 μm, and sessile ARMS sample fractions, spanning five regions of Indonesia based on A) relative abundance of sequence reads and C) numbers of OTUs (Operational Taxonomic Units) based on COI, and B) relative abundance of sequence reads and D) numbers of OTUs based on 18S rRNA, excluding all taxa with <2% relative abundance.

Patterns of Diversity

Total regional OTU diversity from COI (Table 1-1) was highest in Teluk Cenderawasih (4,235 OTUs) and lowest in Aceh (2,063 OTUs). The 100 μm fraction was the most diverse of the three fractions, and was highest in Teluk Cenderawasih (3,215 OTUs) and lowest in Bali

(1,228 OTUs). The 500 μm fraction was the second most diverse, ranging from 1,528 OTUs in Raja Ampat to 873 OTUs in Bali. The sessile fraction was the least diverse, ranging from 1,033 OTUs in Teluk Cenderawasih to 639 OTUs in Aceh. These patterns were consistent, examining total OTU diversity, or excluding any OTU shared across size fractions (Table 1-1), and using the rarefied or non-rarefied data sets (data not shown).

Regional OTU diversity from 18S rRNA was higher than COI, ranging from 5,801 OTUs in Raja Ampat to 4,309 OTUs in Aceh (Table 1-2). As with COI, diversity was highest in the 100 μm fraction, ranging from 4,151 OTUs in Teluk Cenderawasih to 2,835 OTUs in Pulau Seribu. The sessile fraction was slightly more diverse, ranging from 2,867 OTUs in Raja Ampat to 1,753 OTUs in Aceh, than the 500 μm fraction, that ranged from 2,184 in Teluk Cenderawasih to 1,753 in Aceh. These patterns were consistent, examining total OTU diversity, or excluding any OTU shared across size fractions (Table 1-2), and using the rarefied or non-rarefied data sets (data not shown).

Table 1-1. Operational Taxonomic Units (OTU) diversity from COI metabarcoding of autonomous reef monitoring structures spanning five sampled regions of Indonesia, including total diversity and diversity of three individual size fractions. Numbers in parentheses are diversity totals that include only OTUs unique to that size fraction (e.g. excluding any OTUs shared among size fractions).

	Aceh	**Pulau Seribu**	**Bali**	**Cenderawasih**	**Raja Ampat**
Total	2063 (1684)	2090 (1694)	2236 (1810)	4235 (3752)	4198 (3711)
100 μm fraction	1325 (1055)	1244 (991)	1428 (1125)	3215 (2779)	3036 (2638)
500 μm fraction	1021 (732)	939 (655)	873 (589)	1473 (1053)	1528 (1170)
Sessile fraction	639 (326)	788 (447)	729 (418)	1033 (636)	931 (593)

Table 1-2. Operational Taxonomic Units (OTU) diversity based on 18S rRNA metabarcoding of autonomous reef monitoring structures spanning five sampled regions of Indonesia, including total diversity and diversity of three individual size fractions. Numbers in parentheses are diversity totals that include only OTUs unique to that size fraction (e.g. excluding any OTUs shared among size fractions).

	Aceh	**Pulau Seribu**	**Bali**	**Cenderawasih**	**Raja Ampat**
Total	4309 (2663)	4577 (2845)	4567 (2905)	5751 (3875)	5801 (3965)
100 µm fraction	3005 (1655)	2835 (1507)	2959 (1595)	4151 (2530)	3613 (2121)
500 µm fraction	2107 (810)	2002 (779)	1628 (628)	2184 (941)	1829 (733)
Sessile fraction	1753 (711)	2449 (1104)	2127 (1056)	2447 (1105)	2867 (1649)

Spatial Scaling of Diversity

The average percentage of total local diversity captured in a single ARMS unit ranged from 52.4% in Aceh to 46.5% to Raja Ampat for COI (Figure 1-4A) and 58.3% in Aceh to 51.3% in Raja Ampat for 18S (Figure 1-4B), although these differences were not significant differences among any sampling locations (CO, one-way ANOVA, *p-value* = 0.691; 18S rRNA, one-way ANOVA, *p-value* = 0.436) (Supplemental Table S1-4). These patterns were consistent with results from the non-rarefied dataset (Supplemental Figure S1-1).

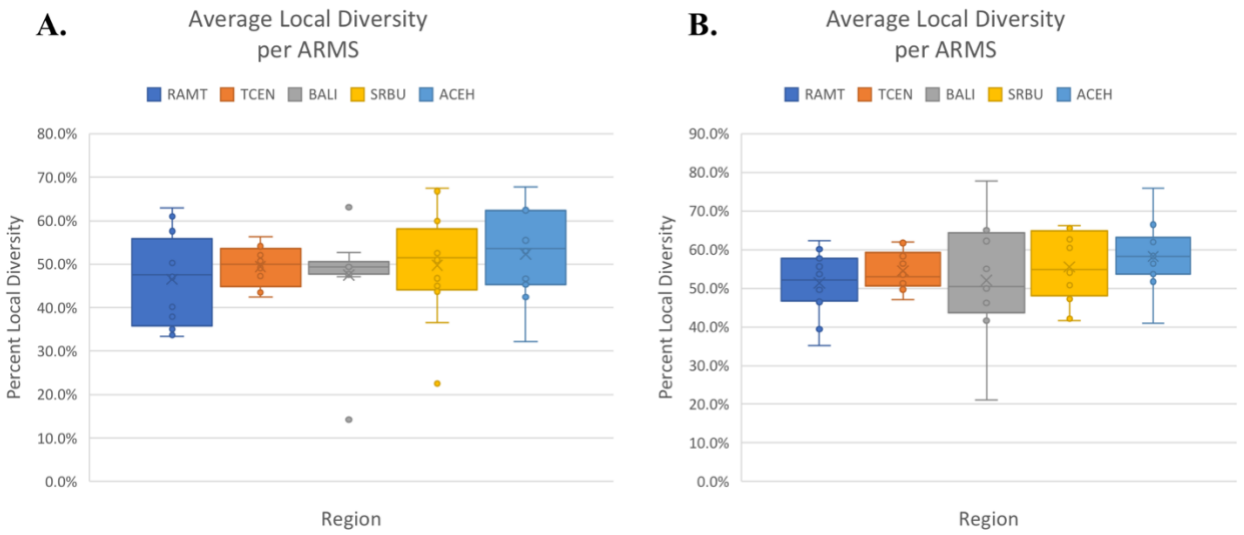


Figure 1-4. Average per site OTU diversity captured in a single ARMS (Autonomous Reef Monitoring Structure) unit across all sites for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Based on COI, absolute diversity per ARMS unit was highest in Teluk Cenderawasih with 889-632 OTUs per ARMS (mean = 732.3 OTU/ARMS) and the lowest in Pulau Seribu with 567-191 OTUs per ARMS (mean = 410.2 OTU/ARMS) (Figure 1-5A). For 18S rRNA, the pattern was similar, with Teluk Cenderawasih having the highest total diversity with 1,667-1,208 OTUs per ARMS (mean = 1,444.5 OTU/ARMS) while Bali had the lowest diversity of 1463-377 OTUs per ARMS (mean = 1,066 OTU/ARMS) (Figure 1-5B). ANOVA showed significant differences in OTU diversity per ARMS for both markers (COI, *p-value* <0.005; 18S rRNA, *p-value* = 0.0014) (Supplemental Table S1-4). Tukey tests showed that Teluk Cenderawasih and Raja Ampat were not significantly different for COI OTU diversity, while both were significantly more diverse than the other locations (Supplemental Table S1-5). For 18S rRNA,

Teluk Cenderawasih had significantly more diversity than Aceh, Bali, and Pulau Seribu, but not Raja Ampat. (Supplemental Table S1-5). The non-rarefied dataset returned similar results (Supplemental Figure S1-2; Supplemental Table S1-6, S1-7).

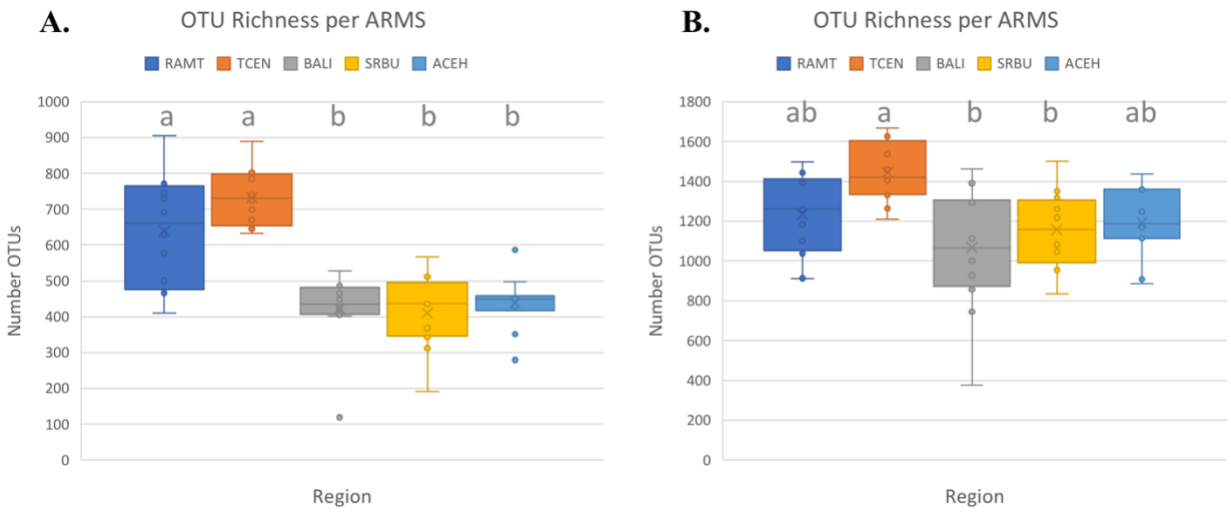


Figure 1-5. Total Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Percent of regional diversity represented in each ARMS unit was highest in Western Indonesian sites outside of the Coral Triangle, and lowest in Eastern Indonesia, the center of the Coral Triangle (Figure 1-6). On average, a single ARMS from any site within Aceh captured 22.1% of regional COI OTU diversity (range 29%-14%). In contrast, a single ARMS from Raja Ampat only captured an average of 15.2% of regional diversity (range from 21%-9.8%) (Figure 1-6A). Patterns were similar for 18S, with a single ARMS capturing an average of 27.7% of

regional OTU diversity in Aceh (range 33.3%-20.5%) compared to 21.3% in Raja Ampat (range 25.8%-15.7%) (Figure 1-6B). ANOVA indicated significant differences for both datasets (COI, *p-value* = 0.00145; 18S, *p-value* = 0.0183) (Supplemental Table S1-4), although Tukey tests only showed significant differences among the extremes of these diversity gradients (Figure 1-6; Supplemental Table S1-5). The non-rarefied dataset returned similar patterns (Supplemental Figure S1-3), but ANOVA were non-significant (Supplemental Table S1-6, S1-7).

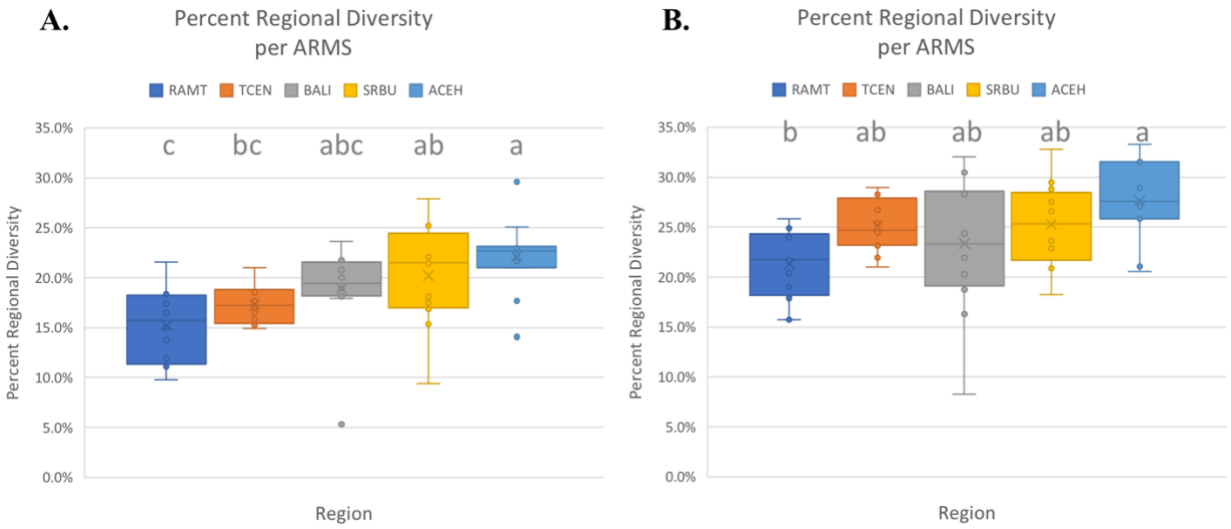


Figure 1-6. Percentage of total regional diversity captured in an individual ARMS unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Total per site OTU diversity was highest in Teluk Cenderawasih (COI range 1,853-1,286, mean = 1,494.5 OTUs/site) and lowest in Pulau Seribu (range 854-760, mean = 826.3 OTUs/ site) (Figure 1-7A). Per site OTU diversity was also highest in Teluk Cenderawasih for 18S (range from 2,831-2,356, mean = 2,653.5 OTUs/site), but lowest in Aceh (range 2,212-1,555,

mean = 2020 OTUs/site) (Figure 1-7B). ANOVA revealed significant differences in total OTU diversity per site across the 5 regions (COI, *p-value* <0.005; 18S, *p-value* = 0.00336) (Supplemental Table S1-4). While Tukey test for COI dataset showed that Teluk Cenderawasih and Raja Ampat were not significantly different, both had significantly more per site OTU diversity than the other three regions (Supplemental Table S1-5). However, for 18S, only Teluk Cenderawasih had significantly more per site OTU diversity than Aceh, Bali and Pulau Seribu (Supplemental Table S1-5). The non-rarefied dataset returned equivalent results, but with higher total numbers of OTUs, and Aceh had the lowest diversity for both markers (Supplemental Figure S1-2; Supplemental Table S1-6, S1-7).

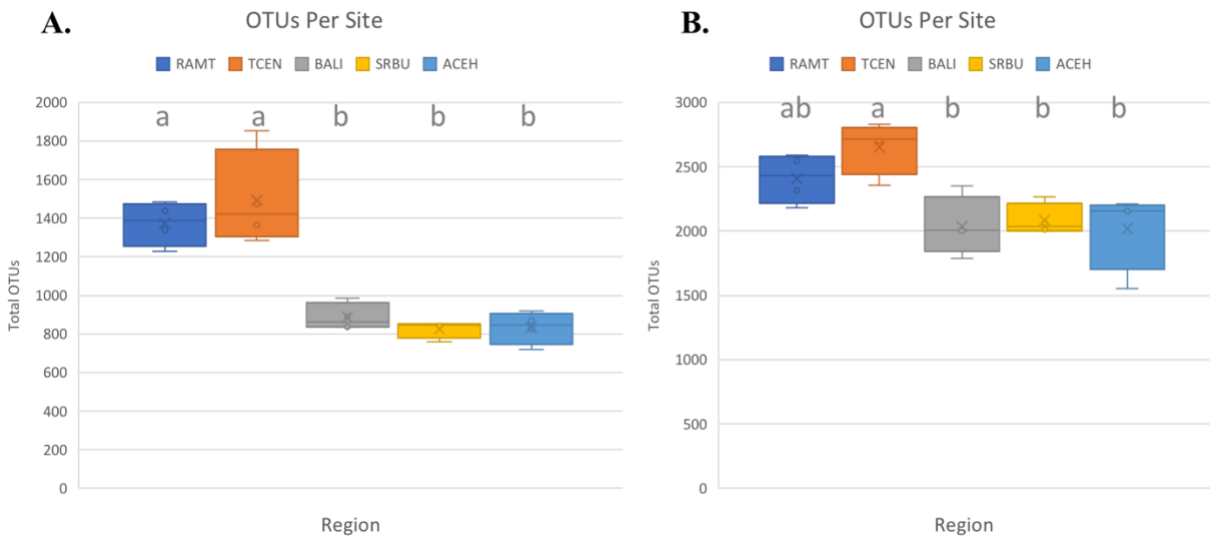


Figure 1-7. Total Operational Taxonomic Units (OTUs) diversity captured in a single sampling site across five sampling regions across Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Percent of regional OTU diversity captured from any sample site was highest in Aceh (range 44.6%-40.1% per site, mean = 42.2%) and lowest in Raja Ampat (range 35.4%-29.3% per site, mean = 32.7%) based on COI and 18S (Aceh; range 51.3%-50% per site, mean 50.5%; Raja Ampat range 44.7%-37.6%, mean = 41.5%) (Figure 1-8). ANOVA results showed significant differences in regional diversity per site across the five sampling location for COI (*p-value* = 0.0205) but not 18S (*p-value* = 0.0623) (Supplemental Table S1-4), with the Tukey test revealing only one significant pairwise comparison, between Aceh with Raja Ampat for COI (Supplemental Table S1-5). The non-rarefied dataset returned equivalent results (Supplemental Figure S1-3A), although the differences were not significant (Supplemental Table S1-6, S1-7).

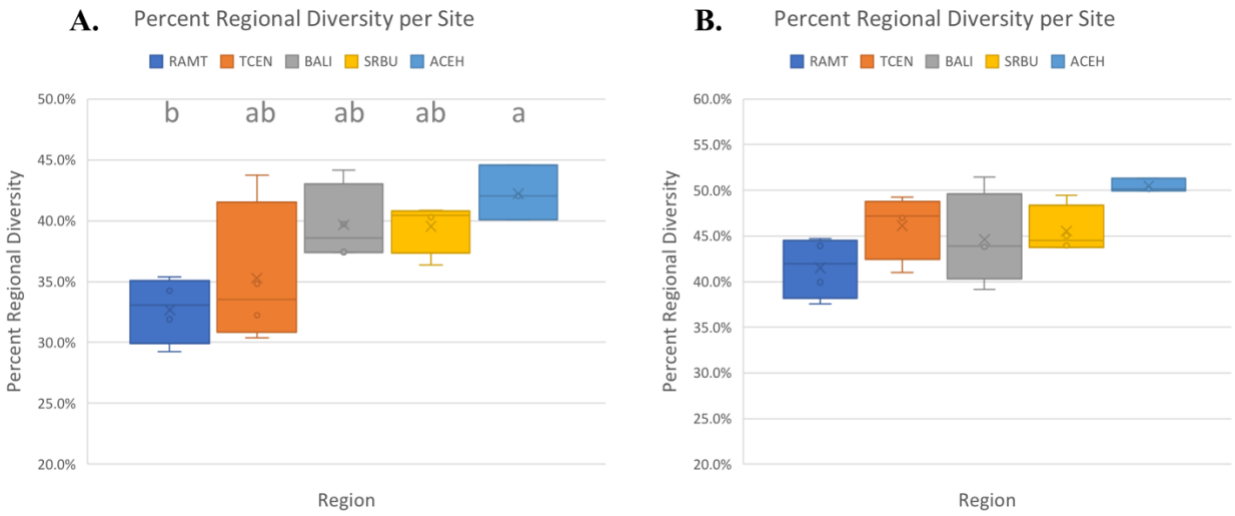


Figure 1-8. Percentage of regional diversity captured in an individual sampling site across five sampled regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Endemic Diversity

As with total diversity, endemic OTU diversity per ARMS unit was highest in Teluk Cenderawasih (range 519-236 OTUs per ARMS, mean = 327.9) and lowest in Pulau Seribu (range 310-63 OTUs per ARMS, mean = 178.4) (Figure 1-9A) based on COI and 18S rRNA (Teluk Cenderawasih, range 688-359 OTUs per ARMS, mean = 514.4 OTUs; Pulau Seribu, range 611-215 OTUs per ARMS, mean = 390.5) (Figure 1-9B). ANOVA revealed a significant difference in endemic OTUs diversity per ARMS unit across sampling location for COI (*p-value* <0.005), but not 18S (*p-value* = 0.25) (Supplemental Table S1-4). Tukey test for COI showed that Teluk Cenderawasih and Raja Ampat were not significantly different, but both had significantly more endemic OTU diversity than the other three locations (Supplemental Table S1-5). The non-rarefied data returned equivalent results, but with overall higher numbers of OTUs (Supplemental Figure S1-2; Supplemental Table S1-6, S1-7).

Endemic diversity per site was highest in Raja Ampat (range 872-774 OTUs per site, mean = 819) and lowest in Aceh (range 395-282 OTUs per site, mean = 317) for COI (Figure 1-10A) and 18S (Raja Ampat, range 1,059-769 OTUs per site, mean = 912.8; Aceh range 663-389 OTUs per site, mean = 589.) (Figure 1-10B). ANOVA revealed significant differences in total endemic diversity per site across the five regions (One-way ANOVA, COI, *p-value* <0.005; One-way ANOVA, 18S rRNA, *p-value* = 0.0397) (Supplemental Table S1-4). Tukey test for COI indicate that Raja Ampat and Teluk Cenderawasih have significantly more endemic diversity per site than the other three regions while for 18S only Teluk Cenderawasih and Aceh were different (Supplemental Table S1-5). Results from the non-rarefied dataset were similar, although endemic OTU numbers were higher, Teluk Cenderawasih had the highest per site diversity, and

only the COI data returned significant ANOVA values (Supplemental Figure S1-2; Supplemental Table S1-6, S1-7).

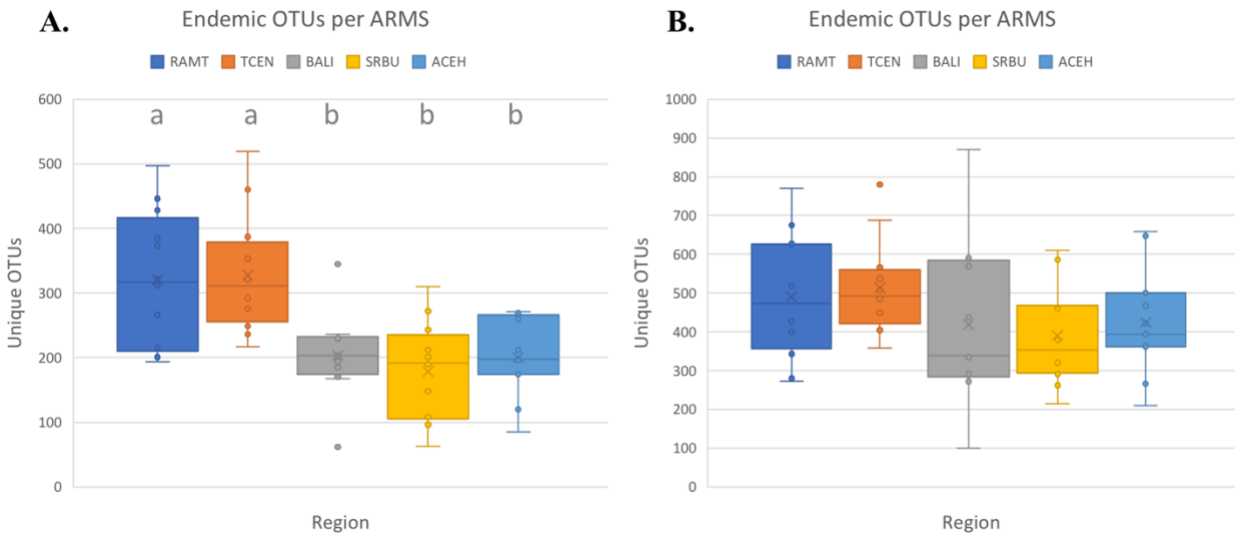


Figure 1-9. Total endemic Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

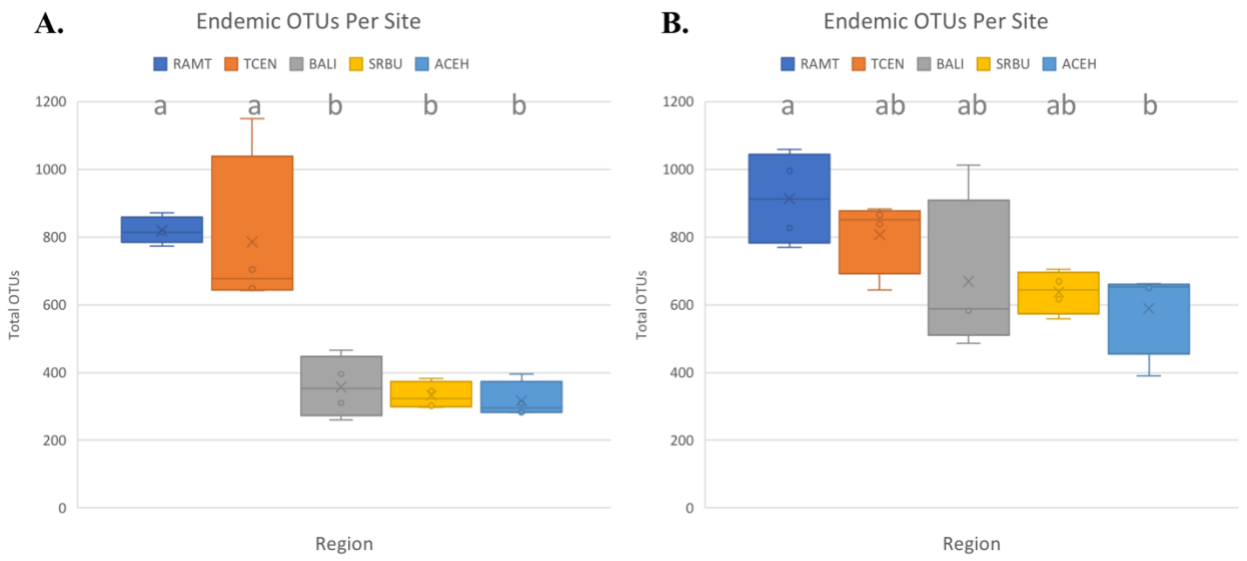


Figure 1-10. Total endemic Operational Taxonomic Units (OTUs) diversity captured per sampling site across five sampling regions across Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers.

Scaling of Diversity Across Space and Size Fractions.

Total OTU diversity per ARMS unit was consistently higher in the Eastern Indonesia (Raja Ampat and Cenderawasih) and lower in Western part of Indonesia (Aceh, Pulau Seribu and Bali) across all size fractions, except for the 500 µm fraction of 18S rRNA, where Aceh had the highest diversity (Figure 1-11). These patterns were consistent using rarefied data sets (including or excluding OTUs shared across fractions) (Supplemental Figure S1-8). Across all size fractions, except 18S sessile, ANOVA showed significant difference among the five regions (Supplemental Table S1-8, S1-10). Tukey tests indicate that these differences were largely driven

by differences between Raja Ampat and/or Teluk Cenderawasih and the remaining three regions (Supplemental Table S1-9, S1-11)

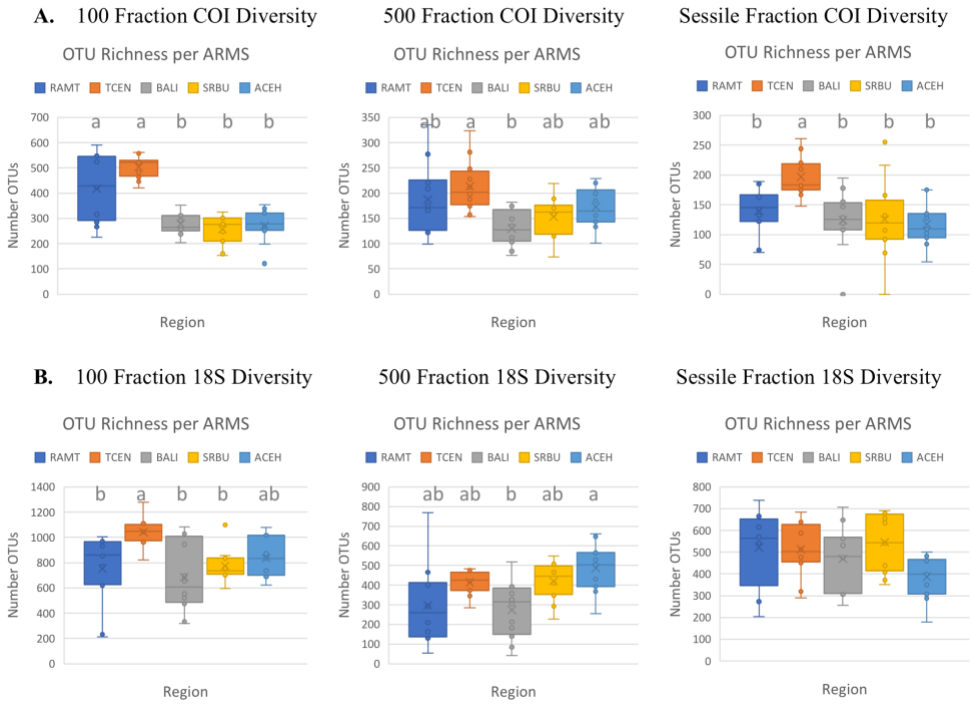


Figure 1-11. Total Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit in the A) 100 µm B) 500 µm and C) Sessile size fractions across five sampled regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers. Scale on Y axes vary.

Per site OTU diversity was also consistently highest in Eastern Indonesia and lowest in Western Indonesia across all size fractions, and both markers, with the exception of the 18S rRNA 500 μm fraction, where Aceh was the highest (Figure 1-12). For all size fractions of COI, ANOVA showed significant differences in per site OTU diversity across the five regions, but for 18S rRNA, only the 100 μm fraction was significant. (Supplemental Table S1-8). Similar results were obtained from the rarefied datasets (excluding OTUs shared among size fractions) (Supplemental Table S1-10). Tukey tests indicate that these differences were largely driven by differences between Raja Ampat and/or Teluk Cenderawasih and the remaining three regions (Supplemental Table S1-9, S1-11).

Endemic OTU diversity per ARMS unit was typically highest in Eastern Indonesia and lowest in Western Indonesia across all COI size fractions, particularly the 100 μm fraction, but was variable for 18S rRNA (Figure 1-13). The only significant ANOVA values were for 100 μm and 500 μm fractions of COI (Supplemental Table S1-8). However, excluding OTUs shared among fractions, all ANOVA results were significant for COI and for the 18S rRNA sessile fraction (Supplemental Table S1-10). Tukey comparisons were only significant for the 100 um fraction of COI, where Raja Ampat and Teluk Cenderawasih were significantly different from the population to the west (Supplemental Table S1-9, S1-11).

Figure 1-12. Total Operational Taxonomic Units (OTUs) diversity captured per sampling site in the A) 100 µm B) 500 µm and C) Sessile size fractions across five sampling regions across Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers. Scales on y axes vary.

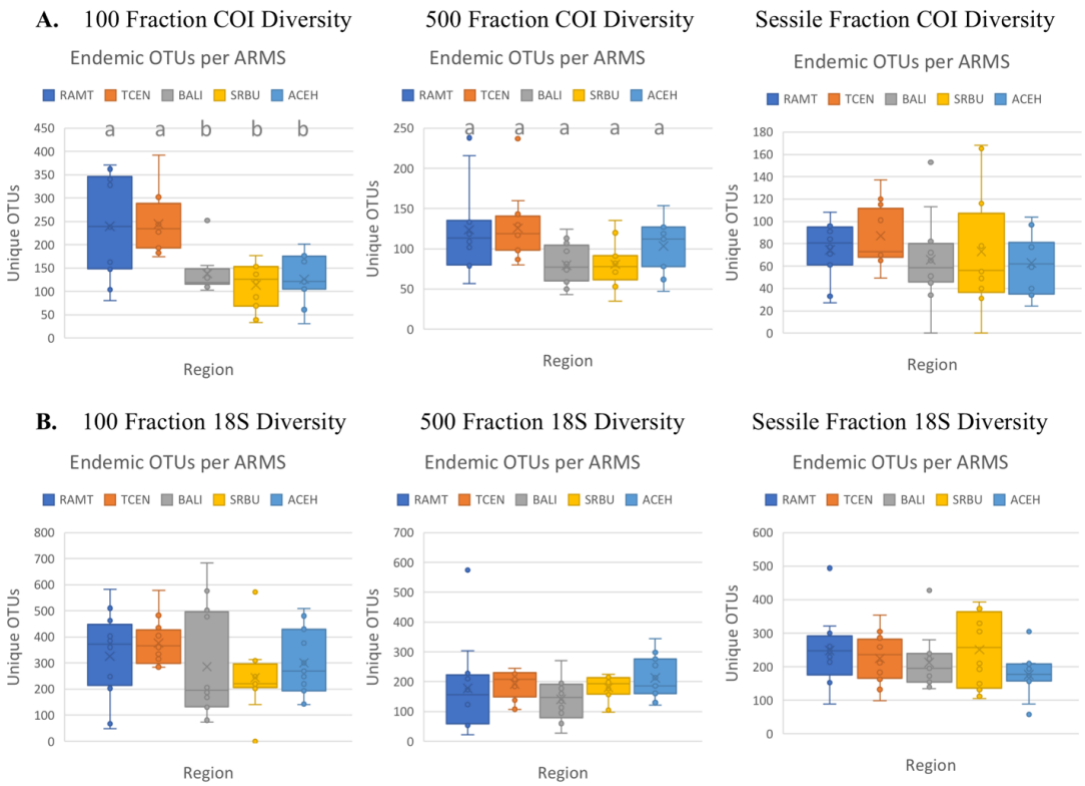


Figure 1-13. Total endemic Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit in the A) 100 µm B) 500 µm and C) Sessile size fractions across five sampled regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers. Scale on y axes vary.

Per site endemic diversity was highest in Eastern Indonesia and lowest in Western Indonesia across all size fractions in both markers, and was most pronounced in the 100 µm fraction (Figure 1-14). ANOVA indicated significant differences in per site endemic diversity in the 100 µm and 500 µm fractions, but not sessile for COI; only the 100 µm fraction was

significant for 18S rRNA (Supplemental Table S1-8)**.** The data set excluding OTUs shared across size fractions returned equivalent patterns (Supplemental Table S1-10), although only the 100 µm and 500 µm fractions from COI were significantly different in the ANOVA analyses. Where observed, significant values were consistently driven by differences among populations in Eastern and Western Indonesia.

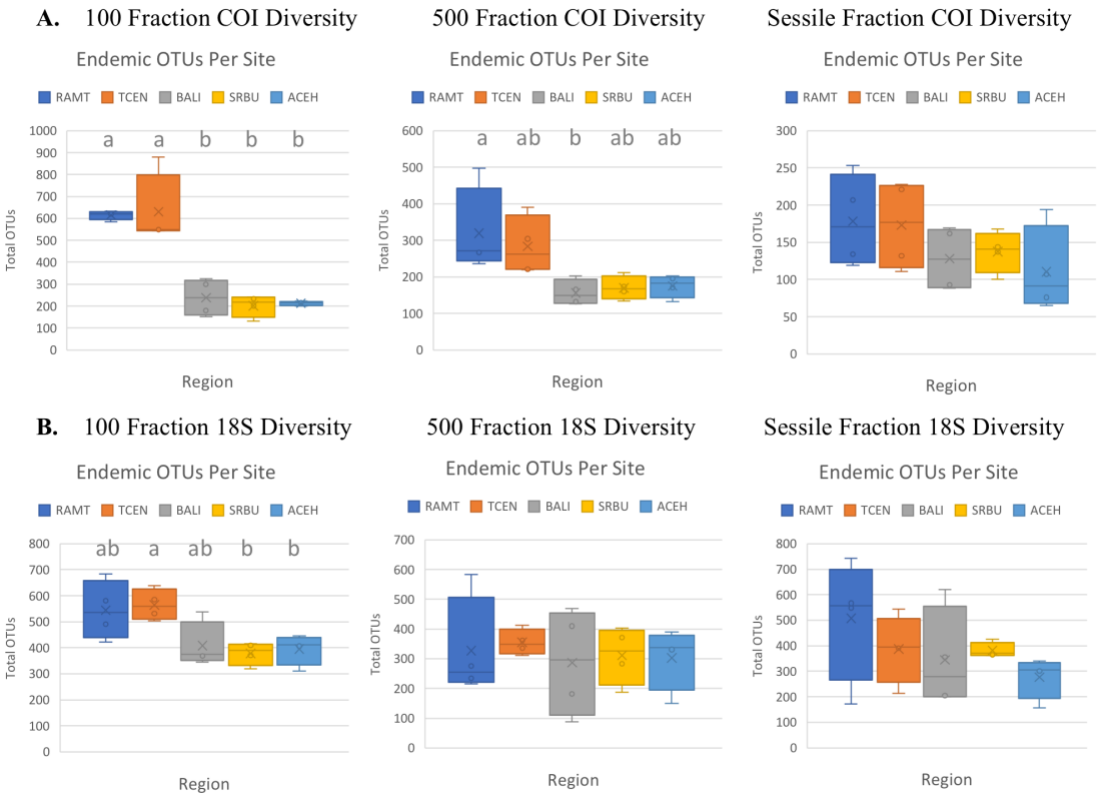


Figure 1-14. Total endemic Operational Taxonomic Units (OTUs) diversity captured per sampling site in the 100 µm, 500 µm and sessile size fractions across five regions of Indonesia for A) COI and B) 18S rRNA. X represents the mean of alpha diversity from each regions, the box represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with points beyond considered outliers. Scales on y axes vary.

Community Similarity and Beta Diversity

Comparing total eukaryote community composition across the 59 ARMS units, ordination plots for both COI and 18S rRNA show clusters of highly similar communities corresponding to each of the five sampled regions and strong differentiation among these regions (Figure 1-15; Supplemental Table S1-13). Aceh and Pulau Seribu were the most dissimilar, with Bali, Raja Ampat and Teluk Cenderawasih clustering together or overlapping. Similarly, UPGMA plot heatmaps show strong similarity within regions and high dissimilarity among regions (Figure 1-16). In both cases, these patterns were stronger using the Jaccard index than the Bray-Curtis index which gives higher weighting to shared taxa.

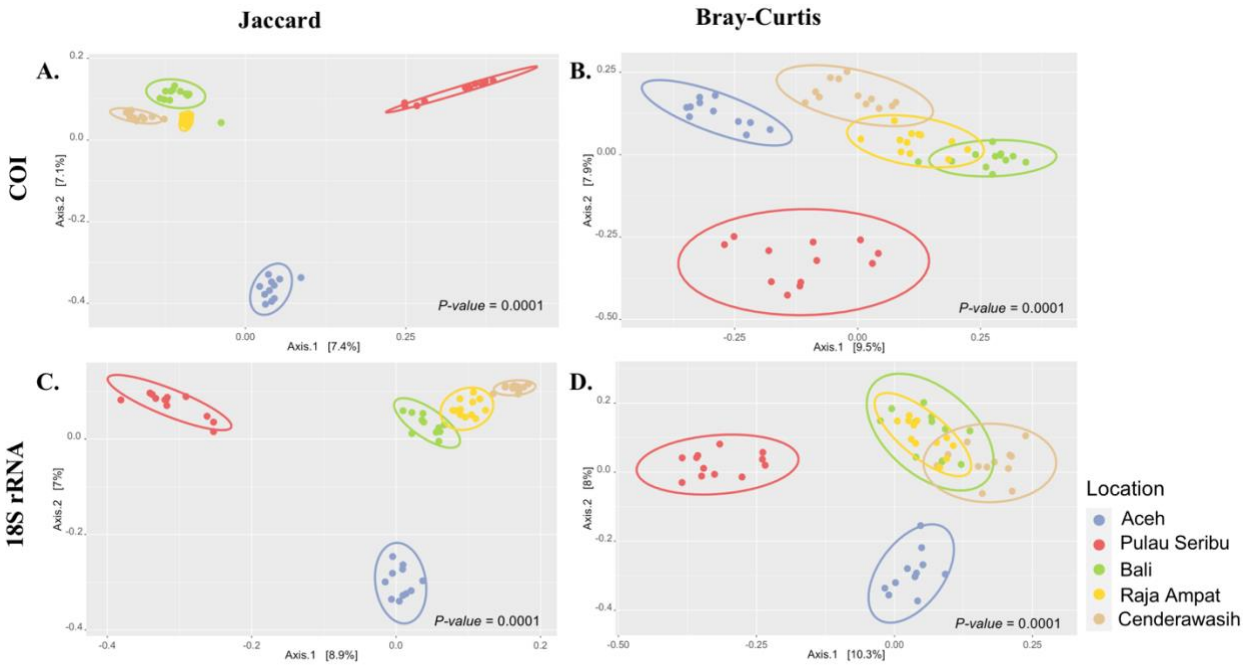


Figure 1-15. Principal Coordinates Analysis (PCoA) analysis illustrating dissimilarities in eukaryote community composition across 59 ARMS units deployed across the Indonesia archipelago based on COI (A and B) and 18S rRNA (C and D) using Jaccard and Bray-Curtis similarities.

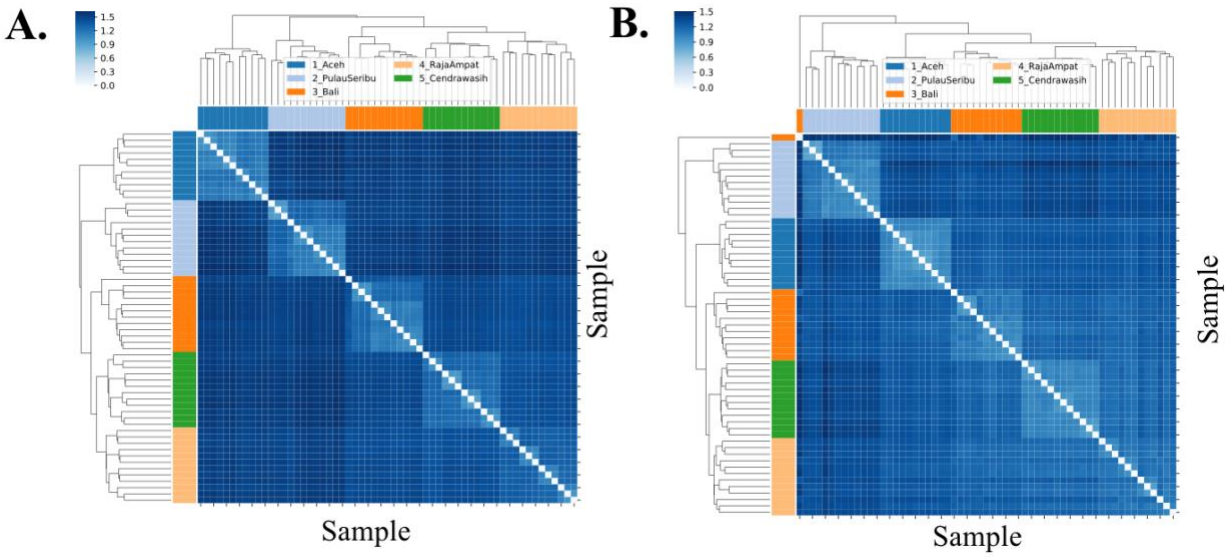


Figure 1-16. Unweighted Pair Group Method with Arithmetic mean (UPGMA) clustering analysis with heatmap and dendrogram illustrating dissimilarities in marine benthic eukaryote community composition based on Jaccard similarity of OTU diversity from 59 ARMS from 5 regions of Indonesia for both A) COI and B) 18S rRNA.

Beta diversity was very high across all five regions, ranging from a high of 0.47 in Raja Ampat to a low of 0.44 in Aceh (Table 1-3) with significant changes in eukaryote community composition across the five regions, sites, and ARMS, and across size fractions within these hierarchical scales for both COI and 18S (PERMANOVA p-value <0.05) (Supplemental Table S1-12). Dispersion homogeneity tests (Betadisper) based on Bray-Curtis, Jaccard, and UniFrac distances were significant (p-value <0.05) across nearly all of the parameters (size fraction, locations, sites and ARMS triplicate), but not at the level of individual ARMS units (Supplemental Table S1-12).

Partitioning COI OTU beta diversity across all ARMS units into replacement (i.e. turnover) and richness differences, turnover accounts for about 75% of the beta diversity, with

richness accounting for 19% (Figure 1-17A). Among sites, turnover of COI OTUs accounted for 76.6% of beta diversity, while richness accounted for 16.2%. At the region level, turnover and richness accounted for 71.2% and 21.6% of beta diversity. Results from 18S were similar, but turnover accounted for ~65% and richness 10-14% of total beta diversity across all spatial scales (Figure 13-7B; Supplemental Figure S1-13). Analysis of each size fraction separately yielded equivalent results (Supplemental Figure S1-12).

Table 1-3. Beta diversity across five sampling locations across Indonesia obtained from analysis of COI and 18S metabarcoding data using the *Adespatial* package in R.

A. COI metabarcoding dataset

Region	Total ß-diversity	Replacement	Richness Difference
Aceh	0.4474	0.3134	0.1340
Pulau Seribu	0.4515	0.3260	0.1255
Bali	0.4560	0.3284	0.1276
Raja Ampat	0.4733	0.3149	0.1584
Cenderawasih	0.4537	0.3209	0.1328

B. 18S rRNA metabarcoding dataset

Region	Total ß-diversity	Replacement	Richness Difference
Aceh	0.3965	0.2627	0.1337
Pulau Seribu	0.4021	0.2901	0.1120
Bali	0.4288	0.2535	0.1754
Raja Ampat	0.4394	0.2495	0.1899
Cenderawasih	0.4051	0.2662	0.1389

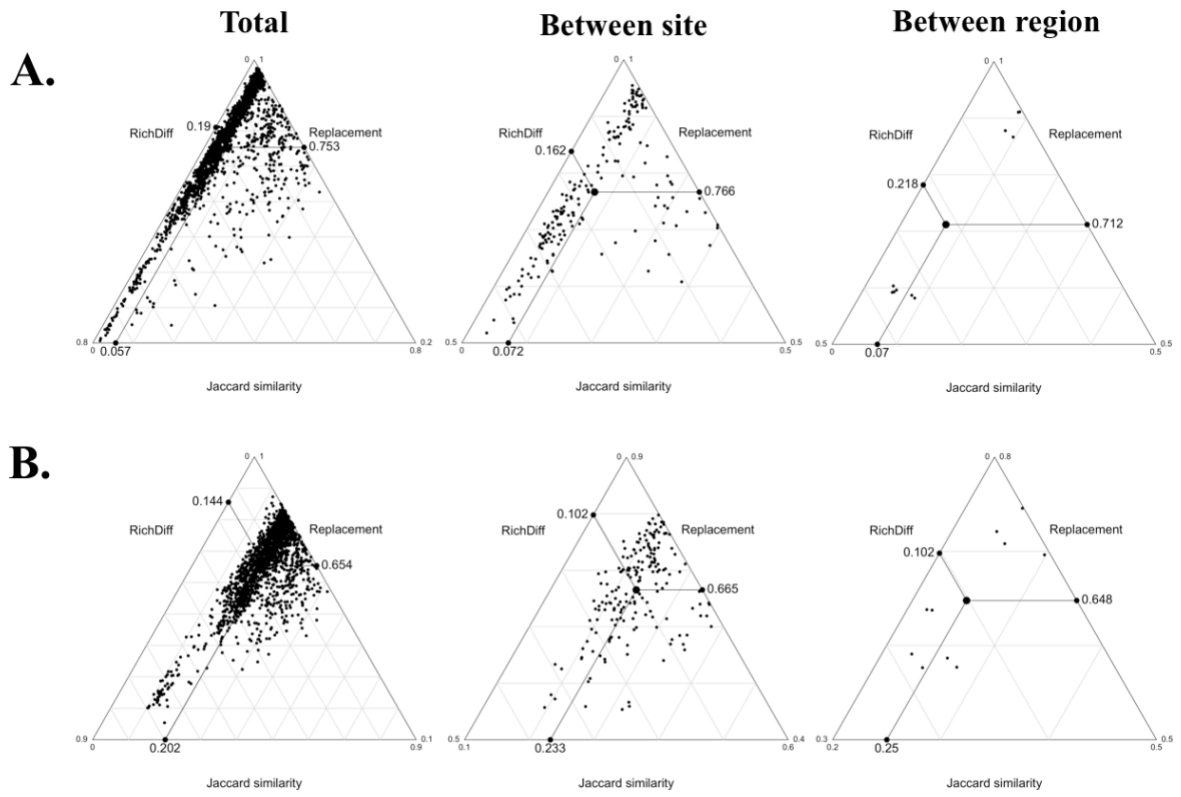


Figure 1-17. Ternary plots of Jaccard similarity and the partitions of beta diversity (replacement, richness difference, and Jaccard similarity) for 59 ARMS (Autonomous Reef Monitoring Structure) deployed across Indonesia for A) COI and B) 18S rRNA metabarcoding.

Discussion

While the Coral Triangle marine biodiversity hotspot is one of the best-known biogeographic patterns in the sea, data used to elucidate this patterns largely comes from fishes, corals and other larger metazoans (Roberts et al., 2002; Bellwood et al., 2005; Carpenter & Springer, 2005; Bellwood & Meyer, 2009) rather than smaller organisms that comprise the majority of marine biodiversity. By combining a standardized sampling approach with metabarcoding facilitated by advancements in high throughput next-generation DNA sequencing, results of this study show that cryptofauna living within the reef matrix also have biodiversity

peaking in Eastern Indonesia, the heart of the Coral Triangle, and significantly less diversity in populations to the west, both inside and outside of the Coral Triangle. These patterns were largely consistent across size fractions, suggesting that the processes structuring diversity in this region act broadly across the diversity of life. While these results show that fishes and coral diversity reflect distribution of marine biodiversity more broadly, the reverse is also true. As such, Autonomous Reef Monitoring Structure (ARMS) can be a powerful, standardized way to explore patterns and drivers of marine biodiversity.

While the origins of the Coral Triangle biodiversity hotspot may never be completely resolved (Bowen et al., 2013; Barber & Meyer, 2015), this study clearly shows that the most diverse parts of the Coral Triangle have a) more diversity per unit area, and b) greater heterogeneity and beta diversity across all spatial scales, including ARMS, sites, and regions. These novel insights are possible because of the highly standardized nature of ARMS, capturing diversity over identical surface areas, in similar habitat types for a set amount of time, something virtually impossible to obtain with other visual survey methods. Interestingly, these results were true across spatial scales and across organism size, and match patterns from large scale biodiversity studies of forests showing that high diversity forests have higher total species richness per unit area as well as higher in beta diversity and species turnover (Hubbell, 2013; Hubbell, 2015). These similarities suggest that the rules governing assembly of biodiversity may be consistent across marine and terrestrial ecosystems, as well as across flora and fauna, large and small.

Given that organisms <1 mm are largely viewed as ubiquitous and cosmopolitan it is surprising that patterns of species richness across all spatial scales and metrics of biodiversity (richness, endemic richness, beta diversity, etc), were always strongest in the 100 µm fraction.

Moreover, these patterns were observed including or excluding OTUs shared among fractions, indicating that there is strong structure and biogeography in the smallest taxa in this study, across a wide diversity of phyla. This results runs counter to the "everything is everywhere" hypothesis that predicts that the large population sizes of organisms < 1mm lead to more cosmopolitan ranges and an absence of biogeography and geographic structure (Fenchel & Finlay, 2004; Finlay, 2014), providing further evidence that this theory is not universally true (Faurby & Funch, 2011; Faurby & Barber, 2015).

Biodiversity across the Indonesian Archipelago

Although the OTU approach used in this study does not identify OTUs to species, this standard approach used in barcoding approximates species diversity (Hebert et al., 2003). As such, by combining ARMS-based marine biodiversity sampling with next generation sequencing and DNA metabarcoding, this study captured an order of magnitude more diversity (12,330 OTUs for COI and 14,350 for 18S rRNA) than traditional approaches to studying marine biodiversity, while revealing equivalent patterns of biodiversity (Roberts et al., 2002; Bellwood & Meyer, 2009; Veron et al., 2009). Discriminant function analyses showed that Aceh and Pulau Seribu were the most different faunas, as expected as Aceh is dominated by Indian Ocean taxa (Allen & Erdmann, 2021) and Pulau Seribu sits on the Sunda Shelf, a shallow marine habitat that was exposed land during the last glacial maximum (Voris, 2000). Raja Ampat and Teluk Cenderawasih, were very similar—often overlapping—as expected since both are part of the Birds Head Seascape. Cryptofauna communities from Bali were more similar to these Eastern Indonesian regions, which is expected, given that it is part of the Coral Triangle (Veron et al., 2009) and directly down in the Indonesian Throughflow (Gordon & Fine, 1996), whereas Aceh and Pulau Seribu are not.

ARMS revealed biodiversity patterns similar to fishes and corals, with 18S rRNA OTU diversity being ~35% higher in Eastern Indonesia than Bali and Western Indonesia, similar to corals (~40% higher) and fishes (~-55% higher). However, while COI had similar patterns overall, there was a much larger disparity in diversity, with Eastern Indonesia having more than twice the total number of OTUs compared to Western Indonesia. One explanation for this difference is that ARMS reveal diversity that is actually present, while other biodiversity can be overinflated due to extrapolation of range maps. Fish distribution data indicate 1,628 species in Raja Ampat, 1,103 in Teluk Cenderawasih, 1,043 in Bali, 1,137 in Pulau Seribu, and 1,050 in Aceh (Allen & Erdmann, 2021). However, numbers from Pulau Seribu values are inflated because they include species that require habitats that do not exist in the shallow, continental shelf habitats of Pulau Seribu; actual fish diversity there is likely closer to 700 species (G. Allen, Pers. Comm.). However, Maduppa et al. (2013) counted and identified over 46,000 individual fishes from Pulau Seribu and only found a total of 216 species. It is thus unclear which figure to use as a point of comparison for our data—216 actually observed, 1,137 based on extrapolated ranges, or ~700 based on ecologically informed extrapolation of ranges.

A similar pattern is seen in scleractinian corals. Western Papua, a province that includes both Raja Ampat and Teluk Cenderawasih boasts 574 scleractinian corals, while Aceh hosts less than 400 (Veron et al., 2009). However, while Veron et al. (2009) would suggest between 401- and 450 corals inhabit the waters of Pulau Seribu based on overlay of range maps, extensive surveys examining 3,500 individual corals only found 158 species in this region (Cleary et al., 2006). This difference between theoretical numbers of species presence and actual numbers, like fish, could result from over extrapolation of range maps. It could also result from challenges in identifying corals in the field with coral taxonomy more broadly (Keshavmurthy et al., 2013). In

either case, it is clear that actual surveys of biodiversity reveal much less diversity than what is predicted based on integration of range map data, providing one explanation for why diversity gradients were more pronounced in COI.

Stronger patterns in COI than 18S could result from the higher taxonomic precision of COI (Guo et al., 2015; Giebner et al., 2020). However, the higher substitution rate of COI does not just provide more power to discriminate among taxa, it also captures intraspecific genetic variation. Many phylogeographic studies in this region (Barber et al., 2002; Kirkendale & Meyer, 2004; Barber et al., 2006; DeBoer at al., 2008; Barber et al., 2011; DeBoer at al., 2014; Simmonds et al., 2018) show structure with levels of intraspecific variation in COI that can exceed our 97% identity threshold. Importantly, structure is commonly seen in Eastern Indonesia, where the Halmahera Eddy limits water and larval transport across the Maluku Sea (Hoeksema, 2007; Barber, 2009; Barber et al., 2011) leading to genetic differentiation as well as admixture of highly differentiated clades in Eastern Indonesia (Barber et al., 2006; Barber et al., 2011; DeBoer et al., 2014; Simmonds et al., 2018). As such, patterns seen in COI may reflect deep population genetic structure within populations from Raja Ampat and Teluk Cenderawasih, elevating diversity estimates.

While the reefs of Eastern Indonesia are known for their exceptional diversity of corals and fish (Turak & Souhoka, 2003; Allen, 2008; Veron et al., 2009; Turak and DeVantier in press), they are also rich with endemic species (Allen, 2008). Here too, ARMS captured endemic biodiversity patterns, with Eastern Indonesia having approximately twice as many endemic taxa. Unfortunately, we do not know what these species actually are. Unidentified OTUs comprised ~30-60% of COI, and while 18S rRNA was better, with only 2% inidentified, this was at the level of *phylum*. The lack of quality reference databases is increasingly being highlighted as an

issue for metabarcoding studies in Indonesia (Juhel et al., 2020; Marwayana et al., 2021), and is a function of limited research investment in this region (Barber et al., 2014). While improved reference databases will expand the utility of metabarcoding approaches in this region by allowing us to determine which species are actually present, the value of the OTU approach applied here is that we can examine biodiversity patterns, even if we cannot attribute a name to every OTU.

Packing of Diversity

Our results indicate that the mechanism underlying high biodiversity in the Coral Triangle include both having more species per unit areas and more variation in community composition among areas. While the Coral Triangle biodiversity hotspot clearly has more marine biodiversity than any in the world, it is unclear how so much diversity is packed into this region because diversity patterns are inferred from species range data, not actual occurrence data. Whether examining OTUs richness per individual ARMS unit or site (a set of three ARMS unit), total diversity and endemic diversity was always highest in the east and lowest in the west. Given the area-standardized nature of ARMS units, this result indicates more diversity per unit area across a broad range of marine taxa. However, beta diversity values were also high, with high taxonomic turnover among individual ARMS and sites. For example, in high diversity regions like Raja Ampat, 78% of OTUs were limited to a single site while only 22% of OTUs were shared among two or more sites. In contrast, Aceh only 61.5% of OTUs were limited to a single site, while 38.5% were shared among two or more sites.

Patterns from ARMS echo results from terrestrial forest ecosystems that sample much larger areas (e.g. 50 ha plots; Hubbell, 2013; Hubbell, 2015). For example, the richer rainforest in Amazonian Ecuador have higher richness per plot as well as higher beta diversity and species

turnover between each plot than lower diversity rainforests (Hubbell, 2013). The source of species richness is largely driven by the extraordinary numbers of very rare species (Hubbell, 2013; Hubbell, 2015), a phenomenon we could not explicitly test, because sequence abundance is not necessarily proportional to numerical abundance of individual taxa. However, the remarkable number of taxa endemic to individual ARMS units, with higher numbers in Eastern Indonesia compared to Western Indonesia suggests that these taxa are rare, and that there are more of them in the most diverse regions of the Coral Triangle.

While studies of marine biodiversity are relatively common, they typically focus on patterns of species richness and on very broad spatial scales, such as across latitude (e.g. Roy et al., 1998; Gray, 2001; Rext et al., 2005). Studies of beta diversity are much less common, but can be important for conservation (Condit et al., 2002), particularly because higher beta diversity is typically associated with greater variation in habitat types (Whittaker, 1960). Studies of beta diversity in marine ecosystems report lower beta diversity in regions with less species richness; however, like most marine biodiversity studies, these are done on broad spatial scales (Anderson et al., 2013). What is unique about the present study is that higher species richness (total and endemic) as well as beta diversity is higher in Eastern Indonesia, whether examining individual ARMS, sites, or entire regions. The similarity of these patterns to well-documented patterns in terrestrial ecosystems, and broadly across marine ecosystems, suggests that the rules that shape the packing of biodiversity act in a consistent fashion regardless of ecosystem, terrestrial or marine, or scale, ranging from a single ARMS unit to entire ocean basins.

Everything isn't everywhere

The Bass-Becking hypothesis for microbial diversity states that "everything is everywhere - the environment select" (Bass Becking, 1934). While developed originally for

microbes, this hypothesis was extended more broadly to microorganism > 1mm in size (Finlay, 2002) driven by the assumption that the exceptionally large population sizes of small organisms should limit differentiation, preventing the formation of population structure or diversity gradients (Fenchel & Finlay, 2004; O'Malley, 2008; Finlay, 2014). Observed heterogeneity in microorganism community diversity and composition is thus a function of selection, resulting from environmental differences across latitude, depth, or habitat types (Pommier et al., 2007; Fuhrman & Steele, 2008; Brown et al., 2012; Ghiglione et al., 2012; Sul et al., 2013; Kelly et al., 2014).

Surprisingly, results from ARMS metabarcoding shows that the strongest patterns of differentiation in benthic marine communities was observed in the 100 μm fraction, with weaker patterns seen in both the 500 μm and sessile fractions. This pattern was seen including all OTUs, as well excluding any OTU that was shared among size fractions, excluding the possibility that this result stems from carryover from one size to another. It was also observed in the full data set that included protists and microalgae as well as in metazoans only (Cahyani et al. Chapter 2). This pattern held examining species richness, endemic species richness beta diversity and species turnover. It held on the scale of individual ARMS, sites, and regions. The physical structure of ARMS are identical, including all parts being made by the same manufacturer. They were deployed in similar habitats and similar depths, and within individual sites, ARMS units were spaced no more than 3-5m apart. As such, it is difficult it to argue that the pronounced diversity patterns observed result from selection based on variation in the environment. One might argue that the larger organisms on ARMS create micro-habitats used by microorganisms driving selection for unique communities. However, given that the 500 μm and sessile fractions have weaker patterns of community diversity and turnover, it is unclear how they could select for

much stronger patterns in smaller organisms. Instead, the most parsimonious explanation is that cryptofauna in the Coral Triangle have pronounced biodiversity patterns.

The difference between the results of this study and previous marine studies supporting "everything is everywhere" is that the latter typically focus on organisms in the water column (Fenchel & Finlay, 2004; Cermeño & Falkowski, 2009) where small organisms are easily advected to more distant areas where environmental variation selects for unique communities. In contrast, microorganisms on ARMS are coastal and benthic. Previous studies on tardigrades in benthic coastal habitats of Southern California found pronounced population structure (Faurby & Barber, 2015). Because of their small size, small benthic marine species may have less ability to escape the boundary layer, which would limit their ability for advection on ocean currents, even in a region like Indonesia with pronounced water transport (Gordon and Fine 1996). While, further work is required to test this hypothesis, and better understand the drivers of differentiation among communities of benthic microorganisms, it is clear that in Indonesia, everything is not everywhere.

Conclusion

By applying standardized sampling methods with high throughput metabarcoding this study confirms that marine cryptofauna biodiversity patterns confirm with that of larger metazoans, showing biodiversity peaking in Eastern Indonesia, the heart of the Coral Triangle, and significantly less diversity in Western Indonesia. The exceptional diversity of Eastern Indonesian reefs, result from higher diversity per unit area as well as greater heterogeneity and beta diversity across all spatial scales, echoing results from tropical forest ecosystems (Hubbell, 2013; Hubbell, 2015). While these results provide novel insights into the distribution of marine biodiversity, there are broader implications. The reefs of the Coral Triangle are among the world's most

threatened marine ecosystems due to a variety of anthropogenic stressors, ranging from pollution to overfishing and destructive fishing practices (Burke et al., 2012). Currently, more than 85% of the reefs in the Coral Triangle are threatened by local stressors, with global threats such as ocean acidification and rising sea surface temperatures caused by climate change (Hoegh-Guldberg et al., 2009; Peñaflor et al., 2009; Hoegh-Guldberg, 2011) exacerbating these threats. Expanding understanding of the distribution of marine biodiversity and the processes shaping these communities, conservation managers will better be able to identify regions and habitat for protection, and monitor for recovery.

Supplemental Tables and Figures

Supplemental Table S1-1. Location and number of metabarcoding samples used on this study.

Country	Region/Reef	Abbr.	Latitude	Longitude	N
Indonesia	**Pulau Weh, Aceh**	**ACEH**			**33**
	Benteng	BTN	05° 50.774' N	095° 22.434' E	9
	Rubiah Sea Garden	RSG	05° 52.608' N	095° 15.596' E	9
	Seulako	SEU	05° 53.658' N	095° 15.176' E	9
	Sumur Tiga	STG	05° 53.370' N	095° 20.683' E	6
Indonesia	**Pulau Seribu, Jakarta**	**SRBU**			**34**
	Pulau Karang Beras	KBS	05° 45.574' S	106° 33.527' E	8
	Pulau Kotok	KOT	05° 41.575' S	106° 32.475' E	9
	Pulau Pramuka	PRM	05° 45.026' S	106° 36.311' E	8
	Pulau Sepa	SEP	05° 34.227' S	106° 34.491' E	9
Indonesia	**Pemuteran, Bali**	**BALI**			**35**
	Close Encouter	CEN	08° 7.675' S	114° 40.084' E	9
	Deep Middle Reef	DMR	08° 8.190' S	114° 39.570' E	9
	Horse Reef	HOR	08° 7.672' S	114° 39.337' E	9
	Napoleon Reef	NAP	08° 7.928' S	114° 40.531' E	8
Indonesia	**Raja Ampat, West Papua**	**RAMT**			**36**
	Kri	KRI	00° 33.284' S	130° 40.712' E	9
	Misool	PEF	02° 14.741' S	130° 33.438' E	9
	Pef	PEF	00° 26' S	130° 26' E	9
	Penemu	PNU	00° 34.664' S	130° 17.039' E	9
Indonesia	**Teluk Cenderawasih, West Papua**	**TCEN**			**36**
	Angromeos Island	ANG	02° 40.828' S	134° 49.515' E	9
	Manguar Cape	MGC	02° 52.866' S	134° 51.411' E	9
	Purup	PRP	02° 03.419' S	134° 09.585' E	9
	Tridacna Atoll	TRI	02° 29.948’S	134° 58.790’E	9

Supplemental Table S1-2. Total samples, sequence reads, and Operational Taxonomic Units (OTUs) for the COI datasets as revealed by DNA metabarcoding of Autonomous Reef Monitoring Structure (ARMS) from Indonesia.

COI metabarcoding dataset	All samples combined	100 µm Fraction	500 µm Fraction	Sessile Fraction
Total no. of individual samples	174	58	59	57
Total no. of individual ARMS	59	58	59	57
Total no. of individual Sites	20	20	20	20
Total no. of individual Location	5	5	5	5
Total no. of sequence reads	19,052,584	5,758,259	6,625,770	6,668,555
- Minimum no. of reads	8,826	8,826	16,711	13,344
- Maximum no. of reads	379,933	362,520	358,556	379,933
- Mean no. of reads	109,498	99,280	112,301	116,992
Total no. of OTUs	22,758	11,694	6,044	4,021
- Mean no. of OTUs	131	202	102	71
Total no. of individual samples	173	57	59	57
Rarefied even depth	11,452	11,452	11,452	11,452
Total rarefied no. of sequence reads	1,981,196	652,764	675,668	652,764
Total rarefied no. of OTUs	12,330	9,059	4,869	3,125
- Mean rarefied no. of OTUs	71	159	83	55

Supplemental Table S1-3. Total samples, sequence reads, and Operational Taxonomic Units (OTUs) for the 18S rRNA datasets as revealed by DNA metabarcoding of Autonomous Reef Monitoring Structure (ARMS) from Indonesia.

18S rRNA metabarcoding dataset	All samples combined	100 µm Fraction	500 µm Fraction	Sessile Fraction
Total no. of individual samples	174	58	59	57
Total no. of individual ARMS	59	58	59	57
Total no. of individual Sites	20	20	20	20
Total no. of individual Location	5	5	5	5
Total no. of sequence reads	46,633,073	13,327,050	12,547,558	20,758,465
- Minimum no. of reads	20,549	59,071	20,549	146,362
- Maximum no. of reads	764,712	400,884	375,973	764,712
- Mean no. of reads	268,006.2	229,776.7	212,670.5	364,183.6
Total no. of OTUs	36,956	13,779	8,522	15,645
- Mean no. of OTUs	212.4	237.6	144.4	274.5
Rarefied even depth	20,549	20,549	20,549	20,549
Total rarefied no. of sequence reads	3,575,526	1,191,842	1,212,391	1,171,293
Total rarefied no. of OTUs	14,350	9,265	5,822	7,083
- Mean rarefied no. of OTUs	82.5	159.7	98.7	124.3

Supplemental Table S1-4. Alpha diversity indices (ANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers, rarefied to a standardized number of OTUs per ARMS to account for variation in sequencing depth.

Diversity	**ANOVA - COI**				
	Df	**Sum Sq**	**Mean Sq**	**F Value**	***p-value***
Average local diversity per ARMS	4	0.0234	0.005859	0.562	0.691
OTUs richness per ARMS	4	1036915	259229	22.41	**6.03E-11**
Percent regional diversity per ARMS	4	0.03254	0.008136	5.122	**0.00142**
OTUs per site	4	1678418	419604	23.27	**2.75E-06**
Percent regional diversity per site	4	0.02141	0.005353	4.127	**0.0205**
Endemic OTUs per ARMS	4	246553	61638	9.224	**9.36E-06**
Endemic OTUs per site	4	1051320	262830	17.75	**1.47E-05**
Diversity	**ANOVA - 18S rRNA**				
	Df	**Sum Sq**	**Mean Sq**	**F Value**	***p-value***
Average local diversity per ARMS	4	0.0357	0.008934	0.961	0.436
OTUs richness per ARMS	4	947562	236891	5.115	**0.00144**
Percent regional diversity per ARMS	4	0.02622	0.006555	3.254	**0.0183**
OTUs per site	4	1253341	313335	6.364	**0.00336**
Percent regional diversity per site	4	0.0142	0.003549	2.879	0.0623
Endemic OTUs per ARMS	4	130869	32717	1.383	0.252
Endemic OTUs per site	4	284781	71195	3.299	**0.0397**

Supplemental Table S1-5. Tukey HSD post hoc tests pairwise comparisons on eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers.

Region	COI				18S rRNA			
	OTUs richness per ARMS				OTUs richness per ARMS			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj
BALI-ACEH	-16.485	-143.187	110.217	0.99601	-125.750	-379.262	127.762	0.63042
RAMT-ACEH	200.348	73.647	327.050	**0.00039**	42.417	-211.095	295.929	0.98954
SRBU-ACEH	-28.568	-155.270	98.134	0.96843	-35.250	-288.762	218.262	0.99484
TCEN-ACEH	293.515	166.813	420.217	**0.00000**	252.500	-1.012	506.012	**0.05140**
RAMT-BALI	216.833	92.917	340.750	**0.00008**	168.167	-79.773	416.106	0.32244
SRBU-BALI	-12.083	-136.000	111.833	0.99870	90.500	-157.439	338.440	0.84033
TCEN-BALI	310.000	186.083	433.917	**0.00000**	378.250	130.311	626.190	**0.00065**
SRBU-RAMT	-228.917	-352.833	-105.000	**0.00003**	-77.667	-325.606	170.273	0.90157
TCEN-RAMT	93.167	-30.750	217.083	0.22597	210.083	-37.856	458.023	0.13339
TCEN-SRBU	322.083	198.167	446.000	**0.00000**	287.750	39.811	535.690	**0.01523**
Region	**Percent regional diversity per ARMS**				**Percent regional diversity per ARMS**			
	diff	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**
BALI-ACEH	-0.032	-0.079	0.015	0.30509	-0.043	-0.096	0.010	0.15955
RAMT-ACEH	-0.069	-0.116	-0.022	**0.00109**	-0.064	-0.117	-0.011	**0.01050**
SRBU-ACEH	-0.019	-0.066	0.028	0.77707	-0.024	-0.077	0.029	0.70709
TCEN-ACEH	-0.048	-0.095	-0.001	**0.04044**	-0.025	-0.078	0.027	0.65620
RAMT-BALI	-0.037	-0.083	0.009	0.17784	-0.021	-0.072	0.031	0.79093
SRBU-BALI	0.013	-0.033	0.059	0.92576	0.019	-0.032	0.071	0.83022
TCEN-BALI	-0.016	-0.062	0.030	0.86281	0.018	-0.034	0.069	0.86903
SRBU-RAMT	0.050	0.004	0.096	**0.02721**	0.040	-0.012	0.092	0.20309
TCEN-RAMT	0.021	-0.025	0.067	0.71153	0.038	-0.013	0.090	0.23741
TCEN-SRBU	-0.029	-0.075	0.017	0.38827	-0.002	-0.053	0.050	0.99999

(continued)

Supplemental Table S1-5 *(continued)*

Region	COI				18S rRNA			
	OTUs per site				OTUs per site			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj
BALI-ACEH	53.250	-239.948	346.448	0.97881	17.000	-467.511	501.511	0.99996
RAMT-ACEH	538.750	245.552	831.948	**0.00036**	389.250	-95.261	873.761	0.14752
SRBU-ACEH	-7.500	-300.698	285.698	0.99999	64.750	-419.761	549.261	0.99326
TCEN-ACEH	660.750	367.552	953.948	**0.00004**	633.500	148.989	1118.011	**0.00809**
RAMT-BALI	485.500	192.302	778.698	**0.00103**	372.250	-112.261	856.761	0.17650
SRBU-BALI	-60.750	-353.948	232.448	0.96595	47.750	-436.761	532.261	0.99791
TCEN-BALI	607.500	314.302	900.698	**0.00010**	616.500	131.989	1101.011	**0.00999**
SRBU-RAMT	-546.250	-839.448	-253.052	**0.00032**	-324.500	-809.011	160.011	0.28296
TCEN-RAMT	122.000	-171.198	415.198	0.70391	244.250	-240.261	728.761	0.54450
TCEN-SRBU	668.250	375.052	961.448	**0.00003**	568.750	84.239	1053.261	**0.01801**
Region	**Percent regional diversity per site**				**Percent regional diversity per site**			
	diff	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**
BALI-ACEH	-0.026	-0.111	0.060	0.87924	-0.059	-0.142	0.025	0.23897
RAMT-ACEH	-0.095	-0.181	-0.010	**0.02614**	-0.089	-0.173	-0.006	0.03348
SRBU-ACEH	-0.027	-0.113	0.059	0.85870	-0.049	-0.133	0.034	0.39202
TCEN-ACEH	-0.069	-0.155	0.016	0.14024	-0.043	-0.127	0.040	0.51113
RAMT-BALI	-0.070	-0.149	0.010	0.09766	-0.031	-0.108	0.047	0.73117
SRBU-BALI	-0.001	-0.081	0.078	1.00000	0.009	-0.068	0.087	0.99500
TCEN-BALI	-0.044	-0.123	0.036	0.45370	0.015	-0.062	0.093	0.96955
SRBU-RAMT	0.068	-0.011	0.148	0.10703	0.040	-0.037	0.118	0.51056
TCEN-RAMT	0.026	-0.053	0.105	0.84250	0.046	-0.031	0.123	0.38267
TCEN-SRBU	-0.042	-0.122	0.037	0.48304	0.006	-0.071	0.083	0.99919

(continued)

Supplemental Table S1-5 *(continued)*

Region	COI				18S rRNA			
	Endemic OTUs per ARMS				Endemic OTUs per ARMS			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj
BALI-ACEH	2.879	-93.416	99.174	0.99999	-5.939	-187.104	175.226	0.99998
RAMT-ACEH	120.045	23.751	216.340	**0.00762**	66.061	-115.104	247.226	0.84082
SRBU-ACEH	-22.038	-118.333	74.257	0.96669	-33.773	-214.938	147.392	0.98430
TCEN-ACEH	127.462	31.167	223.757	**0.00398**	90.144	-91.021	271.309	0.62767
RAMT-BALI	117.167	22.988	211.345	**0.00778**	72.000	-105.183	249.183	0.78097
SRBU-BALI	-24.917	-119.095	69.262	0.94436	-27.833	-205.016	149.349	0.99177
TCEN-BALI	124.583	30.405	218.762	**0.00401**	96.083	-81.099	273.266	0.54772
SRBU-RAMT	-142.083	-236.262	-47.905	**0.00076**	-99.833	-277.016	77.349	0.51010
TCEN-RAMT	7.417	-86.762	101.595	0.99944	24.083	-153.099	201.266	0.99527
TCEN-SRBU	149.500	55.322	243.678	**0.00037**	123.917	-53.266	301.099	0.29256
Region	**OTUs per site**				**OTUs per site**			
	diff	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**
BALI-ACEH	53.250	-239.948	346.448	0.97881	17.000	-467.511	501.511	0.99996
RAMT-ACEH	538.750	245.552	831.948	**0.00036**	389.250	-95.261	873.761	0.14752
SRBU-ACEH	-7.500	-300.698	285.698	0.99999	64.750	-419.761	549.261	0.99326
TCEN-ACEH	660.750	367.552	953.948	**0.00004**	633.500	148.989	1118.011	**0.00809**
RAMT-BALI	485.500	192.302	778.698	**0.00103**	372.250	-112.261	856.761	0.17650
SRBU-BALI	-60.750	-353.948	232.448	0.96595	47.750	-436.761	532.261	0.99791
TCEN-BALI	607.500	314.302	900.698	**0.00010**	616.500	131.989	1101.011	**0.00999**
SRBU-RAMT	-546.250	-839.448	-253.052	**0.00032**	-324.500	-809.011	160.011	0.28296
TCEN-RAMT	122.000	-171.198	415.198	0.70391	244.250	-240.261	728.761	0.54450
TCEN-SRBU	668.250	375.052	961.448	**0.00003**	568.750	84.239	1053.261	**0.01801**

Supplemental Table S1-6. Alpha diversity indices (ANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers, without rarefaction (e.g. raw OTU data).

Diversity	ANOVA - COI				
	Df	**Sum Sq**	**Mean Sq**	**F Value**	**p-value**
Average local diversity per ARMS	4	0.0189	0.004733	0.245	0.911
OTUs richness per ARMS	4	3583484	895871	26.47	**2.90E-12**
Percent regional diversity per ARMS	4	0.02056	0.005139	1.799	0.142
OTUs per site	4	5365211	1341303	38.15	**1.08E-07**
Percent regional diversity per site	4	0.02166	0.005416	0.521	0.722
Endemic OTUs per ARMS	4	835627	208907	11.13	**1.10E-06**
Endemic OTUs per site	4	7820832	1955208	3.675	**0.0281**
Diversity	**ANOVA - 18S rRNA**				
	Df	**Sum Sq**	**Mean Sq**	**F Value**	**p-value**
Average local diversity per ARMS	4	0.0166	0.004157	0.23	0.921
OTUs richness per ARMS	4	5338866	1334716	6.406	**0.000263**
Percent regional diversity per ARMS	4	0.01779	0.004447	1.335	0.269
OTUs per site	4	7203058	1800764	5.868	**0.00477**
Percent regional diversity per site	4	0.01948	0.00487	0.367	0.829
Endemic OTUs per ARMS	4	1228196	307049	2.52	**0.0515**
Endemic OTUs per site	4	2429163	607291	2.413	0.0949

Supplemental Table S1-7. Tukey HSD post hoc tests pairwise comparisons on eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers. The ANOVA were calculated without rarefication (e.g. raw data).

Region	COI				18S rRNA			
	OTUs richness per ARMS				OTUs richness per ARMS			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj
BALI-ACEH	53.917	-157.913	265.747	0.95152	23.917	-501.633	549.466	0.99994
RAMT-ACEH	442.417	230.587	654.247	**0.00000**	457.167	-68.383	982.716	0.11668
SRBU-ACEH	39.083	-172.747	250.914	0.98494	62.750	-462.800	588.300	0.99715
TCEN-ACEH	595.000	383.170	806.830	**0.00000**	753.750	228.200	1279.300	**0.00149**
RAMT-BALI	388.500	176.670	600.330	**0.00003**	433.250	-92.300	958.800	0.15255
SRBU-BALI	-14.833	-226.663	196.997	0.99965	38.833	-486.716	564.383	0.99957
TCEN-BALI	541.083	329.253	752.914	**0.00000**	729.833	204.284	1255.383	**0.00224**
SRBU-RAMT	-403.333	-615.163	-191.503	**0.00002**	-394.417	-919.966	131.133	0.22784
TCEN-RAMT	152.583	-59.247	364.414	0.26498	296.583	-228.966	822.133	0.50905
TCEN-SRBU	555.917	344.087	767.747	**0.00000**	691.000	165.450	1216.550	**0.00427**
Region	**OTUs per site**				**OTUs per site**			
	diff	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**
BALI-ACEH	82.000	-327.427	491.427	0.96983	130.750	-1078.805	1340.305	0.99701
RAMT-ACEH	1006.500	597.073	1415.927	**0.00001**	1156.000	-53.555	2365.555	0.06437
SRBU-ACEH	37.250	-372.177	446.677	0.99847	34.250	-1175.305	1243.805	0.99999
TCEN-ACEH	1173.750	764.323	1583.177	**0.00000**	1380.500	170.945	2590.055	**0.02185**
RAMT-BALI	924.500	515.073	1333.927	**0.00004**	1025.250	-184.305	2234.805	0.11685
SRBU-BALI	-44.750	-454.177	364.677	0.99688	-96.500	-1306.055	1113.055	0.99909
TCEN-BALI	1091.750	682.323	1501.177	**0.00001**	1249.750	40.195	2459.305	**0.04126**
SRBU-RAMT	-969.250	-1378.677	-559.823	**0.00002**	-1121.750	-2331.305	87.805	0.07549
TCEN-RAMT	167.250	-242.177	576.677	0.71727	224.500	-985.055	1434.055	0.97707
TCEN-SRBU	1136.500	727.073	1545.927	**0.00000**	1346.250	136.695	2555.805	**0.02585**

(continued)

Supplemental Table S1-7 (*continued)*

Region	COI				18S rRNA			
	Endemic OTUs per ARMS				Endemic OTUs per ARMS			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj
BALI-ACEH	26.250	-131.510	184.010	0.98979	60.000	-341.948	461.948	0.99325
RAMT-ACEH	253.750	95.990	411.510	**0.00029**	331.083	-70.864	733.031	0.15315
SRBU-ACEH	7.167	-150.593	164.926	0.99994	-11.750	-413.698	390.198	0.99999
TCEN-ACEH	249.000	91.240	406.760	**0.00039**	271.500	-130.448	673.448	0.32692
RAMT-BALI	227.500	69.740	385.260	**0.00139**	271.083	-130.864	673.031	0.32845
SRBU-BALI	-19.083	-176.843	138.676	0.99700	-71.750	-473.698	330.198	0.98669
TCEN-BALI	222.750	64.990	380.510	**0.00182**	211.500	-190.448	613.448	0.57704
SRBU-RAMT	-246.583	-404.343	-88.824	**0.00045**	-342.833	-744.781	59.114	0.12917
TCEN-RAMT	-4.750	-162.510	153.010	0.99999	-59.583	-461.531	342.364	0.99342
TCEN-SRBU	241.833	84.074	399.593	**0.00060**	283.250	-118.698	685.198	0.28565
Region	**Endemic OTUs per site**				**Endemic OTUs per site**			
	diff	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**
BALI-ACEH	56.500	-1536.163	1649.163	0.99996	147.000	-948.477	1242.477	0.99315
RAMT-ACEH	808.750	-783.913	2401.413	0.53787	929.250	-166.227	2024.727	0.11645
SRBU-ACEH	39.000	-1553.663	1631.663	0.99999	39.250	-1056.227	1134.727	0.99996
TCEN-ACEH	1595.750	3.087	3188.413	**0.04945**	478.000	-617.477	1573.477	0.66776
RAMT-BALI	752.250	-840.413	2344.913	0.60229	782.250	-313.227	1877.727	0.23029
SRBU-BALI	-17.500	-1610.163	1575.163	1.00000	-107.750	-1203.227	987.727	0.99793
TCEN-BALI	1539.250	-53.413	3131.913	0.06057	331.000	-764.477	1426.477	0.87964
SRBU-RAMT	-769.750	-2362.413	822.913	0.58224	-890.000	-1985.477	205.477	0.14074
TCEN-RAMT	787.000	-805.663	2379.663	0.56254	-451.250	-1546.727	644.227	0.71127
TCEN-SRBU	1556.750	-35.913	3149.413	0.05690	438.750	-656.727	1534.227	0.73114

Supplemental Table S1-8. Alpha diversity indices (ANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers. Data were analysed per size fraction (100 μm, 500 μm, and sessile fraction) using the rarefied datasets to account for variation in sequencing depth.

Diversity	ANOVA - COI											
	100 μm Fraction				500 μm Fraction				Sessile Fraction			
	Sum Sq	**Mean Sq**	**F Value**	***p-value***	**Sum Sq**	**Mean Sq**	**F Value**	***p-value***	**Sum Sq**	**Mean Sq**	**F Value**	***p-value***
OTUs richness per ARMS	711031	177758	19.51	**4.84E-10**	50047	12512	4.456	**0.00343**	56767	14192	5.992	**0.000447**
OTUs per site	1199422	299855	38.35	**1.04E-07**	108708	27177	3.57	**0.0309**	60568	15142	3.072	**0.0492**
Endemic OTUs per ARMS	234047	58512	11.38	**8.45E-07**	24269	6067	3.779	**0.00872**	5736	1434	1.083	0.374
Endemic OTUs per site	790500	197625	26	**1.36E-06**	90000	22500	4.655	**0.0121**	13712	3428	1.253	0.331
Diversity	**ANOVA - 18S rRNA**											
	100 μm Fraction				**500 μm Fraction**				**Sessile Fraction**			
	Sum Sq	**Mean Sq**	**F Value**	***p-value***	**Sum Sq**	**Mean Sq**	**F Value**	***p-value***	**Sum Sq**	**Mean Sq**	**F Value**	***p-value***
OTUs richness per ARMS	984919	246230	3.961	**0.00678**	296277	74069	3.534	**0.0123**	242389	60597	2.055	0.0992
OTUs per site	1144264	286066	11.27	**2.00E-04**	320553	80138	1.852	0.172	327199	81800	1.514	0.248
Endemic OTUs per ARMS	125943	31486	1.301	0.281	22535	5634	0.702	0.594	69095	17274	1.912	0.121
Endemic OTUs per site	126808	31702	5.369	**0.00691**	10793	2698	0.158	0.956	112306	28077	1.146	0.373

Supplemental Table S1-9. Tukey HSD post hoc tests pairwise comparisons on eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers. Data were analyzed per size fraction (100 µm, 500 µm, and sessile fraction).

Region	OTUs richness per ARMS											
	100 µm Fraction - COI				500 µm Fraction - COI				Sessile Fraction - COI			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj	diff	lwr	upr	p adj
BALI-ACEH	6.50	-103.41	116.41	0.99982	-27.92	-88.93	33.09	0.69808	16.50	-39.53	72.53	0.92001
RAMT-ACEH	170.08	60.18	279.99	**0.00053**	29.58	-31.43	90.59	0.65064	32.08	-23.95	88.12	0.49456
SRBU-ACEH	-13.08	-122.99	96.82	0.99718	-5.00	-66.01	56.01	0.99935	17.42	-38.62	73.45	0.90428
TCEN-ACEH	255.33	145.43	365.24	**0.00000**	55.00	-6.01	116.01	0.09600	89.08	33.05	145.12	**0.00035**
RAMT-BALI	163.58	53.68	273.49	**0.00091**	57.50	-3.51	118.51	0.07377	15.58	-40.45	71.62	0.93409
SRBU-BALI	-19.58	-129.49	90.32	0.98678	22.92	-38.09	83.93	0.82625	0.92	-55.12	56.95	1.00000
TCEN-BALI	248.83	138.93	358.74	**0.00000**	82.92	21.91	143.93	**0.00291**	72.58	16.55	128.62	**0.00505**
SRBU-RAMT	-183.17	-293.07	-73.26	**0.00017**	-34.58	-95.59	26.43	0.50462	-14.67	-70.70	41.37	0.94654
TCEN-RAMT	85.25	-24.66	195.16	0.19972	25.42	-35.59	86.43	0.76548	57.00	0.97	113.03	**0.04432**
TCEN-SRBU	268.42	158.51	378.32	**0.00000**	60.00	-1.01	121.01	0.05605	71.67	15.63	127.70	**0.00579**
Region	**OTUs per site**											
	diff	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**
BALI-ACEH	25.25	-167.81	218.31	0.99379	-65.75	-256.25	124.75	0.82078	34.50	-118.80	187.80	0.95449
RAMT-ACEH	430.00	236.94	623.06	**0.00004**	97.50	-93.00	288.00	0.53065	90.50	-62.80	243.80	0.39702
SRBU-ACEH	-41.75	-234.81	151.31	0.96041	-35.50	-226.00	155.00	0.97673	51.00	-102.30	204.30	0.83896
TCEN-ACEH	544.00	350.94	737.06	**0.00000**	122.00	-68.50	312.50	0.32221	160.75	7.45	314.05	**0.03773**
RAMT-BALI	404.75	211.69	597.81	**0.00009**	163.25	-27.25	353.75	0.11114	56.00	-97.30	209.30	0.78967
SRBU-BALI	-67.00	-260.06	126.06	0.81795	30.25	-160.25	220.75	0.98709	16.50	-136.80	169.80	0.99706
TCEN-BALI	518.75	325.69	711.81	**0.00000**	187.75	-2.75	378.25	**0.05431**	126.25	-27.05	279.55	0.13277
SRBU-RAMT	-471.75	-664.81	-278.69	**0.00002**	-133.00	-323.50	57.50	0.24827	-39.50	-192.80	113.80	0.92790
TCEN-RAMT	114.00	-79.06	307.06	0.39682	24.50	-166.00	215.00	0.99417	70.25	-83.05	223.55	0.62797
TCEN-SRBU	585.75	392.69	778.81	**0.00000**	157.50	-33.00	348.00	0.13053	109.75	-43.55	263.05	0.22826

(continued)

Supplemental Table S1-9 (*continued)*

Region	Endemic OTUs per ARMS							
	100 µm Fraction - COI				500 µm Fraction - COI			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj
BALI-ACEH	10.58	-71.98	93.15	0.99624	-15.25	-61.38	30.88	0.88309
RAMT-ACEH	124.08	41.52	206.65	**0.00080**	28.50	-17.63	74.63	0.41726
SRBU-ACEH	-11.75	-94.31	70.81	0.99437	-14.50	-60.63	31.63	0.90070
TCEN-ACEH	128.42	45.85	210.98	**0.00049**	30.67	-15.47	76.80	0.34285
RAMT-BALI	113.50	30.94	196.06	**0.00254**	43.75	-2.38	89.88	0.07096
SRBU-BALI	-22.33	-104.90	60.23	0.94008	0.75	-45.38	46.88	1.00000
TCEN-BALI	117.83	35.27	200.40	**0.00159**	45.92	-0.22	92.05	**0.05165**
SRBU-RAMT	-135.83	-218.40	-53.27	**0.00021**	-43.00	-89.13	3.13	0.07894
TCEN-RAMT	4.33	-78.23	86.90	0.99989	2.17	-43.97	48.30	0.99993
TCEN-SRBU	140.17	57.60	222.73	**0.00012**	45.17	-0.97	91.30	**0.05775**
Region	**Endemic OTUs per site**							
	diff	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**
BALI-ACEH	25.75	-164.62	216.12	0.99294	-18.75	-170.56	133.06	0.99501
RAMT-ACEH	402.00	211.63	592.37	**0.00008**	144.25	-7.56	296.06	0.06641
SRBU-ACEH	-11.00	-201.37	179.37	0.99974	-5.00	-156.81	146.81	0.99997
TCEN-ACEH	417.75	227.38	608.12	**0.00005**	108.75	-43.06	260.56	0.22779
RAMT-BALI	376.25	185.88	566.62	**0.00017**	163.00	11.19	314.81	**0.03258**
SRBU-BALI	-36.75	-227.12	153.62	0.97356	13.75	-138.06	165.56	0.99850
TCEN-BALI	392.00	201.63	582.37	**0.00011**	127.50	-24.31	279.31	0.12179
SRBU-RAMT	-413.00	-603.37	-222.63	**0.00006**	-149.25	-301.06	2.56	0.05508
TCEN-RAMT	15.75	-174.62	206.12	0.99895	-35.50	-187.31	116.31	0.94805
TCEN-SRBU	428.75	238.38	619.12	**0.00004**	113.75	-38.06	265.56	0.19406

(continued)

Supplemental Table S1-9 (*continued)*

Region	OTUs richness per ARMS							
	100 µm Fraction - 18S rRNA				500 µm Fraction - 18S rRNA			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj
BALI-ACEH	-82.58	-369.66	204.50	0.92606	-172.00	-338.70	-5.30	0.03997
RAMT-ACEH	-14.50	-301.58	272.58	0.99990	-148.08	-314.78	18.62	0.10422
SRBU-ACEH	-57.08	-344.16	230.00	0.98011	-25.33	-192.03	141.37	0.99277
TCEN-ACEH	273.17	-13.91	560.25	0.06948	-32.00	-198.70	134.70	0.98254
RAMT-BALI	68.08	-219.00	355.16	0.96225	23.92	-142.78	190.62	0.99420
SRBU-BALI	25.50	-261.58	312.58	0.99910	146.67	-20.03	313.37	0.10982
TCEN-BALI	355.75	68.67	642.83	**0.00807**	140.00	-26.70	306.70	0.13952
SRBU-RAMT	-42.58	-329.66	244.50	0.99341	122.75	-43.95	289.45	0.24477
TCEN-RAMT	287.67	0.59	574.75	0.04930	116.08	-50.62	282.78	0.29708
TCEN-SRBU	330.25	43.17	617.33	0.01646	-6.67	-173.37	160.03	0.99996
Region	**OTUs per site**				**Endemic OTUs per site**			
	diff	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**
BALI-ACEH	-81.50	-429.32	266.32	0.94769	13.75	-154.04	181.54	0.99899
RAMT-ACEH	87.75	-260.07	435.57	0.93280	149.75	-18.04	317.54	0.09160
SRBU-ACEH	-195.00	-542.82	152.82	0.44561	-16.25	-184.04	151.54	0.99805
TCEN-ACEH	503.50	155.68	851.32	0.00350	170.25	2.46	338.04	**0.04594**
RAMT-BALI	169.25	-178.57	517.07	0.57624	136.00	-31.79	303.79	0.14212
SRBU-BALI	-113.50	-461.32	234.32	0.84791	-30.00	-197.79	137.79	0.97999
TCEN-BALI	585.00	237.18	932.82	**0.00088**	156.50	-11.29	324.29	0.07327
SRBU-RAMT	-282.75	-630.57	65.07	0.14038	-166.00	-333.79	1.79	**0.05315**
TCEN-RAMT	415.75	67.93	763.57	**0.01585**	20.50	-147.29	188.29	0.99521
TCEN-SRBU	698.50	350.68	1046.32	**0.00014**	186.50	18.71	354.29	**0.02607**

Supplemental Table S1-10. Alpha diversity indices (ANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers. Data were analysed per size fraction (100 μm, 500 μm, and sessile fraction). ANOVA were generated from rarefied data after excluding OTUs among size fraction to account for the possibility of carryover during ARMS sample processing.

Diversity	ANOVA - COI											
	100 µm Fraction				500 µm Fraction				Sessile Fraction			
	Sum Sq	Mean Sq	F Value	*p-value*	Sum Sq	Mean Sq	F Value	*p-value*	Sum Sq	Mean Sq	F Value	*p-value*
OTUs richness per ARMS	546217	136554	21.69	**8.77E-11**	29111	7278	4.379	**0.00382**	16814	4204	6.174	**0.000354**
OTUs per site	949593	237398	28.8	**6.99E-07**	74355	18589	3.337	**0.0383**	27383	6846	3.54	**0.0317**
Endemic OTUs per ARMS	188228	47057	11.78	**5.58E-07**	15549	3887	3.042	**0.0245**	4975	1244	2.569	**0.0479**
Endemic OTUs per site	696681	174170	23.68	**2.47E-06**	64353	16088	3.891	**0.0232**	11395	2849	1.943	0.156
Diversity	**ANOVA - 18S rRNA**											
	100 µm Fraction				**500 µm Fraction**				**Sessile Fraction**			
	Sum Sq	**Mean Sq**	**F Value**	***p-value***	**Sum Sq**	**Mean Sq**	**F Value**	***p-value***	**Sum Sq**	**Mean Sq**	**F Value**	***p-value***
OTUs richness per ARMS	234034	58509	5.009	**0.00163**	16428	4107	2.497	**0.0531**	67391	16848	3.682	**0.00999**
OTUs per site	389856	97464	10.4	**0.000307**	27705	6926	1.142	0.375	150168	37542	2.216	0.116
Endemic OTUs per ARMS	68289	17072	2.444	0.0573	5164	1291	1.118	0.358	39540	9885	3.119	**0.022**
Endemic OTUs per site	104597	26149	7.969	**0.00119**	7700	1925	0.51	0.729	89324	22331	1.848	0.172

Supplemental Table S1-11. Tukey HSD post hoc tests pairwise comparisons on eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using COI and 18S rRNA markers. Data were analysed per size fraction (100µm, 500µm, and sessile fraction). Post hoc tests were generated from rarefied data after excluding the shared OTUs between size fraction.

Region	OTUs richness per ARMS											
	100 µm Fraction - COI				500 µm Fraction - COI				Sessile Fraction - COI			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj	diff	lwr	upr	p adj
BALI-ACEH	4.08	-87.27	95.44	0.99994	-22.33	-69.27	24.61	0.66663	19.75	-10.29	49.79	0.35409
RAMT-ACEH	158.50	67.15	249.85	**0.00009**	35.17	-11.77	82.11	0.22939	33.33	3.29	63.38	**0.02256**
SRBU-ACEH	-18.92	-110.27	72.44	0.97691	-11.92	-58.86	35.02	0.95196	15.08	-14.96	45.13	0.62025
TCEN-ACEH	213.33	121.98	304.69	**0.00000**	26.42	-20.52	73.36	0.51183	49.33	19.29	79.38	**0.00021**
RAMT-BALI	154.42	63.06	245.77	**0.00013**	57.50	10.56	104.44	**0.00907**	13.58	-16.46	43.63	0.70737
SRBU-BALI	-23.00	-114.35	68.35	0.95334	10.42	-36.52	57.36	0.97028	-4.67	-34.71	25.38	0.99214
TCEN-BALI	209.25	117.90	300.60	**0.00000**	48.75	1.81	95.69	**0.03809**	29.58	-0.46	59.63	0.05559
SRBU-RAMT	-177.42	-268.77	-86.06	**0.00001**	-47.08	-94.02	-0.14	**0.04896**	-18.25	-48.29	11.79	0.43455
TCEN-RAMT	54.83	-36.52	146.19	0.44673	-8.75	-55.69	38.19	0.98435	16.00	-14.04	46.04	0.56567
TCEN-SRBU	232.25	140.90	323.60	**0.00000**	38.33	-8.61	85.27	0.15944	34.25	4.21	64.29	**0.01785**
Region	**OTUs per site**											
	diff	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**
BALI-ACEH	15.00	-183.25	213.25	0.99926	-53.00	-215.97	109.97	0.84944	41.00	-55.02	137.02	0.68453
RAMT-ACEH	395.00	196.75	593.25	**0.00016**	94.75	-68.22	257.72	0.41136	74.00	-22.02	170.02	0.17438
SRBU-ACEH	-37.25	-235.50	161.00	0.97602	-45.50	-208.47	117.47	0.90625	38.75	-57.27	134.77	0.72584
TCEN-ACEH	471.50	273.25	669.75	**0.00002**	76.00	-86.97	238.97	0.61321	110.00	13.98	206.02	**0.02130**
RAMT-BALI	380.00	181.75	578.25	**0.00023**	147.75	-15.22	310.72	0.08474	33.00	-63.02	129.02	0.82295
SRBU-BALI	-52.25	-250.50	146.00	0.92230	7.50	-155.47	170.47	0.99990	-2.25	-98.27	93.77	0.99999
TCEN-BALI	456.50	258.25	654.75	**0.00003**	129.00	-33.97	291.97	0.15681	69.00	-27.02	165.02	0.22534
SRBU-RAMT	-432.25	-630.50	-234.00	**0.00006**	-140.25	-303.22	22.72	0.10898	-35.25	-131.27	60.77	0.78677
TCEN-RAMT	76.50	-121.75	274.75	0.75603	-18.75	-181.72	144.22	0.99620	36.00	-60.02	132.02	0.77414
TCEN-SRBU	508.75	310.50	707.00	**0.00001**	121.50	-41.47	284.47	0.19769	71.25	-24.77	167.27	0.20111

(continued)

Supplemental Table S1-11 (*continued)*

Region	Endemic OTUs per ARMS											
	100 µm Fraction - COI				500 µm Fraction - COI				Sessile Fraction - COI			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj	diff	lwr	upr	p adj
BALI-ACEH	7.75	-65.03	80.53	0.99817	-12.92	-54.07	28.24	0.90119	9.08	-16.25	34.42	0.84911
RAMT-ACEH	112.25	39.47	185.03	**0.00055**	24.33	-16.82	65.49	0.46210	17.92	-7.42	43.25	0.28220
SRBU-ACEH	-6.75	-79.53	66.03	0.99894	-9.67	-50.82	31.49	0.96353	11.25	-14.08	36.58	0.72084
TCEN-ACEH	116.25	43.47	189.03	**0.00033**	24.00	-17.16	65.16	0.47610	27.25	1.92	52.58	**0.02905**
RAMT-BALI	104.50	31.72	177.28	**0.00147**	37.25	-3.91	78.41	0.09385	8.83	-16.50	34.17	0.86161
SRBU-BALI	-14.50	-87.28	58.28	0.97996	3.25	-37.91	44.41	0.99944	2.17	-23.17	27.50	0.99923
TCEN-BALI	108.50	35.72	181.28	**0.00089**	36.92	-4.24	78.07	0.09875	18.17	-7.17	43.50	0.26912
SRBU-RAMT	-119.00	-191.78	-46.22	**0.00023**	-34.00	-75.16	7.16	0.15104	-6.67	-32.00	18.67	0.94552
TCEN-RAMT	4.00	-68.78	76.78	0.99987	-0.33	-41.49	40.82	1.00000	9.33	-16.00	34.67	0.83608
TCEN-SRBU	123.00	50.22	195.78	**0.00013**	33.67	-7.49	74.82	0.15818	16.00	-9.33	41.33	0.39467
Region	**Endemic OTUs per site**											
	diff	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**
BALI-ACEH	20.00	-167.25	207.25	0.99715	-22.50	-162.91	117.91	0.98664	14.50	-69.12	98.12	0.98211
RAMT-ACEH	374.75	187.50	562.00	0.00015	120.50	-19.91	260.91	0.11040	60.75	-22.87	144.37	0.21692
SRBU-ACEH	-10.00	-197.25	177.25	0.99981	-8.25	-148.66	132.16	0.99973	24.50	-59.12	108.12	0.89067
TCEN-ACEH	392.50	205.25	579.75	0.00009	83.75	-56.66	224.16	0.38739	57.25	-26.37	140.87	0.26433
RAMT-BALI	354.75	167.50	542.00	0.00027	143.00	2.59	283.41	0.04495	46.25	-37.37	129.87	0.45832
SRBU-BALI	-30.00	-217.25	157.25	0.98665	14.25	-126.16	154.66	0.99766	10.00	-73.62	93.62	0.99559
TCEN-BALI	372.50	185.25	559.75	0.00016	106.25	-34.16	246.66	0.18705	42.75	-40.87	126.37	0.53165
SRBU-RAMT	-384.75	-572.00	-197.50	0.00011	-128.75	-269.16	11.66	0.08002	-36.25	-119.87	47.37	0.67283
TCEN-RAMT	17.75	-169.50	205.00	0.99821	-36.75	-177.16	103.66	0.92406	-3.50	-87.12	80.12	0.99993
TCEN-SRBU	402.50	215.25	589.75	0.00007	92.00	-48.41	232.41	0.30197	32.75	-50.87	116.37	0.74626

(continued)

Supplemental Table S1-11 (*continued)*

Region	OTUs richness per ARMS											
	100 µm Fraction - 18S rRNA				500 µm Fraction - 18S rRNA				Sessile Fraction - 18S rRNA			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj	diff	lwr	upr	p adj
BALI-ACEH	-35.08	-159.53	89.36	0.93095	-41.50	-88.19	5.19	0.10389	50.67	-27.22	128.55	0.36463
RAMT-ACEH	21.92	-102.53	146.36	0.98735	-32.00	-78.69	14.69	0.31258	104.58	26.70	182.47	**0.00336**
SRBU-ACEH	-29.00	-153.44	95.44	0.96455	-17.83	-64.52	28.86	0.81739	47.92	-29.97	125.80	0.42152
TCEN-ACEH	136.92	12.47	261.36	**0.02419**	-0.75	-47.44	45.94	1.00000	63.75	-14.14	141.64	0.15774
RAMT-BALI	57.00	-67.44	181.44	0.69727	9.50	-37.19	56.19	0.97835	53.92	-23.97	131.80	0.30281
SRBU-BALI	6.08	-118.36	130.53	0.99992	23.67	-23.02	70.36	0.61160	-2.75	-80.64	75.14	0.99998
TCEN-BALI	172.00	47.56	296.44	**0.00238**	40.75	-5.94	87.44	0.11466	13.08	-64.80	90.97	0.98942
SRBU-RAMT	-50.92	-175.36	73.53	0.77707	14.17	-32.52	60.86	0.91160	-56.67	-134.55	21.22	0.25571
TCEN-RAMT	115.00	-9.44	239.44	0.08312	31.25	-15.44	77.94	0.33604	-40.83	-118.72	37.05	0.58047
TCEN-SRBU	165.92	41.47	290.36	**0.00365**	17.08	-29.61	63.77	0.83949	15.83	-62.05	93.72	0.97842
Region	**OTUs per site**				**Endemic OTUs per site**				**Endemic OTUs per ARMS**			
	100 µm Fraction - 18S rRNA				**100 µm Fraction - 18S rRNA**				**Sessile Fraction - 18S rRNA**			
	diff	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**	**diff**	**lwr**	**upr**	**p adj**
BALI-ACEH	-25.25	-236.60	186.10	0.99561	0.50	-124.58	125.58	1.00000	26.50	-38.32	91.32	0.77754
RAMT-ACEH	126.25	-85.10	337.60	0.38599	112.00	-13.08	237.08	0.09013	79.67	14.85	144.48	**0.00877**
SRBU-ACEH	-43.50	-254.85	167.85	0.96675	-29.75	-154.83	95.33	0.94494	36.00	-28.82	100.82	0.52484
TCEN-ACEH	330.25	118.90	541.60	**0.00177**	155.00	29.92	280.08	0.01218	33.67	-31.15	98.48	0.58918
RAMT-BALI	151.50	-59.85	362.85	0.22727	111.50	-13.58	236.58	0.09213	53.17	-11.65	117.98	0.15617
SRBU-BALI	-18.25	-229.60	193.10	0.99876	-30.25	-155.33	94.83	0.94172	9.50	-55.32	74.32	0.99370
TCEN-BALI	355.50	144.15	566.85	**0.00088**	154.50	29.42	279.58	0.01248	7.17	-57.65	71.98	0.99788
SRBU-RAMT	-169.75	-381.10	41.60	0.14768	-141.75	-266.83	-16.67	0.02292	-43.67	-108.48	21.15	0.32952
TCEN-RAMT	204.00	-7.35	415.35	**0.06099**	43.00	-82.08	168.08	0.82280	-46.00	-110.82	18.82	0.27892
TCEN-SRBU	373.75	162.40	585.10	**0.00054**	184.75	59.67	309.83	0.00294	-2.33	-67.15	62.48	0.99998

Supplemental Table S1-12. Beta diversity summary (PERMANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Cenderawasih Bay) in Indonesia for A) COI and B) 18S rRNA.

A. COI dataset

Indices	Factors	Adonis		Betadispers		100 µm Fraction				500 µm Fraction				Sessile Fraction			
						Adonis		Betadispers		Adonis		Betadispers		Adonis		Betadispers	
		R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*
Jaccard	Region	0.12103	**1.00E-04**	16.665	**1.00E-04**	0.25978	**1.00E-04**	16.915	**1.00E-04**	0.19038	**1.00E-04**	8.5128	**1.00E-04**	0.20854	**1.00E-04**	8.4792	**2.00E-04**
	Site	0.2652	**1.00E-04**	3.2602	**2.00E-04**	0.58598	**1.00E-04**	2.4667	**0.0103**	0.4875	**1.00E-04**	3.1503	**0.002**	0.52849	**1.00E-04**	2.0111	**0.0375**
	ARMS	0.45098	**1.00E-04**	2.1413	**4.00E-04**												
	Size fraction	0.04247	**1.00E-04**	86.492	**1.00E-04**												
Bray-Curtis	Region	0.11833	**1.00E-04**	8.6567	**1.00E-04**	0.30348	**1.00E-04**	10.409	**1.00E-04**	0.2199	**1.00E-04**	2.1827	0.086	0.24823	**1.00E-04**	1.5079	0.2194
	Site	2.9288	**1.00E-04**	2.6391	**7.00E-04**	0.65892	**1.00E-04**	1.9039	**0.0419**	0.54383	**1.00E-04**	1.6419	0.0933	0.60744	**1.00E-04**	0.6947	0.804
	ARMS	0.43014	**1.00E-04**	0.785	0.8495												
	Size fraction	0.08042	**1.00E-04**	83.308	**1.00E-04**												
UniFrac	Region	0.11627	**1.00E-04**	6.9621	**1.00E-04**	0.28269	**1.00E-04**	8.9327	**1.00E-04**	0.19361	**1.00E-04**	3.2364	**0.0184**	0.20319	**1.00E-04**	5.9358	**5.00E-04**
	Site	0.2467	**1.00E-04**	2.4032	**0.0018**	0.59949	**1.00E-04**	1.904	**0.0463**	0.48998	**1.00E-04**	3.2152	**5.00E-04**	0.51621	**1.00E-04**	1.3225	0.231
	ARMS	0.41807	**1.00E-04**	1.5363	**0.0278**												
	Size fraction	0.09587	**1.00E-04**	23.483	**1.00E-04**												

B. 18S rRNA dataset

Indices	Factors	Adonis		Betadispers		100 µm Fraction				500 µm Fraction				Sessile Fraction			
						Adonis		Betadispers		Adonis		Betadispers		Adonis		Betadispers	
		R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*
Jaccard	Region	0.11718	**1.00E-04**	18.895	**1.00E-04**	0.2572	**1.00E-04**	10.812	**2.00E-04**	0.17995	**1.00E-04**	11.93	**1.00E-04**	0.185	**1.00E-04**	2.3568	0.0615
	Site	0.24863	**1.00E-04**	5.1346	**1.00E-04**	0.5499	**1.00E-04**	2.1661	**0.0216**	0.47319	**1.00E-04**	4.7099	**1.00E-04**	0.49185	**1.00E-04**	2.4244	**0.0129**
	ARMS	0.42852	**1.00E-04**	1.078	0.3627												
	Size fraction	0.08208	**1.00E-04**	26.368	**1.00E-04**												
Bray-Curtis	Region	0.10711	**1.00E-04**	2.6927	**0.035**	0.33303	**1.00E-04**	4.265	**0.004**	0.21136	**1.00E-04**	0.6791	0.6182	0.24209	**1.00E-04**	3.0351	**0.0246**
	Site	0.23116	**1.00E-04**	0.7612	0.738	0.64575	**1.00E-04**	3.1199	**0.001**	0.51127	**1.00E-04**	2.0382	**0.0251**	0.5864	**1.00E-04**	0.8912	0.5992
	ARMS	0.37724	**1.00E-04**	0.2732	1												
	Size fraction	0.19603	**1.00E-04**	1.2511	0.2905												
UniFrac	Region	0.06103	**1.00E-04**	14.426	**1.00E-04**	0.15273	**1.00E-04**	7.6599	**2.00E-04**	0.11342	**1.00E-04**	13.297	**1.00E-04**	0.10783	**1.00E-04**	2.907	**0.0291**
	Site	0.1646	**1.00E-04**	2.9341	**1.00E-04**	0.42998	**1.00E-04**	2.668	**0.0051**	0.38515	**1.00E-04**	3.4336	**6.00E-04**	0.38236	**1.00E-04**	1.4384	0.1527
	ARMS	0.37363	**1.00E-04**	1.1332	0.2858												
	Size fraction	0.0393	**1.00E-04**	10.931	**2.00E-04**												

Supplemental Table S1-13. Beta diversity indices (PERMANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Cenderawasih Bay) in Indonesia. Analysis were calculated using total diversity per ARMS unit (e.g. summing diversity across all size fractions from a single ARMS).

Indices	**Factors**	**COI**				**18S rRNA**			
		Adonis		**Betadispers**		**Adonis**		**Betadispers**	
		R2	***p-value***	**F**	***p-value***	**R2**	***p-value***	**F**	***p-value***
Jaccard	**Region**	0.2449	**1.00E-04**	8.2385	**1.00E-04**	0.24919	**1.00E-04**	7.6072	**1.00E-04**
	Site	0.56239	**1.00E-04**	1.4833	0.1475	0.54774	**1.00E-04**	3.1812	**0.0014**
Bray-Curtis	**Region**	0.27231	**1.00E-04**	3.3898	**0.0133**	0.27728	**1.00E-04**	1.9632	0.1129
	Site	0.61509	**1.00E-04**	1.6799	0.0886	0.60414	**1.00E-04**	1.7766	0.0635
UniFrac	**Region**	0.25716	**1.00E-04**	3.5416	**0.0059**	0.12094	**1.00E-04**	6.6623	**0.0005**
	Site	0.56966	**1.00E-04**	1.3399	0.205	0.3962	**1.00E-04**	1.9507	**0.0389**

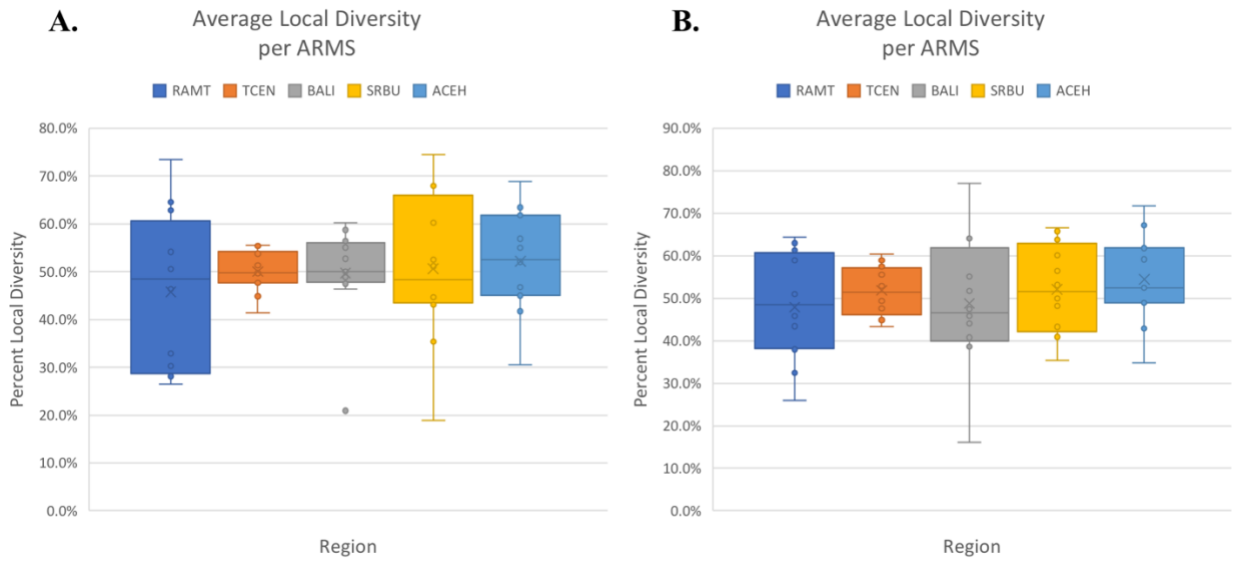
Supplemental Table S1-14. Beta diversity indices (PERMANOVA) of eukaryote diversity based on site and ARMS within locations (Aceh, Pulau Seribu, Bali, Raja Ampat and Cenderawasih Bay) in Indonesia for A) COI and B) 18S rRNA.

A. COI metabarcoding dataset

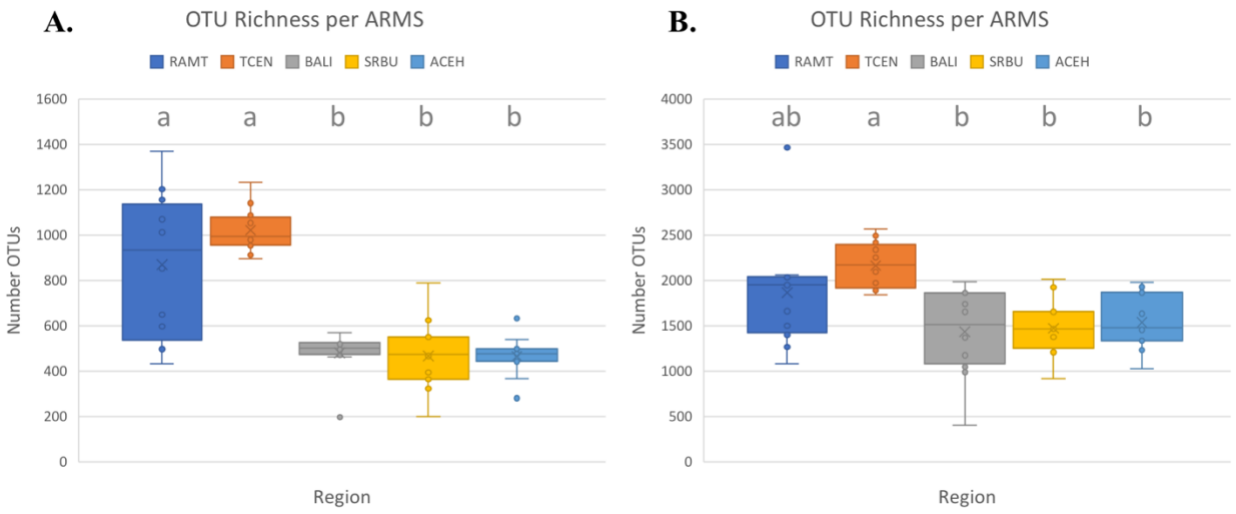
Indices	Factors	Aceh		Pulau Seribu		Bali		Raja Ampat		Cenderawasih	
		R2	*p-value*	R2	*p-value*	R2	*p-value*	R2	*p-value*	R2	*p-value*
Jaccard	**Site**	0.15698	**1.00E-04**	0.17791	**1.00E-04**	0.15636	**1.00E-04**	0.15919	**1.00E-04**	0.16988	**1.00E-04**
	ARMS	0.373	**1.00E-04**	0.392	**2.00E-04**	0.37263	**0.0028**	0.36293	**1.00E-04**	0.37813	**1.00E-04**
Bray-Curtis	**Site**	0.15805	**0.0012**	0.18105	**1.00E-04**	0.19201	**1.00E-04**	0.15464	**1.00E-04**	0.15821	**1.00E-04**
	ARMS	0.3568	**0.037**	0.36175	0.1046	0.3927	0.011	0.32927	0.1616	0.33372	0.1544

B. 18S rRNA metabarcoding dataset

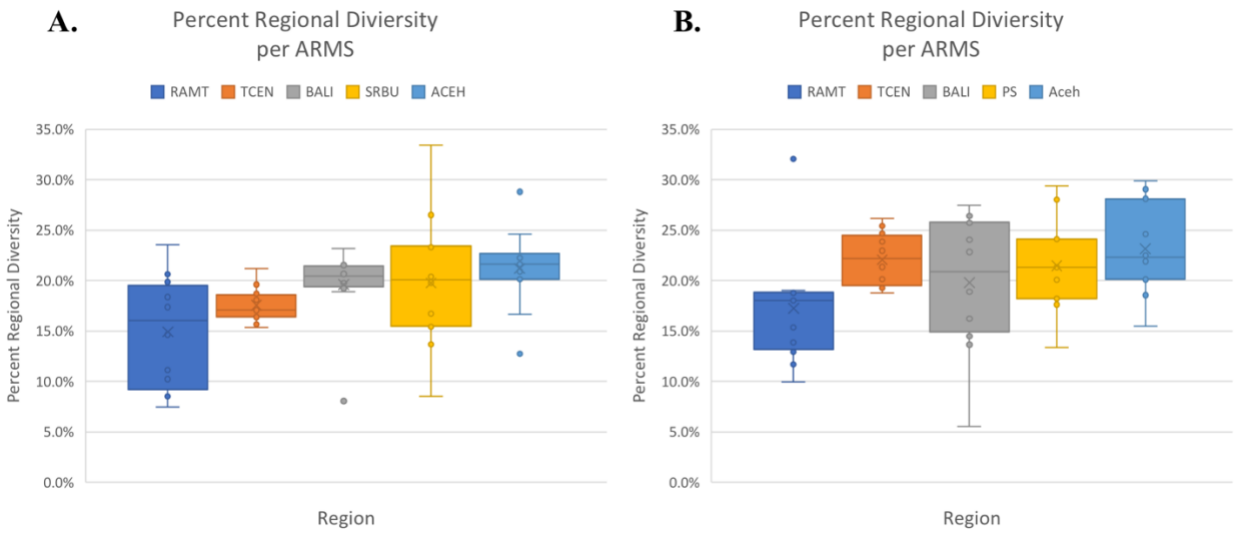
Indices	Factors	Aceh		Pulau Seribu		Bali		Raja Ampat		Cenderawasih	
		R2	*p-value*	R2	*p-value*	R2	*p-value*	R2	*p-value*	R2	*p-value*
Jaccard	**Site**	0.16167	**1.00E-04**	0.1879	**1.00E-04**	0.14215	**1.00E-04**	0.12664	**2.00E-04**	0.13494	**2.00E-04**
	ARMS	0.35952	**0.0038**	0.39931	**2.00E-04**	0.35046	**0.0278**	0.32582	0.1585	0.33728	**0.0385**
Bray-Curtis	**Site**	0.14806	**0.0039**	0.16363	**5.00E-04**	0.14291	**0.0079**	0.12386	**0.0165**	0.11956	**0.0481**
	ARMS	0.33397	0.1881	0.33312	0.4872	0.30841	0.6826	0.28042	0.9334	0.26252	0.9695



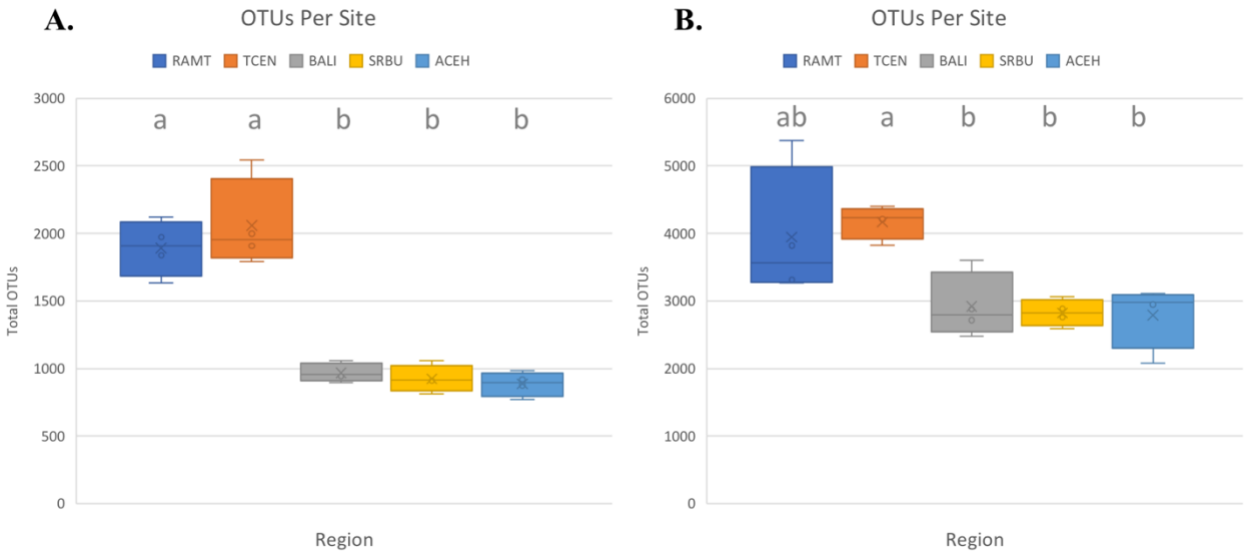
Supplemental Figure S1-1. Average site diversity captured in a single ARMS (Autonomous Reef Monitoring Structure) unit across five sampling regions for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.



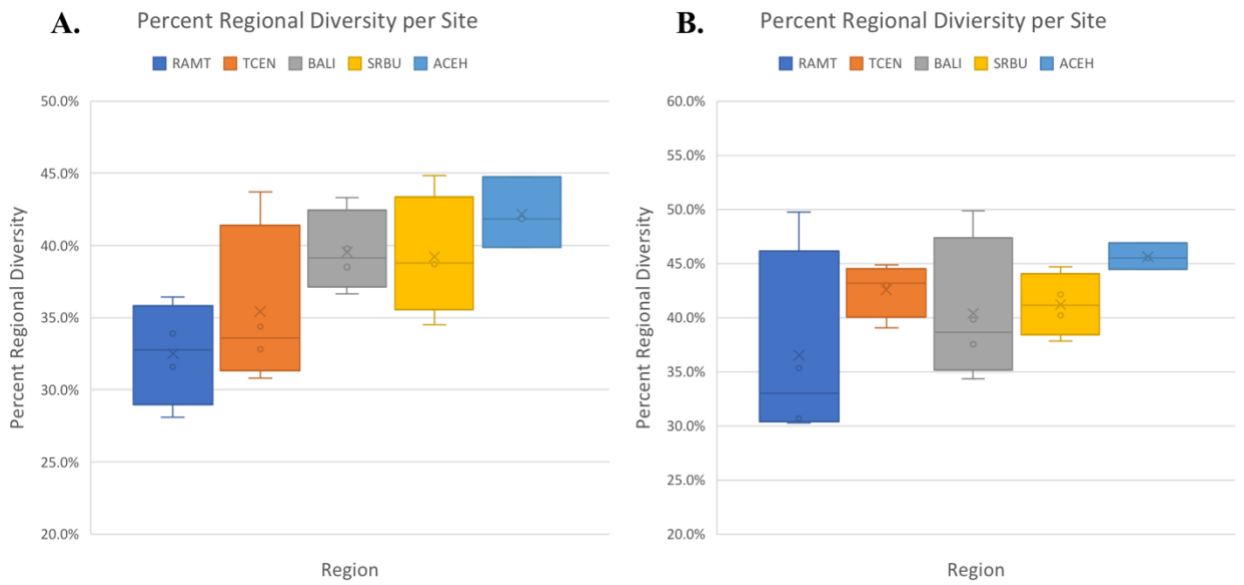
Supplemental Figure S1-2. Total Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.



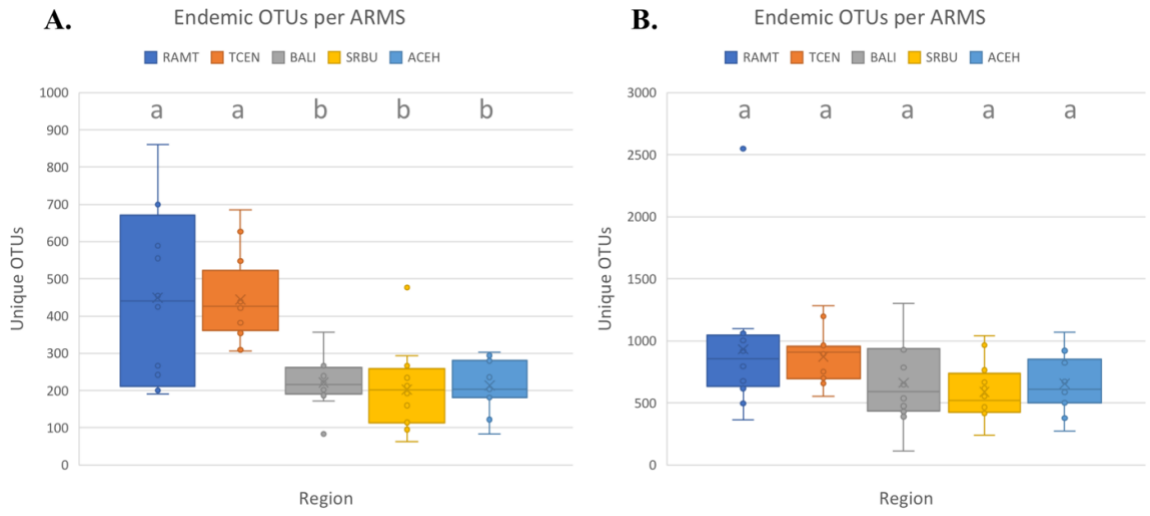
Supplemental Figure S1-3. Percentage of regional diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.



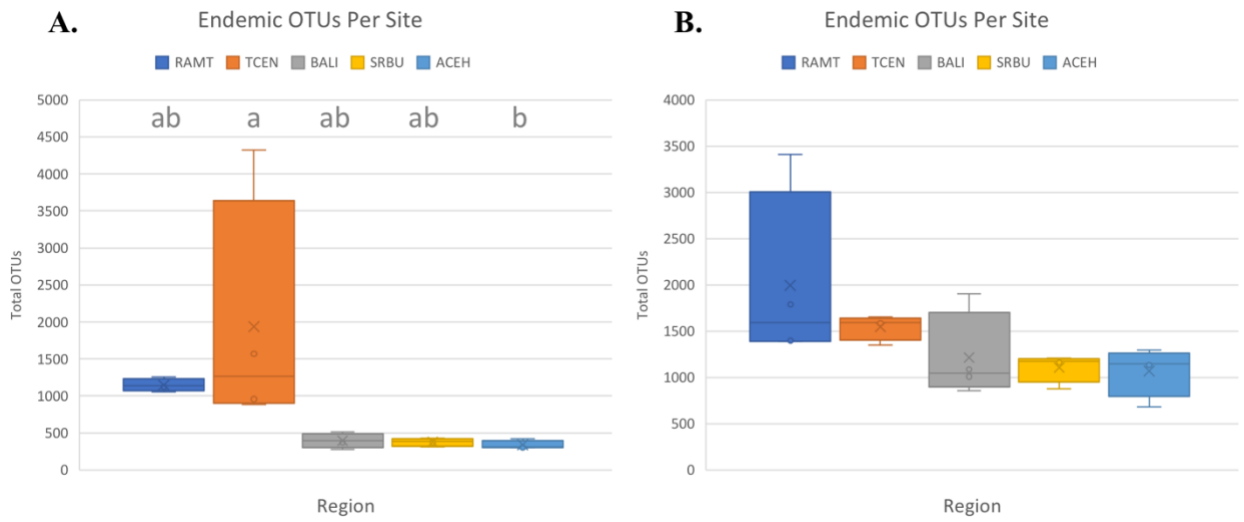
Supplemental Figure S1-4. Total Operational Taxonomic Units (OTUs) diversity captured in a single sampling site across five sampling regions across Indonesia for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.



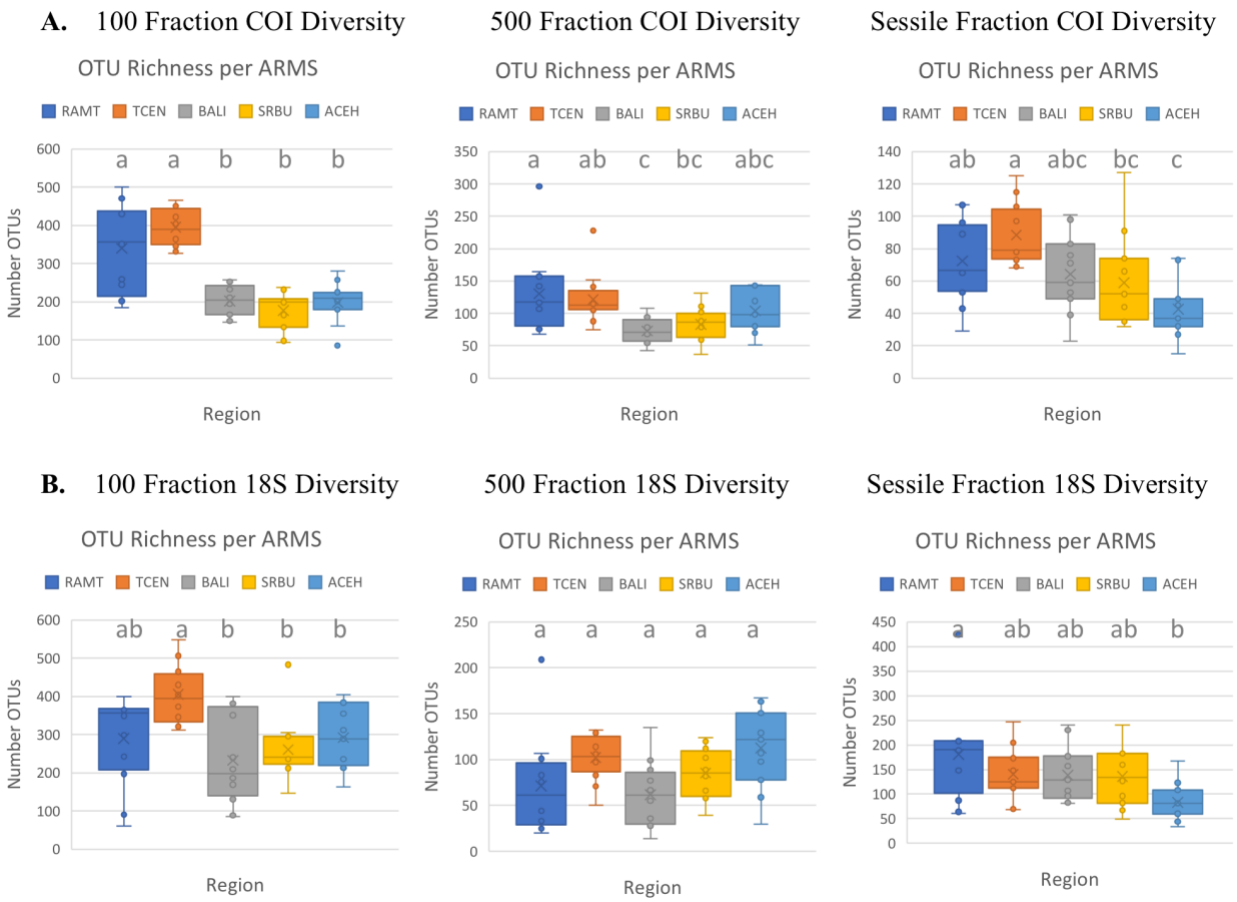
Supplemental Figure S1-5. Percentage of regional diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.



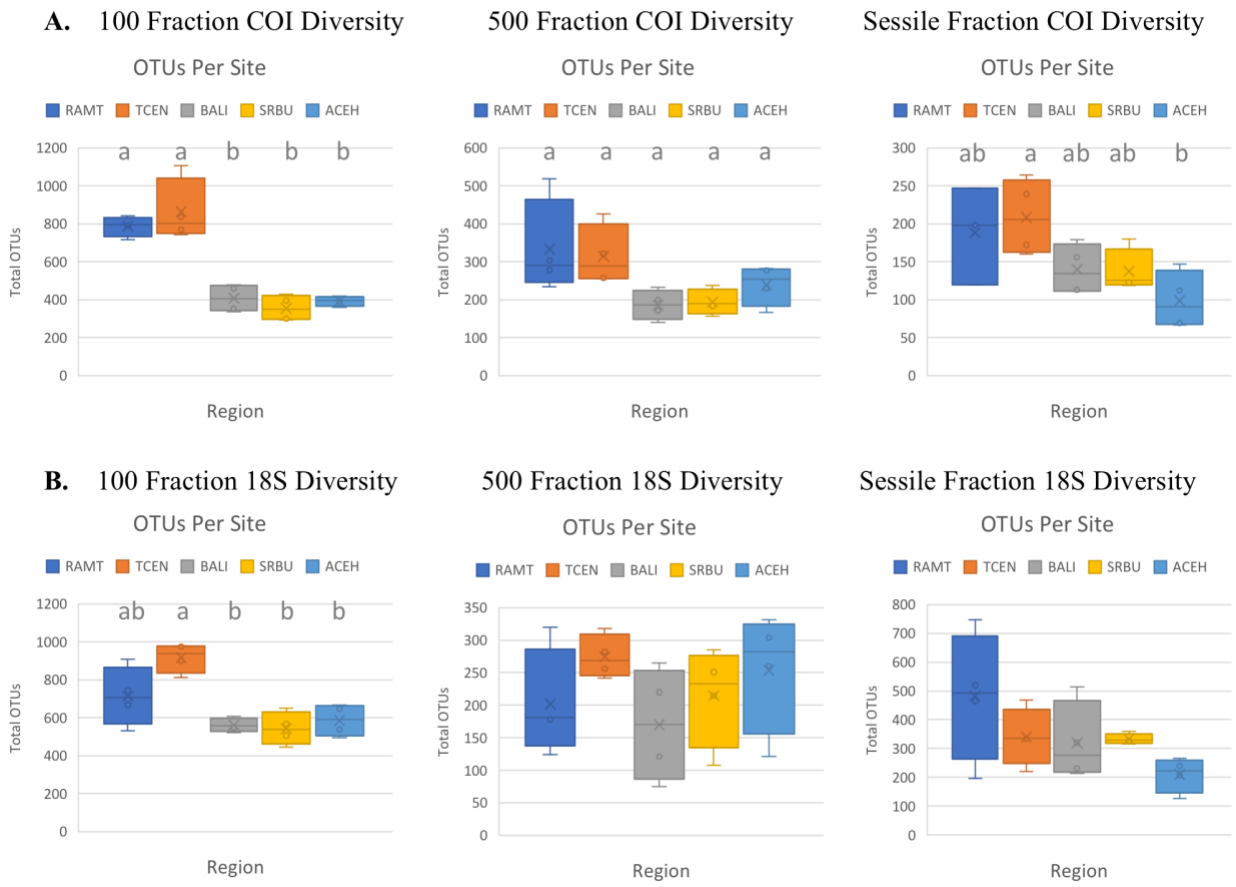
Supplemental Figure S1-6. Total endemic Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit across five sampled regions of Indonesia for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.



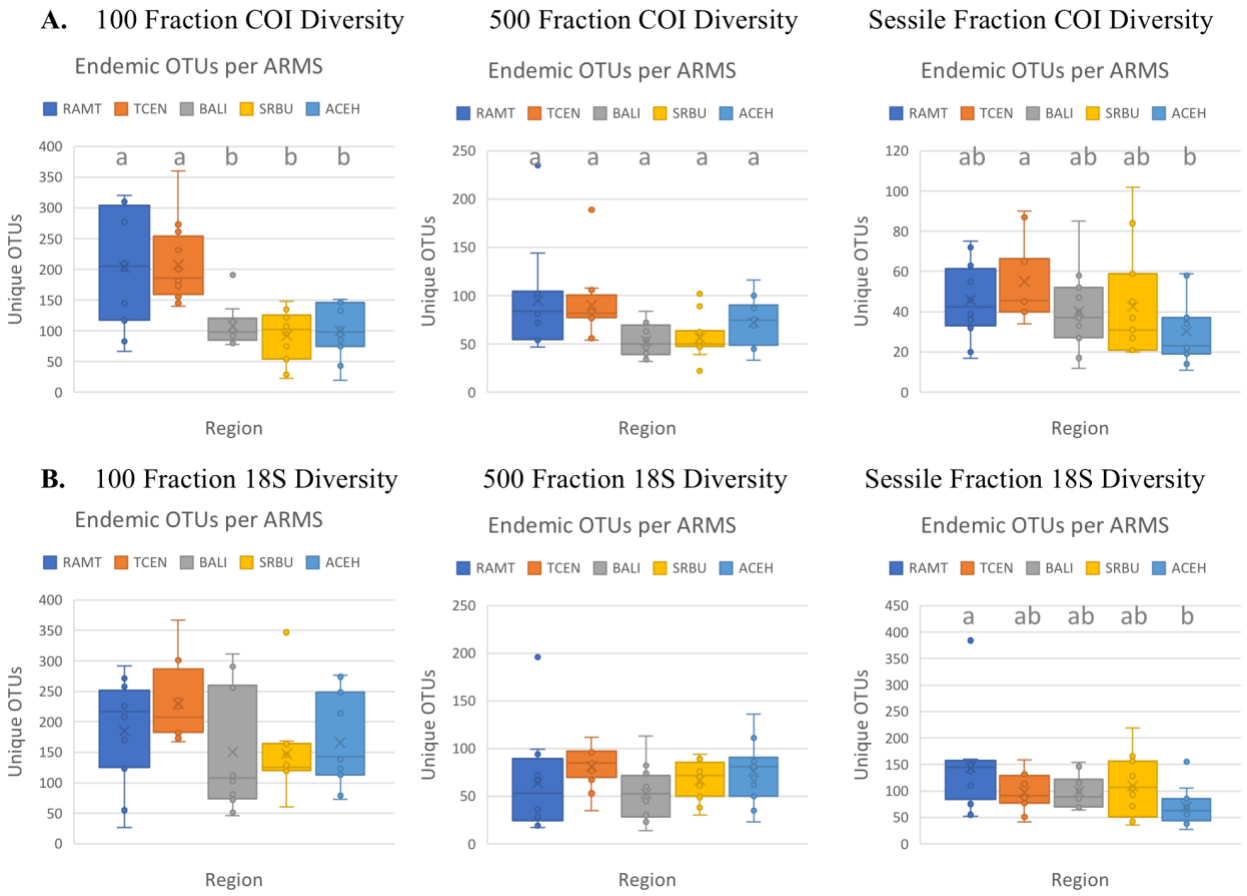
Supplemental Figure S1-7. Total endemic Operational Taxonomic Units (OTUs) diversity captured in a single sampling site across five sampling regions across Indonesia for A) COI and B) 18S rRNA. Box plots were generated from raw data prior to rarefaction.



Supplemental Figure S1-8. Total Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit in the A) 100 µm B) 500 µm and C) Sessile size fractions across five sampled regions of Indonesia for A) COI and B) 18S rRNA. Box plots were generated from rarefied data after excluding OTUs shared among size fractions.

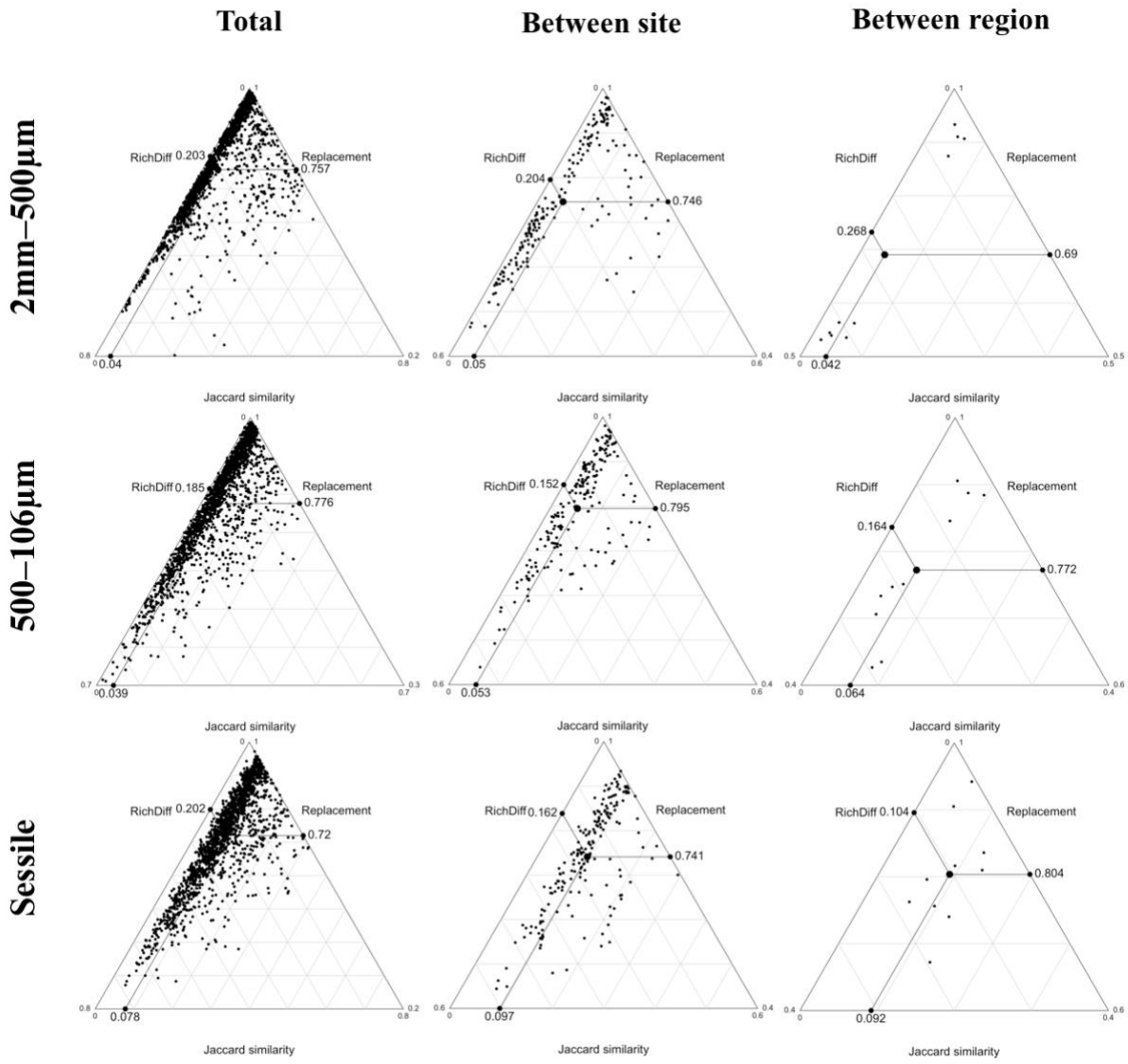


Supplemental Figure S1-9. Total Operational Taxonomic Units (OTUs) diversity captured in a single sampling site in the A) 100 μm B) 500 μm and C) Sessile size fractions across five sampling regions across Indonesia for A) COI and B) 18S rRNA. Box plots were generated from rarefied data after excluding OTUs shared among size fractions.

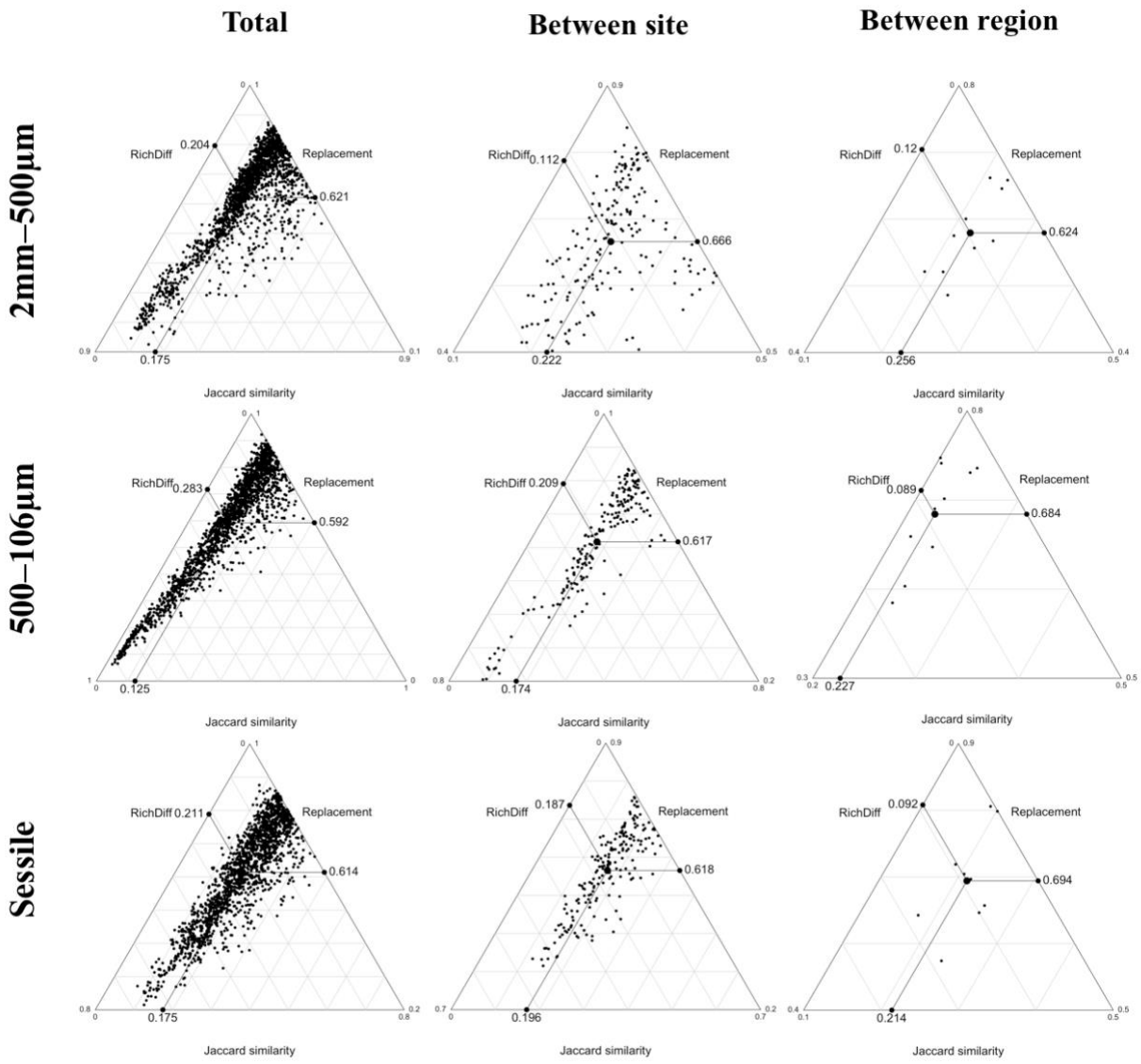


Supplemental Figure S1-10. Total endemic Operational Taxonomic Units (OTUs) diversity captured in an individual ARMS (Autonomous Reef Monitoring Structure) unit in the A) 100 µm B) 500 µm and C) Sessile size fractions across five sampled regions of Indonesia for A) COI and B) 18S rRNA. Box Box plots were generated from rarefied data after excluding OTUs shared among size fractions.

Supplemental Figure S1-11. Total endemic Operational Taxonomic Units (OTUs) diversity captured in a single sampling site in the A) 100 μm B) 500 μm and C) Sessile size fractions across five sampling regions across Indonesia for A) COI and B) 18S rRNA. Box plots were generated from rarefied data after excluding OTUs shared among size fractions.



Supplemental Figure S1-12. Ternary plots of Jaccard similarity and the partitions of beta diversity (replacement, richness difference, and Jaccard similarity) for the 100 µm, 500 µm, and sessile fraction obtained from the COI metabarcoding. Ternary plots are shown for the total experiment as well as between sites and between locations. The mean values are represented by numbers in the bigger circle.



Supplemental Figure S1-13. Ternary plots of Jaccard similarity and the partitions of beta diversity (replacement, richness difference, and Jaccard similarity) for the 100 µm, 500 µm, and sessile fraction obtained from the 18S rRNA metabarcoding. Ternary plots are shown for the total experiment as well as between sites and between locations. The mean values are represented by numbers in the bigger circle

References

Al-Rshaidat, M. M. D., Snider, A., Rosebraugh, S., Devine, A. M., Devine, T. D., Plaisance, L., Knowlton, N. & Leray, M. (2016). Deep COI sequencing of standardized benthic samples unveils overlooked diversity of Jordanian coral reefs in the northern Red Sea. *Genome,* 59, 724–737.

Allen, G. R. (2008). Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conserv: Mar. Freshw. Ecosyst.* 18: 541–556. https://doi.org/10.1002/aqc

Allen, G. R. & Erdmann, M. V. (2021) Reef Fishes of the East Indies. Apple App Store, Vers. 2.1.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*. https://doi.org/10.1016/S0022-2836(05)80360-2

Andersen, K. S., Kirkegaard, R. H., Karst, S. M. & Albertsen, M. (2018). "ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. *bioRxiv*. http://dx.plos.org/10.1371/journal.pone.0132783

Anderson, M. J., Tolimieri. N. & Millar, R. B. (2013) Beta Diversity of Demersal Fish Assemblages in the North-Eastern Pacific: Interactions of Latitude and Depth. *PLoS ONE* 8(3): e57918. doi:10.1371/journal.pone.0057918

Appeltans, W., Ahyong, S. T., Anderson, G., Angel, M. V., Artois, T., Bailly, N., Bamber, R., Barber, A., Bartsch, I., Berta, A., Błazewicz-Paszkowycz, M., Bock, P., Boxshall, G., Boyko, C. B., Brandão, S. N., Bray, R. A., Bruce, N. L., Cairns, S. D., Chan, T. Y., Cheng, L., Collins, A.G., Cribb, T., Curini-Galletti, M., Dahdouh-Guebas, F., Davie, P. J. F., Dawson, M. N., Clerck, O. De., Decock, W., Grave, S. De., Voogd ,N. J .De., Domning, D. P., Emig, C. C., Erséus, C., Eschmeyer, W., Fauchald, K., Fautin, D. G., Feist, S. W, Fransen, C. H. J. M., Furuya, H., Garcia-Alvarez, O., Gerken, S., Gibson, D., Gittenberger, A., Gofas, S., Gómez-Daglio, L., Gordon, D. P., Guiry, M. D., Hernandez, F., Hoeksema, B. W., Hopcroft, R. R., Jaume, D., Kirk, P., Koedam, N., Koenemann, S., Kolb J. B., Kristensen, R. M., Kroh A., Lambert, G., Lazarus, D. B., Lemaitre, R., Longshaw, M., Lowry, J., MacPherson, E., Madin, L.P., Mah, C., Mapstone, G., McLaughlin, P.A., Mees, J., Meland, K., Messing, C. G., Mills, C. E., Molodtsova, T. N., Mooi, R., Neuhaus, B., Ng, P. K. L., Nielsen, C., Norenburg, J., Opresko, D. M., Osawa, M., Paulay, G., Perrin, W, Pilger, J. F., Poore, G. C. B., Pugh, P., Read, G. B., Reimer, J. D., Rius, M., Rocha, R. M., Saiz-Salinas, J. I., Scarabino, V., Schierwater, B., Schmidt-Rhaesa, A., Schnabel, K. E., Schotte, M., Schuchert, P., Schwabe, E., Segers, H., Self-Sullivan, C., Shenkar, N., Siegel, V., Sterrer, W., Stöhr, S., Swalla, B., Tasker, M. L., Thuesen, E. V., Timm, T., Todaro, M. A., Turon, X., Tyler, S., Uetz, P., Land, J Van Der., Vanhoorne, B., Ofwegen, L. P. Van., Soest, R. W. M. Van., Vanaverbeke, J., Walker-Smith G., Walter, T. C., Warren, A., Williams, G. C., Wilson, S. P. & Costello M. J. (2012) The magnitude of global marine

species diversity. *Curr Biol* 22:2189–2202

Barber, P. H. (2009) The challenge of understanding the Coral Triangle biodiversity hotspot. *J Biogeogr* 36:1845–1846

Barber, P. H., Ablan-lagman, M. C. A., Cahyani, D., Crandall, E. D., Ravago-gotanco, R., Juinio-meñez, M. A., Mahardika, I.G.N., Shanker, K., Starger, C.J., Toha, A.H.A., Anggoro, A.W. and Willette, D.A. (2014). Advancing biodiversity research in developing countries : the need for changing paradigms. *Bulletin of Marine Science*. *90*(1). 2014 Rosenstiel School of Marine & Atmospheric Science of the University of Miami

Barber, P. H. & Bellwood, D.R. (2005) Biodiversity hotspots: Evolutionary origins of biodiversity in wrasses (Halichoeres: Labridae) in the Indo-Pacific and new world tropics. *Mol Phylogenet Evol* 35: 235-253

Barber, P & Boyce, S. L. (2006) Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. *Proc R Soc B Biol Sci* 273(1597):2053-61. doi: 10.1098/rspb.2006.3540.

Barber, P. H., Cheng, S. H., Erdmann, M. V., Tengardjaja, K. & Ambariyanto (2011) Evolution and conservation of marine biodiversity in the Coral Triangle: insights from stomatopod Crustacea. In: Crustacean Issues 19 Phylogeography and Population Genetics in Crustacea (eds Held C, Koenemann S and Schubart CD), pp. 129–156; CRC Press, Boca Raton.

Barber, P. H. & Meyer, C. P. (2015) Pluralism explains diversity in the Coral Triangle. *Ecol Fishes Coral Reefs*, 258–263

Bellwood, D. R. (2001). Regional-Scale Assembly Rules and Biodiversity of Coral Reefs. *Science*, *292*(5521):1532–1535. https://doi.org/10.1126/science.1058635

Bellwood, David R, Hughes, T. P., Connolly, S. R., & Tanner, J. (2005). Environmental and geometric constraints on Indo-Pacific coral reef biodiversity. *Ecology Letters*, *8*:643–651. https://doi.org/10.1111/j.1461-0248.2005.00763.x

Bellwood, D. R., Hughes, T. P., Folke, C. & Nyström, M. (2004) Confronting the coral reef crisis. *Nature* 429:827–833

Bellwood, David R., & Meyer, C. P. (2009). Searching for heat in a marine biodiversity hotspot. *Journal of Biogeography*, *36*(4): 569–576. https://doi.org/10.1111/j.1365-2699.2008.02029.x

Bouchet, P., Lozouet, P., & Maestrati, P. (2002). Assessing the magnitude of species richness in tropical marine environments : exceptionally high numbers of molluscs at a New Caledonia site. *Biological Journal of the Linnean Society*, 75:421–436.

Bowen, B. W., Rocha, L. A., Toonen, R. J. & Karl, S. A. (2013) The origins of tropical marine biodiversity. *Trends Ecol Evol* 28:359-366.

Brandt, M. I., Trouche, B., Quintric, L., Wincker, P., Poulain, J., & Arnaud-Haond, S. (2020). A flexible pipeline combining clustering and correction tools for prokaryotic and eukaryotic metabarcoding. *Peer Community In Ecology*, 100043. https://doi.org/10.24072/pci.ecology.100043

Brown, M. V., Lauro, F. M., Demaere, M. Z., Muir, L., Wilkins, D., Thomas, T., Riddle, M. J., Fuhrman, J. A., Andrews-Pfannkoch, C., Hoffman, J. M., McQuaid, J. B., Allen, A., Rintoul, S. R. & Cavicchioli, R. (2012) Global biogeography of SAR11 marine bacteria. *Mol Syst Biol* 8:1–13

Bucklin, A., Steinke, D., & Blanco-bercial, L. (2011). DNA Barcoding of Marine Metazoa. *Annu. Rev. Mar. Sci*, *3*, 471–508. https://doi.org/10.1146/annurev-marine-120308-080950

Burke, L., Reytar, K., Spalding, M., & Perry, A. (2012). *Reefs at risk, Revisited In the Coral Triangle*. *National Geographic*. https://doi.org/10.1016/0022-0981(79)90136-9

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*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Carpenter, K. E. & Springer, V. G. (2005) The center of the center of marine shore fish biodiversity: The Philippine Islands. *Environ Biol Fishes* 72:467–480

Carugati, L., Corinaldesi, C., Dell'Anno, A., & Danovaro, R. (2015). Metagenetic tools for the census of marine meiofaunal biodiversity: An overview. *Marine Genomics*, *24*, 11–20. https://doi.org/10.1016/j.margen.2015.04.010

Cermeño, P. & Falkowski, P. G. (2009) Controls on diatom biogeography in the ocean. *Science* 325(5947):1539-1541.

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*. https://doi.org/10.1890/13-0133.1

Chao, A. & Chiu, C. (2016) Species Richness: Estimation and Comparison. *Stat Ref Online*:1–11

Cleary DFR, Suharsono, & Hoeksema BW (2006) Coral diversity across a disturbance gradient in the Pulau Seribu reef complex off Jakarta, Indonesia. *Biodiversity & Conservation* 15:3653-3674.

Condit, R., Pitman, N., Leigh, E. G., Chave, J., Terborgh, J., Foster, R. B., Núñez, P. V., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, He. C., Losos, E. & Hubbell, S. P. (2002) Beta-diversity in tropical forest trees. *Science* 295: 666–669.

Connolly, S. R., Bellwood, D. R. & Hughes, T. P. (2003). Indo-Pacific biodiversity of coral reefs: Deviations from a mid-domain model. *Ecology*, 84:2178-2190.

Curd, E. E., Gold, Z., Kandlikar, G. S., Gomer, J., Ogden, M., O'Connell, T., Pipes, L., Schweizer, T., Rabichow, L., Lin, M., Shi, B., Barber, P., Kraft, N., Wayne, R. & Meyer, R. S. (2018). Anacapa Toolkit: an environmental DNA toolkit for processing multilocus metabarcode datasets. *BioRxiv*, 488627. https://doi.org/10.1101/488627

De Boer, T. S., Naguit, M. R. A., Erdmann, M. V., Ablan-Lagman, M. C. A., Ambariyanto, A., Carpenter, K. E., Toha, A. H. A. & Barber, P. H. (2014). Concordance between phylogeographic and biogeographic boundaries in the Coral Triangle: Conservation implications based on comparative analyses of multiple giant clam species. *Bulletin of Marine Science*, *90*(1). https://doi.org/10.5343/bms.2013.1003

Dray, A. S., Blanchet, G., Borcard, D., Guenard. G., Jombart, T., Larocque, G., Legendre. P. & Wagner, H. H. (2016) Package ' adespatial .'

Droettboom, M., Caswell, T. A., Hunter, J., Firing, E., Nielsen, J. H., Varoquaux, N., Root, B., Elson, P., Dale, D., Lee, J. -J., de Andrade, E. S., Seppänen, J. K., McDougall, D., May, R., Lee, A., Straw, A., Stansby, D., Hobson, P., Yu, T. S., Ma, E., Gohlke, C., Silvester, S., J. Moad, J., Schulz, J., Vincent, A. F., Würtz, P., Ariza, F., Cimarron, T., Hisch, N. & Kniazev (2017) Matplotlib v2.0.2. http://dx.doi.org/10.5281/zenodo.573577.

Ekman, S. (1953) Zoogeography of the sea. The Macmillan Company, 60 Fifth Avenue, New York 11, N. Y. 417 pp

Faurby, S. & Barber, P. H. (2015) Extreme population subdivision despite high colonization ability: Contrasting regional patterns in intertidal tardigrades from the west coast of North America. *J Biogeogr* 42:1006–1017

Faurby S, Funch P (2011) Size is not everything: A meta-analysis of geographic variation in microscopic eukaryotes. *Global Ecology and Biogeography* 20(3):475 - 485

Fenchel, T. & Finlay, B. J. (2004) The Ubiquity of Small Species: Patterns of Local and Global Diversity. *Bioscience* 54:777

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C. M., Boyer, F., Pompanon, F., Rayé, G. & Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.12338

Finlay, B. J. (2002) Global dispersal of free-living microbial eukaryote species. *Science* 296(5570):1061-3

Finlay, B. J. (2014) Global Dispersal of Eukaryote Species Microbial. *Science* 296:1061–1063

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01312-x

Fuhrman, J. A. & Steele, J. A. (2008) Community structure of marine bacterioplankton: Patterns, networks, and relationships to function. *Aquat Microb Ecol* 53:69–81

Gaither, M. R., Bowen, B. W. & Toonen, R. J. (2013) Population structure in the native range predicts the spread of introduced marine species. *Proceedings of the Royal Society B: Biological Sciences* 280(1760):20130409

Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, *13*(5), 851–861. https://doi.org/10.1111/1755-0998.12138

Ghiglione, J. F., Galand, P. E., Pommier, T., Pedros-Alio, C., Maas, E. W., Bakker, K., Bertilson, S., Kirchman, D. L., Lovejoy, C., Yager, P. L. & Murray, A. E. (2012) Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proc Natl Acad Sci* 109:17633–17638

Giebner, H., Langen, K., Bourlat, S. J., Kukowka, S., Mayer, C., Astrin, J. J., Misof, B. & Fonseca, V. G. (2020). Comparing diversity levels in environmental samples: DNA sequence capture and metabarcoding approaches using 18S and COI genes. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13201

Gordon, A. L. & Fine, R. A. (1996) Pathways of water between the Pacific and Indian oceans in the Indonesian seas. *Nature* 379: 146–149.

Gray, J.S. (2001) Marine diversity: the paradigms in patterns of species richness examined. *Sci Mar* (suppl 2): 41–56.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., Del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A. L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P. & Christen, R. (2013). The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*(D1). https://doi.org/10.1093/nar/gks1160

Guo, L., Sui, Z., Zhang, S., Ren, Y., & Liu, Y. (2015). Comparison of potential diatom 'barcode' genes (The 18S rRNA gene and ITS, COI, rbcL) and their effectiveness in discriminating and determining species taxonomy in the Bacillariophyta. *International Journal of Systematic and Evolutionary Microbiology*. https://doi.org/10.1099/ijs.0.000076

Halas, D. & Winterbottom, R. (2009) A phylogenetic test of multiple proposals for the origins of the East Indies coral reef biota. *Journal of Biogeography* 36:10, 1847–1860

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences, 270(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Hellberg, M. E. (2009) Gene flow and isolation among populations of marine animals. *Annu Rev Ecol Evol Syst* 40:291-310

Hoegh-Guldberg, O. (2011). The Impact of Climate Change on Coral Reef Ecosystems. In Z. Dubinsky & N. Stambler (Eds.), *Coral Reefs: An Ecosystem in Transition*. Springer.

Hoegh-Guldberg, O., Hoegh-Guldber, H., Veron, J. E. (Charlie), Green, A., Gomez, E. D., Lough, J., King, M., Ambariyanto, Hansen, L., Cinner, J., Dews, G., Russ, G., Schuttenberg, H. Z., Peñaflor, E. L., Eakin, C.M., Christensen, T. R. L., Abbey, M., Areki, F., Kosaka, R. A., Tewfik, A. & Oliver, J. (2009). *The Coral Triangle and Climate Change: Ecosystems, People and Societies at Risk*.

Hoeksema, B. W. (2007). Delineation of the Indo-Malayan centre of maximum marine biodiversity: the Coral Triangle. In: Renema W (ed) Biogeography, time, and place: distribu- tions, barriers, and islands. Springer, Netherlands, pp 117- 178

Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.12613

Hubbell, S. P. (2013) Tropical rain forest conservation and the twin challenges of diversity and rarity. *Ecol Evol* 3:3263–3274

Hubbell, S. P. (2015) Estimating the global number of tropical tree species, and Fisher's paradox. *Proc Natl Acad Sci U S A* 112:7343–7344

Jokiel, P, & Martinelli, F. J. (1992) The vortex model of coral reef biogeography. *J Biogeogr* 19: 449-458

Juhel, J. B., Utama, R. S., Marques, V., Vimono, I. B., Sugeha, H. Y., Kadarusman, Pouyaud, L., Dejean, T., Mouillot, D. & Hocdé, R. (2020). Accumulation curves of environmental DNA sequences predict coastal fish diversity in the coral triangle: EDNA predict fish diversity. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1930). https://doi.org/10.1098/rspb.2020.0248rspb20200248

Kandlikar, G. S., Gold, Z. J., Cowen, M. C., Meyer, R. S., Freise, A. C., Kraft, N. J. B., Moberg-Parker, J., Sprague, J., Kushner, D. J. & Curd, E. E. (2018). Ranacapa: An R package and shiny web app to explore environmental DNA data with exploratory statistics and interactive visualizations [version 1; referees: 1 approved, 2 approved with reservations]. *F1000Research*. https://doi.org/10.12688/f1000research.16680.1

Karlson, R. H., Cornell, H. V. & Hughes, T. P. (2004) Coral communities are regionally enriched along an oceanic biodiversity gradient. *Nature* 429:867–870

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4). https://doi.org/10.1093/molbev/mst010

Kay, E. A. (1984) Patterns of speciation in the Indo-West Pacific. *B.P. Bishop Museum Special Publications* 72: 15–31.

Kelly, L. W., Williams, G. J., Barott, K. L., Carlson, C. A., Dinsdale, E. A.,, Edwards, R. A.,, Haas, A. F., Haynes, M., Lim, Y. W., McDole, T., Nelson, C. E., Sala, E., Sandin, S. A., Smith, J. E., Vermeij, M. J. A., Youle, M. & Rohwer, F. (2014) Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors. *Proc Natl Acad Sci* 111:10227–10232

Keshavmurthy, S., Yang, S., Alamaru, A., Chuang, Y., Pichon, M., Macdonald, A., Noreen, A. M. E., Chen, C. & Wallace, C. C. (2013) DNA barcoding reveals the coral '"laboratory-rat"', Stylophora pistillata encompasses multiple identities. *Scientific Reports* 2:1–7

Knowlton N (1986) Cryptic and Sibling Species among the Decapod Crustacea. *J Crustac Biol* 6:3, 356-363.

Knowlton N (1993) Sibling species in the sea. *Annu Rev Ecol Syst* 24:189-216

Knowlton, N., Brainard, R. E., Fisher, R., Moews, M., Plaisance, L., & Caley, M. J. (2010). Coral Reef Biodiversity. In A. D. McIntyre (Ed.), *Life in the World's Oceans* (pp. 65–77). Blackwell Publishing Ltd.

Ladd, H. S. (1960) Origin of the Pacific island molluscan fauna. *Am. J. Sci*. 258(A), 137-150.

Larsson, J., Godfrey, A. J. R., Gustafsson, P., Eberly, D. H., Huber, E., Slowikowski, K., Privé, F. & Maintainer (2020) Area-Proportional Euler and Venn Diagrams with Ellipses. Package 'eulerr'. R Package

Legendre, P. (2014) Interpreting the replacement and richness difference components of beta diversity. *Glob Ecol Biogeogr* 23:1324–1334

Legendre, P. & De Cáceres, M. (2013) Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecol Lett* 16:951–963

Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, *112*(7), 2076–2081. https://doi.org/10.1073/pnas.1424997112

Leray, M., & Knowlton, N. (2016). Censusing marine eukaryotic diversity in the twenty-first century. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1702), 20150331. https://doi.org/10.1098/rstb.2015.0331

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T. & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*(1). https://doi.org/10.1186/1742-9994-10-34

Madduppa, H. H., Subhan, B., Suparyani, E. N. Y. & Siregar, A. M. (2013) Dynamics of fish diversity across an environmental gradient in the Seribu Islands reefs off Jakarta. *Biodiversitas Journal of Biological Diversity* 14(1):17-24

Malay, M. C. D. & Paulay, G. (2010). Peripatric speciation drives diversification and distributional pattern of reef hermit crabs (Decapoda: Diogenidae: Calcinus). *Evolution*, 64, 634-662.

Marko, P. B. & Moran, A. L. (2009) Out of sight, out of mind: High cryptic diversity obscures the identities and histories of geminate species in the marine bivalve subgenus Acar. *J Biogeogr* 36:10.

Marwayana, O.N., Gold, Z. & Barber, P. (2021). Environmental DNA in a Global Biodiversity Hotspot: Lessons from Coral Reef Fish Diversity Across the Indonesian Archipelago. *bioRxiv* 2021.02.19.432056; doi: https://doi.org/10.1101/2021.02.19.432056

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061217

Meyer, C. P., & Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, *3*(12), e422. https://doi.org/10.1371/journal.pbio.0030422

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. & Worm, B. (2011) How many species are there on earth and in the ocean? *PLoS Biol* 9(8):e1001127

Mustika, P. L. K., Gunawan, T., & Erdmann, M. V. (2013). *A Marine Rapid Assessment (MRAP) of the Anambas Islands Marine Tourism Park, 3-31 May 2012*. Denpasar.

Mustika, P. L. K., Ratha, I. M. J., & Purwanto, S. (2012). *Bali Marine Rapid Assessment Program 2011*. Denpasar.

Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., Green, L. E. & Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.12895

Obst, M., Exter, K., Allcock, A.L., Arvanitidis, C., Axberg, A., Bustamante, M., Cancio, I., Carreira-Flores, D., Chatzinikolaou, E., Chatzigeorgiou, G., Chrismas, N., Clark, M., Comtet, T., Dailianis, T., Davies, N., Deneudt, K., Diaz de Cerio, O., Fortič, A., Gerovasileiou, V., Hablützel, P., Keklikoglou, K., Kotoulas, G., Lasota, R., Leite, B., Loisel, S., Lévêque, L., Levy, L., Malachowicz, M., Mavrič, B., Meyer, C., Mortelmans, J., Norkko, J., Pade, N., Power, A., Ramšak, A., Reiss, H., Solbakken, J., Staehr, P., Sundberg, P., Thyrring, J., Troncoso, J., Viard, F., Wenne, R., Yperifanou, E.I., Zbawicka, M., Pavloudi, C. (2020) A marine biodiversity observation network for genetic monitoring of hard-bottom communities (ARMS-MBON). *Frontiers in Marine Science.* 7:572680. doi: 10.3389/fmars.2020.572680

O'Malley, M. A. (2008) "Everything is everywhere: but the environment selects": ubiquitous distribution and ecological determinism in microbial biogeography. *Stud Hist Philos Sci Part C Stud Hist Philos Biol Biomed Sci* 39:314–325

Ogle, D. H., P. Wheeler. & A. Dinno. 2018. FSA: Fisheries Stock Analysis. R package version 0.8.22, https://github.com/droglenc/FSA).

Oksanen, J. (2017). Vegan: ecological diversity. *R Package Version 2.4-4*. https://doi.org/10.1029/2006JF000545

Pavan-Kumar, A., Gireesh-Babu, P., & Lakra, W. S. (2015). DNA Metabacoding: a new approach for rapid biodiversity assessment. *Journal of Cell Science and Molecular Biology*. Volume 2. Issues 1.

Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., & Esling, P. (2016). Protist metabarcoding and environmental biomonitoring: Time for change. *European Journal of Protistology*, *55*, 12–25. https://doi.org/10.1016/j.ejop.2016.02.003

Pearman, J. K., Anlauf, H., Irigoien, X., & Carvalho, S. (2016). Please mind the gap - Visual census and cryptic biodiversity assessment at central Red Sea coral reefs. *Marine Environmental Research*, *118*, 20–30. https://doi.org/10.1016/j.marenvres.2016.04.011

Peñaflor, E. L., Skirving, W. J., Strong, A. E., Heron, S. F., & David, L. T. (2009). Sea-surface temperature and thermal stress in the Coral Triangle over the past two decades. *Coral Reefs*, *28*(4), 841–850. https://doi.org/10.1007/s00338-009-0522-8

Plaisance, L., Brainard, R., Caley, M. J., & Knowlton, N. (2011a). Using DNA Barcoding and Standardized Sampling to Compare Geographic and Habitat Differentiation of Crustaceans: A Hawaiian Islands Example. *Diversity*, (4), 581–591. https://doi.org/10.3390/d3040581

Plaisance, L., Caley, M. J., Brainard, R. E., & Knowlton, N. (2011b). The Diversity of Coral Reefs : What Are We Missing ?, *6*(10). https://doi.org/10.1371/journal.pone.0025026

Podani, J. & Schmera, D. (2011) A new conceptual and methodological framework for exploring and explaining pattern in presence - absence data. *Oikos* 120:1625–1638

Pommier, T., Canbäck, B., Riemann, L., Boström, K. H., Simu, K., Lundberg, P., Tunlid, A. & Hagström, Å. (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* 16:867–880

Rex, M. A., Crame, J. A., Stuart, C. T. & Clarke, A. (2005) Large-scale biogeographic patterns in marine mollusks: a confluence of history and productivity? *Ecology* 2005: 2288–2297.

Roy, K., Jablonski, D., Valentine, J. W. & Rosenberg, G. (1998) Marine latitudinal gradients: tests of causal hypotheses. *P Natl Acad Sci* USA 95: 3699–3702.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0009490

Ransome, E., Geller, J. B., Timmers, M., Leray, M., Mahardini, A., Sembiring, A., Collins, A. G. & Meyer, C. P. (2017). The importance of standardization for biodiversity comparisons: A case study using autonomous reef monitoring structures (ARMS) and metabarcoding to measure cryptic diversity on Mo'orea coral reefs, French Polynesia. *PLoS ONE*, *12*(4), 1–19. https://doi.org/10.1371/journal.pone.0175066

Roberts, C. M., McClean, C. J., Veron, J. E., Hawkins, J. P., Allen, G. R., McAllister, D. E., Mittermeier, C. G., Schueler, F. W., Spalding, M., Wells, F., Vynne, C. & Werner, T. B. Marine biodiversity hotspots and conservation priorities for tropical reefs. Science. 2002 Feb 15;295(5558):1280-4. doi: 10.1126/science.1067728. PMID: 11847338.

Rocha LA, Rocha CR, Robertson DR, & Bowen BW (2008) Comparative phylogeography of Atlantic reef fishes indicates both origin and accumulation of diversity in the Caribbean. *BMC Evolutionary Biology* 16:1–16

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*. https://doi.org/10.7717/peerj.2584

Simmonds, S. E., Chou, V., Cheng, S. H., Rachmawati, R., Calumpong, H. P., Ngurah Mahardika, G, & Barber, P. H. (2018) Evidence of host-associated divergence from coral-eating snails (genus Coralliophila) in the Coral Triangle. *Coral Reefs* 37:355–371

Souter, P. (2010) Hidden genetic diversity in a key model species of coral. *Marine Biology* 157(4):875–885

Sul, W. J., Oliver, T. A., Ducklow, H. W., Amaral-Zettlera, L. A. & Sogin, M. L. (2013) Marine bacteria exhibit a bipolar distribution. *Proc Natl Acad Sci U S A* 110(6):2342-7

Tittensor, D. P., Mora, C. Jetz, W., Lotze, H. K., Ricard, D., Berghe, E., Vanden & Worm, B. (2010) Global patterns and predictors of marine biodiversity across taxa. *Nature* 466:1098–101

Turak E, DeVantier L (in press) Biodiversity and conservation priorities of reef-building corals in the Papuan Bird's Head Seascape. In: Katz LS, Firman A, Erdmann MV (eds) A Rapid Marine Biodiversity Assessment of Teluk Cenderawasih and the FakFak-Kaimana Coastline of the Papuan Bird's Head Seascape, Indonesia. RAP Bulletin of Biological Assessment. Conservation International, Washington, D.C.

Turak E, Souhoka J (2003) Coral diversity and the status of coral reefs in the Raja Ampat Islands. In: Donnelly R, Neville D, Mous P (eds) Report on a rapid ecological assessment of the Raja Ampat Islands, Papua, Eastern Indonesia, held October 30 – November 22, 2002. The Nature Conservancy Southeast Asia Center for Marine Protected Areas, Sanur, Bali Indonesia.

Veron, J. E. N., Devantier, L. M., Turak, E., & Green, A. L. (2009). Delineating the Coral Triangle. *Galaxea, Journal of Coral Reef Studies*, *11*, 91–100.

Voris, H. K. (2000). Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems and time durations. *Journal of Biogeography*, *27*(5), 1153–1167. https://doi.org/10.1046/j.1365-2699.2000.00489.x

Wangensteen, O. S., & Turon, X. (2016). Metabarcoding Techniques for Assessing Biodiversity of Marine Animal Forests. In *Marine Animal Forests* (pp. 1–34). https://doi.org/10.1007/978-3-319-17001-5

Waskom, M., Botvinnik, O., Y. Drewokane, Y., Hobson David, Y., Halchenko, S., Lukauskas, J.B., Cole, J., Warmenhoven de Ruiter, J. S., Hoyer, J., Vanderplas, S., Villalba, G., Kunter, E., Quintero, M., Martin, A., Miles, T., Meyer, T., Augspurger, T., Yarkoni, P., Bachant, M., Williams, C., Evans, C., Fitzgerald, B. D., Wehner, G., Hitz, E., Ziegler, A., Qalieh, A. &Lee. (2017) Seaborn. http://dx.doi.org/10.5281/zenodo.592845.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York. Media*. https://doi.org/10.1007/978-0-387-98141-3

Whittaker, R. H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol Monogr* 30: 279–338.

Williams, S. T. & Reid, D. G. (2004) Speciation and diversity on tropical rocky shores: A global phylogeny of snails of the genus Echinolittorina. *Evolution* 58(10):2227-2251

Woodland, D. J. (1983) Zoogeography of the Siganidae (Pisces): an interpretation of distribution and richness patterns (Indo-Pacific). *Bull Mar Sci* 33:713–717

Woodland, D. J. (1986). Wallace's line and the distribution of marine inshore fishes. In Uyeno, T. et al. (eds.) Indo-Pacific Fish Biology, Second International Conference, Tokyo, Japan, July 29-Aug. 3, Tokyo, Japan: *The Ichthyological Society of Japan*, pp.453-460.

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

CHAPTER 2

The Diversity of Indonesian Marine Bacteria: Assessing Baas Becking's Hypothesis "Everything Is Everywhere" Using Autonomous Reef Monitoring Structures and Metabarcoding

Abstract

The Coral Triangle is a global marine biodiversity hotspot with pronounced biodiversity gradients in fishes, corals, and other invertebrates with biodiversity peaking in Eastern Indonesia and declining in western parts of the Coral Triangle. While theory predicts that for organisms smaller than 1mm "everything is everywhere", increasingly research shows that eukaryotic taxa have strong microbial associations, suggesting that microbial communities could display similar patterns of biodiversity and regional differentiation as their hosts. Such direct comparisons are challenging, however, as they require simultaneous collection of a wide diversity of benthic marine taxa and their associated microbial communities. In this study, we employ Autonomous Reef Monitoring Structure (ARMS) and DNA metabarcoding to conduct the first systematic study of marine bacterial diversity across Indonesia, the heart of the Coral Triangle. Results showed substantial regional differentiation in microbial diversity. Consistent with larger metazoans, diversity was highest in Raja Ampat, the most biodiverse region of the Coral Triangle, and lower in other regions, microbial diversity was correlated with both eukaryote and metazoan diversity. Microbial communities also showed a highly significant pattern of isolation by distance, indicating that limits to dispersal are influencing geographic differentiation, a pattern observed across all three size classes of organisms inhabiting the ARMS. Results indicate that associations with larger eukaryotes and physical limitations to dispersal differentiate microbial communities in the Coral Triangle, as seen in other marine organisms. These results

are counter to the Baas Becking's hypothesis that "everything is everywhere, but the environment selects", and provide novel insights into the processes shaping marine microbial diversity in the world's most diverse marine ecosystem.

Introduction

Discerning how organisms are structured over geographic space is essential to understanding how populations evolve. For example, allopatric speciation (divergence due to geographic isolation) is typically regarded as the most common mechanism of speciation (Mayr, 1963; Lande, 1980; Coyne & Orr, 2004), migration can either facilitate or hinder local adaptation (Lenormand, 2002; Kawecki & Ebert, 2004), and population structure congruence shapes the coevolutionary process (Lively, 1999; Thompson & Cunningham, 2002; Thompson, 2005; Thompson, 2009). In larger metazoans, it is well established that limits to dispersal create structured populations (Bohonak, 1999); however, for many microorganisms (e.g. < 1mm), the basic question of whether a species is dispersal limited can be ambiguous (Finlay & Esteban, 2004; Hedlund & Staley, 2004; Martiny et al., 2006).

Baas Becking's (1934) hypothesis that "everything is everywhere, but the environment selects, (EEBES)" introduced a cosmopolitan view of microorganisms where dispersal is unlimited. Indeed, the concept of ubiquitous dispersal has persisted for over eighty-five years and continues to shape views of microbial ecology and evolution (Finlay, 2002; Fenchel & Finlay, 2004; De Wit & Bouvier, 2006; O'Malley, 2007; Thurber, 2009; Whittaker & Rynearson, 2017). In 2002, Finlay expanded on the theoretical expectations of EEBES by proposing that a size threshold of approximately 1mm determines ubiquity (i.e. organisms smaller than 1mm are predicted to have cosmopolitan distributions whereas organisms larger than 1mm experience limits to dispersal) (Figure 2-1; Finlay, 2002).

Early evidence for cosmopolitan dispersal was founded on species morphology; however, the axiom that "everything is everywhere" continued to gain support in the genetic era – 16S rRNA genotypes are found globally (e.g., Glöckner et al., 2000; Bolch & Reynolds, 2002; Zwart 2002), which is often cited as evidence for ubiquitous dispersal (Finlay & Esteban, 2004). When microorganisms do show geographic structuring of genotypes, there are difficulties in determining whether the structure reflects limited dispersal or environmental selection – the latter being consistent with the second component Baas Becking's hypothesis "but the environment selects" (Wise et al.,1995; Fenchel, 2003; Antony-Babu, 2008). The "environmental selection" component of EEBES as a force shaping biogeographic patterns is often ignored, as many studies refute the hypothesis upon discovering spatial structure in microbial taxa. For example, a premier microbiology text book (Madigan et al., 2003) spends only a single paragraph discussing microbial dispersal, and dismisses EEBES, with the singular statement "The distribution of microorganisms in nature resembles that of macroorganisms in the sense that a given species resides in certain places but not others; that is, everything is not everywhere."

Isolation by distance (IBD), a correlation between geographic and genetic distance among populations (Slatkin, 1993) is a common method for identifying limited dispersal within species. A similar approach was used to refute ubiquitous dispersal in the hyperthermophilic archaeon *Solfolobus islandicus* (Whitaker et al., 2003; Whitaker et al., 2005; Whitaker 2006). This example, however, may be a rare exception that proves the rule, given the extreme chemical environment of the archaea (Pommier et al., 2007; Fuhrman & Steele 2008; Brown et al., 2012; Ghiglione et al, 2012; Kelly et al., 2014). Similar patterns of isolation by distance have been observed in symbiotic bacteria, where symbiont population structure is shaped by coevolved

interactions with pathogens and/or animal hosts (Wirth et al., 2005; Vollmer et al., 2011; Caldera & Currie, 2012; Caldera et al., 2019). These population-scale IBD analyses can be expanded to assess whether microbial communities exhibit biogeography by correlating community dissimilarity indices (rather than genetic distance) with geographic distance (e.g. Hillebrand et al., 2001).

Another potential driver of microbial biogeography could be symbiotic associations with larger eukaryotic organisms. On coral reefs, microbes form complex associations with other organisms, including coral, sponges, giant clam, and algae (Ashen & Goff, 2000; Webster et al., 2001), associations known as holobionts (Wegley et al., 2007; Barott & Rohwer, 2012). These animal-microbial associations can be obligate and co-evolved (Nishiguchi et al., 1998) to such an extent that microbial communities can be shaped by the phylogeography of their host (Coryell et al., 2018). As such, processes shaping biogeography and phylogeography in marine ecosystems could shape patterns of microbial community diversity by shaping the composition of benthic eukaryote communities (Kelly et al., 2014).

The Coral Triangle is the world's largest and most diverse marine ecosystem (Allen & Werner, 2002; Bellwood et al., 2005; Veron et al., 2009). This area hosts 76% of coral species in the world and 37% of the world's reef fish, 8% of which are endemic or locally restricted species (Allen, 2008; Veron et al., 2009). Defined by the presence of ≥500 scleractinian coral species (Veron et al., 2009), the Coral Triangle displays pronounced biodiversity gradients in fishes, corals and other invertebrates (Meyer, 2003; Roberts et al., 2002; Allen, 2008; Bellwood & Meyer, 2009; Gaither & Rocha, 2013). Along this gradient marine biodiversity peaks in Eastern Indonesia and the Philippines and declines in western parts of the Coral Triangle. In addition, species within this region can display strong phylogeographic patterns, particularly among

populations in Eastern, Central and Western Indonesia (Barber & Erdmann, 2006; Barber et al., 2011; Carpenter et al., 2011; DeBoer et al., 2014). If the above taxa have strong microbial associations, microbial communities could display similar patterns of biodiversity and regional differentiation. However, such direct comparisons are challenging, as they require simultaneous collection of a wide diversity of benthic marine taxa and their associated microbial communities.

Autonomous Reef Monitoring Structures (ARMS) are artificial structures designed to mimic the structure of the coral reef ecosystem and provide a substrate for marine biota to colonize providing a standardized method to survey coral reef diversity (Knowlton et al., 2010). Documenting marine diversity with ARMS relies on DNA metabarcoding, a technique where entire communities of organisms can be identified through species-specific DNA sequences obtained through high-throughput DNA sequencing of community DNA (Knowlton et al., 2010; Plaisance et al., 2011; Leray & Knowlton, 2015; Leray & Knowlton, 2016; Pearman et al., 2016; Al-Rshaidat et al., 2016; Ransome et al., 2017). DNA metabarcoding expands on traditional surveys of biodiversity (e.g., biological collections, visual counts) by capturing a wider range of taxa, including rare or cryptic species that may be overlooked through traditional survey methods. Importantly, metabarcoding also allows for simultaneous comparison of co-distributed eukaryotic and prokaryotic diversity. Moreover, standard ARMS processing fractionates biological material by sizes that span the Finlay (2002) proposed theoretical 1mm threshold for dispersal (Figure 2-1), thereby allowing for testing the prediction that bacteria associating with larger sized holobiont should have greater biogeographic signal.

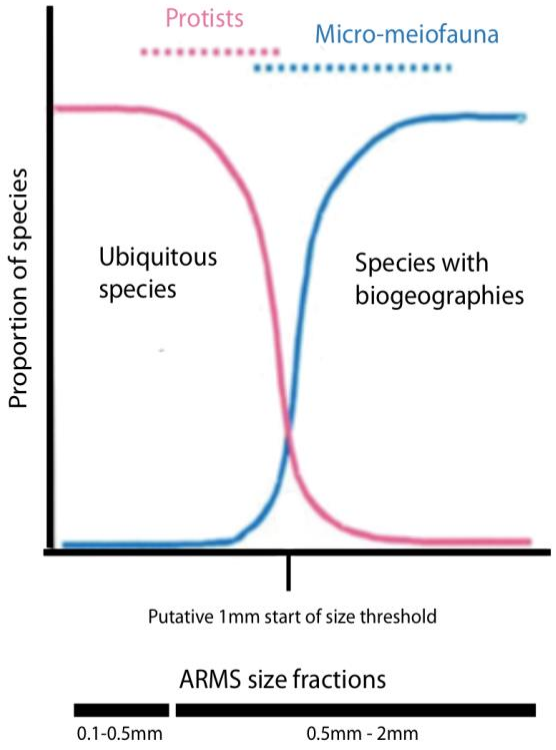


Figure 2-1. Theoretical size threshold determining ubiquity (modified from Finlay, 2002), including where ARMS size fractions fall along the size threshold.

Cahyani et al. (Chapter 1) showed that benthic eukaryotic communities living on ARMS display strong patterns of regional variation in biodiversity across the Indonesian archipelago, with the highest richness, species turnover, and endemic diversity in Eastern Indonesia, similar to patterns for corals (Veron et al., 2009) and fishes (Roberts et al., 2002; Allen, 2008; Bellwood & Meyer, 2009). In this study, we employ ARMS and metabarcoding to conduct the first systematic study of marine bacterial diversity across Indonesia, the heart of the Coral Triangle. Specifically, we test 1) whether bacterial associates conform to the "everything is everywhere" size threshold hypothesis (Finlay 2002) and 2) whether the bacterial diversity comports with the well-known West to East marine biodiversity gradient across the Indonesian archipelago.

Materials and Methods

ARMS Deployment, Collection, and Sampling

We deployed ARMS across multiple biogeographic regions in Indonesia to capture well-established biodiversity gradients (Supplemental Table S2-1) (Meyer, 2003; Allen, 2008; DeBoer et al., 2008; Bellwood & Meyer, 2009; Veron et al., 2009; Gaither & Rocha, 2013). These sites ranged from Aceh in Western Indonesia, a site that is officially outside of the Coral Triangle (Hoeksema, 2007; Bellwood & Meyer, 2009) to Raja Ampat, a region of Eastern Indonesia that is known for having the highest coral (Veron et al., 2009) and fish diversity (Allen, 2008) in the world (Figure 2-2). We employed a spatially hierarchical sampling design, deploying ARM in sets of three (spaced 3-5 m apart), at four different reefs within five marine ecoregions defined by Spalding et al. (2007) totalling 60 ARMS. To minimize habitat variation across deployments, we anchored ARMS to the seafloor, in forereef environments at a standardized depth of 10m. To ensure that chemical variation in the materials used to build the ARMS would not impact our results, we had a single supplier manufacture ARMS components from a single batch of source materials.

We deployed ARMS underwater for three years (2013-2016) to be colonized with benthic reef communities (Figure 1-2A, 2B.; https://www.oceanarms.org/protocols/arms-assembly). We then recovered the ARMS and the associated organisms, following the methods of Ransome et al. (2017). Briefly, we first encapsulated ARMS in crates lined with 40 µm nitex mesh to prevent the escape of motile organisms during recovery. We then transported ARMS in crates of aerated filtered seawater back to the lab where we disassembled and photographed the top and bottom of every ARMS plate. Next, to prevent organisms with higher biomass from swamping the signal of smaller organisms, we separated motile organisms into three size classes using geological sieves.

First, we removed all organisms >2 mm, and saved these for future taxonomic study. We then isolated a "500 µm fraction" (all motile organisms from 500 µm-2 mm) and a "100 µm fraction" (all motile organisms between 106-500 µm). Fractions were isolated sequentially, from largest to smallest, and fractions were washed with filtered seawater at each step. After washing, we consolidated each sample fraction with a final wash in filtered seawater in a 45 µm nitex net, preserved the samples in 95% ethanol, and stored the samples at -20 °C until DNA extraction.

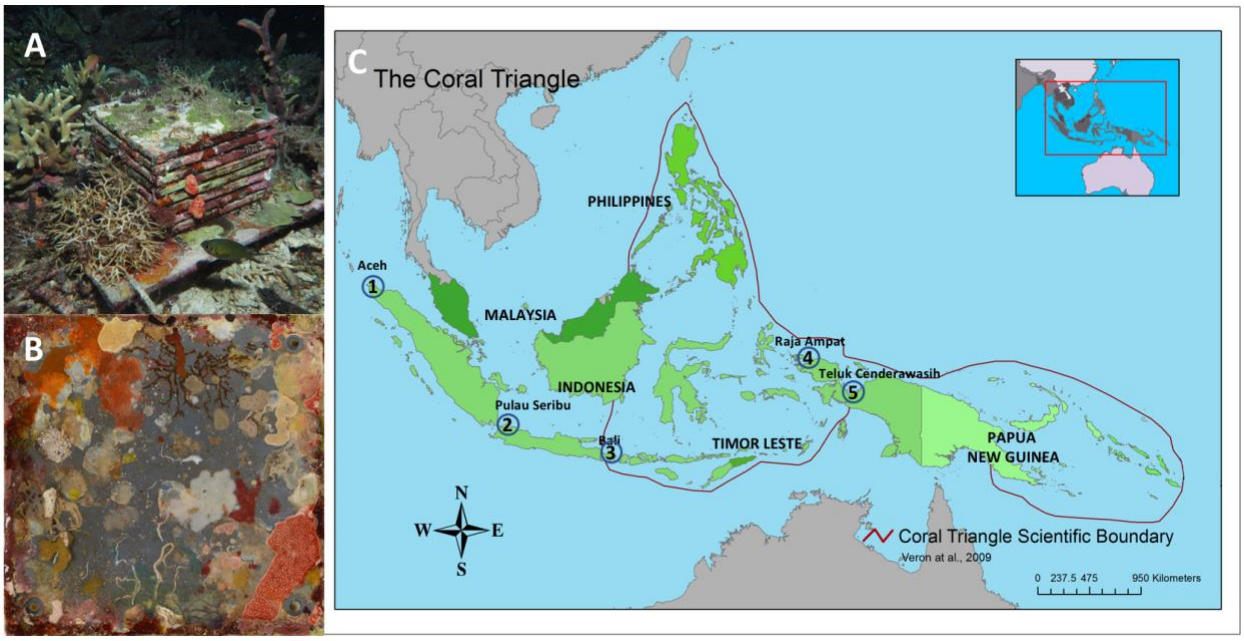


Figure 2-2. A) Autonomous Reef Monitoring Structure (ARMS) photographed underwater. B) ARMS plate colonized by organisms. C) Map of the Coral Triangle and sampling locations: (1) Pulau Weh, Aceh, (2) Kepulauan Seribu, Jakarta, (3) Pemuteran, Bali, (4) Raja Ampat, West Papua, and (5) Teluk Cenderawasih, West Papua. The Coral Triangle Scientific Boundary (red line) is based on Veron et al. (2009).

To document non-motile fractions, we scraped all encrusted or sessile biota from the ARMS plates into filtered seawater and then homogenized it with a kitchen blender for 30 s at maximum speed. We rinsed the homogenate with filtered sea water taken from the ARMS recovery site into a 45 μm Nitex mesh collection net until the water ran clear, then placed approximately 10 g of the homogenate into a 50 ml falcon tube filled with DMSO, stored at -20 °C until DNA extraction.

DNA Preparation and Extraction

To remove inorganic material (e.g. sediment) that could inhibit DNA amplification, we performed a series of decantations (see Leray and Knowlton, 2015; https://www.oceanarms.org/protocols/molecular-analysis/bulk-dna-extractions/sample-preparation). Next, we extracted DNA from the 100 μm, 500 μm, and sessile samples using the MO-BIO Powermax® Soil DNA Isolation Kit according to the manufacturer's protocol with the addition of 400 μg/ml Proteinase K. To remove potential PCR inhibitors, we purified the DNA extractions using MO-BIO PowerClean® DNA Clean-Up Kits and quantified DNA concentrations using Qubit dsDNA HS Kit. The decantation and DNA extraction were performed at Yayasan Biodiversitas Indonesia (Bionesia), Denpasar, Bali, Indonesia and Laboratory of Marine Molecular Genetics, Research Center for Oceanograpy, Indonesian Institute of Sciences, Jakarta, Indonesia.

To assess microbial diversity within each of these three size fractions, we amplified the 16S rRNA gene using primers 515f and 806r, which target the V4 region (Caporaso et al., 2012; Walters et al., 2015). Library preparation for 16S rRNA amplicons followed a single indexing approach where barcodes were incorporated into the forward primer to facilitate multiplexing of up to 96 samples per run. To account for potential PCR bias (Ficetola et al., 2015; Nichols et al.,

2018) we performed PCR in triplicate following standard Earth Microbiome procedures. Each PCR reaction was 25 μL in volume, consisting of 0.5 μL of 0.2 μM each forward or reverse primer, 10 ml Platinum Hot Start PCR Master Mix (Thermo Fisher), using 5 ng of extracted DNAas template. The PCR cycling profile was: initial denature at 94 °C for 3 min, 35 cycles each at 94°C for 45 s, 50°C for 60 s, and 72 °C for 90 s, followed by a final extension step at 72 °C for 10 min. Triplicate PCR products were electrophoresed on a 1.2% agarose gel and then pooled. We then cleaned successful PCR reactions using the Qiagen UltraClean PCR Cleanup Kit. The library was validated via qPCR using the KAPA library quantification kit and diluted to a final concentration of 2nM. The pooled PCR products were sequenced on an Illumina MiSeq using V2 300-cycles kit with 20% PhiX DNA added to each run to improve data quality. All DNA sequencing was performed at the Laboratories of Analytical Biology, Smithsonian Institution National Museum of Natural History, Washington DC.

Data analyses

We demultiplexed raw paired-end FASTQ reads and then imported the resulting sequences into *QIIME2*, ver. 2017.8.0 (the Quantitative Insights Into Microbial Ecology 2 program, https://qiime2.org/) (Bolyen et al., 2018). Sequences were quality filtered, paired-end, and analysed using *QIIME2* version 2017.8.0. We used the Divisive Amplicon Denoising Algorithm 2 (*DADA2*) software (Callahan et al., 2016), wrapped in *QIIME2*, for quality filtering, trimming, de-noising, and merging the data, and then removed chimeric sequences using the consensus method. Next, we used the *LULU* algorithm (Frøslev et al., 2017) to filter out spurious sequences that may originate from PCR and/or sequencing errors, intra-individual variability (pseudogenes, heteroplasmy). *LULU* filters based on sequence similarity and co-occurrence rate with more abundant clusters, allowing us to curate the datasets while avoiding arbitrary

abundance filters (Frøslev et al., 2017; Brandt et al., 2020). We ran *LULU* with minimum relative co-occurrence of 0.95, using a minimum similarity threshold (minimum match) at 84% (default).

To assign taxonomy to the resulting Amplicon Sequence Variants (ASVs), we trained a feature classifier in *QIIME2* against the SILVA SSU non-redundant database (132 release) (https://github.com/qiime2/q2-feature-classifier), employing a default confidence threshold of 0.7. To limit results to microbes, we filtered out all mitochondrial and chloroplast sequences from the resulting feature table. We then aligned the remaining sequences with *MAFFT* (Katoh & Standley, 2013) and used this alignment for phylogenetic reconstruction in *FastTree* (Price et al., 2010).

Statistical analyses

To test for saturation of ASV discovery, we created rarefaction curves with *iNEXT* (Chao et al., 2014; Hsieh et al., 2016) and *Ranacapa* package using *ggrare* command (Kandlikar et al., 2018). Then, we summarized the taxonomic composition of each sample with *phyloseq* (McMurdie & Holmes, 2013), generating stacked bar plots summarizing taxonomic composition and sequence abundance using *ggplot2* (Wickham, 2009) in R. Because of paucity of microbial barcoding data from the Coral Triangle, we merged taxa at the phylum level and removed groups that represented less than 2% total abundance of the community.

To compare community composition across the five regions and reefs therein, we first examined patterns of microbial diversity, calculating ASV richness for individual ARMS, sites, and regions. Next, we calculated alpha and beta diversity statistics using the *phyloseq* package in R (R development core team) (McMurdie & Holmes, 2013), testing for significance in one way ANOVA framework that examined the richness from total dataset and per size fraction (100 μm,

500 μm, and sessile) across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Cenderawasih Bay). To further explore patterns of microbial community composition and turnover, we used the *VennDiagram* package (Chen & Boutros, 2011) in R to visualize unique and shared microbial diversity across spatial scales, and across size fraction. We then compared microbial diversity patterns to eukaryote diversity obtained from the same ARMS units (Cahyani et al. Chapter 1) using linear regression and ANOVA to test significance. To test whether microbial associations with metazoans drive patterns in eukaryotes, we repeated the above including only metazoans.

To test for regional differentiation of microbial communities, we conducted multivariate analyses (PERMANOVA) based on Jaccard dissimilarity distances in the *vegan* package (Oksanen, 2017) in R (R development core team) testing statistical significance using 9999 permutations and a significance level of $\alpha = 0.05$. We calculate the compositional dissimilarity using '*adonis*' command and the homogeneity of group dispersion using '*betadisper*' command in *vegan* package (Oksanen, 2017). Then we conducted Principles Coordinates Analyses (PCoA) using the *Ampvis2* package (Andersen et al., 2018) with the ordination function of *phyloseq* for both Jaccard and Bray Curtis dissimilarity matrices and generate the ordination plot using *ggplot2* (Wickham, 2009). Then, to test whether differentiation resulted from isolation by distance, we performed Mantel tests between matrices of the natural log of geographic distance and Jaccard distance (dissimilarity) for each of three ARMS fragments (100 μm, 500 μm, and sessile). Geographic distances among ARMS were calculated using Haversine's formula for determining the great-circle distance between two points on a sphere given their longitudes and latitudes. Mantel tests were performed in Arlequin (Excoffier et al., 2005) and significance tests used 1000 permutations.

Results

DNA sequencing

We recovered a total of 59 of 60 ARMS; one ARMS unit from Sumur Tiga in Aceh could not be recovered. Due to one amplification failure (the sessile fraction of one ARMS from Bali we obtained 176 individual samples from 20 sites using 16S rRNA. We generated a total of 23,205,524 bacterial sequence reads from these samples, ranging from 36,719 to 364,814 reads per sample (Supplemental Table S2-2).

Rarefaction plots on each ARMS unit showed that sequencing depth was sufficient to capture all diversity (Figure 2-3). However, at the region level, ASV discovery had less saturation across sequencing depth (Figure 2-3). To ensure that downstream diversity analyses were not impacted by variation in sequencing depth, we rarefied all samples to an even depth of 36,719 sequences per sample. Following rarefaction, quality filtering, and exclusion of chimeras, we obtained a total of 6,462,544 reads and 39,358 Amplicon Sequence Variants (ASVs) (Supplemental Table S2-2).

Microbial Community Composition

Microbial communities displayed very similar community composition across the five regions when examining ARMS summed across individual units in each region (i.e. combining all three size fractions; (Figure 2-4A). Summed across all regions, *Proteobacteria* was numerically dominant in terms of ASVs (29.3%) and sequence reads (49.7%), with *Gammaproteobacteria* (24.2% of all sequences) and *Alphaproteobacteria* (19.3% of all sequences) being the most common *Proteobacteria*. The next most abundant phyla were

Chloroflexi (9.7%), *Bacteroidetes* (7.7%), *Cyanobacteria* (6.7%), *Acidobacteria* (6.6%), and *Planctomycetes* (5.3%)

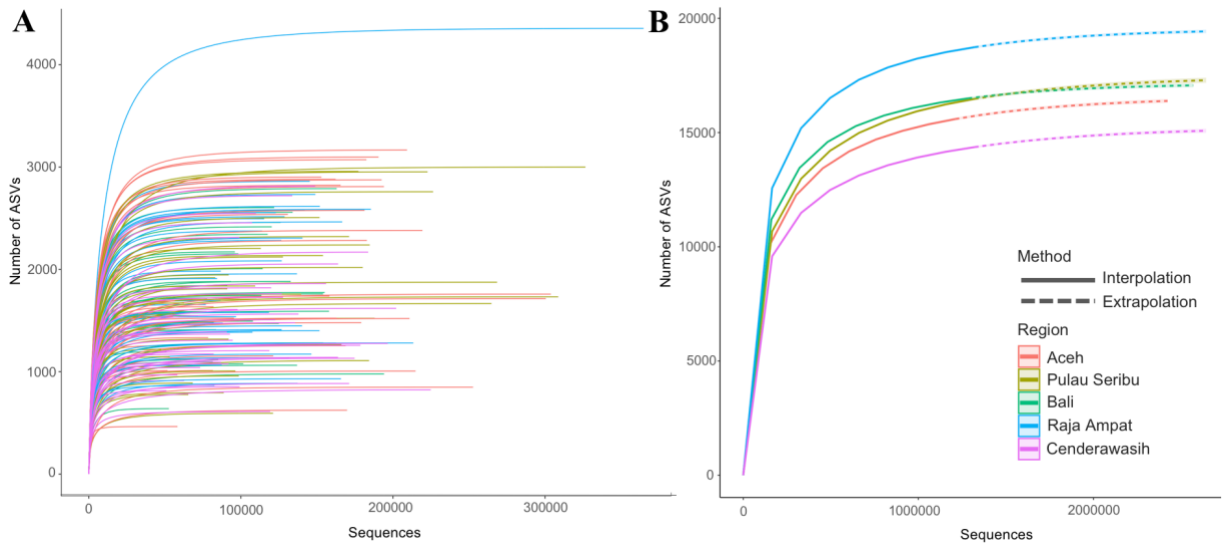


Figure 2-3. Alpha diversity rarefaction plot of full dataset generated with Ranacapa (Kandlikar et al., 2018) and iNEXT packages (Chao et al., 2014; Hsieh et al., 2016) in R (R development core team). Number of amplified sequence variants (ASVs) (left axis) plotted against sequencing depth (bottom axis) for A) each individual ARMS unit, and B) each of the five regions.

Taxonomic composition of microbial communities varied across each of the size fractions (Figure 2-4B). The 100 µm and 500 µm fractions were the most similar in composition, and were dominated by *Proteobacteria* (51% and 55%, respectively), *Bacteroidetes* (8% and 13%, respectively), *Cyanobacteria* (10% and 6%, respectively), and *Planctomycetes* (6% and 7%, respectively). The community from the sessile fraction was distinctly different, being *Proteobacteria* (43%), but with higher proportions of *Chloroflexi* (20%), *Acidobacteria* (10.5%), *Actinobacteria* (4%) and *PAUC34f* (2.6%).

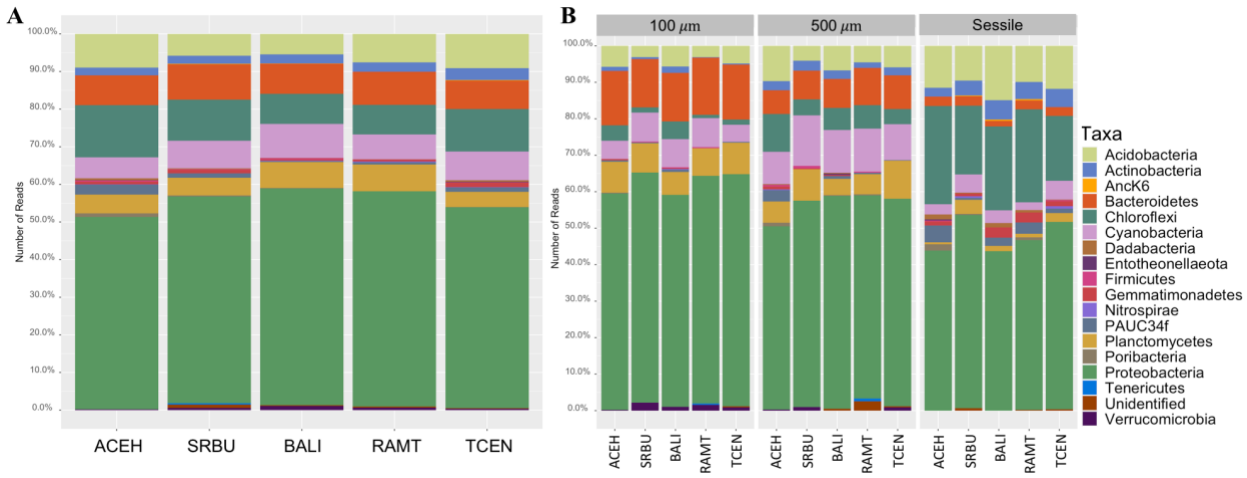


Figure 2-4. Taxonomic composition of eukaryote diversity at Phylum level for A) microbial communities for entire ARMS units, and B) for each of the three size fractions. Bar plot showing taxa relative abundance of the sample across five different location in Indonesia. The bar plot constructed based on phyla contribute more than 2% of the relative abundance of each sample.

Patterns of Microbial Diversity

On average, each ARMS unit captured a maximum of 54.4% (Aceh) and a minimum of 50.0% (Raja Ampat) of microbial diversity at each site (Figure 2-5A). In turn, each site captured a maximum of 45.3% (Raja Ampat) and minimum of 43.9% (Teluk Cenderawasih) of regional microbial diversity within each region (Figure 2-5B). Abasolute ASV richness per ARMS was highest in Raja Ampat (mean = 4,249 ASVs) and lowest in Teluk Cenderawasih (mean = 3,166 ASVs) (Figure 2-5C), a pattern that was repeated examining only local endemic ASVs (Figure 2-5D). Similarly, absolute ASV richness per site was highest in Raja Ampat (mean = 8,484 ASVs) and lowest in Teluk Cenderwasih (mean = 6,309 ASVs) (Figure 2-5E), a pattern that was repeated examining only local endemic ASVs (Figure 2-5F; Table 2-1, 2-2).

ANOVA analysis showed that total per ARMS ASVs richness and endemic per ARMS ASV richness are significantly different between regions (*p-value* = 0.0154 and 0.00319, respectively) (Supplemental Table S2-3). Tukey test for per ARMS ASVs richness and per ARMS endemic ASVs showed that significance was driven by differences between Teluk Cenderawasih and Raja Ampat (Supplemental Table S2-4).

Table 2-1. Amplicon Sequence Variants (ASVs) diversity captured in a set of three ARMS based on 16S rRNA metabarcoding.

Region	Maximum ASVs per 3-ARMS Set	Minimum ASVs per 3-ARMS Set	Average ASVs per 3-ARMS Set	S.D.	% Regional ASV Diversity
Aceh	8140	4489	7201.5	1367.8	46.20%
Pulau Serbu	8042	6407	7275.3	606.8	44.20%
Bali	8018	6798	7375.3	511.4	44.70%
Raja Ampat	9040	7848	8483.75	573	45.30%
Teluk Cendrawasih	7612	4606	6308.8	1170.8	43.90%

Table 2-2. Amplicon Sequence Variants (ASVs) diversity captured in a single ARMS based on 16S rRNA metabarcoding.

Region	Total ASVs per region	Average total ASVs per ARMS	S.D.	Average local diversity in 1 ARMS	% Regional ASV Diversity
Aceh	15591	3881	695.8	54.4%	24.9%
Pulau Seribu	16474	3642	495.5	50.1%	22.1%
Bali	16498	3744	422.4	50.7%	22.7%
Raja Ampat	18736	4249	491.5	50.0%	22.7%
Teluk Cendrawasih	14363	3166.3	585.1	50.4%	22.0%
Average	16332.4	3736.5	538.1	51.1%	22.9%

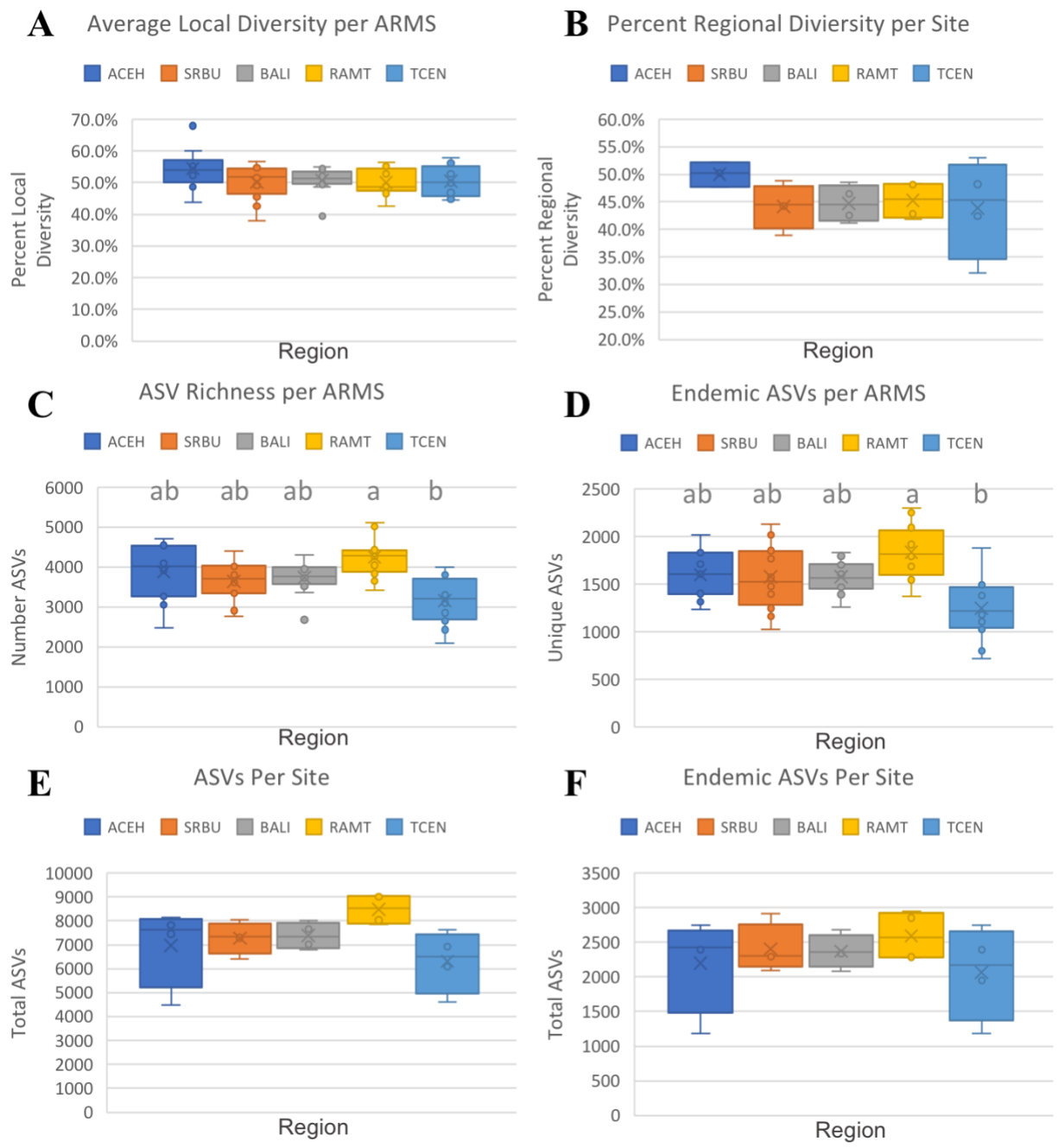


Figure 2-5. Microbial diversity patterns from autonomous reef monitoring structures across the Indonesian archipelago, including A) the average diversity at a site captured in a single ARMS unit, B) the average regional diversity captured in a single sampling site, C) ASV richness per ARMS, D) endemic ASV richness per ARMS, E) ASV richness per site and F) endemic ASV richness per site based on 16S rRNA metabarcoding.

Comparison of diversity indices across the five sampling regions also indicated that Raja Ampat had the highest diversity, and Teluk Cenderawasih the lowest, but these differences were not significant based on the Chao1 (One-way ANOVA *p-value* = 0.104) and Shannon indices, (One-way ANOVA *p-value* = 0.504; Figure 2-6A). However, when separated into size fractions, ANOVA revealed significant differences in diversity within the 500 µm fraction for both Chao1 (One-way ANOVA *p-value* = 0.0002) and Shannon (One-way ANOVA *p-value* = 0.0395) (Figure 2-6B, Supplemental Table S2-5). Tukey tests indicate that Teluk Cenderawasih is significantly different with Aceh and Pulau Seribu in 100 um fraction and with Raja Ampat in both 100 um and 500 um fraction (Supplemental Table S2-6).

Microbial community beta diversity was significantly different among regions (PERMANOVA, *p-value* = 0.0001), based on the abundance of reads (Bray-Curtis) and ASVs presence-absence (Jaccard) (Supplemental Table S2-8A). While there was a similar trend in beta dispersion (betadisper), it was non-significant (Betadisper, *p-value* = 0.0782 for Jaccard; *p-value* = 0.443 for Bray-Curtis), indicating regions have same variance (Supplemental Table S2-8A). Across size fractions, PERMANOVA again showed a significant differences in beta diversity among regions but dispersion analyses were non-significant for all size fractions (Supplemental Table S2-8A).

Linear regression showed that metazoan community diversity did not predict microbial diversity across all five regions, for both COI and 18S rRNA. However, after excluding Teluk Cenderawasih, an apparent outlier, COI ASV diversity was a significant predictor of microbial community diversity across the remaining 47 ARMS units (R^2 = 0.22, *p-value* = 0.0008), and endemic COI ASV diversity was a significant predictor of endemic microbial diversity (R^2 = 0.22, *p-value* = 0.0004; Figure 2-7). Comparing all 18S rRNA ASV diversity to 16S rRNA

diversity was also significant both for total diversity (R^2 = 0.155, *p-value* = 0.0076) and endemic diversity (R^2 = 0.134, *p-value* = 0.011), and for total metazoan diversity (R^2 = 0.148, *p-value* = 0.006) and endemic metazoan diversity (R^2 = 0.144, *p-value* = 0.008; Figure 2-7).

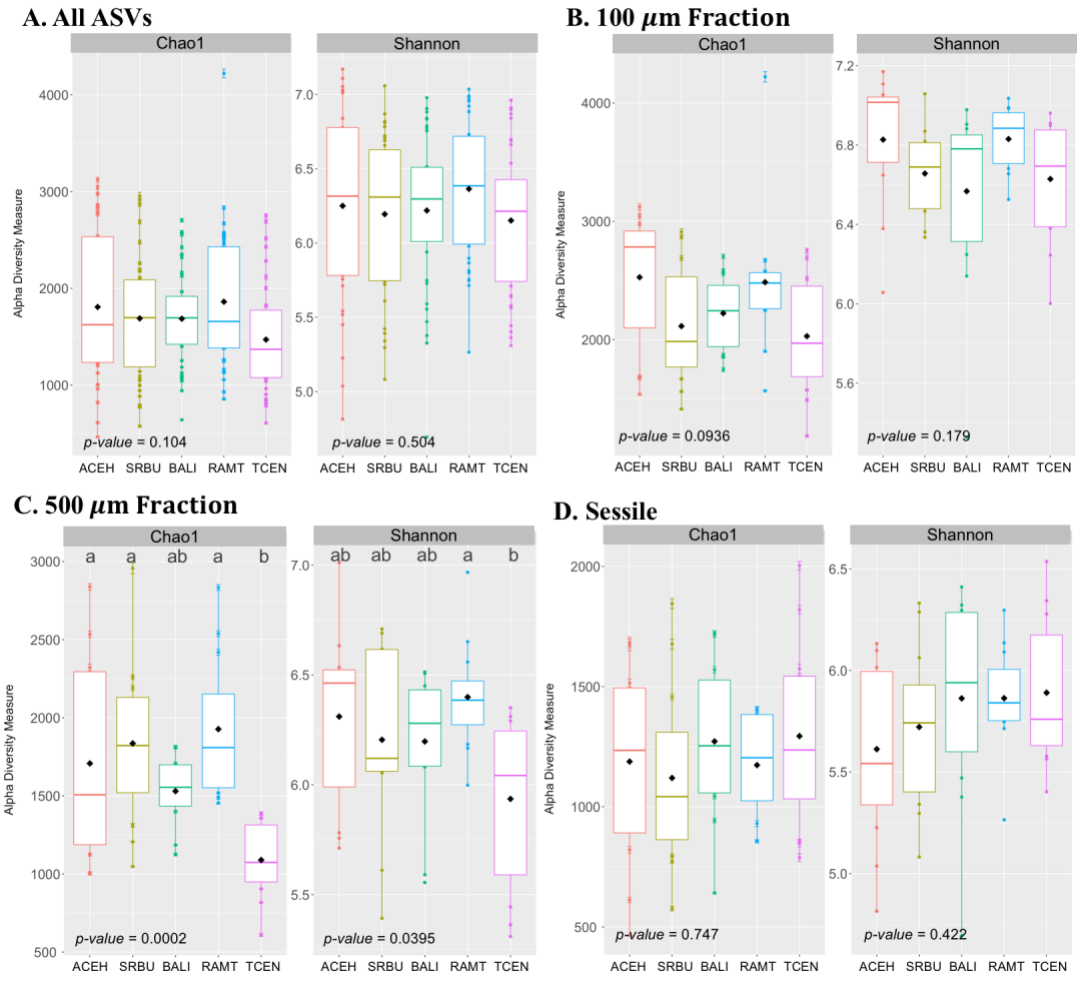


Figure 2-6. Boxplot showing the diversity indices (Chao1 and Shannon) of microbial community composition across Indonesia. Analyses include A) the full dataset, and the different fractions, B) 100 µm–500 µm, C) 500 µm-2 mm, and D) sessile fraction. Black diamonds represent mean alpha diversity from each location. Colored boxes represent the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The letters at the top of the box are the results of Tukey test of multiple comparisons.

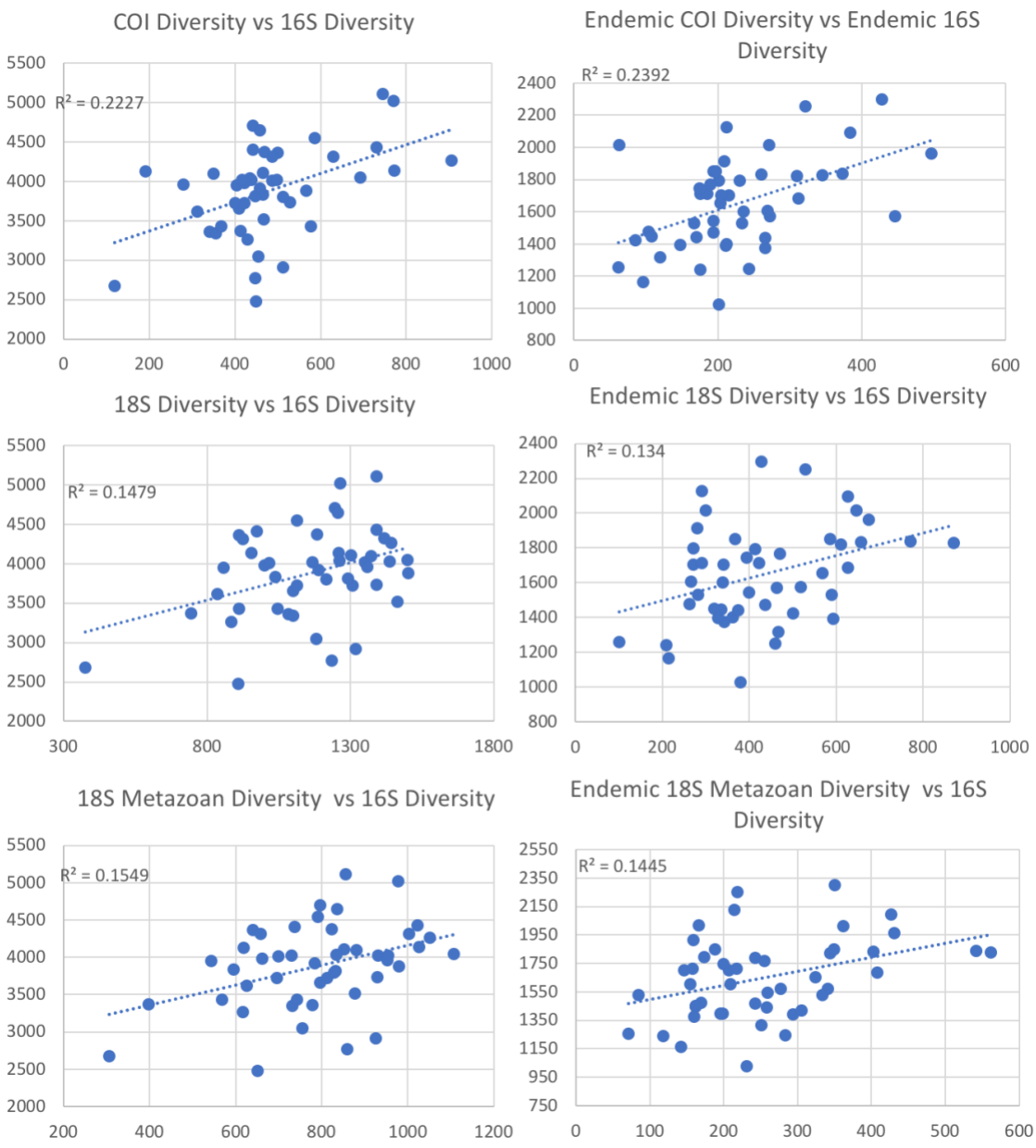


Figure 2-7. Plot of microbial diversity as a function of eukaryote/metazoan diversity based on A) total COI ASVs and microbial ASVs, B) endemic COI ASVs and endemic microbial ASVs, D) total 18S rRNA ASVs and microbial ASVs, D) endemic 18S rRNA ASVs and endemic microbial ASVs, E) metazoan 18S rRNA ASVs and microbial ASVs, D) endemic metazoan 18S rRNA ASVs and endemic microbial ASVs, including best fit line and R^2 values, all of which are significant.

Regional Differentiation

Ordination plots of microbial communities show significant regional differentiation using both Jaccard (*p-value* = 0.001) and Bray-Curtis (*p-value* = 0.001) distances (Figure 2-8). This pattern held true across all size fractions, although the100 µm fraction using Jaccard index displayed the most prominent pattern of geographical clustering of microbial diversity (Figure 2-9).

Mantel tests show a significant pattern of isolation by distance (Figure 2-10, Table 2-3). The correlation was strongest in the 100 µm fragment (r=0.6364, R^2=0.4049), with a decreasing percentage of community structure explained by geographic distance in the 500 µm (r=0.4916, R^2=0.2416) and sessile (r=0.3561, R^2=0.1268) ARMS fragments respectively. All tests had significant P-values (<0.0001).

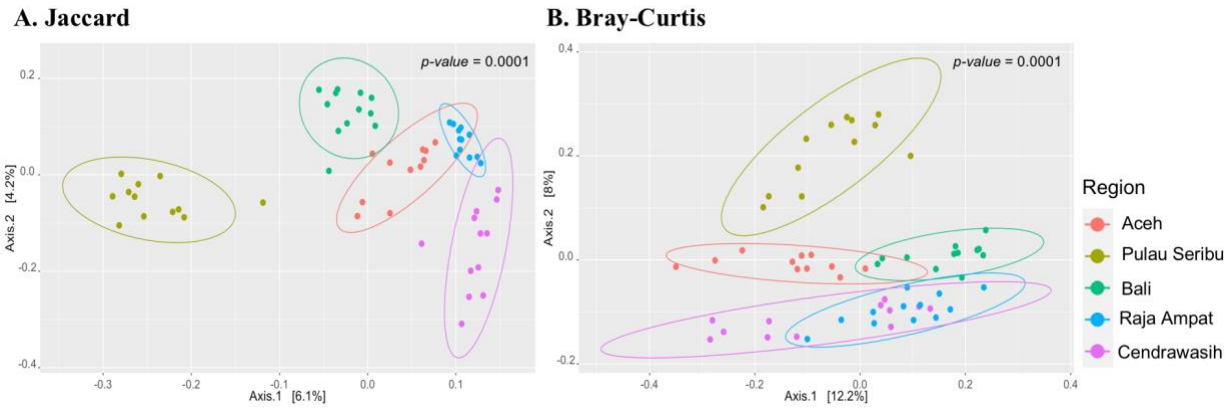


Figure 2-8. Principal Coordinates Analysies (PCoA) illustrating dissimilarities in microbial community composition across Indonesia caculated on total microbial diversity for individual ARMS unit (e.g. summed across all three size fractions). PCoA was performed using Bray-Curtis and Jaccard similarity on the A) full dataset, and B) the dataset without shared amplified sequence variants (ASVs) across different locations.

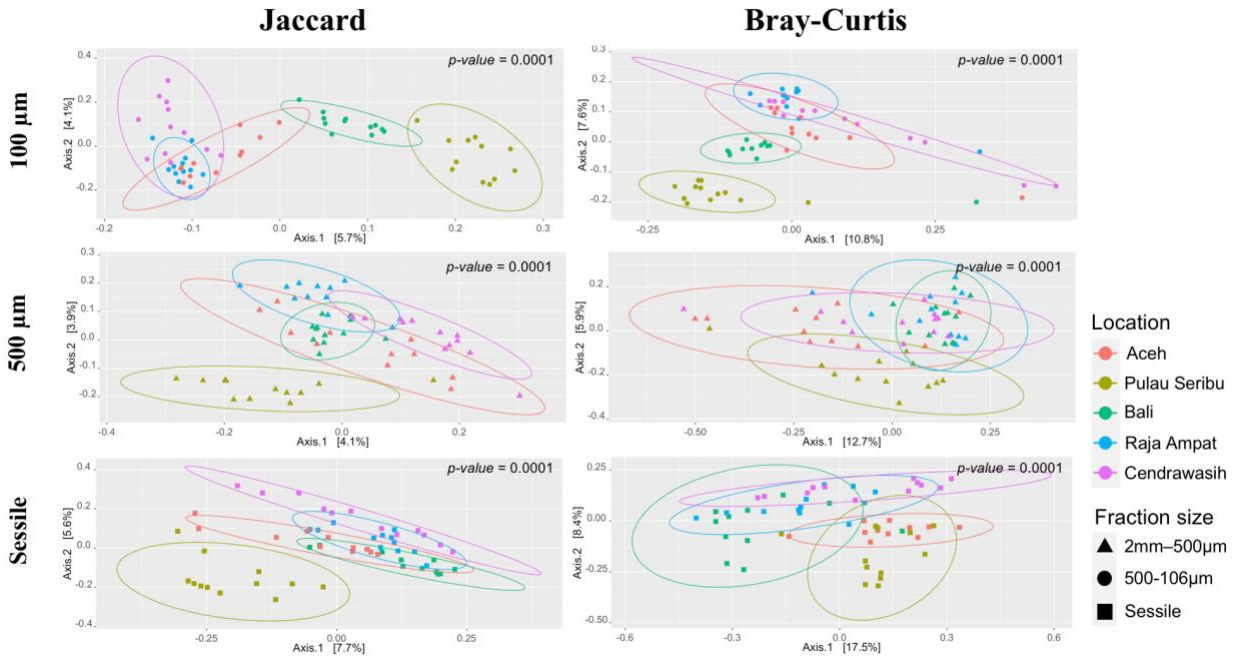


Figure 2-9. Principal Coordinates Analysis (PCoA) analysis illustrating dissimilarities in microbial community composition across Indonesia. Shared amplified sequence variants (ASVs) between size fraction were excluded from this dataset and were rarefied even depth to 1,038 reads per samples. Analyses using Bray-Curtis and Jaccard similarity were undertaken on the different fractions (106–500 µm, 500 µm -2mm, and sessile) across the five sampling regions.

Table 2-3. Isolation by distance correlation (Mantel's *r* and correlation coefficient) between matrices of geographic distance and Jaccard dissimilarity distance among ARMS size fractions.

ARMS fragment	Mantel's *r*	R^2	P-value
106-500 µm	0.6364	0.4049	<0.0001
500 µm -2mm	0.4916	0.2416	<0.0001
Sessile	0.3561	0.1268	<0.0001

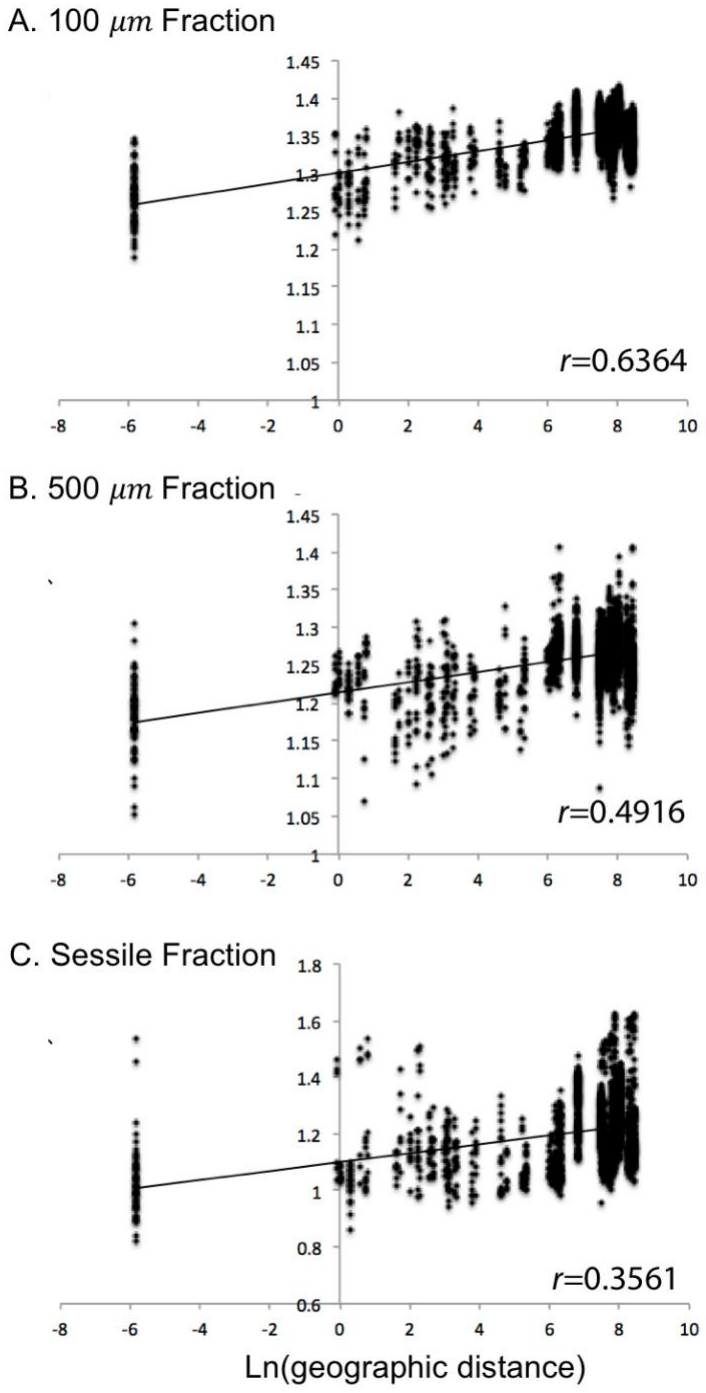


Figure 2-10. Isolation by distance correlation (Mantel's *r*) between matrices of the natural log of geographic distance and Jaccard community dissimilarity distance among ARMS size fractions: A) 100 µm, B) 500 µm, and C) sessile.

Shared ASVs Across Size Fractions

Examining the distribution of ASVs across the three size fractions showed that 18% (7,082) ASVs were present in all three, 52% were unique to individual size fractions, while the remaining 30% were shared among two size fractions. The 100 µm fraction had the highest proportion of unique ASVs (22.6%) followed by the 500 µm fraction with 19.9% unique ASVs; the sessile fraction had 9.43% unique ASVs (Figure 2-11).

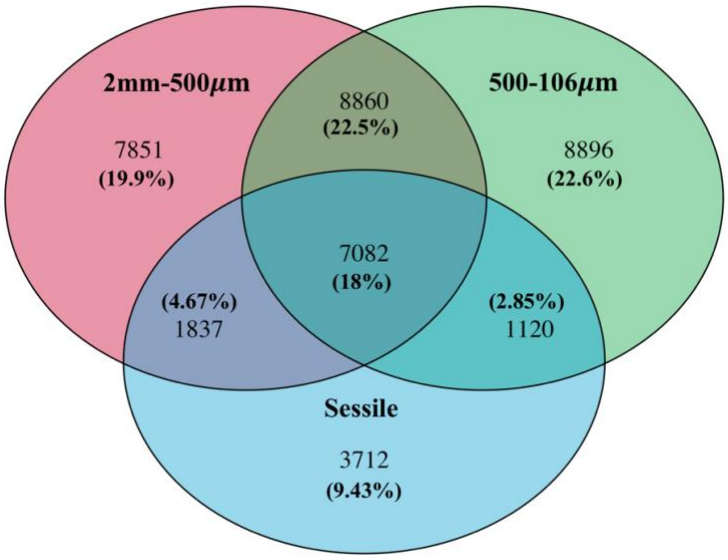


Figure 2-11. Number and distribution of microbial ASVs revealed from16S rRNA metabarcoding of 100 µm, 500 µm, and sessile size fractions obtained from autonomous reef monitoring structures from across the Indonesian archipelago. Plot represents data rarefied to an even depth of 36,719 reads per ARMS unit.

To account for the possibility that the ARMS processing pipelines could lead to carryover among the three separate size fractions, we repeated the analyses, above, excluding any ASVs shared among the size fractions . In total, 5,404,926 sequence reads (23.3% of total) representing

7,082 microbial ASVs (13.7% of total ASVs) were shared among the 100 µm, 500 µm, and sessile fraction. These ASVs included 48 Phylum and 114 Classes, and was dominated by *Proteobacteria*, 37% ASVs, and 51% of sequence reads. Other common taxa include *Chloroflexi* (10.2%), *Acidobacteria* (7.2%), *Cyanobacteria* (6.5%), *Bacteroidetes* (6.3%), and *Planctomycetes* (4.6%). Microbial community composition in the three size fractions excluding ASVs shared across fractions were highly similar to patterns from the full dataset (Supplemental Figure S2-1).

As with the full dataset, alpha diversity was highest in Raja Ampat and lowest in Teluk Cenderawasih (Supplemental Figure S2-2), but ANOVA revealed no significant differences (ANOVA, *p-value* = 0.105 for Chao1*; p-value* = 0.794 for Shannon) across sampling locations (Supplemental Table S2-5B). When separated by size fraction, 500 fraction showed a significantly different in richness (ANOVA, *p-value* = 0.0000208 for Chao1*; p-value* = 0.0299 for Shannon) as well as 100 fraction with Chao1 (ANOVA, *p-value* = 0.0388) (Supplemental Table S2-5B; S2-7).

Beta diversity analyses excluding shared ASVs returned equivalent results to the full dataset (Supplemental Table S2-8A), except the 500 µm - 2 mm fraction was significant different among locations (Supplemental Table S2-8B) (Supplemental Figure S2-3). Similarly, across size fractions, the PERMANOVA p-value for both datasets showed a significant result based on region (Supplemental Table S2-8A, S2-8B). The dispersion analysis showed that non-significant results for all 100 µm, 500 µm, and sessile sample for the Bray-Curtis and Jaccard index, except for a sessile sample (Jaccard), that showed a non-significant p-value between site (Supplemental Table S2-8B).

Excluding shared ASVs resulted in more prominent differentiation among regions in the ordination plots for both Jaccard and Bray-Curtis indices. In contrast to the full dataset where only the 100 µm fraction was strongly differentiated (Figure 2-9), clustering by region was seen in every size fraction when excluding shared ASVs (Supplemental Figure S2-3).

Discussion

The Indonesian archipelago is known for strong biodiversity gradients (Roberts et al., 2002; Allen, 2008; Bellwood & Meyer, 2009; Veron et al., 2009) and phylogeographic structure (Lourie et al., 2005; Barber et al., 2006; Barber et al., 2011; Carpenter et al., 2011; DeBoer et al 2014; Jackson et al., 2014) in large metazoans like fish, gastropods and crustacea. Results from16S rRNA metabarcoding of samples from autonomous reef monitoring structures reveal substantial regional differentiation in microbial communities across the Indonesian archipelago. Consistent with larger metazoans, diversity was consistently highest in Raja Ampat, with lower diversity in other regions, whether examining total diversity or individual size fractions, a pattern that held examining only endemic microbial diversity.

Kelly et al. (2014) found that benthic community structure shaped microbial communities. Here, we show a correlation between eukaryote/metazoan diversity and microbial diversity across the Indonesian archipelago, a pattern that held for endemic eukaryote/metazoan diversity and endemic microbial diversity. While other studies highlight associations between microbes and marine metazoans (Nishiguchi et al., 1998; Ashen & Goff, 2000; Webster et al., 2001; Wegley et al., 2007; Barott & Rohwer, 2012; Al-Rshaidat et al., 2016; Pearman et al., 2016), the results of this study suggest that these associations could influence patterns of microbial diversity. Indeed, bacterial ASV diversity in Indonesia was nearly twice that of diversity from 56 ARMS deployed in the Red Sea (Pearman et al., 2019) a region with much less

marine metazoan diversity. Combined, these results, suggesting that the Coral Triangle may be a global hotspot for benthic marine microbial biodiversity as well as marine metazoan diversity.

While the results above demonstrate a significant relationship between eukaryote/metazoan diversity and microbial diversity, there is also a highly significant pattern of isolation by distance. Isolation by distance indicates that limits to dispersal are also influencing geographic differentiation (Slatkin, 1993) of Indonesian microbial communities, a pattern rarely observed in microbes (Whitaker et al., 2003; Whitaker et al., 2005; Whitaker, 2006). While spurious patterns of isolation by distance could result from sampling broadly across latitude where geographic distance and environmental variation (e.g. sea surface temperature) are highly correlated, the hierarchical sampling design of our study over a small range of tropical latitudes makes such a correlation highly unlikely. As such, microbial patterns in this study are unlikely the result of environmental selection, and cannot be explained by the hypothesis that "everything is everywhere, but the environment selects" (Becking's. 1934; Finlay, 2002). Instead, microbial communities in the Coral Triangle are likely differentiated by associations with larger eukaryotes and by physical limits to dispersal as seen in other marine organisms (reviewed in Barber et al., 2011; Carpenter et al., 2011).

Fine-scale partitioning of microbial communities on coral reefs

Benthic marine organisms host incredibly diverse microbial communities that are often species-specific (Rohwer et al., 2002). Among the most extensively studied microbial communities in tropical marine environments are those associated with corals, where individual species can host 300-400 microbial taxa, many not related to any others known to science (Rohwer et al., 2002; Barott et al., 2011). Benthic algae also host diverse bacterial communities, and can be species-specific, with up to 3,500 different bacterial species on a single algal species

(Barott et al., 2011). While diversity of individual ARMS was relatively high, averaging approximately 3,000-4,000 bacterial ASVs, this number seems relatively modest given the diversity of eukaryote communities on ARMS. Averaged across all ARMS, richness of microbial ASVs was only 8 times higher (range 3-22) than richness of eukaryote ASVs based on COI metabarcoding. This number seems surprisingly low, particularly given that *Porifera* was a substantial portion of the sessile ARMS community, and are known to harbor very diverse microbial communities (Moitinho-Silva et al., 2017).

The metabarcoding approach used in this study does not allow us to examine microbial communities associated with individual taxa. However, each size fraction had distinctly different eukaryotic communities allowing us to investigate how the composition of marine microbial communities vary with eukaryote community composition, broadly. For example, the eukaryotic 100 µm and 500 µm fractions were dominated by *Arthropoda* and *Annelida* while the sessile fraction was dominated by B*ryozoa, Rhodophyta,* and *Porifera* (Cahyani et al. Chapter 1). Not surprisingly, the 100 µm and 500 µm motile fractions had the most similar microbial communities. It was dominated by *Proteobacteria*, one of the most common bacteria phyla observed in the marine system (Rusch et al., 2009; Kelly et al., 2014; Ziegler et al., 2016; Bakenhus et al., 2017; Hernandez-Agreda et al., 2018; Pearman et al., 2019). Classes *Gammaproteobacteria* and *Alphaproteobacteria* were also common in the 100 µm and 500 µm motile fractions, similar to findings from other studies of coral reef microbial diversity (Rusch et al., 2009; Kelly et al., 2014; Ziegler et al., 2016; Bakenhus et al., 2017; Hernandez-Agreda et al., 2018; Pearman et al., 2019). *Bacteroidetes*, was also common (~6% of the sequence reads) in the 100 µm and 500 µm motile fraction, an important bacterioplankton known for degrading particulate matter, especially proteins, in marine ecosystem (Fernández-Gómez et al., 2013).

Microbial communities from the sessile fraction were distinctly different. While still dominated by *Proteobacteria* (43%), the sessile fraction had much higher proportions of *Chloroflexi* (20%), *Acidobacteria* (10.5%), *Actinobacteria* (4%) and *PAUC34f* (2.6%). he most likely explanation for the differences in microbial communities among the motile and sessile size fractions is the strong differences in the eukaryote/metazoan communities within these fractions. The high abundance of *Chloroflexi* is particularly notable as Pearman et al. (2019) only reported 5% abundance on ARMS from the Red Sea.

While community abundance at the Phylum level provides a coarse view of differentiation of microbial communities across size fractions, Venn diagrams provide detail at the level of individual ASVs. Even though the process of size-fractioning organismal communities living on ARMS could result in DNA or small pieces of larger organisms being captured in smaller size fractions (Leray & Knowlton, 2015; Wangensteen & Turon, 2016), nearly 52% of all microbial ASVs were found in only one size fraction, indicating highly differentiated communities. Moreover, the PCoA plots that incorporate information about genetic distance among ASV sequences also show strong differentiation among the three size fractions. Combined, the above indicates finely partitioned microbial communities associated with eukaryotic communities living within the matrix of coral reefs.

Biogeography of Microbes in the Coral Triangle

The Coral Triangle biodiversity hotspot is one of the best-known patterns in marine biogeography, and is seen in fish (Allen, 2008), corals (Veron et al., 2009), and other invertebrates (Roberts et al., 2002; Bellwood & Meyer, 2009). Data from COI and 18S rRNA metabarcoding (Cahyani et al. Chapter 1) show that this biodiversity pattern holds over a wide diversity of metazoan/eukaryotic taxa, with ARMS from Eastern Indonesia (Raja Ampat and

Teluk Cenderawasih) having more diversity within each ARMS unit, each reef, and each region than reefs on the border (Bali) or outside of the Coral Triangle (Pulau Seribu, Aceh). Surprisingly, microbial communities largely followed this same pattern; Raja Ampat has the highest richness and evenness of microbial diversity compared to reef ecosystems to the west, and all microbial communities were highly differentiated based on PCoA results.

Typically studies showing biogeography of marine microbes do so on the scale of ocean basins, with differences largely driven by environmental factors. For example, Pearman et al. (2019) showed strong geographical differentiation among ARMS deployed across 16 degrees latitude in the Red Sea, variation interpreted to result from environmental variation. Our deployments spanned 13 degrees latitude, but also spanned 8 degrees south and 5 degrees north of the equator, representing minimal temperature variation. While environmental variation almost certainly contributes to shaping microbial communities across the Indonesian archipelago, metazoan diversity inferred from COI explained more than 24% of the variance in microbial communities from Aceh to Raja Ampat. Given that eukaryote/metazoan diversity was predictive of microbial diversity suggests that eukaryote/metazoan diversity is playing a major role in shaping microbial communities.

Unlike metazoans/eukaryotes, Teluk Cenderawasih had the lowest microbial diversity at all scales—ARMS, site, or region. It is unclear how to interpret this puzzling result. In many comparisons of eukaryote/metazoan diversity with COI and 18S rRNA, Teluk Cenderawasih was the most diverse (Cahyani et al. Chapter 1). Given that benthic community composition influences microbial communities (Kelly et al., 2014) and the tight associations between microbes and many marine eukaryotes (Nishiguchi et al., 1998; Ashen & Goff, 2000; Webster et al., 2001; Wegley et al., 2007; Coryell et al., 2018), higher eukaryotic diversity should lead to

higher microbial diversity. Indeed, across 47 ARMS units over the four other regions, higher eukaryotic diversity is a highly significant predictor of higher microbial diversity, both for total diversity and for endemic diversity. Given that COI, 18S rRNA and 16S rRNA amplifications used the exact same DNA extractions, low microbial diversity in Teluk Cenderawasih cannot be explained away as degraded samples. Sequencing depths were as high or higher in Teluk Cenderawasih than other populations, and was sufficient to achieve saturation of ASV discovery, so this result isn't a sequencing artifact. Moreover, the pattern is seen in diversity patterns at all scales—per ARMS and per reef—suggesting that this pattern is real. Thus, while results broadly support the conclusion that high eukaryote/metazoan diversity leads to more microbial diversity, likely through species-species interactions in holobionts (Wegley et al., 2007; Barott & Rohwer, 2012), further work is needed to understand the origins of this perplexing anomaly.

Everything is not Everywhere

Ordination plots showed highly differentiated microbial communities on ARMS, patterns that are particularly strong examining the different size fractions. Such variation in microbial community composition has been reported across latitude, depth, or different habitat (Pommier et al., 2007; Fuhrman & Steele, 2008; Brown et al., 2012; Ghiglione et al., 2012; Kelly et al., 2014). However, the strong pattern of isolation by distance seen in our data set indicates significant barriers to dispersal (e.g. Hillebrand et al., 2001; Whitaker et al., 2005) in bacteria associated with Indonesian coral reefs, and thus do not conform to the axiom that "everything is everywhere". Incredibly, Mantel's *r* for the combined ARMS fragments for bacteria in The Coral Triangle (r=0.4947) is greater than reports in corals (r=0.396; Hillebrand et al., 2001). Furthermore, the association between community structure and distance appears greater than that

of benthic diatoms (r=0.591) and ciliates (r=0.256) but not polychaetes (r=0.925) where the correlation is much stronger.

A potential caveat to the IBD approach is that environmental variables may also impose population and/or community structure (i.e. "but the environment selects") and these environmental variables may themselves corelate with geographic distance. For example, a study examining putatively cosmopolitan diatoms yielded a significant correlation coefficient of (R^2=0.297) between genetic distance and Euclidian distance of environmental variables such as temperature, salinity and chlorophyll, but lacked significant correlation with geographic distance (Whittaker & Rynearson 2017). Similarly, a study using ARMS on coral reefs in The Red Sea reported that 67 % of bacterial community structure was determined by variable environmental selection across sites (Pearman et al., 2019). Our study, however, lacks major shifts in similar environmental variables and further controls for habitat standardization through the ARMS deployment method. In a stark contrast to the results we present here, Pearman et al. (2019) claim that only 3% of bacterial community structure in Red Sea coral reef ARMS was determined by dispersal limitation. Another important utility of our IBD results is that the significant correlations ensure that the patterns of community structure across locations cannot be attributed to stochastic assemblages on the ARMS units and/or under-sequencing, as there is no reason why these phenomena would correlate with physical distance.

Considering Finlay's (2002) model predicting a size threshold for ubiquitous dispersal beginning at 1mm, and the broader holobiont's potential to impose dispersal barriers on otherwise free-living bacteria, we predicted an "everything is everywhere" signal in the smaller 100 µm fraction (i.e. no correlation between geographic and community distance) and a positive geographic-community distance correlation in the 500 µm fraction. We did not, however,

observe this pattern. In fact, both fractions had positive correlations, with the smaller fraction yielding a stronger correlation. The sessile ARMS fraction also yielded a positive correlation, albeit to a lesser degree than either of the other fractions

The robust IBD correlation in the 100 µm fraction might suggest that the true size threshold for ubiquitous dispersal is below 100 µm. For example, a study of *Testate amoebae* species between the Arctic and Antarctic polls found cosmopolitan species below 100 µm while also finding endemic, range restricted species at a size of 230 µm (Wilkinson, 2001). The decrease in Mantel's *r* from the 100 µm fraction to the 500 µm fraction may suggest that the larger fraction is capturing the holobiont of organisms with greater dispersal capacity, such as dinoflagellates with resting spore formation that assists with movement across greater distances (Whittaker & Rynearson 2017). The sessile fraction's smaller still correlation may result from the fraction being dominated by organisms like sponges and bryozoan that filter large amounts of seawater, thereby introducing a greater number of transient bacteria from the seawater versus the other two fractions, which may contain a greater proportion of holobiont symbiotic bacteria.

Caveats

Seawater contains an abundance of microbial life (Venter et al., 2004; Fuhrman et al., 2015) and many of the microbial taxa we detected on ARMS are common in marine ecosystems from seawater to coral reefs (Rusch et al., 2009; Kelly et al., 2014; Pearman et al., 2019). Given that ARMS processing involves rinsing with seawater, and the fact that 18% of all microbial taxa were associated with all three fractions raises the possibility that microbial communities in seawater might affect the patterns of microbial community composition observed in this study. We tested this hypothesis by excluding microbial taxa shared across size fractions and found that exclusion of shared taxa did not affect patterns of microbial community richness or composition

across the locations. Exclusion of taxa shared among fractions, however, did result in stronger patterns of spatial structuring of microbial communities across Indonesia, both considering ARMS as a single unit, and as individual size fraction. Microbial communities from sample sites in Aceh and Pulau Seribu were always distinct, while the other three clustered together.

It is impossible to tell if the microbial taxa seen in all size fractions represent a community of generalist bacteria with broad associations with benthic eukaryotes, or whether these taxa represent contamination from microbial communities in sea water. This hypothesis could be tested by collecting and sequencing microbial communities from seawater samples from the water column at ARMS deployment sites, and comparing these communities to microbial communities on ARMS. However, the fact that we see such clear patterns of regional structure in benthic microbial communities when microbes shared among fractions are included, and these pattern strengthen when they are excluded indicates that freely associated water column microbes may weaken our observed patterns of regional structure in benthic microbial communities across Indonesia, but these microbes do not drive these patterns.

Conclusions

Counter to the "everything is everywhere" hypothesis, and studies of pelagic microbial communities, benthic marine microbial communities on ARMS showed substantial regional differentiation across Indonesia. Consistent with larger metazoans like fish, corals, and cryptic eukaryotic diversity from ARMS, microbial diversity was highest in Raja Ampat in Eastern Indonesia, lower in other regions, and displayed a highly significant pattern of isolation by distance, indicating that limits to dispersal are influencing geographic differentiation of these microbial communities.

While results clearly indicate that these microbial communities vary over space, due both to associations with eukaryotes, particularly metazoans, and by limits to dispersal, it is unclear whether these differences are functional. The composition of these microbial communities are common taxa in marine benthic communities, especially on coral reefs (Rusch et al., 2009; Kelly et al., 2014; Campbell et al., 2015; Pearman et al., 2019) and can be involved in processes such as photosynthesis or stress responses (Thurber et al., 2009). Given the covariance of eukaryote/metazoan diversity and microbial diversity, it is highly probable that members of this microbial community are involved in symbioses. However, studies at finer resolution are needed to begin to answer these questions. In the future, data collected using the ARMS and DNA metabarcoding should be expanded to address these questions of ecologically relevant shifts in marine microbial community structure and function. Given how quickly microbes respond to environmental change, such data could provide an important baseline for resource managers and government agencies to develop policies for the sustainability of Indonesia's remarkable marine biodiversity.

Supplemental Tables and Figures

Supplemental Table S2-1. Location and number of metabarcoding samples used on this study.

Country	Region/Reef	Abbr.	Latitude	Longitude	N
Indonesia	**Pulau Weh, Aceh**	**ACEH**			**33**
	Benteng	BTN	05° 50.774' N	095° 22.434' E	9
	Rubiah Sea Garden	RSG	05° 52.608' N	095° 15.596' E	9
	Seulako	SEU	05° 53.658' N	095° 15.176' E	9
	Sumur Tiga	STG	05° 53.370' N	095° 20.683' E	6
Indonesia	**Pulau Seribu, Jakarta**	**SRBU**			**34**
	Pulau Karang Beras	KBS	05° 45.574' S	106° 33.527' E	8
	Pulau Kotok	KOT	05° 41.575' S	106° 32.475' E	9
	Pulau Pramuka	PRM	05° 45.026' S	106° 36.311' E	8
	Pulau Sepa	SEP	05° 34.227' S	106° 34.491' E	9
Indonesia	**Pemuteran, Bali**	**BALI**			**35**
	Close Encouter	CEN	08° 7.675' S	114° 40.084' E	9
	Deep Middle Reef	DMR	08° 8.190' S	114° 39.570' E	9
	Horse Reef	HOR	08° 7.672' S	114° 39.337' E	9
	Napoleon Reef	NAP	08° 7.928' S	114° 40.531' E	8
Indonesia	**Raja Ampat, West Papua**	**RAMT**			**36**
	Kri	KRI	00° 33.284' S	130° 40.712' E	9
	Misool	PEF	02° 14.741' S	130° 33.438' E	9
	Pef	PEF	00° 26' S	130° 26' E	9
	Penemu	PNU	00° 34.664' S	130° 17.039' E	9
Indonesia	**Teluk Cenderawasih, West Papua**	**TCEN**			**36**
	Angromeos Island	ANG	02° 40.828' S	134° 49.515' E	9
	Manguar Cape	MGC	02° 52.866' S	134° 51.411' E	9
	Purup	PRP	02° 03.419' S	134° 09.585' E	9
	Tridacna Atoll	TRI	02° 29.948’S	134° 58.790’E	9

Supplemental Table S2-2. Total samples, sequence reads, and Amplicon Sequence Variants (ASVs) for the 16S rRNA datasets as revealed by DNA metabarcoding of Autonomous Reef Monitoring Structure (ARMS) from Indonesia.

A. Full dataset	**All samples combined**	**2 mm-500 um**	**500-106 um**	**Sessile**
Total no. of individual samples	176	59	59	58
Total no. of individual ARMS	59	59	59	58
Total no. of individual site	20	20	20	20
Total no. of individual location	5	5	5	5
Total no. of sequence reads	23,205,524	6,463,160	7,781,148	8,961,216
- Minimum no. of reads	36,719	36,719	50,416	52,659
- Maximum no. of reads	364,814	326,593	364,814	308,720
- Mean no. of reads	131,849.6	109,545.1	131,883.9	154,503.7
Total no. of ASVs	43,327	27,153	28,508	15,523
- Mean no. of ASVs	246.2	460.2	483.2	267.6
Rarefied even depth	36,719	36,719	36,719	36,719
Total rarefied no. of sequence reads	6,462,544	2,166,421	2,166,421	2,129,702
Total rarefied no. of ASVs	39,358	25,630	25,958	13,751
- Mean rarefied no. of ASVs	223.6	434.4	440.0	237.1
B. Dataset without shared ASVs	**All samples combined**	**2 mm-500 um**	**500-106 um**	**Sessile**
Total no. of individual samples	176	59	59	58
Total no. of individual ARMS	59	59	59	58
Total no. of individual site	20	20	20	20
Total no. of individual location	5	5	5	5
Total no. of sequence reads	1,057,618	386,844	442,695	228,079
- Minimum no. of reads	918	918	2,690	1,038
- Maximum no. of reads	13,210	13,210	10,389	10,098
- Mean no. of reads	6,009.2	6,556.7	7,503.3	3,932.4
Total no. of ASVs	32,276	18,548	18,876	6,669
- Mean no. of ASVs	183.4	314.4	319.9	115.0
Total no. of individual samples	175.0	59.0	58.0	58.0
Rarefied even depth	1,038	1,038	1,038	1,038
Total rarefied no. of sequence reads	181,650	60,204	61,242	60,204
Total rarefied no. of ASVs	21,015	11,769	11,039	4,573
- Mean rarefied no. of ASVs	120.1	199.5	190.3	78.8

Supplemental Table S2-3. Alpha diversity indices (ANOVA) of microbial diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using 16S rRNA markers, rarefied to a standardized number of Amplicon Sequence Variants (ASVs) per ARMS to account for variation in sequencing depth.

Diversity	**ANOVA**				
	Df	**Sum Sq**	**Mean Sq**	**F Value**	***p-value***
Average local diversity per ARMS	4	0.0005	0.000128	0.017	0.999
Percent regional diversity per site	4	0.01601	0.004002	0.267	0.895
ASVs richness per ARMS	4	7292948	1823237	3.372	**0.0154**
Endemic ASVs per ARMS	4	2176702	544176	4.51	**0.00319**
ASVs per site	4	9976350	2494088	2.198	0.119
Endemic ASVs per site	4	652840	163210	0.657	0.631

Supplemental Table S2-4. Tukey HSD post hoc tests pairwise comparisons on microbial diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using 16S rRNA.

Region	ASVs richness per ARMS			
	diff	**lwr**	**upr**	**p adj**
BALI-ACEH	186.167	-660.493	1032.826	0.97125
RAMT-ACEH	691.167	-155.493	1537.826	0.15968
SRBU-ACEH	84.167	-762.493	930.826	0.99860
TCEN-ACEH	-391.500	-1238.160	455.160	0.68979
RAMT-BALI	505.000	-341.660	1351.660	0.45315
SRBU-BALI	-102.000	-948.660	744.660	0.99705
TCEN-BALI	-577.667	-1424.326	268.993	0.31699
SRBU-RAMT	-607.000	-1453.660	239.660	0.26936
TCEN-RAMT	-1082.667	-1929.326	-236.007	**0.00581**
TCEN-SRBU	-475.667	-1322.326	370.993	0.51349
Region	**Endemic ASVs per ARMS**			
	diff	**lwr**	**upr**	**p adj**
BALI-ACEH	112.333	-287.633	512.300	0.93184
RAMT-ACEH	371.667	-28.300	771.633	0.08041
SRBU-ACEH	111.667	-288.300	511.633	0.93321
TCEN-ACEH	-216.583	-616.550	183.383	0.54963
RAMT-BALI	259.333	-140.633	659.300	0.36796
SRBU-BALI	-0.667	-400.633	399.300	1.00000
TCEN-BALI	-328.917	-728.883	71.050	0.15432
SRBU-RAMT	-260.000	-659.967	139.967	0.36536
TCEN-RAMT	-588.250	-988.217	-188.283	**0.00107**
TCEN-SRBU	-328.250	-728.217	71.717	0.15579

Supplemental Table S2-5. Alpha diversity indices (ANOVA) of microbial diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using 16S rRNA markers. Data were analysed per size fraction (100 μm, 500 μm, and sessile fraction) using a rarefied datasets to account for variation in sequencing depth.

A. Full dataset

Indices	ANOVA							
	All sample		100 μm Fraction		500 μm Fraction		Sessile	
	F-value	*p-value*	F-value	*p-value*	F-value	*p-value*	F-value	*p-value*
Chao1	1.952	0.104	2.099	0.0936	6.472	**0.000249**	0.485	0.747
Shannon	0.835	0.504	1.633	0.179	2.71	**0.0395**	0.988	0.422

B. Dataset without shared ASVs

Indices	ANOVA							
	All sample		100 μm Fraction		500 μm Fraction		Sessile	
	F-value	*p-value*	F-value	*p-value*	F-value	*p-value*	F-value	*p-value*
Chao1	1.949	0.105	2.722	**0.0388**	8.565	**2.08E-05**	1.429	0.237
Shannon	0.42	0.794	2.176	0.084	2.912	**0.0299**	0.571	0.685

Supplemental Table S2-6. Tukey HSD post hoc tests pairwise comparisons on microbial diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using 16S rRNA.

Region	Chao1 - 500 µm Fraction				Shannon - 500 µm Fraction			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj
SRBU-ACEH	128.09	-402.50	658.68	0.95968	-0.11	-0.54	0.32	0.95693
BALI-ACEH	-177.54	-708.13	353.05	0.87821	-0.11	-0.54	0.32	0.94494
RAMT-ACEH	219.56	-311.03	750.15	0.76945	0.09	-0.34	0.52	0.97745
TCEN-ACEH	-618.41	-1149.00	-87.82	**0.01465**	-0.37	-0.80	0.06	0.11533
BALI-SRBU	-305.63	-824.55	213.30	0.46554	-0.01	-0.43	0.41	1.00000
RAMT-SRBU	91.47	-427.46	610.39	0.98727	0.19	-0.23	0.61	0.69058
TCEN-SRBU	-746.50	-1265.43	-227.58	**0.00145**	-0.27	-0.69	0.15	0.38111
RAMT-BALI	397.09	-121.83	916.02	0.21081	0.20	-0.22	0.62	0.65891
TCEN-BALI	-440.88	-959.80	78.05	0.13161	-0.26	-0.68	0.16	0.41060
TCEN-RAMT	-837.97	-1356.90	-319.04	**0.00028**	-0.46	-0.88	-0.04	**0.02400**

Supplemental Table S2-7. Tukey HSD post hoc tests pairwise comparisons on microbial diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using 16S rRNA markers. Post hoc test were generated from rarefied data after excluding Amplicon Sequence Variants (ASVs) shared among size fraction.

Region	Chao1 - 100 µm Fraction				Chao1 - 500 µm Fraction				Shannon - 500 µm Fraction			
	diff	lwr	upr	p adj	diff	lwr	upr	p adj	diff	lwr	upr	p adj
SRBU-ACEH	-52.40	-244.22	139.41	0.93783	33.93	-125.63	193.50	0.97441	-0.03	-0.54	0.47	0.99971
BALI-ACEH	-50.72	-242.53	141.09	0.94447	38.36	-121.20	197.93	0.96016	-0.01	-0.52	0.49	1.00000
RAMT-ACEH	81.88	-109.93	273.69	0.74866	165.50	5.93	325.07	**0.03849**	0.24	-0.27	0.74	0.67954
TCEN-ACEH	-128.92	-320.73	62.89	0.33145	-161.68	-324.68	1.32	**0.05288**	-0.37	-0.89	0.15	0.26772
BALI-SRBU	1.68	-185.91	189.28	1.00000	4.43	-151.63	160.49	0.99999	0.02	-0.47	0.52	0.99994
RAMT-SRBU	134.29	-53.31	321.88	0.27048	131.57	-24.49	287.63	0.13662	0.27	-0.22	0.76	0.53982
TCEN-SRBU	-76.51	-264.11	111.08	0.77865	-195.61	-355.18	-36.04	**0.00906**	-0.34	-0.84	0.17	0.33940
RAMT-BALI	132.60	-54.99	320.20	0.28239	127.14	-28.92	283.20	0.16083	0.25	-0.25	0.74	0.62083
TCEN-BALI	-78.20	-265.79	109.40	0.76466	-200.04	-359.61	-40.47	**0.00722**	-0.36	-0.86	0.15	0.27678
TCEN-RAMT	-210.80	-398.40	-23.21	**0.02027**	-327.18	-486.75	-167.61	**0.00000**	-0.61	-1.11	-0.10	**0.01109**

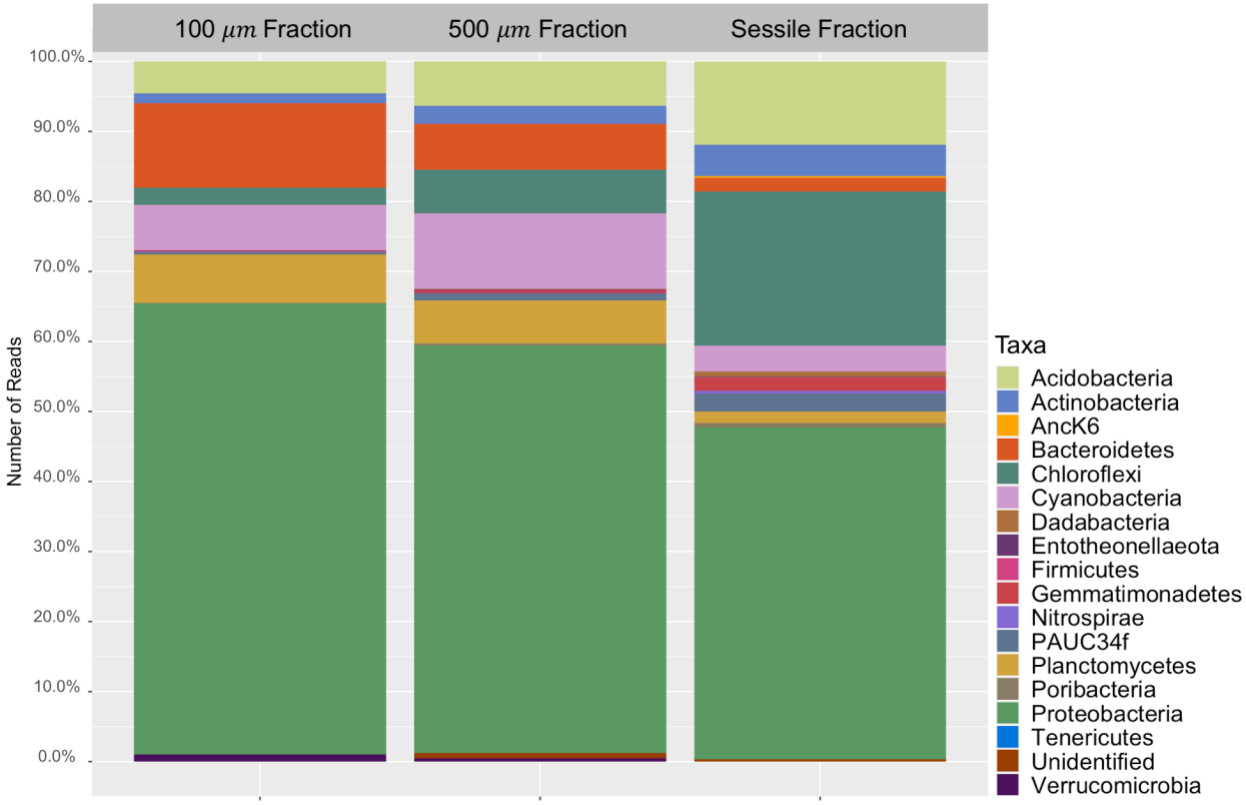
Supplemental Table S2-8. Beta diversity summary (PERMANOVA) of microbial diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Cenderawasih Bay) in Indonesia using 16S rRNA.

A. Full dataset

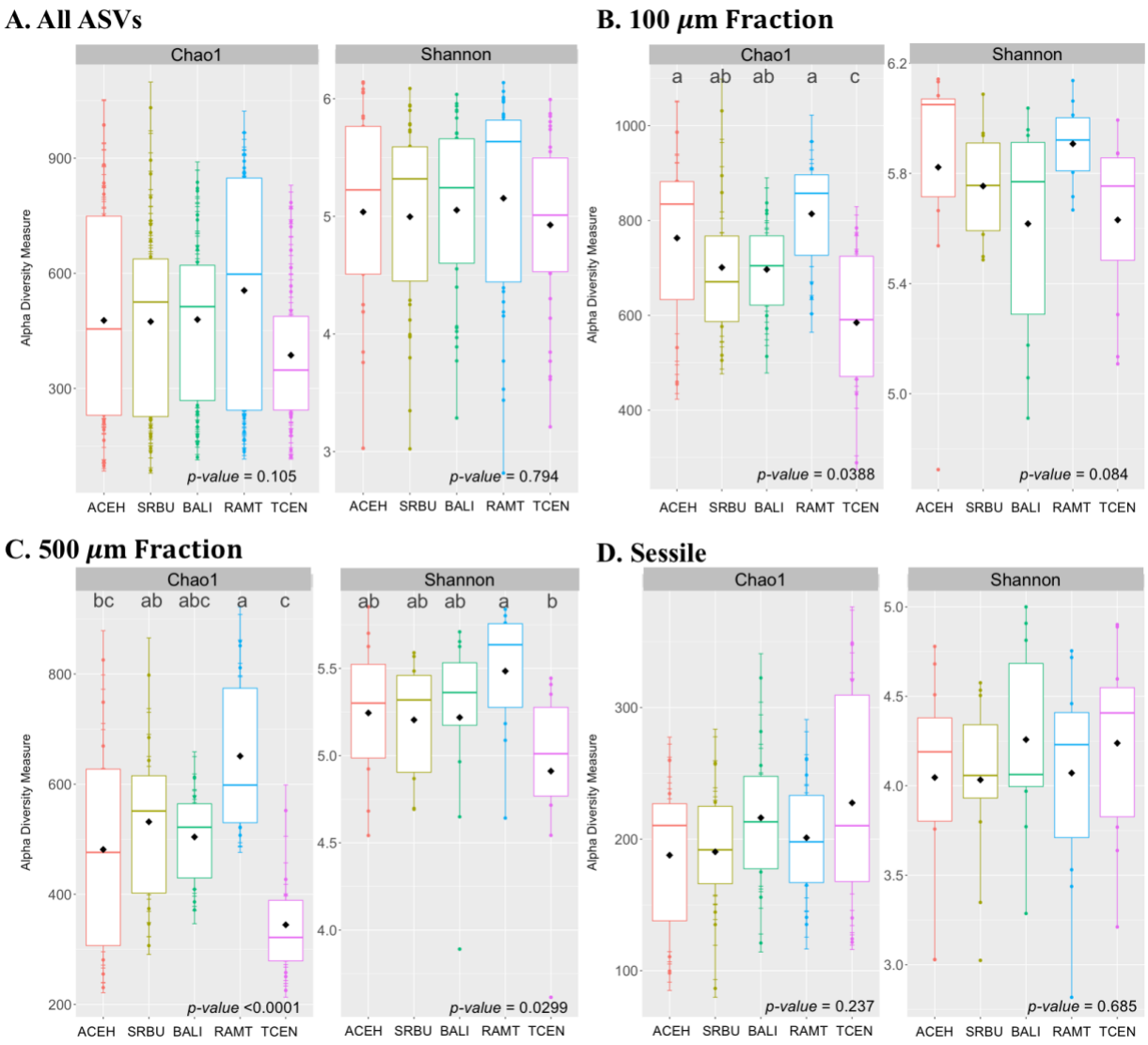
Indices	Adonis		Betadisper		100 µm Fraction				500 µm Fraction				Sessile			
					Adonis		Betadisper		Adonis		Betadisper		Adonis		Betadisper	
	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*
Jaccard	0.07049	**1.00E-04**	1.5511	0.1884	0.15602	**1.00E-04**	1.5725	0.1919	0.12233	**1.00E-04**	2.2135	0.0782	0.1664	**1.00E-04**	2.2001	0.0669
Bray-Curtis	0.09758	**1.00E-04**	0.6176	0.6514	0.21772	**1.00E-04**	0.2354	0.9243	0.17478	**1.00E-04**	0.9553	0.443	0.23634	**1.00E-04**	2.032	0.0975

B. Dataset without shared ASVs

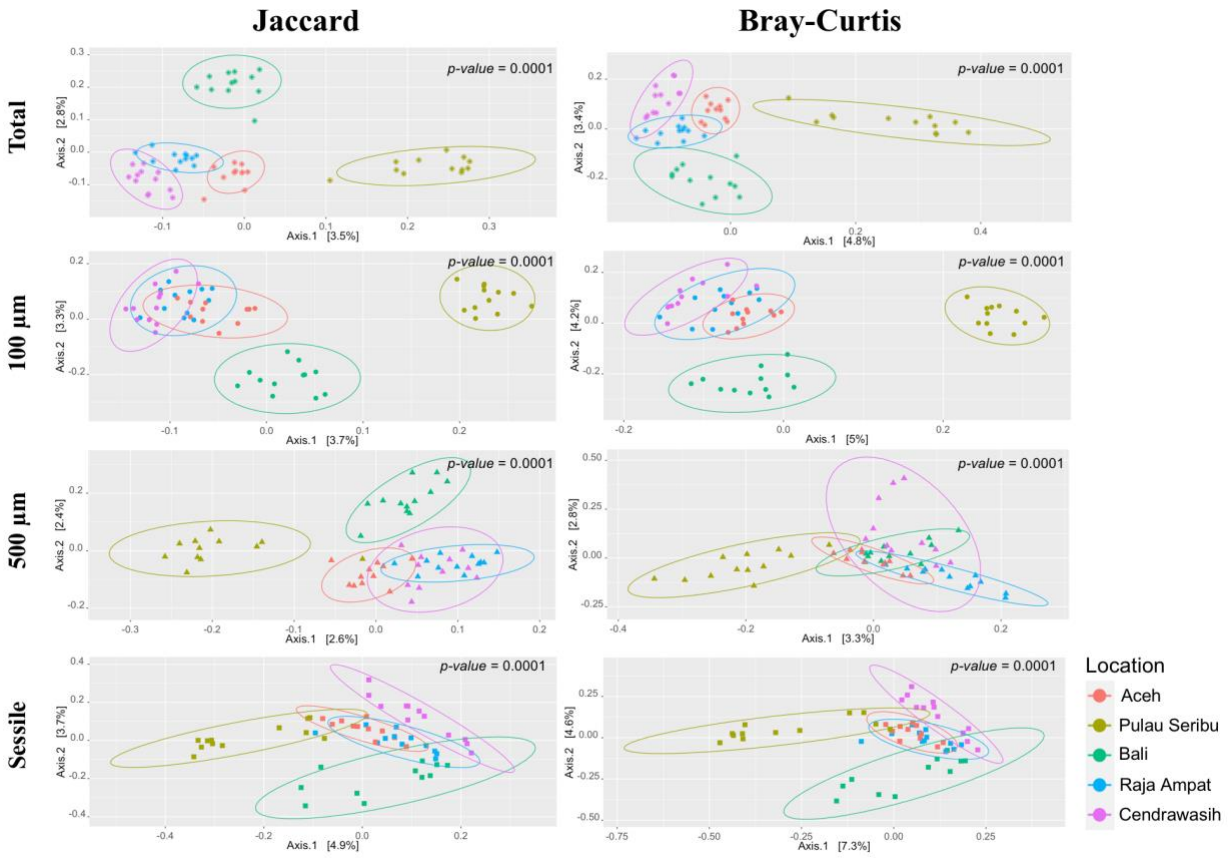
Indices	Adonis		Betadisper		100 µm Fraction				500 µm Fraction				Sessile			
					Adonis		Betadisper		Adonis		Betadisper		Adonis		Betadisper	
	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*
Jaccard	0.04241	**1.00E-04**	0.3468	0.8475	0.11806	**1.00E-04**	1.9214	0.1159	0.09145	**1.00E-04**	0.2891	0.8858	0.12757	**1.00E-04**	2.9449	**0.0268**
Bray-Curtis	0.05428	**1.00E-04**	0.1031	0.9812	0.14783	**1.00E-04**	1.4546	0.2276	0.10718	**1.00E-04**	0.5134	0.7314	0.15932	**1.00E-04**	1.6953	0.1633



Supplemental Figure S2-1. Taxonomic composition of 7,082 shared microbial amplified sequence variants (ASVs) between 500 µm - 2 mm, 106-500 µm, and sessile size fraction. Bar plots showing taxa relative abundance of different size fraction from the sample across five different location in Indonesia. The bar plots were constructed based on phyla contribute more than 2% of the relative abundance of each sample.



Supplemental Figure S2-2. Boxplots showing the microbial diversity indices (Chao1 and Shannon) across Indonesia. Analysis was undertaken on A) full dataset after removing the ASVs shared among the different fractions B) 106-500 µm, C), 500 µm - 2 mm, and D) Sessile fraction. The diamonds represent mean alpha diversity from each location, the box represents the 1st and 3rd quartiles and the vertical line is the median of the dataset. The letters at the top of the box are the results of Tukey test of multiple comparisons.



Supplemental Figure S2-3. Principal Coordinates Analysis (PCoA) analysis illustrating dissimilarities in microbial community composition across Indonesia. Shared amplified sequence variants (ASVs) between size fraction were excluded from this dataset and were rarefied even depth to 1,038 reads per samples. Analysis was undertaken using Bray-Curtis and Jaccard similarity on the different fractions (106-500 µm, 500 µm-2 mm, and Sessile) across all sampling locations.

References

Andersen, K. S., Kirkegaard, R. H., Karst, S. M.& Albertsen, M (2018). "ampvis2: an R package to analyse and visualise 16S rRNA amplicon data." *bioRxiv*. http://dx.plos.org/10.1371/journal.pone.0132783

Antony-Babu, S., Stach, J. E. M. & Goodfellow, M. (2008) Genetic and phenotypic evidence for *Streptomyces griseus* ecovars isolated from a beach and dune sand system. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* 94:63-74.

Al-Rshaidat, M. M. D., Snider, A., Rosebraugh, S., Devine, A. M., Devine, T. D., Plaisance, L., Knowlton, N. & Leray, M. (2016). Deep COI sequencing of standardized benthic samples unveils overlooked diversity of Jordanian coral reefs in the northern Red Sea. *Genome*, 59, 724–737.

Allen, G. R. (2008). Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes, *556*(August 2007), 541–556. https://doi.org/10.1002/aqc

Allen, G. R., & Werner, T. B. (2002). Coral reef fish assessment in the ' coral triangle ' of southeastern Asia. *Environmental Biology of Fishes*, *65*, 209–214.

Ashen, J. O. N. B., & Goff, L. J. (2000). Molecular and Ecological Evidence for Species Specificity and Coevolution in a Group of Marine Algal-Bacterial Symbioses, *66*(7), 3024–3030.

Baas Becking, L. G. M. (1934) Geobiologie of inleiding tot de milieukunde. The Hague, the Netherlands, W. P. Van Stockum and Zoon.

Bakenhus, I., Dlugosch, L., Billerbeck, S., Giebel, H. A., Milke, F., & Simon, M. (2017). Composition of total and cell-proliferating bacterioplankton community in early summer in the North Sea - roseobacters are the most active component. *Frontiers in Microbiology*, *8*(SEP), 1–14. https://doi.org/10.3389/fmicb.2017.01771

Barber, P. H. (2009). The challenge of understanding the Coral Triangle biodiversity hotspot. *Journal of Biogeography*, *36*, 1845–1846. https://doi.org/10.1111/j.1365-2699.2009.02198.x

Barber, P. H., & Erdmann, M. V. (2006). Comparative Phylogeography of three Codistributed Stomatopods: Origins and Timing of Regional Lineage Diversification in the Coral Triangle, *60*(9), 1825–1839.

Barber, P. H., Cheng, S. H., Erdmann, M. V., Tengardjaja, K. & Ambariyanto (2011) Evolution and conservation of marine biodiversity in the Coral Triangle: insights from stomatopod Crustacea. In: Crustacean Issues 19 Phylogeography and Population Genetics in Crustacea (eds Held C, Koenemann S and Schubart CD), pp. 129–156; CRC Press, Boca Raton.

Barott, K. L., Rodriguez-Mueller, B., Youle, M., Marhaver, K. L., Vermeij, M. J. A., Smith, J. E., & Rohwer, F. L. (2012). Microbial to reef scale interactions between the reef-building coral Montastraea annularis and benthic algae. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2011.2155

Barott, K. L., & Rohwer, F. L. (2012). Unseen players shape benthic competition on coral reefs. *Trends in Microbiology*, *20*(12), 621–628. https://doi.org/10.1016/j.tim.2012.08.004

Bellwood, D. R., Hughes, T. P., Connolly, S. R., & Tanner, J. (2005). Environmental and geometric constraints on Indo-Pacific coral reef biodiversity. *Ecology Letters*, *8*, 643–651. https://doi.org/10.1111/j.1461-0248.2005.00763.x

Bellwood, D. R., & Meyer, C. P. (2009). Searching for heat in a marine biodiversity hotspot. *Journal of Biogeography*, *36*(4), 569–576. https://doi.org/10.1111/j.1365-2699.2008.02029.x

Bohonak, A. J. (1999) Dispersal, Gene Flow, and Population Structure. T*he Quarterly Review of Biology* 74(1):21-45

Bolch, C. J. S. & Reynolds, M. J. (2002) Species resolution and global distribution of microreticulate dinoflagellate cysts. *Journal of Plankton Research* 24:565-578.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Ghalith, G. A. Al, … Naimey, A. T. (2018). QIIME 2 : Reproducible , interactive , scalable , and extensible microbiome data science. *PeerJ*. https://doi.org/10.7287/peerj.preprints.27295v2

Brandt, M. I., Trouche, B., Quintric, L., Wincker, P., Poulain, J., & Arnaud-Haond, S. (2020). A flexible pipeline combining clustering and correction tools for prokaryotic and eukaryotic metabarcoding. *Peer Community In Ecology*, 100043. https://doi.org/10.24072/pci.ecology.100043

Brown, M. V., Lauro, F. M., Demaere, M. Z., Muir, L., Wilkins, D., Thomas, T., Riddle, M. J., Fuhrman, J. A., Andrews-Pfannkoch, C., Hoffman, J. M., McQuaid, J. B., Allen, A., Rintoul, S. R. & Cavicchioli, R. (2012) Global biogeography of SAR11 marine bacteria. *Mol Syst Biol* 8:1–13

Caldera, E. J., Chevrette, M. G., McDonald, B. R., & Currie, C. R. (2019). Local adaptation of bacterial symbionts within a geographic mosaic of antibiotic coevolution. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.01580-19

Caldera, E. J., & Currie, C. R. (2012). The population structure of antibiotic-producing bacterial symbionts of apterostigma dentigerum ants: Impacts of coevolution and multipartite symbiosis. *American Naturalist*. https://doi.org/10.1086/667886

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*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Campbell, A. M., Fleisher, J., Sinigalliano, C., White, J. R., & Lopez, J. V. (2015). Dynamics of marine bacterial community diversity of the coastal waters of the reefs, inlets, and wastewater outfalls of southeast Florida. *MicrobiologyOpen*, *4*(3), 390–408. https://doi.org/10.1002/mbo3.245

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A.,Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, *6*(8), 1621–1624.

Carpenter, K. E., Barber, P. H., Crandall, E. D., Ablan-Lagman, M. C. A., Ambariyanto, Mahardika, G. N., Manjaji-Matsumoto, B. M., Juinio-Meñez, M. A., Santos, M. D., Starger, C.J. & Toha, A. H. A. (2011). Comparative Phylogeography of the Coral Triangle and Implications for Marine Management. *Journal of Marine Biology*, *2011*, 1–14. https://doi.org/10.1155/2011/396982

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*. https://doi.org/10.1890/13-0133.1

Chen, H., & Boutros, P. C. (2011). VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-12-35

Cheng, S. H., Anderson, F. E., Bergman, A., Mahardika, G. N., Muchlisin, Z. A., Dang, B. T., Calumpong, H. P., Mohamed, K. S., Sasikumar, G., Venkatesan, V. & Barber, P. H. (2013). Molecular evidence for co-occurring cryptic lineages within the Sepioteuthis cf . lessoniana species complex in the Indian and Indo-West Pacific Oceans. https://doi.org/10.1007/s10750-013-1778-0

Coryell, R., Turnham, K. E., Ayson, E. G. J., Pitogo, C., Alacala, A. C., Sotto, F. B., Gonzales, B.J. & Nishiguchi, M. K. (2018) Phylogeographic patterns in the Philippine archipelago influence symbiont diversity in the bobtail squid- Vibrio mutualism. *J. Evol. Ecol.* DOI: 10.1002/ece3.4266.

Coyne, J. A. & Orr, A. H. (2004) Speciation. Sunderland, Massachusetts, Sinauer Associates.

Craig, M. T., Eble, J. A., Bowen, B. W., & Robertson, D. R. (2007). High genetic connectivity across the Indian and Pacific Oceans in the reef fish Myripristis berndti (Holocentridae). *Marine Ecology Progress Series*, *334*, 245–254.

De Wit, R., & Bouvier, T. (2006). “Everything is everywhere, but, the environment selects”; what did Baas Becking and Beijerinck really say? *Environmental Microbiology*, *8*(4), 755–758. https://doi.org/10.1111/j.1462-2920.2006.01017.x

De Boer, T. S., Naguit, M. R. A., Erdmann, M. V., Ablan-Lagman, M. C. A., Ambariyanto, A., Carpenter, K. E., Toha, A. H. A. & Barber, P. H. (2014). Concordance between phylogeographic and biogeographic boundaries in the Coral Triangle: Conservation implications based on comparative analyses of multiple giant clam species. *Bulletin of Marine Science*, *90*(1). https://doi.org/10.5343/bms.2013.1003

DeBoer, T. S., Subia, M. D., Erdmann, M. V, Kovitvongsa, K., & Barber, P. H. (2008). Phylogeography and Limited Genetic Connectivity in the Endangered Boring Giant Clam across the Coral Triangle, *22*(5), 1255–1266. https://doi.org/10.1111/j.1523-1739.2008.00983.x

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* 1:47-50.

Fenchel, T., & Finlay, B. J. (2004). The Ubiquity of Small Species: Patterns of Local and Global Diversity. *BioScience*, *54*(8), 777. https://doi.org/10.1641/0006-3568(2004)054[0777:TUOSSP]2.0.CO;2

Fenchel, T. (2003) Biogeography for bacteria. *Science* 301:925-926.

Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S. G., González, J. M., & Pedrós-Alió, C. (2013). Ecology of marine bacteroidetes: A comparative genomics approach. *ISME Journal*, *7*(5), 1026–1037. https://doi.org/10.1038/ismej.2012.169

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C. M., Boyer, F., Pompanon, F., Rayé, G. & Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.12338

Finlay, B. J. (2002). Global dispersal of free-living microbial eukaryote species. *Science*. https://doi.org/10.1126/science.1070710

Finlay, B. J. & Esteban, G. F. (2004) Ubiquitous dispersal of free-living microorganisms *in* B. Alan, T, ed. Microbial diversity and bioprospecting. Washington, ASM.

Finlay, B. J. (2014). Global Dispersal of Eukaryote Species Microbial, *296*(5570), 1061–1063.

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01312-x

Fuhrman, J. A., & Steele, J. A. (2008). Community structure of marine bacterioplankton: Patterns, networks, and relationships to function. *Aquatic Microbial Ecology*, *53*(1), 69–81. https://doi.org/10.3354/ame01222

Gaither, M. R., & Rocha, L. A. (2013). Origins of species richness in the Indo-Malay-Philippine biodiversity hotspot: Evidence for the centre of overlap hypothesis. *Journal of Biogeography*. https://doi.org/10.1111/jbi.12126

Ghiglione, J.-F., Galand, P. E., Pommier, T., Pedros-Alio, C., Maas, E. W., Bakker, K., Bertilson, S., Kirchman, D. L., Lovejoy, C., Yager, P. L., & Murray, A. E. (2012). Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proceedings of the National Academy of Sciences*, *109*(43), 17633–17638.

Glockner, F. O., Zaichikov, E., Belkova, N., Denissova, L., Pernthaler, J., Pernthaler, A. & Amann. R. (2000) Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of Actinobacteria. *Applied and Environmental Microbiology* 66:5053-5065.

Hedlund, B. P. & Staley, J. T. (2004) Microbial Endemism and Biogeography *in* A. T. Bull, ed. Microbial diversity and bioprospecting. Washington, ASM.

Hernandez-Agreda, A., Leggat, W., Bongaerts, P., Herrera, C., & Ainsworth, T. D. (2018). Rethinking the Coral Microbiome: Simplicity Exists within a Diverse Microbial Biosphere. *MBio*, *9*(5), 1–14. https://doi.org/10.1128/mbio.00812-18

Hillebrand, H., Watermann, F., Karez, R., & Berninger, U. G. (2001). Differences in species richness patterns between unicellular and multicellular organisms. *Oecologia*. https://doi.org/10.1007/s004420000492

Hoeksema, B. W. (2007). Delineation of the Indo-Malayan Centre of Maximum Marine Biodiversity : The Coral Triangle, 117–178.

Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.12613

Jackson, A. M., Erdmann, M. V, Toha, A. H. A., Stevens, L. A., & Barber, P. H. (2014). Phylogeography of commercial tuna and mackerel in the Indonesian Archipelago, *90*(1), 1–22.

Kandlikar, G. S., Gold, Z. J., Cowen, M. C., Meyer, R. S., Freise, A. C., Kraft, N. J. B., Moberg-Parker, J., Sprague, J., Kushner, D. J. & Curd, E. E. (2018). Ranacapa: An R package and shiny web app to explore environmental DNA data with exploratory statistics and interactive visualizations [version 1; referees: 1 approved, 2 approved with reservations]. *F1000Research*. https://doi.org/10.12688/f1000research.16680.1

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4). https://doi.org/10.1093/molbev/mst010

Kawecki, T. J. & Ebert, D. (2004) Conceptual issues in local adaptation. *Ecology Letters* 7:1225-1241.

Kelly, L. W., Williams, G. J., Barott, K. L., Carlson, C. A., Dinsdale, E. A., Edwards, R. A., Haas, A. F., Haynes, M., Lim, Y. W., McDole, T., Nelson, C. E., Sala, E., Sandin, S. A., Smith, J. E., Vermeij, M. J. A., Youle, M. & Rohwer, F. (2014). Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(28), 10227–10232. https://doi.org/10.1073/pnas.1403319111

Knowlton, N., Brainard, R. E., Fisher, R., Moews, M., Plaisance, L., & Caley, M. J. (2010). Coral Reef Biodiversity. In A. D. McIntyre (Ed.), *Life in the World's Oceans* (pp. 65–77). Blackwell Publishing Ltd.

Ladd, H.S. (1960) Origin of the Pacific Island molluscan fauna. American Journal of Science, 258A, 137–150.

Lande, R. (1980) Genetic variation and phenotypic evolution during allopatric speciation. *American Naturalist* 116:463-479.

Leray, M., & Knowlton, N. (2015a). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, *112*(7), 2076–2081. https://doi.org/10.1073/pnas.1424997112

Leray, M., & Knowlton, N. (2015b). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, *112*(7), 2076–2081. https://doi.org/10.1073/pnas.1424997112

Leray, M., & Knowlton, N. (2016). Censusing marine eukaryotic diversity in the twenty-first century. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1702), 20150331. https://doi.org/10.1098/rstb.2015.0331

Lenormand, T. (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* 17:183-189.

Lively, C. M. (1999) Migration, virulence, and the geographic mosaic of adaptation by parasites. *American Naturalist* 153:S34-S47.

Lourie, S. A., Green, D. M. & Vincent, A. C. J. (2005) Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (Syngnathidae: *Hippocampus*). *Mol Ecol.* 14:1073–1094. PMid:15773937. http://dx.doi.org/10.1111/j.1365-294X.2005.02464.x

Moitinho-Silva, L., Nielsen, S., Amir, A., Gonzalez, A., Ackermann, G. L., Cerrano, C., Astudillo-Garcia, C., Easson, C., Sipkema, D., Liu, F., Steinert, G., Kotoulas, G., McCormack, G. P., Feng, G., Bell, J. J., Vicente, J., Björk, J. R., Montoya, J. M., Olson, J. B., Reveillaud, J., Steindler, L., Pineda, M. C., Marra, M. V., Ilan, M., Taylor, M. W., Polymenakou, P., Erwin, P. M., Schupp, P. J., Simister, R. L., Knight, R., Thacker, R. W., Costa, R., Hill, R. T., Lopez-Legentil, S., Dailianis, T., Ravasi, T., Hentschel, U., Li, Z., Webster, N. S. & Thomas, T. The sponge microbiome project, *GigaScience*, Volume 6, Issue 10, October 2017, gix077, https://doi.org/10.1093/gigascience/gix077

Madigan, M. T., Martinko, J. M., & Parker, J. (2003). Brock biology of microorganisms. Upper Saddle River, NJ: Prentice Hall/Pearson Education.

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Krumins, J. A., Kuske, C. R., Morin, P. J., Naeem, S., Øvreås, L., Reysenbach, A. L., Smith, V. H. & Staley, J. T. (2006). Microbial biogeography: Putting microorganisms on the map. *Nature Reviews Microbiology*, *4*(2), 102–112. https://doi.org/10.1038/nrmicro1341

Mayr, E. (1963) Animal species and evolution. Cambridge, MA, Harvard University Press.

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061217

Meyer, C. P. (2003). Molecular systematics of cowries (Gastropoda : Cypraeidae) and diversification patterns in the tropics, 401–459.

Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., Green, L. E. & Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.12895

Nishiguchi, M. K., Ruby, E. G. & McFall-Ngai, M. J. (1998) Competitive dominance during colonization is an indicator of coevolution in an animal-bacterial symbiosis. *Appl. Environ. Microbiol.* 64 (9):3209-13.

Ogle, D. H., P. Wheeler. & A. Dinno. 2018. FSA: Fisheries Stock Analysis. R package version 0.8.22, https://github.com/droglenc/FSA).

O'Malley, M. A. (2007) The nineteenth century roots of 'everything is everywhere'. *Nature Reviews Microbiology* 5:647-651.

O'Malley, M. A. (2008). "Everything is everywhere: but the environment selects": ubiquitous distribution and ecological determinism in microbial biogeography. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*, *39*(3), 314–325. https://doi.org/10.1016/j.shpsc.2008.06.005

Oksanen, J. (2017). Vegan: ecological diversity. *R Package Version 2.4-4*. https://doi.org/10.1029/2006JF000545

Pearman, J., Aylagas, E., Voolstra, C. R., Anlauf, H., Villalobos, R., & Carvalho, S. (2019). Disentangling the complex microbial community of coral reefs using standardized Autonomous Reef Monitoring Structures (ARMS). *Molecular Ecology*. https://doi.org/10.1111/mec.15167

Pearman, J. K., Anlauf, H., Irigoien, X., & Carvalho, S. (2016). Please mind the gap - Visual census and cryptic biodiversity assessment at central Red Sea coral reefs. *Marine Environmental Research*, *118*, 20–30. https://doi.org/10.1016/j.marenvres.2016.04.011

Plaisance, L., Brainard, R., Caley, M. J., & Knowlton, N. (2011). Using DNA Barcoding and Standardized Sampling to Compare Geographic and Habitat Differentiation of Crustaceans: A Hawaiian Islands Example. *Diversity*, (4), 581–591. https://doi.org/10.3390/d3040581

Pommier, T., Canbäck, B., Riemann, L., Boström, K. H., Simu, K., Lundberg, P., Tunlid, A., & Hagström, Å. (2007). Global patterns of diversity and community structure in marine bacterioplankton. *Molecular Ecology*, *16*(4), 867–880.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0009490

Ransome, E., Geller, J. B., Timmers, M., Leray, M., Mahardini, A., Sembiring, A., Collins, A. G. & Meyer, C. P. (2017). The importance of standardization for biodiversity comparisons: A case study using autonomous reef monitoring structures (ARMS) and metabarcoding to measure cryptic diversity on Mo'orea coral reefs, French Polynesia. *PLoS ONE*, *12*(4), 1–19. https://doi.org/10.1371/journal.pone.0175066

Roberts, C. M., McClean, C. J., Veron, J. E., Hawkins, J. P., Allen, G. R., McAllister, D. E., Mittermeier, C. G., Schueler, F. W., Spalding, M., Wells, F., Vynne, C. & Werner, T. B. Marine biodiversity hotspots and conservation priorities for tropical reefs. Science. 2002 Feb 15;295(5558):1280-4. doi: 10.1126/science.1067728. PMID: 11847338.

Rohwer, F., Seguritan, V., Azam, F., & Knowlton, N. (2002). Diversity and distribution of coral-associated bacteria. *Marine Ecology Progress Series*. https://doi.org/10.3354/meps243001

Rusch, A., Hannides, A. K., & Gaidos, E. (2009). Diverse communities of active Bacteria and Archaea along oxygen gradients in coral reef sediments. *Coral Reefs*, *28*(1), 15–26. https://doi.org/10.1007/s00338-008-0427-y

Slatkin, M. (1993). Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution*. https://doi.org/10.2307/2410134

Spalding, M. D., Fox, H. E., Allen, G. R., Davidson, N., Ferdaña, Z. A., Finlayson, M., Halpern, B. S., Jorge, M. A., Lombana, A., Lourie, S. A., Martin, K. D., McManus, E., Molnar, J., Recchia, C. A., Robertson, J. & Robertson, J. (2007). Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. *BioScience*, *57*(7), 573–583. https://doi.org/10.1641/B570707

Thompson, J. N. & Cunningham, B. M. (2002) Geographic structure and dynamics of coevolutionary selection. *Nature* 417:735-738.

Thompson, J. N. (2005) The geographic mosaic of coevolution. Chicago, University of Chicago Press.

Thompson, J. N. (2009) The Coevolving Web of Life. *American Naturalist* 173:125-140.

Thurber, R. V., Willner-Hall, D., Rodriguez-Mueller, B., Desnues, C., Edwards, R. A., Angly, F., Dinsdale, E., Kelly, L. & Rohwer, F. (2009). Metagenomic analysis of stressed coral holobionts. *Environmental Microbiology*. https://doi.org/10.1111/j.1462-2920.2009.01935.x

Thurber, R. V. (2009) Current insights into phage biodiversity and biogeography. *Current Opinion in Microbiology* 12:582-587.

Veron, J. E. N., Devantier, L. M., Turak, E., & Green, A. L. (2009). Delineating the Coral Triangle. *Galaxea, Journal of Coral Reef Studies*, *11*, 91–100.

Vollmer, S. A., Bormane, A., Dinnis, R. E., Seelig, F., Dobson, A. D. M., Aanensen, D. M., James, M. C., Donaghy, M., Randolph, S. E., Feil, E. J., Kurtenbach, K. & Margos, G. (2011). Host migration impacts on the phylogeography of Lyme Borreliosis spirochaete species in Europe. *Environmental Microbiology*. https://doi.org/10.1111/j.1462-2920.2010.02319.x

Walters, W., Hyde, E. R., Berg-lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J. A. & Jansson, J. K. (2015). Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *MSystems*, *1*(1), 1–15. https://doi.org/10.1128/mSystems.00009-15.Editor

Wangensteen, O. S., & Turon, X. (2016). Metabarcoding Techniques for Assessing Biodiversity of Marine Animal Forests. In *Marine Animal Forests* (pp. 1–34). https://doi.org/10.1007/978-3-319-17001-5

Webster, N. S., Wilson, K. J., & Blackall, L. L. (2001). Phylogenetic Diversity of Bacteria Associated with the Marine Sponge Rhopaloeides odorabile. *Applied and Environmental Microbiology 67*(1), 434–444. https://doi.org/10.1128/AEM.67.1.434

Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., & Rohwer, F. (2007). Metagenomic analysis of the microbial community associated with the coral Porites astreoides. *Environmental Microbiology*, *9*(11), 2707–2719.

Whitaker, R. J. (2006). Allopatric origins of microbial species. In *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2006.1927

Whitaker, R. J., Grogan, D. W., & Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science*. https://doi.org/10.1126/science.1086909

Whitaker, R. J., Grogan, D. W., & Taylor, J. W. (2005). Recombination shapes the natural population structure of the hyperthermophilic archaeon Sulfolobus islandicus. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msi233

Whittaker, K. A., & Rynearson, T. A. (2017). Evidence for environmental and ecological selection in a microbe with no geographic limits to gene flow. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1612346114

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York. Media*. https://doi.org/10.1007/978-0-387-98141-3

Wilkinson (2001) What is the upper size limit for cosmopolitan distribution in free-living microorganisms? *Journal of Biogeography* 28(3)

Wirth, T., Meyer, A. & Achtman, M. (2005) Deciphering host migrations and origins by means of their microbes. *Molecular Ecology* 14:3289-3306.

Wise, M. G., Shimkets, L. J. & McArthur, J. V. (1995) Genetic structure of a lotic population of *Burkholderia (Psuedomonas) cepacia*. Applied and *Environmental Microbiology* 61:1791-1798.

Ziegler, M., Roik, A., Porter, A., Zubier, K., Mudarris, M. S., Ormond, R., & Voolstra, C. R. (2016). Coral microbial community dynamics in response to anthropogenic impacts near a major city in the central Red Sea. *Marine Pollution Bulletin*, *105*(2), 629–640. https://doi.org/10.1016/j.marpolbul.2015.12.045

Zwart, G., Crump, B. C., Agterveld, M., Hagen, F. & Han. S. K. (2002) Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquatic Microbial Ecology* 28:141-155.

CHAPTER 3

Biodiversity Monitoring in High Diversity Marine Ecosystems

Abstract

Indonesia is the center of the Coral Triangle, the global epicenter of marine biodiversity. Key to promoting ecosystem sustainability and preserving the benefits derived from this diversity is monitoring changes (positive and negative) from baseline data on ecosystem health and assessing how local ecosystems respond to management practices**.** While metabarcoding is increasingly used for monitoring of marine ecosystems, its costs can be prohibitive for scientists in developing counties. Here, we re-examine DNA metabarcoding data from Autonomous Reef Monitoring Structure (ARMS) presented in chapter one to determine how marine monitoring efforts in Indonesia could be constructed to yield the best results for the least cost. Both Cytochrome Oxidase 1 (COI) and 18S rRNA returned diversity patterns that correspond to biodiversity patterns from common monitoring targets, fish and corals. However, there were tradeoffs associated with marker choice. While 18S rRNA captured the greatest absolute number of OTUs, 18S rRNA species accumulation curves saturated slower than COI. Similarly, while 18S rRNA had far fewer unidentified OTUs than COI (2% vs 35%), identifications were only at higher taxonomic levels, indicating greater precision in COI. Comparison of size fractions indicated that the 106-500 µm fraction captured substantially more total diversity and endemic diversity compared to the 500 µm-2 mm, and sessile fractions. Combined, results suggests that metabarcoding only the 106-500 µm size fraction with COI could provide the most accurate and economical approach to monitoring diversity in megadiverse regions where limited research investment may preclude sequencing multiple size fractions with multiple metabarcoding markers. By integrating metabarcoding data with traditional ecological survey methods, we can

expand our understanding of coral reef ecosystems' diversity and function, positioning us better to manage them sustainably and better target conservation interventions.

Introduction

Coral reefs are among the world's most important marine ecosystems; they support economically valuable fisheries, promote food security, are a source of pharmaceutical products, and sustain jobs and other businesses through tourism and recreation (Hoegh-Guldberg et al., 2009). Among the world's coral reef ecosystems, the Coral Triangle is one of the most important. Defined by the presence of 500 or more scleractinian coral species (Veron et al., 2009), the Coral Triangle spans six countries including Indonesia, Malaysia, the Philippines, Papua New Guinea, Timor-Leste, and the Solomon Islands. The Coral Triangle is the global epicenter of marine biodiversity, hosting 76% of the world's coral species (Veron et al., 2009), 37% of the world's reef fishes (Allen, 2008), 8% of which are endemic or locally restricted species. This diversity is the primary protein source and supports the livelihoods of more than 120 million people (Cruz-Trinidad et al., 2014).

Despite this biological and economic importance, the Coral Triangle is among the world's most threatened marine ecosystems. Currently, more than 85% of the reefs in the Coral Triangle are threatened and degraded due to anthropogenic stressors ranging from pollution to overfishing and destructive fishing practices (Burke et al., 2012). This dramatic loss of reef ecosystems is exacerbated by global environmental threats such as ocean acidification and rising sea surface temperatures caused by climate change (Hoegh-Guldberg et al. 2009; Peñaflor et al. 2009; Hoegh-Guldberg, 2011; Hughes et al., 2018), threatening the very future of the Coral Triangle and the goods and services it provides the 360 million people inhabiting this region.

A key aspect of promoting ecosystem sustainability is monitoring changes (positive and negative) from baseline data on ecosystem health and assessing how local ecosystems respond to management practices (van der Meij et al., 2010). Typically, biodiversity monitoring of coral reefs focuses on macrofauna such as fishes, corals, and molluscs, because they are large and relatively well known taxonomically (Bellwood, 2001; Bellwood et al., 2005; Mustika et al., 2012; Mustika et al., 2013). Moreover, the above methods typically employ visual surveys that require specialized taxonomic expertise and are challenging to repeat (Bouchet et al., 2002; Knowlton et al., 2010). This is particularly true in the developing countries of the Coral Triangle where resources and taxonomic experts are very limited (Barber et al., 2014).

Another important issue with current monitoring practices is that monitored taxa, such as fish and corals, only comprise a small fraction of marine life and are likely not representative of all marine biodiversity, the majority of which are "cryptofauna", small, cryptic marine species that are largely undocumented (Knowlton et al., 2010; Plaisance et al., 2011a). Cryptofauna are critical component of the coral reef food webs. They capture and recycle nutrients from plankton and detritus, and act as the primary food source for many reef organism (e.g., wrasses, snappers, groupers, and moray eels) (Enochs & Manzello, 2012; Leray et al., 2015), with larvae of cryptobenthic fish representing >50% of consumed reef fish biomass (Brandl et al., 2019). Additionally, coral-associated cryptofauna (e.g., trapeziid crabs and alpheid shrimps) protect their hosts from corallivores (Seabird McKeon & Moore 2014; Counsell et al., 2018) or remove sediment from the coral tissue, increasing the growth and survival of the host coral (Stewart et al., 2006; Stier et al., 2012). As such, coral reef cryptofauna a critical part of coral reef ecosystem function, making them essential to monitor for sustainability management.

DNA metabarcoding (Carugati et al., 2015; Wangensteen & Turon 2016; Pavan-Kumar et al., 2015) is a technique where entire communities of organisms can be identified through species-specific DNA sequences obtained through high-throughput DNA sequencing of community DNA (Kelly et al., 2017; Leray & Knowlton 2015; Leray & Knowlton 2016a; Al-Rshaidat et al., 2016). Autonomous Reef Monitoring Structures (ARMS) are artificial structures that mimic the structure of the coral reef ecosystem (Figure 3-1A, 3-1B), providing a standardized tool for sampling benthic marine biodiversity that can be subsequently identified through DNA metabarcoding (Knowlton et al., 2010; Plaisance, et al., 2011a; Leray & Knowlton 2015; Leray & Knowlton 2016a; Pearman et al., 2016; Al-Rshaidat et al., 2016; Ransome et al., 2017). The ARMS approach dramatically expands the number and taxonomic breadth of species captured in biodiversity surveys, and does it in a standardized way ideally suited to the monitoring of marine ecosystems.

While ARMS hold much promise for improving the power of marine ecosystem monitoring in imperilled high diversity ecosystems like the Coral Triangle, recent studies suggest that high diversity, itself, can pose a significant challenge for metabarcoding approaches. For example, Raja Ampat, Indonesia, is the global epicenter of reef fish diversity (Allen, 2008), and a recent study using metabarcoding of environmental DNA in this region shows that it is difficult to sample enough or sequence deep enough to saturate species (or Operational Taxonomic Units; OTU) accumulation curves (Juhel et al., 2020). Another environmental DNA metabarcoding study spanning the entire Indonesian Archipelago returned similar results, even in locations like Bali or Sumatera, where fish diversity is much lower than Raja Ampat (Marwayana, 2018). As such, it is unclear whether metabarcoding-based monitoring approaches like ARMS have the power to effectively capture local diversity in megadiverse regions like the Coral Triangle.

Another challenge for ARMS based metabarcoding in regions like the Coral Triangle is cost. Typical ARMS workflows separate samples into three size fractions (106-500 µm, 500 µm -2mm, sessile encrusting) to prevent DNA from larger organisms from swamping the signal of smaller ones (Ransome et al. 2017). Moreover, metabarcoding typically uses two markers. *Cytochrome C Oxidase Subunit I* (COI) gene is a common metabarcoding marker (Leray & Knowlton 2015b; Al-Rshaidat et al., 2016; Pearman et al., 2018) with great utility across a wide range of taxa, including marine organisms (Hebert et al., 2003; Leray et al., 2013). However, because COI metabarcoding primers were designed to target metazoans, 18S rRNA (V9 and V4 region) is usually paired with COI to distinguish eukaryotes and protists (Caporaso et al., 2011; Hadziavdic et al., 2014). Kelly (2016) shows that these two markers only have partial taxonomic overlap, suggesting that they be used in tandem to provide the most complete coverage of marine biodiversity. While using two markers on three size fractions may be ideal in terms of biodiversity monitoring, this approach substantially increases costs, potentially putting metabarcoding out of reach for developing countries like Indonesia where government investment is relatively low (Barber et al., 2014).

In this study, we re-examine data from Chapter 1 to determine the most efficient and cost effective way to use ARMS where resources are limited. Specifically, we examining data from each of the two metabarcoding markers and size fractions to determine whether each can independently capture the diversity of this global marine biodiversity hotspot, from the most diverse reefs in the world in Eastern Indonesia to regions of Western Indonesia that fall outside of the Coral Triangle. We also specifically compare results from COI and 18S rRNA, to determine whether they differ in performance across this biodiversity gradient. Importantly, we do this in a geographically nested design to determine how marine monitoring efforts in

Indonesia could be constructed to yield the best results for the least cost, allowing resource managers to better monitor this region's biodiversity by providing detailed assessments of marine biodiversity across a broad range of taxonomic groups.

Materials and Methods

ARMS Deployment, Collection, and Sampling

We deployed ARMS across the Indonesian archipelago (Figure 3-1). Deployments ranged from reefs in Western Indonesia (Aceh, Sumatera and the Pulau Seribu, Java) that sit outside of the Coral Triangle (Hoeksema 2007; Bellwood & Meyer 2009) to the western edge of the Coral Triangle (Pemuteran, Bali), to Eastern Indonesia (Raja Ampat and Teluk Cenderawasih, Western Papua), a region known for having the highest coral (Veron et al., 2009) and fish (Allen, 2008) diversity in the world.

In each of these five areas, we deployed ARMS at four different reefs and each deployment consisted of three individual ARMS units, totalling 60 ARMS units (Supplemental Table S3-1). To minimize biodiversity differences related to habitat variation, we deployed ARMS on the seafloor at ~10 m depth (min 9.7m, max 13m) in similar fore reef environments. After being deployed underwater for three years, we recovered the ARMS and the associated organisms that had colonized them, following the methods of Ransome et al. (2017). Briefly, we first encapsulated ARMS in crates lined with 40 µm nitex mesh to prevent the escape of motile organisms during recovery. We then transported ARMS in crates of aerated filtered seawater back to the lab where we disassembled and photographed the top and bottom of every ARMS plate.

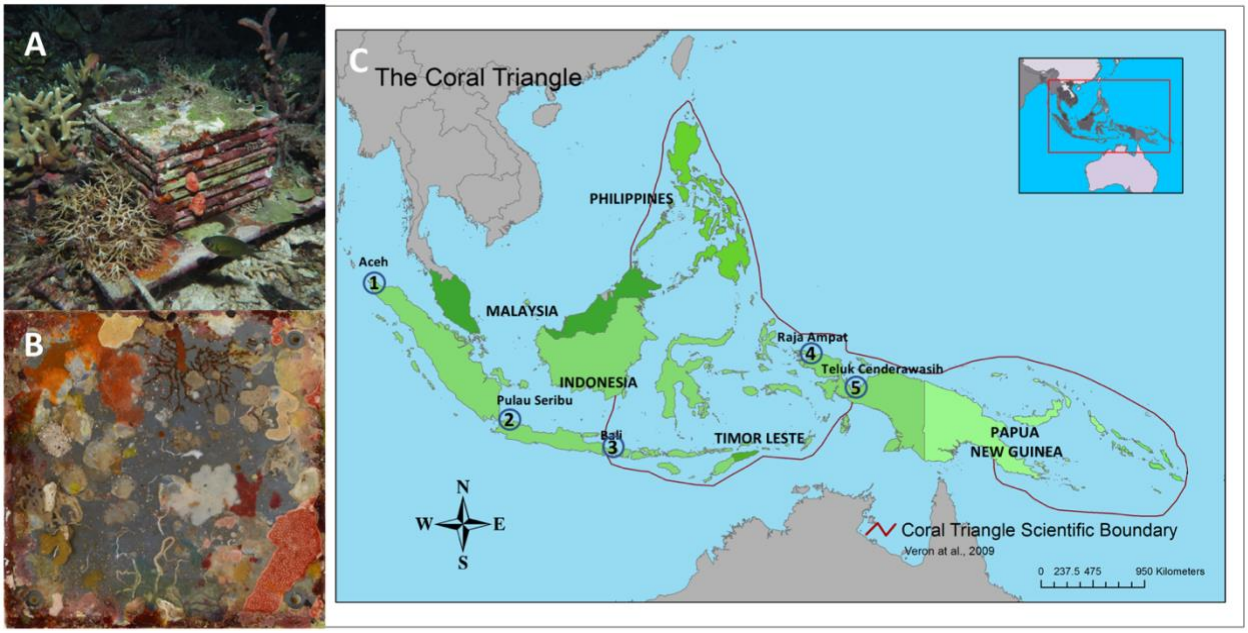


Figure 3-1. A) ARMS structure photographed underwater. B) Autonomous Reef Monitoring Structure (ARMS) plate colonized by organisms. C) Map of the Coral Triangle with five sampling locations: (1) Pulau Weh, Aceh, (2) Kepulauan Seribu, Jakarta, (3) Pemuteran, Bali, (4) Raja Ampat, West Papua, and (5) Teluk Cenderawasih, West Papua. The Coral Triangle Scientific Boundary (red line) was based on Veron et al. (2009).

Next, using standard ARMS protocols that separate organisms by size class to prevent organisms with higher biomass from swamping the signal of smaller organisms (Ransome et al., 2017), we separated motile organisms into three size classes using geological sieves. First, we removed all organisms >2 mm, and saved these for future taxonomic study. We then isolated all motile organisms from 500 µm-2 mm (hereafter "500 µm fraction") and all motile organisms between 106-500 µm (hereafter "100 µm fraction"). We isolated these sequentially, from largest to smallest, washing the fractions with filtered seawater. After washing, we consolidated each

sample fraction with a final wash in filtered seawater into a 45 μm nitex net, preserved the samples in 95% ethanol, and stored the samples at -20 °C until DNA extraction.

To document non-motile fractions, we scraped all encrusted biota from the ARMS plates (hereafter "sessile fraction") into filtered seawater and then homogenized it with a kitchen blender for 30 s at maximum speed. We rinsed the homogenate with filtered sea water taken from the ARMS recovery site into a 45 μm Nitex mesh collection net until the water ran clear, then placed approximately 10 g of the homogenate into a 50 ml falcon tube filled with DMSO, stored at -20 °C until DNA extraction, following standard ARMS protocols (Ransome et al. 2017).

DNA Preparation and Extraction

To remove inorganic material (e.g. sediment) that could inhibit DNA amplification, we performed a series of decantations (see Leray and Knowlton, 2015; https://www.oceanarms.org/protocols/molecular-analysis/bulk-dna-extractions/sample-preparation). Next, we extracted DNA from the 500 μm, 100 μm, and sessile samples using the MO-BIO Powermax® Soil DNA Isolation Kit according to the manufacturer's protocol with the addition of 400 μg/ml Proteinase K. To remove potential PCR inhibitors, we purified the DNA extractions using MO-BIO PowerClean® DNA Clean-Up Kits and quantified DNA concentrations using Qubit dsDNA HS Kit. The decantation and DNA extraction were performed at Yayasan Biodiversitas Indonesia (Bionesia), Denpasar, Bali, Indonesia and Laboratory of Marine Molecular Genetics, Research Center for Oceanograpy, Indonesian Institute of Sciences, Jakarta, Indonesia.

We amplified *Cytochrome c Oxidase Subunit I* (COI) using a dual-indexing approach with seven pairs of tagged COI PCR primers (mlCOIintF/jgHCO2198) (Geller et al., 2013;

Leray et al., 2013). To account for potential PCR bias (Ficetola et al., 2015; Nichols et al., 2018) and maximize probability of amplification of low copy templates, we performed PCR reactions in triplicate. Each PCR reaction was 20 μL in volume, consisting of 1 μL of 10 μM each forward or reverse primer, 1.4 μL of dNTPs, 0.4 μL of Taq Polymerase (CIontech), 2 μL of PCR buffer (CIontech), using 10 ng of extracted DNA. Thermocycling utilized a touchdown profile beginning with 16 initial cycles of denaturing at 95 °C for 10 s, annealing for 30 s at 62 °C (−1 °C per cycle), and extension for 60 s at 72 °C, followed by 20 cycles at an annealing temperature of 46 °C. We then pooled and visualized triplicate PCR products via electrophoreses in a 1.2% agarose gel.

To prepare the sequencing libraries, we pooled 1 ug of each tagged PCR sample, each comprised of a series of the seven tailed-primer pairs, into 12 samples and purified with Agencourt AMPure XP beads. We used a total of 1 μg of these pooled amplicons for end repair, A-tailing, and adaptor ligation using the TruSeq PCR-free kit (Illumina) following manufacturer protocols. We then validated the libraries via qPCR using the KAPA library quantification kit and diluted to a final concentration of 4nM before sequencing on an Illumina MiSeq using the 600 cycle reagent kit v3 (Illumina, San Diego, CA).

To assess ARMS diversity using the V4 region of 18S rRNA gene (Luddington et al., 2012), we conducted 18S rRNA PCRs using the V4_18SNext.For and V4_18SNext.Ref primer pair (Manzari et al., 2015), using 25uL reaction volumes consisting of 1.25 μL of 0.5 μM each forward or reverse primer, 2.5 μL of 0.2 mM dNTPs, 0.5 μL of Taq Polymerase (Phusion), 5 μL of PCR buffer (Phusion), and 2.5 ng DNA. Thermocycling parameters used an initial denaturing at 98 °C for 30 s, 10 cycles each at 98°C for 10 s, 44°C for 30 s, and 72 °C for 15 s, followed by 15 cycles each at 98°C for 10 s, 62°C for 30 s, and 72 °C for 15 s and finish with a final

extension step at 72 °C for 7 min. We then confirmed and visualized successful PCRs through electrophoresis on a 1.2% agarose gel.

To create the 18S sequencing library, we employed a dual-indexing approach using the Nextera® index kit (Illumina), confirming successful indexing through electrophoresis on a 1.2% agarose gel. We then cleaned the libraries with Agencourt AMPure XP beads, pooled and diluted to a final concentration of 2nM. The sequencing was performed with MiSeq Illumina using a V2 500-cycle kit with 20% PhiX DNA added to each run to improve data quality. All DNA sequencing was performed at the Laboratories of Analytical Biology (LAB), Smithsonian Institution National Museum of Natural History, Washington DC.

Data analyses

COI sequences were pre-processed, quality filtered and analyzed using *QIIME2* version 2017.8.0. (the Quantitative Insights Into Microbial Ecology 2 program, https://qiime2.org/). We used the Divisive Amplicon Denoising Algorithm 2 (*DADA2*) software, wrapped in *QIIME2*, to filter, trim, de-noise, and merge the data, removing chimeric sequences using the consensus method (Callahan et al., 2016).

For 18S rRNA, we merged the forward and reverse reads using *PEAR* (Zhang et al., 2014) allowing for a maximum of 10 differences in the overlap (default in *PEAR*) and only keeping aligned reads between 380 and 440 base pairs before quality filtering and analysis using *QIIME2* version 2017.8.0 (the Quantitative Insights Into Microbial Ecology 2 program, https://qiime2.org/) and the Divisive Amplicon Denoising Algorithm 2 (*DADA2*) software to filter, trim, de-noise, and removing chimeric sequences using the consensus method (Callahan et al., 2016).

Next, we used the *LULU* algorithm (Frøslev et al., 2017) to filter out spurious sequences that may originate from PCR and/or sequencing errors, intra-individual variability (pseudogenes, heteroplasmy). *LULU* filters based on sequence similarity and co-occurrence rate with more abundant clusters, allowing us to curate the datasets while avoiding arbitrary abundance filters (Frøslev et al., 2017; Brandt et al., 2020). We ran *LULU* with a minimum relative co-occurrence of 0.95 for both COI and 18S rRNA dataset, using a minimum similarity threshold (minimum match) at 84% (default) for COI and 18S rRNA.

Following quality filtering, trimming, de-noising, and chimera removal, we performed Operational Taxonomic Unit (OTU) clustering using the *Vsearch* (Rognes et al., 2016) plug-in in *QIIME2*. OTU clustering employed a 97% sequence identity threshold for COI and 99% sequence identity threshold for 18S rRNA using de novo clustering with "QIIME vsearch cluster-features-de-novo" command. This *de novo* OTU picking process clusters sequence reads by comparing sequences against one another without any external reference sequence collection.

To assign taxonomy to COI OTUs, we performed taxonomy assignment using *BlastN* (Altschul et al., 1990) against the *CRUX* (Curd et al., 2018) database, using a cutoff of 85% sequence identity. All assigned sequences were then aligned with *MAFFT* (Katoh & Standley, 2013), and used for phylogenetic reconstruction in *FastTree* (Price et al., 2010). We excluded all COI sequences that *BLAST* assigned to bacteria, keeping metazoan and unidentified taxa. To assign taxonomy for 18S rRNA sequences, we used a feature classifier in *QIIME2* trained against the PR2 database (Guillou et al., 2013), adopting a default confidence threshold of 0.7. To summarize the taxonomic composition of each sample we used *phyloseq* (McMurdie & Holmes 2013) to generate stacked bar plots summarizing taxonomic composition and sequence abundance using *ggplot2* (Wickham, 2009) in R. Because of paucity of barcoding data from the

Coral Triangle, we merged taxa at the phylum level and removed groups that represented less than 2% total abundance of the community.

Statistical Analyses

To test for saturation of OTU discovery, we created rarefaction curves with *iNEXT* (Chao et al., 2014; Hsieh et al., 2016) and *Ranacapa* package using *ggrare* command (Kandlikar et al., 2018). Next, we calculated alpha and beta diversity statistics using the *phyloseq* package in R (R development core team) (McMurdie & Holmes 2013) and tested for significant differences in an one way ANOVA framework that examined the richness from total dataset and per size fraction (100 μm, 500 μm, and sessile) across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cendrawasih). In addition, we combined all OTUs from the three size fractions from each ARMS, analyzing diversity of ARMS as a single unit, capturing the diversity of the entire ARMS community.

To examine patterns in species richness, we calculated Chao and Shannon diversity indices separately for all locations and size fractions and conducted one-way ANOVAs. To account for tests for multiple comparisons, we employed a Tukey's test implemented in the package *FSA* (Ogle and Dinno. 2018). Because early life stages or dissociated pieces of larger organisms could potentially be recovered in smaller size fractions, we explored the degree of taxonomic overlap among the three sample fractions using *VennDiagram* package (Chen & Boutros, 2011) in R.

In addition, we examined OTU richness in a hierarchical fashion. We used Venn diagrams to visualize similarities and differences in community composition among the different locations through *http://bioinformatics.psb.ugent.be/webtools/Venn/*, and to calculate the numbers of shared and unique OTUs across 1) all five regions, and 2) all four reefs within each

region. We did this analysis for total OTU diversity (e.g. diversity from all three fractions) as well as for each individual size fraction.

To compare community composition across the five regions and reefs therein, we employed multiple approaches. First, we conducted multivariate analyses (PERMANOVA) based on Jaccard distances in the *vegan* package (Oksanen, 2017) in R (R development core team) testing statistical significance using 9999 permutations and a significance level of $\alpha = 0.05$. We calculate the compositional dissimilarity using '*adonis*' command and the homogeneity of group dispersion using '*betadisper*' command in *vegan* package (Oksanen, 2017). Second, we conducted Principles Coordinates Analyses (PCoA) using the *Ampvis2* package (Andersen et al.,, 2018) with the ordination function of *phyloseq* for both Jaccard and Bray Curtis dissimilarity matrices and generate the ordination plot using *ggplot2* (Wickham, 2009).

To explore the most efficient way to sample and monitor marine biodiversity in Indonesia, we used a hierarchical approach, examining OTU richness on various scales. First, using the Venn diagram data, above, we examined the percentage of total regional diversity that can be captured in each of the three fractions. Next, we examined the degree to which the data from an individual size fraction from an individual ARMS captured total site diversity (e.g. total diversity captured from all size fractions from a set of 3 ARMS deployed on a given reef) and total regional diversity (e.g. total diversity captured from all size fractions from 4 sets of 3 ARMS deployed in each of 5 sampling regions). We then repeated these comparisons, but examining how total diversity (e.g. all size fractions) from a single ARMS captured total site and regional diversity.

Results

We recovered a total of 59 of 60 ARMS as one ARMS unit from Sumur Tiga in Aceh could not be recovered. From the 500 µm, 100 µm, and sessile fractions, we amplified a total of 174 samples with both COI and 18S rRNA primers. Due to PCR amplification failure, we obtained data from 58 ARMS units using COI and 59 units using 18S rRNA.

DNA sequencing returned 19,052,584 reads and 46,633,073 reads from COI and 18S rRNA, respectively. OTU clustering yielded a total of 12,330 OTUs for COI and 14,350 OTUs for 18S rRNA (Supplemental Table S3-2, S3-3). The mean number of OTUs per size fraction per ARMS unit was 71 for COI and 82 for 18S rRNA. While rarefaction curves for both COI and 18S approached the asymptote, curves for COI were flatter than 18S, showing that COI saturates OTU diversity faster (Figure1-2A, 1-2B). In addition, both COI and 18S showed less saturation in OTU diversity in high diversity regions of Eastern Indonesia, compared to lower diversity regions to the west. Although sequencing depth did not saturate at the regional scale, sequencing depth was sufficient to saturate in most individual ARMS units (Figure 3-2C, 1-2D), although less saturation was observed in high diversity regions compared to low diversity regions.

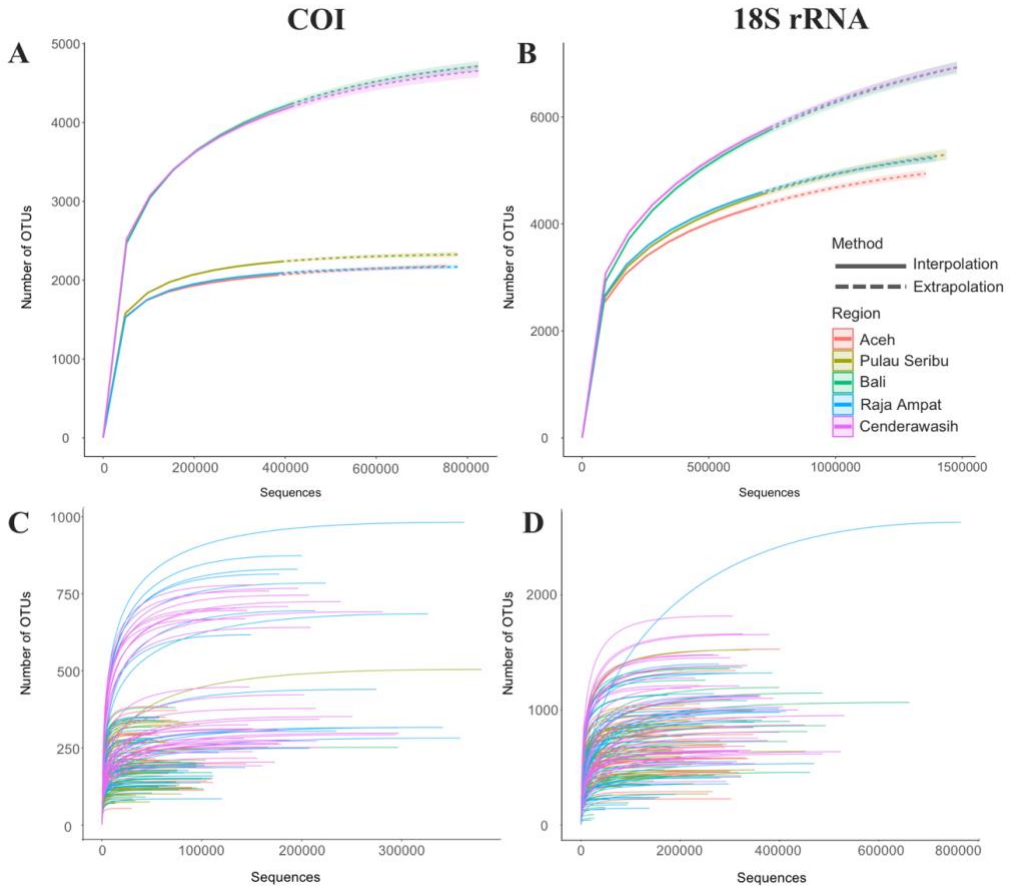


Figure 3-2. Rarefaction plot showing numbers of Operational Taxonomic Units (OTUs) as a function of sequencing depth for the five regions based on A) COI and B) 18S rRNA, as well as for each individual ARMS unit for C) COI and D) 18S rRNA.

Of the 12,330 OTUs obtained from COI, the 100 µm fraction contained 9,059 (73.5%) the 500 µm fraction contained 4,869 (39.5%) and the sessile fraction contained 3,125 (25.3%). Of the 14,350 OTUs identified from 18S rRNA, the 100 µm fraction contained 9,265 (64.6%), the 500 µm fraction contained 5,822 (40.6%) and the sessile fraction contained 7,083 (49.4%). Similarly, excluding all OTUs shared among the three size fractions, the 100 µm fraction of COI had the highest number of unique OTUs for COI (5,697; 46.2%) followed by the 500 µm fraction with 4,869 OTUs (14.0%) and the sessile fraction with 1,242 OTUs (10.1%;

Supplemental Table S3-2, Figure 3-3A). The 100 µm fraction from 18S also had the highest number of unique OTUs (4,216, 29.4%), although the number of unique OTUs in the sessile fraction (3,279; 22.9%) was higher than the 1,563 (10.9%) unique OTUs in the 500 µm (Supplemental Table S3-3) (Figure 3-3B). Measures of Alpha-diversity differed significantly among the size fractions (ANOVA-Chao1 = 70.21, *p-value* < 0.001 for COI; ANOVA-Chao1 = 61.05, *p-value* < 0.001 for 18S rRNA) (Supplemental Table S3-4). For both primer sets, the 100 µm size fraction had the highest mean richness based on Chao1 and Shannon indices (Figure 3-4).

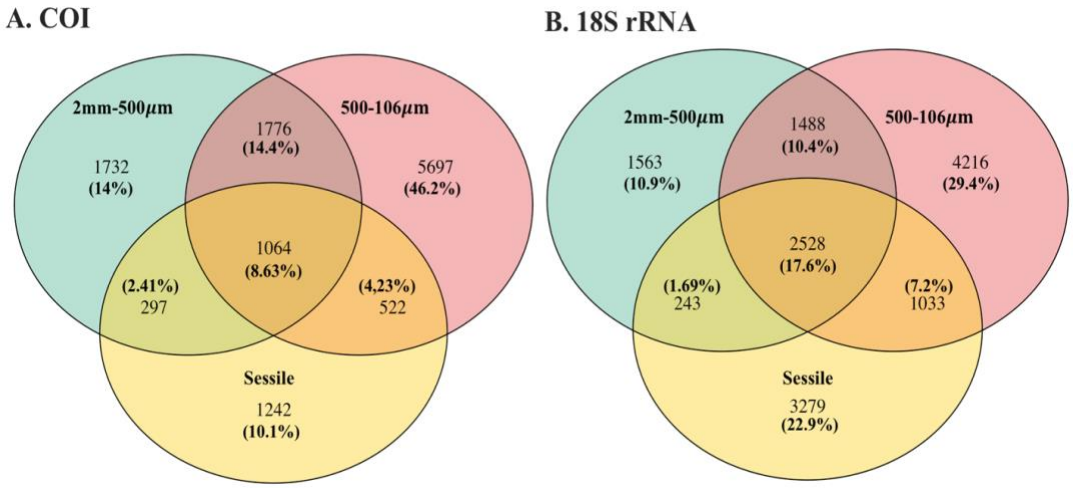


Figure 3-3. Venn diagram showing overlap in Operational Taxonomic Units (OTUs) among the 500µm, 100µm, and sessile fractions based on A) COI and B) 18S rRNA.

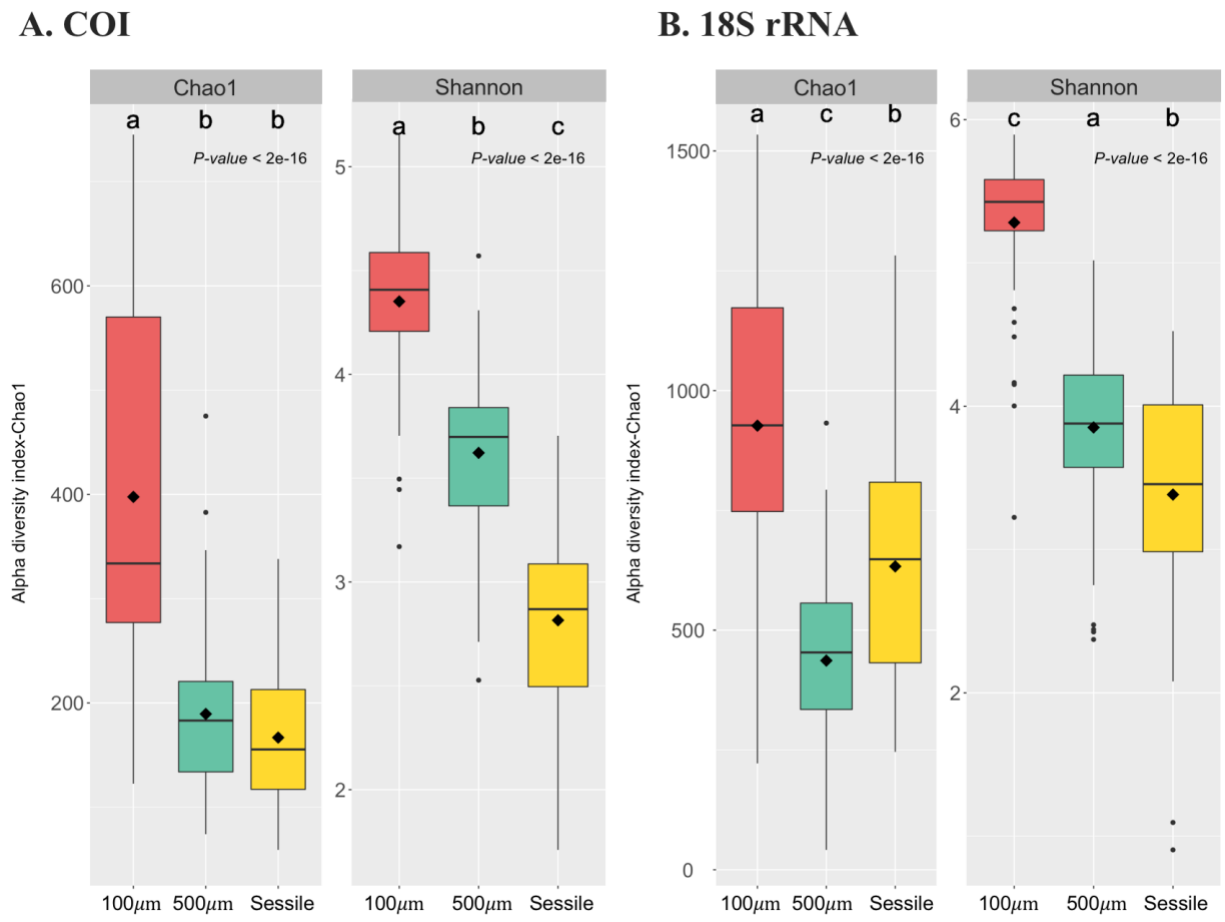


Figure 3-4. Boxplot of Chao1 and Shannon diversity indices of eukaryote communities based on A) COI and B) 18S rRNA summarized across each of three Autonomous Reef Monitoring Structure (ARMS) fractions. The diamond shapes represent the mean of alpha diversity from each fraction size. The box in the boxplot represent the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with the points outside considered outliers. The letters at the top of the box are the results of Tukey test of multiple comparisons.

Diversity and community composition across size fractions.

Because of variation in sequencing depth, we rarefied all samples to 11,452 sequences per size fraction per ARMS unit for COI and 20,549 sequences per sample for 18S rRNA, using this standardized dataset for all downstream applications (Supplemental Table S3-2, S3-3). One sample from Napoleon Reef, Bali was excluded from COI dataset because it has less than 10,000 reads.

The 500 µm and 100 µm motile fractions of the COI dataset were both dominated by *Arthropoda* (18% and 24%), *Rhodophyta* (6% and 10%), *Annelida* (5% and 8%), and unidentified taxa (55% and 40%). In contrast, the sessile fraction was dominated by *Porifera* (40%), *Rhodophyta* (30%), *Cnidaria* (10%), and unidentified species (11%) (Figure 3-5A). In the 18S rRNA dataset, both 500 µm and 100 µm fractions were dominated by *Annelida* (36% and 24%) *Arthropoda* (27% and 40%), *Porifera* (6% and 8%), and *Rhodophyta* (6% and 5%). As with COI, 18S rRNA showed that *Rhodophyta* (35%), *Porifera* (33%), and *Cnidaria* (13%) dominated the sessile fraction (Figure 3-5B).

COI revealed significant differences in beta diversity between three sample fractions (PERMANOVA, *p-value = 0.0001*), based on OTUs presence-absence (Jaccard), read abundance (Bray-Curtis), and phylogenetic distance (UniFrac) (Supplemental Table S3-5A). Homogeneity tests showed significant dispersion effects for Jaccard (*p-value* = 0.0001), Bray-Curtis (*p-value* = 0.0001), and UniFrac (*p-value* = 0.0001). Similarly, the 18S rRNA dataset revealed significant differences in beta diversity among sample fractions based on OTUs presence-absence (Jaccard), the abundance of reads (Bray-Curtis), and phylogenetic distance (UniFrac) (PERMANOVA, *p-value = 0.0001*). Homogeneity tests were non-significant indicating similar variance across size fractions (Supplemental Table S3-5B).

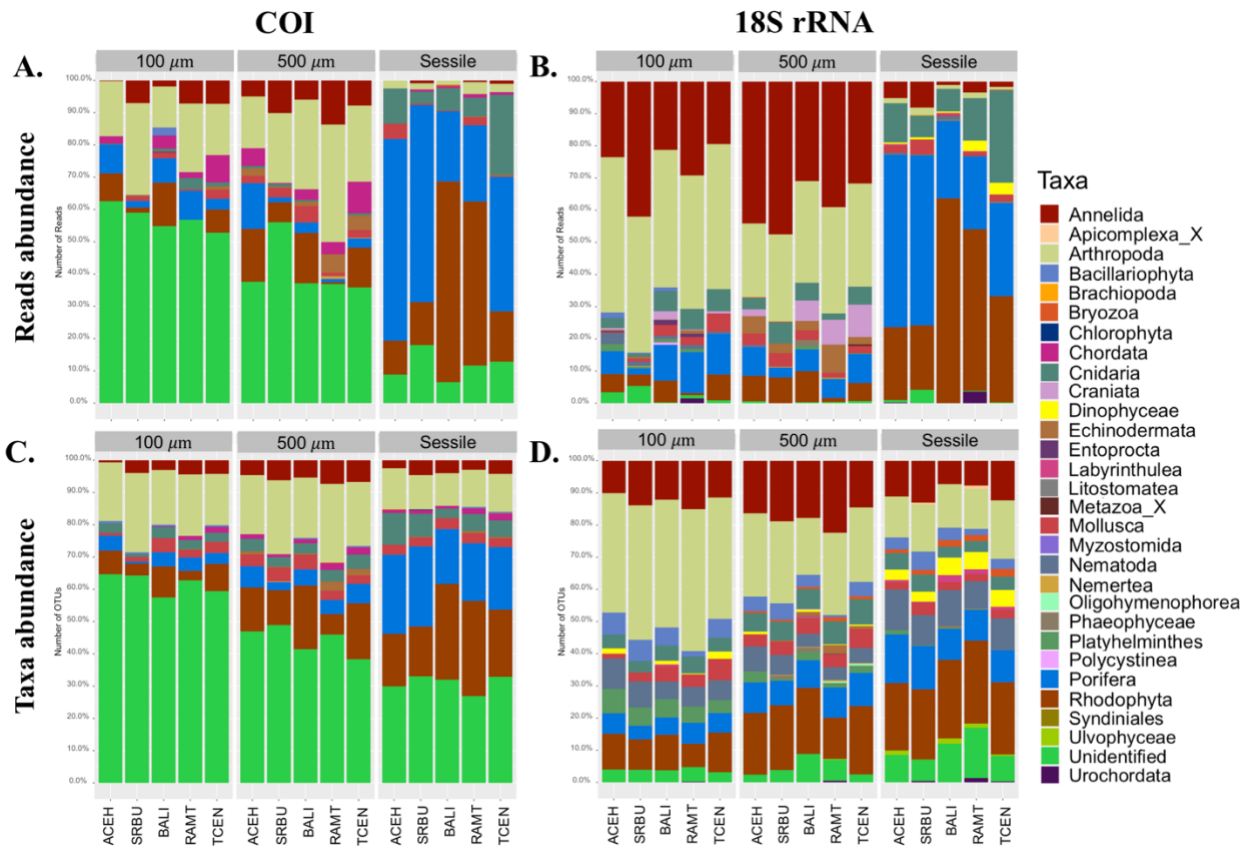


Figure 3-5. Taxonomic composition of eukaryote communities identified across 500μm, 100μm, and sessile sample fractions across five regions of Indonesia based on A) relative abundance of sequence reads and C) numbers of OTUs (Operational Taxonomic Units) based on COI, and B) relative abundance of sequence reads and D) numbers of OTUs based on 18S rRNA, excluding all taxa with <2% relative abundance.

Diversity and community composition across geography

In total, we recovered 2,063 OTUs from Aceh, 2,090 from Pulau Seribu, 2,236 from Bali, 4,198 from Raja Ampat, and 4,235 from Teluk Cenderawasih based on COI. Based on 18S rRNA there were 4,309 OTUs from Aceh, 4,577 from Pulau Seribu, 4,567 from Bali, 5,801 from Raja Ampat, and 5,751 from Teluk Cenderawasih. Examining only OTUs unique to each of the

five sampled region Raja Ampat had the highest number of unique OTUs with 3,204 COI and 2,356 18S rRNA OTUs, followed by Teluk Cenderawasih (3,204 COI; 2,356 18S rRNA) (Figure 3-6A). In contrast, Aceh (1,476 COI; 1,559 18S rRNA), Pulau Seribu (1,432 COI; 1,813 18S rRNA), and Bali (1,321 COI; 1,461 18S rRNA) all had many fewer unique OTUs (Figure 1.6B). Interestingly, only 96 COI OTUs (0.8%) were recovered in all five regions whereas 980 of 18S rRNA OTUs (7%) were shared among all 5 locations (Figure 3-6). Highest levels of taxon sharing were between Raja Ampat and Teluk Cenderawasih for both markers (293 COI; 503 18S rRNA). Lowest levels of sharing were between Raja Ampat and Aceh for COI (42), and Bali and Aceh (126) for 18S rRNA.

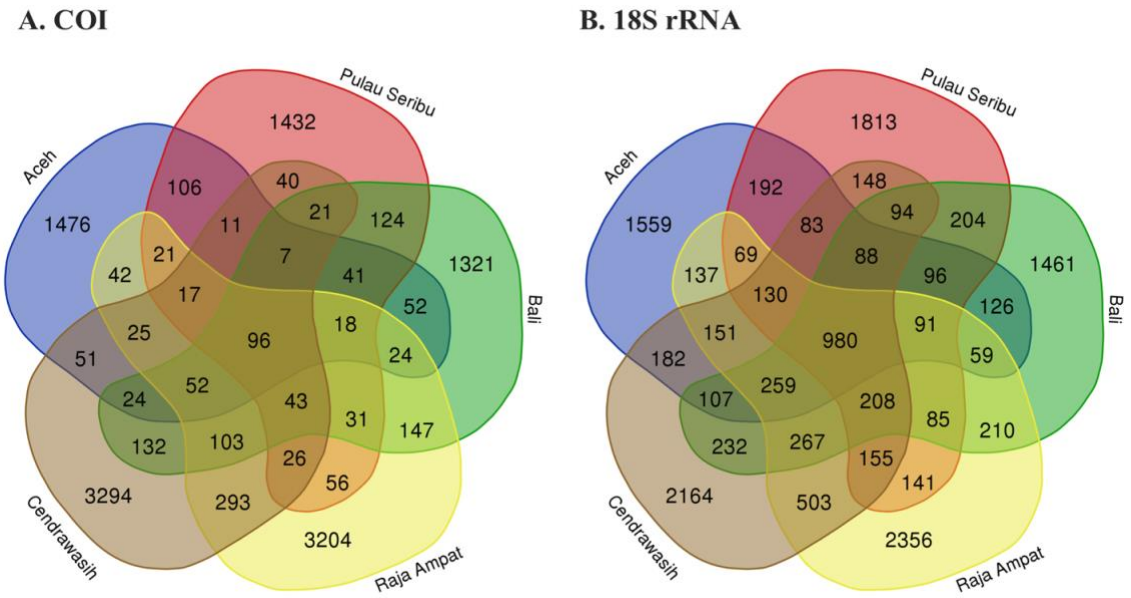


Figure 3-6. Numbers of unique and shared Operational Taxonomic Units (OTUs) among five sampled regions of Indonesia based on A) COI and B) 18S rRNA.

Alpha-diversity was significantly different between the five sampled regions for COI (One-way ANOVA, Chao1 = 11.83, *p-value* <0.001) and 18S rRNA (One-way ANOVA, Chao1 = 4.392, *p-value* = 0.0021) (Supplemental Table S3-4), although comparisons of the Shannon index were not significant for both markers. Richness (Chao1 and Shannon) of COI OTUs was highest in Teluk Cenderawasih (Figure 3-7A). While 18S rRNA OTU richness was highest in Teluk Cenderawasih based on Chao, Pulau Seribu had the highest richness based on the Shannon index (Figure 3-7B).

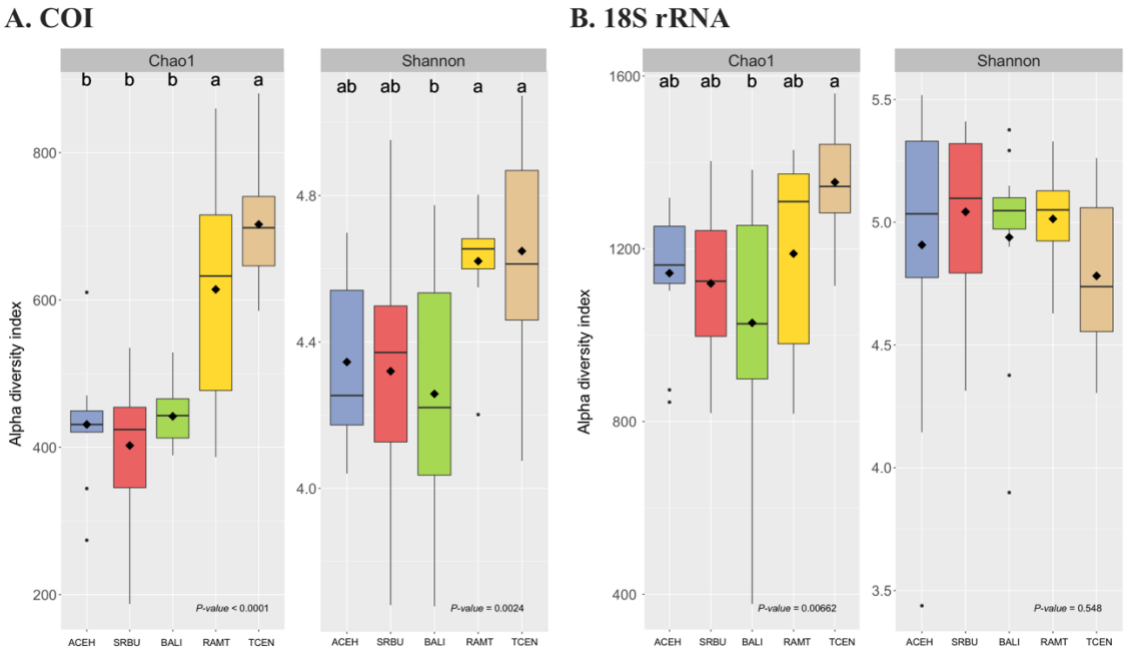


Figure 3-7. Boxplot showing the diversity indices (Chao1 and Shannon) of eukaryote community composition from 59 ARMS units across five regions of Indonesia. The diamond shapes represent the mean of alpha diversity from each location. The box in the boxplot represents the 1st and 3rd quartiles and the horizontal line is the median of the dataset. The whiskers (vertical line) extend from the end of the box to the minimum and maximum value, with point outside consider outliers. The letters at the top of the box are the results of Tukey test of multiple comparisons.

The PCoA plot for both COI and 18S rRNA dataset showed a clear pattern of clustering by location for both Bray-Curtis and Jaccard indices (Figure 3-8), with Pulau Seribu and Aceh clustering separately, and Bali, Raja Ampat and Teluk Cenderawasih clustered more closely together. Beta diversity for both datasets showed significant differences in eukaryote community composition across the five regions (PERMANOVA, p-value = 0.0001 for COI and 18S rRNA) (Supplemental Table S3-6). For COI dataset, the beta dispersion showed a significant difference across the five regions based on Jaccard (p-value = 0.0001) and Bray Curtis (p-value = 0.002) indices. Meanwhile, for 18S rRNA dataset, the beta dispersion p-value showed a significant difference across location based on Jaccard index (p-value = 0.0001) (Supplemental Table S3-6).

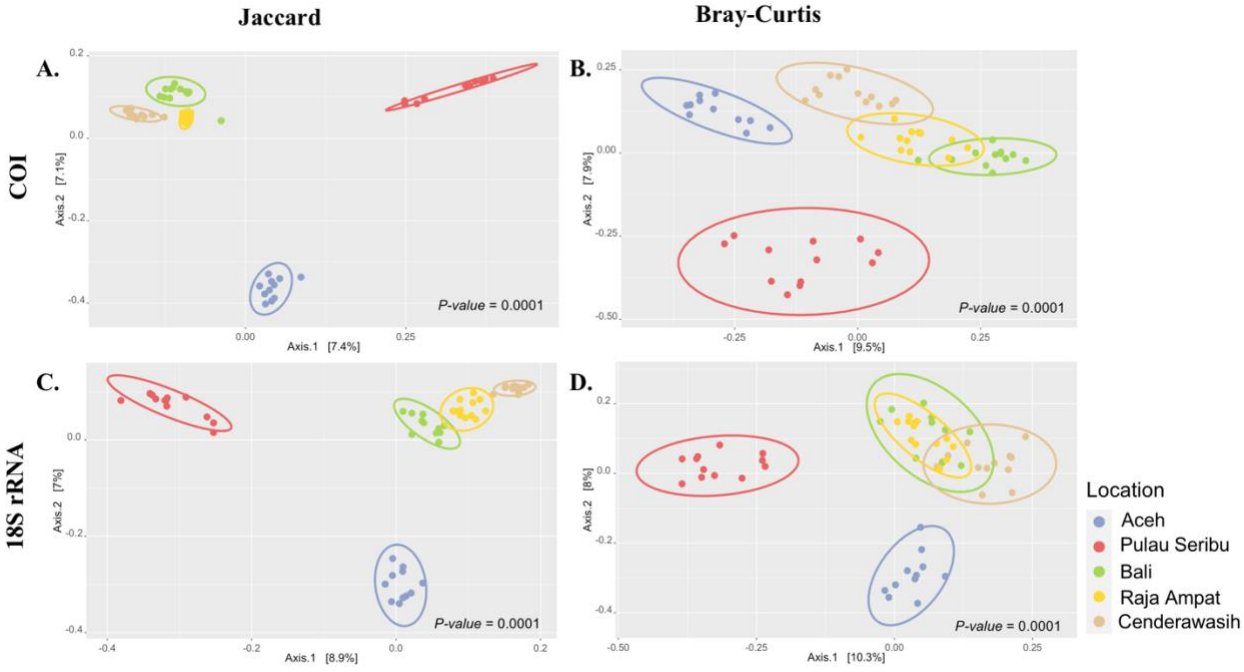


Figure 3-8. Principal Coordinates Analysis (PCoA) analysis illustrating dissimilarities in eukaryote community composition from 59 ARMS representing 5 regions of Indonesia. Analysis was undertaken using Jaccard and Bray-Curtis similarities on the full dataset of COI (A and B) and 18S rRNA (C and D) across different all sampling locations.

Detecting diversity across size fractions and diversity gradients

The 100 µm fraction of CO1 captured the highest percentage of regional OTU diversity, ranging from a low of 59.5% in Pulau Seribu to a high of 75.9% in Teluk Cenderawasih, averaging 67.2% of OTU diversity across all regions (Figure 3-9, Figure 3-10, Supplemental Table S3-7). Diversity in the 500 µm fraction was much less, ranging from a low of 34.8% of total OTU diversity in Teluk Cenderawasih to a high of 49.5% in Aceh, averaging 40.9% across all five regions. The sessile fraction had the lowest diversity, ranging from a low of 22.2% of total OTU diversity in Raja Ampat to a high of 37.7% in Pulau Seribu, averaging 29.6% across all five regions. Similar patterns were seen examining only OTUs unique to a region (Figure 3-9, Figure 3-10, Supplemental Table S3-8).

The 100 µm fraction of 18S rRNA also captured the highest percentage of regional OTU diversity, ranging from a low of 61.9% in Pulau Seribu to a high of 72.2% in Teluk Cenderawasih, averaging 66.2% of OTU diversity across all regions (Supplemental Table S3-7, Figure 3-9, Figure 3-10). Diversity in the 500 µm fraction was much less, ranging from a low of 31.5% of total OTU diversity in Teluk Cenderawasih to a high of 48.9% in Aceh, averaging 39.6% across all five regions. The sessile fraction had the lowest diversity, ranging from a low of 42.5% of total OTU diversity in Teluk Cenderawasih to a high of 53.5% in Pulau Seribu, averaging 46.5% across all five regions. Similar patterns were seen examining only OTUs unique to a region (Supplemental Table S3-8, Figure 3-9, Figure 3-10).

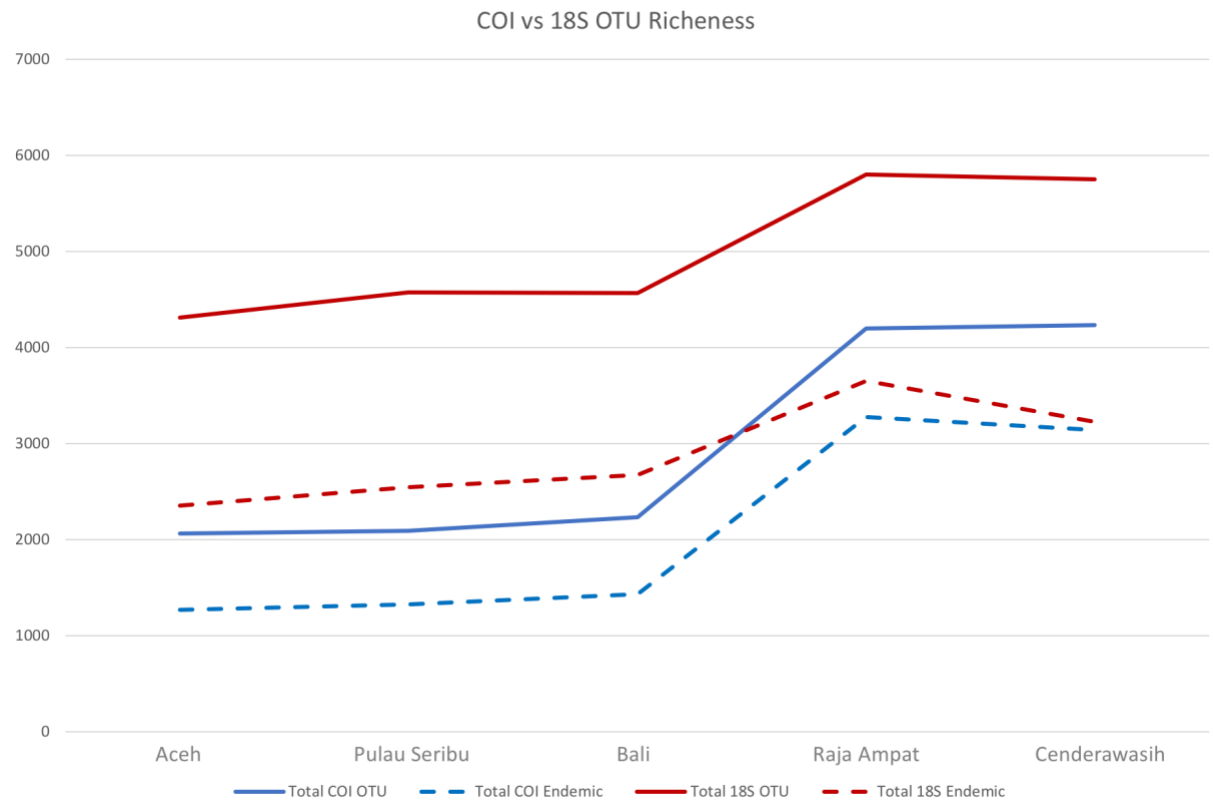


Figure 3-9. Operational Taxonomic Units (OTU) richness across five regions of Indonesia for 18S rRNA (red) and COI (blue). Solid lines are total OTU diversity, while dashed lines are OTUs that are unique to a single region.

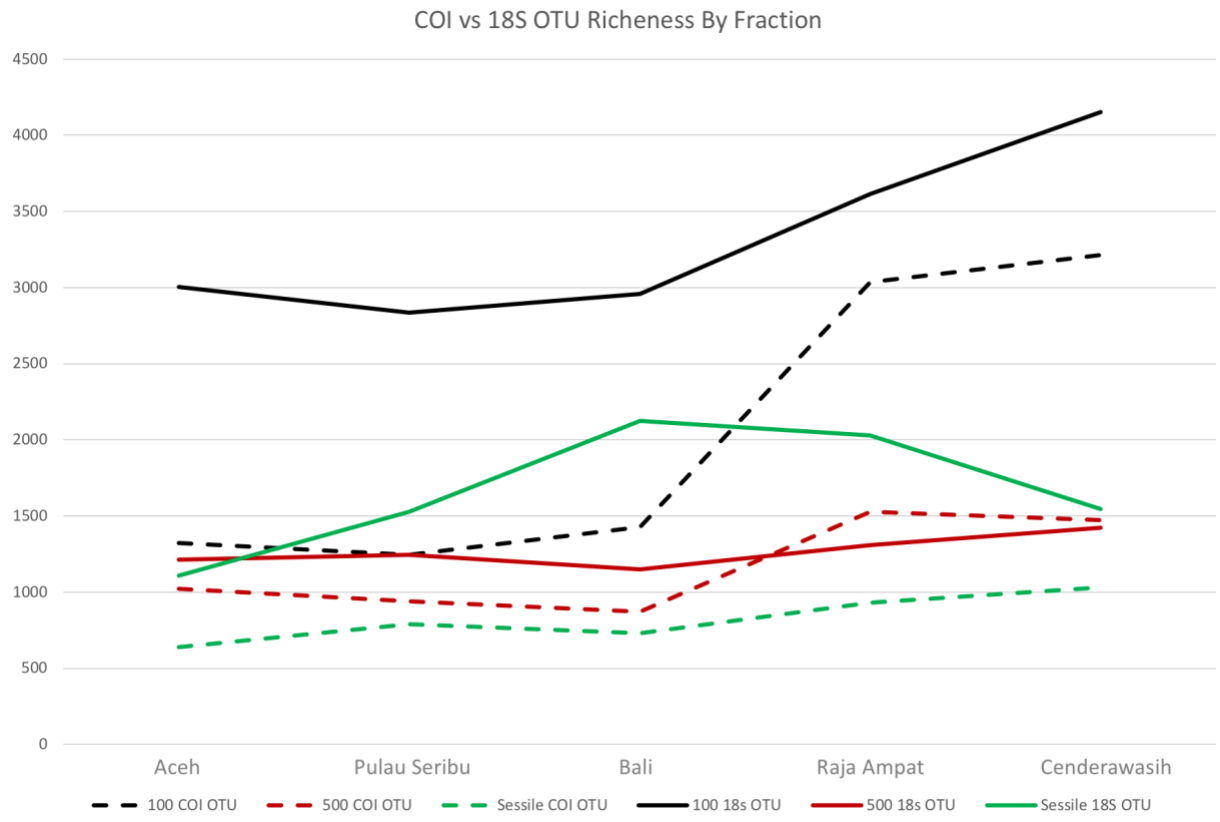


Figure 3-10. Operational Taxonomic Units (OTU) richness by fraction across five regions of Indonesia for 18S rRNA (solid line) and COI (dashed line).

In contrast, COI data from all three fractions of a standard 3 ARMS deployment captured a maximum of 45.8% of regional diversity in Aceh, to a minimum of 32.7% regional diversity in Raja Ampat, averaging 38.6% of total OTUs across all sites (Supplemental Table S3-9A). 18S rRNA performed slightly better with a minimum of 41.5% of total OTUs in Raja Ampat to a maximum of 47.9% of diversity in Aceh, averaging 45.1% of total OTU richness across all sites (Supplemental Table S3-9B). These values dropped by half for both markers when examining only OTUs unique to each of the five regions (data not shown).

For COI, a single ARMS unit captured a low of 46.5% of local diversity in Raja Ampat to a high of 52.4% in Aceh, averaging 49.1% across all five regions. However, a single ARMS

unit only captures a maximum of 22.1% of total diversity in Aceh and a low of 15.2% of total diversity in Raja Ampat, averaging 18.7% across all five regions (Supplemental Table S3-10A). Similar results were obtained examining only OTUs unique to each region averaging 49.1% of endemic regional diversity, although total diversity was lower, averaging 12.7% (data not shown). Results from 18S rRNA were similar with a single ARMS unit capturing a low of 51.3% of local diversity in Raja Ampat to a high of 58.3% in Aceh, averaging 54.3% across all five regions. However, a single ARMS unit only captures a maximum of 27.7% of total diversity in Aceh and a low of 21.3% of total diversity in Raja Ampat, averaging 24.5% across all five regions (Supplemental Table S3-10B). Similar results were obtained examining only OTUs unique to each region at the local, averaging 53.4% of endemic regional diversity, although total diversity was lower, averaging 15.7% (data not shown).

Discussion

Achieving sustainability of the threatened marine ecosystems in the Coral Triangle will require more effective monitoring approaches, such as DNA metabarcoding, yet previous metabarcoding studies in Indonesia failed to effectively capture marine biodiversity in the heart of this global marine biodiversity hotspot (Juhel et al., 2020; Marwayana et al., 2021). This study shows that metabarcoding of Autonomous Reef Monitoring Structures (ARMS) successfully captures patterns of biodiversity across the Indonesian archipelago. Total OTU (Operational Taxonomic Units) richness was highest in Eastern Indonesia and decreased moving towards the western margins of the Coral Triangle and to reefs in Western Indonesia outside the Coral Triangle. This result shows that biodiversity in lesser studied cryptofauna largely tracks well-studied biodiversity patterns reported from fish, molluscs, and corals (Roberts et al., 2002;

Bellwood & Meyer 2009; Veron et al., 2009), and that ARMS can be a useful monitoring tool, even in megadiverse regions of the ocean.

Data from 18S rRNA returned more total diversity than COI, although OTU recovery saturated slightly faster in COI than 18S rRNA, suggesting a trade off in maximizing numbers of OTUs recovered and precision of documenting community composition. For both markers, the 100 µm fraction captured the greatest total diversity, capturing an average of nearly 70% of total regional diversity. The 100 µm fraction of COI also captured up to 80% of unique regional biodiversity, whereas 18S rRNA captured a maximum of 70%. In comparison, sequencing all size fractions from a set of three ARMS units captured less than 50% of total OTU diversity for both COI and 18S rRNA. Combined, these results indicate that the most cost-effective approach to ecosystem monitoring would be achieved by metabarcoding COI from the 100 µm fraction, which would maximize precision and recovery of endemic diversity, while sacrificing only a small portion of total diversity.

Indonesian marine eukaryote diversity and composition

Central to any marine ecosystem monitoring approach is the ability for that approach to reflect actual biodiversity patterns. Coral cover and fish diversity/biomass are common metrics of coral reef health used in monitoring of coral reef ecosystems (Allen & Werner 2002; Burke et al., 2012; Cleary et al., 2014). Biodiversity patterns of these taxa are well known (Roberts et al., 2002, Bellwood and Meyer 2009; Veron et al., 2009), facilitating a direct comparison between visual monitoring efforts and local biodiversity. In contrast, metabarcoding approaches, particularly ARMS, capture biodiversity that is typically not included in traditional visual monitoring efforts. As such, it is important to establish that data from ARMS reflects local biodiversity. While it is not possible to directly compare ARMS to a visual census of the

thousands of taxa recovered in this study, it is instructive to compare diversity data from ARMS to the well-known biodiversity patterns, above.

Whether examining total OTU diversity, or the diversity of any of the three size fractions, there was a general pattern of increasing richness from Aceh and Pulau Seribu, lower diversity marine ecosystems in Western Indonesia that are outside of the Coral Triangle, to the reefs of Raja Ampat, and Teluk Cenderawasih Bay, regions of Eastern Indonesia known for their exceptional diversity of corals and fish (Turak & Souhoka 2003; Allen, 2008; Veron, 2009; Turak and DeVantier in press). Moreover, this pattern was seen across all size fractions with the exception of the 18S rRNA sessile fraction where diversity peaked in Bali. Importantly, ARMS metabarcoding captures patterns both in total biodiversity as well as endemic biodiversity. Raja Ampat and Teluk Cenderawasih are rich with endemic species (Allen, 2008), a pattern reflected in ARMS where unique OTUs were lowest in Western Indonesia and peaked in Eastern Indonesia. Combined, these results lend confidence that ARMS are capturing local diversity and that diversity of cryptofauna tracks diversity of larger metazoans like corals and fish.

In addition to recovering general patterns of eukaryotic marine biodiversity across Indonesia, PCA (Principal Component Analysis) results on genetic similarity indicate that eukaryote diversity was geographically clustered, with significantly different compositions across the five regions. Each of the five distinct clusters revealed by PCA correspond to distinct marine ecoregions as defined by (Spalding et al., 2007). While these ecoregions were defined including data like geomorphology, ocean currents, and ocean temperature, they were largely based on biogeographic data, including species range discontinuities. As such, biogeography would predict that all five regions should be distinct, as found in our analyses.

In addition to differential distribution of OTUs driving differentiation among the five sampled regions, genetic differentiation among populations of the same species could also contribute to this pattern. The exposure of the Sunda Shelf during Pleistocene sea-level fluctuations created a barrier to marine larval dispersal in Western Indonesia (Voris, 2000), resulting in phylogeographic differentiation in many marine taxa in this region, particularly Aceh (Kirkendale & Meyer, 2004; Barber & Erdmann 2006; Kochzius & Nuryanto, 2008; Carpenter et al., 2011; Vogler et al., 2012; Ackiss et al., 2013; Cheng et al., 2013; De Boer et al., 2014; Tornabene et al., 2015). Similarly, the Halmahera Eddy isolates Eastern Indonesia by limiting water and larval transport across the Maluku Sea (Barber et al 2006; Hoeksema 2007; Barber et al., 2011). And while populations in Teluk Cenderwasih can be genetically different from Raja Ampat (Barber et al., 2006, DeBoer et al., 2008), these populations can also be genetically similar (Ackiss et al., 2014; DeBoer et al., 2014), potentially explaining why Raja Ampat and Teluk Cenderawasih are distinct, but often overlapping in the PCA plots. Given that the genetic differentiation among these phylogeographic regions can exceed 3% in COI (e.g. Barber et al., 2006; Barber et al., 2011; DeBoer et al., 2014), it is possible that phylogeographic structure within taxa captured by ARMS contributes to the limited number OTUs shared between locations (especially in the COI dataset) and the prominent regional differentiation revealed by Jaccard and Bray-Curtis distances.

COI vs 18S rRNA

COI is a standard DNA barcoding marker for most animals groups, especially marine metazoans (Ivanova et al., 2007; Bucklin et al., 2011). It is commonly used because of the broad utility of the primers developed by Folmer et al. (1994) that amplifies a 658 base pair long barcode, as well as the early adoption of COI for DNA barcoding (Hebert et al., 2003). While

effective across a broad range of taxa, the smaller 313bp fragment targeted in next-generation sequencing based meta-barcoding (Leray et al., 2013) performs poorly in many taxa, particularly corals (Reimer et al., 2006; Aguilar & Reimer 2010; Grajales et al., 2007), gastropods (Meyer & Paulay 2005; Puillandre et al., 2009), and echinoderms (Clouse et al., 2005; Ward et al., 2008) among others. As such, Leray and Knowlton (2016a) recommend using 18S rRNA to capture eukaryotes broadly, especially protists (Lovejoy et al., 2007; Guillou et al., 2013; Chain et al., 2016), paired with a hypervariable barcoding marker such as COI for maximum taxonomic resolution among metazoans.

As expected, based on the design of these metabarcoding markers, 18S rRNA captured much more diversity (>20,000 OTUs) across Indonesia than did COI (~12,330 OTUs), yielding very different taxonomic composition. Overall, 18S rRNA captured almost all taxa identified in the COI dataset, except for *Chordata*. However, the 18S rRNA recovered additional taxa like *Dinophyceae, Nematoda,* and *Platyhelminthes,* each representing >2% abundance of OTUs on ARMS, while these taxa were absent from the COI dataset. There were also differences in the relative abundance of taxa recovered with both markers. For instance, *Annelida* made up 22% of the reads observed in the 18S rRNA dataset, but less than 9% of the total number of OTUs in the COI dataset. In contrast, COI recovered a higher abundance of *Porifera.* Overall, taxonomic assignments were better with 18S rRNA, with only 2% of OTUs unidentified at the phylum level, compared to 35% for COI. These results show how COI and 18S rRNA metabarcoding data are complementary, recovering the maximum number and diversity of marine eukaryotes, which is essential for a biodiversity hotspot like Indonesian coral reefs.

Maximizing monitoring, minimizing cost.

While combining COI and 18S rRNA metabarcoding data is ideal to recover the maximum diversity, the use of both these metabarcoding markers doubles lab and sequencing costs. Given that many Coral Triangle countries have relatively limited resources to invest in research, particularly Indonesia (Barber et al., 2014), it is essential to develop strategies that can maximize the ability to monitor high diversity marine ecosystems, while minimizing cost. Using only one marker reduces lab costs by nearly half. Choosing which marker to use, however, depends on the goal of the monitoring.

In terms of capturing the largest amount of diversity, 18S rRNA outperforms COI by more than 50%. Similarly, 18S rRNA yields far fewer unknown taxa. As such, 18S rRNA is likely the best choice for studies seeking to maximize the recovery of biodiversity and minimize unknowns. However, while 18S rRNA identifies more OTUs, it is only at higher taxonomic levels. Because 18S rRNA evolves much more slowly that COI, it lacks the precision of COI in discriminating among closely related taxa, a highly desirable feature for biodiversity monitoring, particularly in a global biodiversity hotspot. Indeed, the major reason that COI has such high numbers of unidentified taxa is the lack of reference barcodes in public databases, a point noted by Juhel et al. (2020) and Gold et al. (2021) for the mitochondrial 12S barcodes used for fish metabarcoding. In terms of capturing the maximum diversity of local endemics, which could be critical to local monitoring efforts, COI outperforms 18S rRNA. As such, COI is likely the marker of choice if taxonomic precision is required. Financial considerations also support the use of COI. Because 18S rRNA captures more taxa, it is necessary to sequence deeper to recover all of the taxa present, potentially reducing the cost savings of using a single metabarcoding marker. Moreover, because 18S rRNA recovers more protists than COI, the additional diversity

recovered by 18S rRNA may not add value to monitoring efforts, despite greater sequencing expense.

A second strategy for reducing costs, which is not mutually exclusive to using a single metabarcoding marker, is to only sequence a single size fraction, reducing lab costs by two thirds. The rationale for size fractioning ARMS samples is that larger organisms, particularly sessile taxa, have many orders of magnitude more biomass than small taxa. As such, if size fractionation is not employed, this higher biomass could swamp out the signal of smaller taxa during PCR and DNA sequencing. Such concerns are well-founded. A recent eDNA metabarcoding study (Gold et al., 2021) showed that high biomass of kelp forest fishes inside a marine protected area swamped out the signal of non-kelp forest taxa, whereas non-kelp forest taxa were recovered outside the MPA where biomass of kelp forest fishes was very low. Given that the sessile fraction had the least diversity in our study, and the 100 µm fraction had the most diversity, eliminating size fractionation could result in the sessile fraction swamping out the signal of the 100 µm fraction, resulting in the recovery of less total biodiversity.

Given that pooling all three size fractions is inadvisable, an alternate strategy would be to sequence only one of the three size fractions. Results show that the 100 µm fraction had the highest diversity, followed by the 500 µm fraction, with the sessile fraction having the least diversity. This pattern was largely consistent across both markers for total diversity as well as endemic diversity. One likely explanation for the high diversity of the 100 µm fraction is that smaller life stages, or smaller pieces of the larger 500 µm and sessile fractions are captured in the 100 µm faction. Indeed, excluding all taxa shared among the three size fractions reveals that 26-51% (mean 35%) of all diversity in the 100 µm was also seen in one or both of the larger size fractions. However, even discounting this potential carryover from larger size fractions, the 100

μm size fraction had the highest diversity, indicating the presence of many small and cryptic species. The fact that the 100 μm size fraction integrates over all three size fractions, makes it ideal for monitoring the largest amount of diversity while reducing the costs of sequencing a single size fraction.

Combined, the above suggests that an economical approach to monitoring diversity in megadiverse regions with limited research investment would be to use COI for metabarcoding of the 100 μm size fraction. This approach would reduce lab costs by over 80%, making biodiversity monitoring with metabarcoding more accessible in developing countries like Indonesia. However, this does not mean that there aren't other approaches to consider. For example, extracting DNA from each size fraction separately, quantifying these DNA concentrations, and then adjusting them equal molar concentrations could reduce the impacts of swamping during metabarcoding. Moreover, given that fieldwork to deploy and recover ARMS is a fixed cost, it would be advisable to keep the 500 μm and sessile fractions, even if they are not sequenced, so that they could be studied in the future as techniques become less expensive, or research investment increases.

Implications for biodiversity research and monitoring in Indonesia

Given that Indonesia has already lost a substantial percentage of its coral reefs, and the remainder are highly threatened (Burke et al., 2012), it is essential that the remaining reef ecosystems of this global biodiversity hotspot are managed sustainably, employing monitoring to ensure the success of these efforts. Metabarcoding of ARMS using COI and 18S rRNA markers identified a wide range of eukaryote taxa. While these taxa do not represent the taxa that are the typical focus of biomonitoring efforts (e.g. fish and corals), metabarcoding captures much more

diversity at much lower costs. It also provides these results without the need for specialized taxonomic training.

Another potential advantage of metabarcoding approaches is that by capturing the diversity of very small organisms, it may provide a more sensitive metric of change. Corals are very sensitive to temperatures above a thermal threshold, resulting in visually obvious bleaching (Ainsworth et al., 2016; Gardner et al., 2019). Similarly, eutrophication of reefs can lead to visually obvious phase shifts from coral to algal dominated reefs (Karcher et al., 2020). However, much of the environmental changes and stressors on reef environments may not result in such obvious changes. Moreover, by the time bleaching or phase shifts occur, it may be too late to mitigate environmental conditions. In contrast, protists are very sensitive to a wide range of abiotic stressors (Slaveykova et al., 2016). As such it is possible that they may be better measures of subtle environmental stressors than fish or corals.

The major drawback of metabarcoding as a monitoring tool is the numerous unidentified OTUs, particularly COI. This failure can result from the scarcity of COI sequences in reference databases, misidentifications within the database, or methodological artifacts (e.g. PCR and sequencing errors) or sequencing of pseudogenes (Leray & Knowlton 2015a; Cowart et al., 2015). The lack of comparative sequences is particularly acute in high biodiversity regions like Indonesian compared to well-studied, low diversity regions such as North America where DNA barcoding was first widely implemented (Hebert et al., 2003). Given the power of metabarcoding approaches to inform marine management (e.g. Gold et al., 2021), it is critical to develop better reference databases in mega-diverse regions like the Coral Triangle.

Molecular approaches such as metabarcoding will be a significant part of biodiversity assessment in the future, particularly with the expansion of standardized sampling protocols such

as ARMS and the expansion of reference databases. The results of this study show how ARMS and DNA metabarcoding can capture the diversity within the megadiverse reefs of the Coral Triangle, providing data that can help manage these critically important resources. However, metabarcoding should not be looked at as a wholesale replacement of traditional monitoring approaches. While metabarcoding data can tell us about changes in invertebrate communities in a reef ecosystem, more work will be required to show how these changes correspond to changes in the ecosystem that impacts its function, such as loss of live coral cover or phase shifts from coral to algal dominated reefs. By integrating metabarcoding with traditional ecological methods, we can expand our understanding of the diversity and function of coral reef ecosystems, positioning us better to manage them sustainably.

Supplemental Tables and Figures

Supplemental Table S3-1. Location and number of metabarcoding samples used on this study.

Country	Region/Reef	Abbr.	Latitude	Longitude	N
Indonesia	**Pulau Weh, Aceh**	**ACEH**			**33**
	Benteng	BTN	05° 50.774' N	095° 22.434' E	9
	Rubiah Sea Garden	RSG	05° 52.608' N	095° 15.596' E	9
	Seulako	SEU	05° 53.658' N	095° 15.176' E	9
	Sumur Tiga	STG	05° 53.370' N	095° 20.683' E	6
Indonesia	**Pulau Seribu, Jakarta**	**SRBU**			**34**
	Pulau Karang Beras	KBS	05° 45.574' S	106° 33.527' E	8
	Pulau Kotok	KOT	05° 41.575' S	106° 32.475' E	9
	Pulau Pramuka	PRM	05° 45.026' S	106° 36.311' E	8
	Pulau Sepa	SEP	05° 34.227' S	106° 34.491' E	9
Indonesia	**Pemuteran, Bali**	**BALI**			**35**
	Close Encouter	CEN	08° 7.675' S	114° 40.084' E	9
	Deep Middle Reef	DMR	08° 8.190' S	114° 39.570' E	9
	Horse Reef	HOR	08° 7.672' S	114° 39.337' E	9
	Napoleon Reef	NAP	08° 7.928' S	114° 40.531' E	8
Indonesia	**Raja Ampat, West Papua**	**RAMT**			**36**
	Kri	KRI	00° 33.284' S	130° 40.712' E	9
	Misool	PEF	02° 14.741' S	130° 33.438' E	9
	Pef	PEF	00° 26' S	130° 26' E	9
	Penemu	PNU	00° 34.664' S	130° 17.039' E	9
Indonesia	**Teluk Cenderawasih, West Papua**	**TCEN**			**36**
	Angromeos Island	ANG	02° 40.828' S	134° 49.515' E	9
	Manguar Cape	MGC	02° 52.866' S	134° 51.411' E	9
	Purup	PRP	02° 03.419' S	134° 09.585' E	9
	Tridacna Atoll	TRI	02° 29.948’S	134° 58.790’E	9

Supplemental Table S3-2. Total samples, sequence reads, and Operational Taxonomic Units (OTUs) for the COI datasets as revealed by DNA metabarcoding of Autonomous Reef Monitoring Structure (ARMS) from Indonesia.

COI metabarcoding dataset	All samples combined	100 um Fraction	500 um Fraction	Sessile
Total no. of individual samples	174	58	59	57
Total no. of individual ARMS	59	58	59	57
Total no. of individual Sites	20	20	20	20
Total no. of individual Location	5	5	5	5
Total no. of sequence reads	19,052,584	5,758,259	6,625,770	6,668,555
- Minimum no. of reads	8,826	8,826	16,711	13,344
- Maximum no. of reads	379,933	362,520	358,556	379,933
- Mean no. of reads	109,498	99,280	112,301	116,992
Total no. of OTUs	22,758	11,694	6,044	4,021
- Mean no. of OTUs	131	202	102	71
Total no. of individual samples	173	57	59	57
Rarefied even depth	11,452	11,452	11,452	11,452
Total rarefied no. of sequence reads	1,981,196	652,764	675,668	652,764
Total rarefied no. of OTUs	12,330	9,059	4,869	3,125
- Mean rarefied no. of OTUs	71	159	83	55

Supplemental Table S3-3. Total samples, sequence reads, and Operational Taxonomic Units (OTUs) for the 18S rRNA datasets as revealed by DNA metabarcoding of Autonomous Reef Monitoring Structure (ARMS) from Indonesia.

18S rRNA metabarcoding dataset	All samples combined	100 um Fraction	500 um Fraction	Sessile
Total no. of individual samples	174	58	59	57
Total no. of individual ARMS	59	58	59	57
Total no. of individual Sites	20	20	20	20
Total no. of individual Location	5	5	5	5
Total no. of sequence reads	46,633,073	13,327,050	12,547,558	20,758,465
- Minimum no. of reads	20,549	59,071	20,549	146,362
- Maximum no. of reads	764,712	400,884	375,973	764,712
- Mean no. of reads	268,006.2	229,776.7	212,670.5	364,183.6
Total no. of OTUs	36,956	13,779	8,522	15,645
- Mean no. of OTUs	212.4	237.6	144.4	274.5
Rarefied even depth	20,549	20,549	20,549	20,549
Total rarefied no. of sequence reads	3,575,526	1,191,842	1,212,391	1,171,293
Total rarefied no. of OTUs	14,350	9,265	5,822	7,083
- Mean rarefied no. of OTUs	82.5	159.7	98.7	124.3

Supplemental Table S3-4. Alpha diversity indices (ANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia using A) COI and B) 18S rRNA.

A. COI metabarcoding dataset

Indices	Factors	ANOVA							
		All data		100 µm Fraction		500 µm Fraction		Sessile	
		F-value	*p-value*	F-value	*p-value*	F-value	*p-value*	F-value	*p-value*
Observed	Region	8.301	**3.91E-06**	23.77	**3.25E-11**	4.751	**0.00234**	6.466	**0.000266**
	Site	1.777	**0.03**	4.212	**9.05E-05**	2.401	0.0102	3.61	**0.000408**
	Size fraction	101.2	**<2e-16**						
Chao1	Region	11.83	**1.73E-08**	25.48	**1.02E-11**	6.391	**0.000276**	10.63	**2.27E-06**
	Site	2.475	**0.00122**	4.4	**5.77E-05**	2.606	**0.0056**	3.785	**0.000261**
	Size fraction	70.21	**<2e-16**						
Shannon	Region	1.12	0.349	1.134	0.351	0.77	0.55	2.497	**0.0539**
	Site	0.579	0.917	0.836	0.655	1.368	0.199	2.722	**0.00447**
	Size fraction	192	**<2e-16**						

B. 18S rRNA metabarcoding dataset

Indices	Factors	ANOVA							
		All data		100 µm Fraction		500 µm Fraction		Sessile	
		F-value	*p-value*	F-value	*p-value*	F-value	*p-value*	F-value	*p-value*
Observed	Region	2.581	**0.0391**	5.363	**0.00106**	6.98	**0.000132**	2.25	0.0762
	Site	0.81	0.693	1.309	0.234	3.739	**0.000241**	4.394	**5.85E-05**
	Size fraction	92.86	**<2e-16**						
Chao1	Region	4.392	**0.0021**	7.19	**0.000106**	7.401	**7.91E-05**	2.845	**0.033**
	Site	1.166	0.294	1.596	0.108	3.234	**0.000942**	4.084	**0.000124**
	Size fraction	61.05	**<2e-16**						
Shannon	Region	0.466	0.761	0.362	0.834	2.264	0.0741	2.711	**0.0398**
	Site	0.636	0.874	1.426	0.172	2.515	**0.00729**	2.202	**0.0196**
	Size fraction	141.2	**<2e-16**						

Supplemental Table S3-5. Beta diversity summary (PERMANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia for A) COI and B) 18S rRNA.

A. COI dataset

Indices	Factors	Adonis		Betadispers		100 µm Fraction				500 µm Fraction				Sessile			
						Adonis		Betadisper		Adonis		Betadisper		Adonis		Betadisper	
		R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*
Jaccard	Region	0.12103	**1.00E-04**	16.665	**1.00E-04**	0.25978	**1.00E-04**	16.915	**1.00E-04**	0.19038	**1.00E-04**	8.5128	**1.00E-04**	0.20854	**1.00E-04**	8.4792	**2.00E-04**
	Site	0.2652	**1.00E-04**	3.2602	**2.00E-04**	0.58598	**1.00E-04**	2.4667	**0.0103**	0.4875	**1.00E-04**	3.1503	**0.002**	0.52849	**1.00E-04**	2.0111	**0.0375**
	Size fraction	0.04247	**1.00E-04**	86.492	**1.00E-04**												
Bray-Curtis	Region	0.11833	**1.00E-04**	8.6567	**1.00E-04**	0.30348	**1.00E-04**	10.409	**1.00E-04**	0.2199	**1.00E-04**	2.1827	0.086	0.24823	**1.00E-04**	1.5079	0.2194
	Site	2.9288	**1.00E-04**	2.6391	**7.00E-04**	0.65892	**1.00E-04**	1.9039	**0.0419**	0.54383	**1.00E-04**	1.6419	0.0933	0.60744	**1.00E-04**	0.6947	0.804
	Size fraction	0.08042	**1.00E-04**	83.308	**1.00E-04**												
UniFrac	Region	0.11627	**1.00E-04**	6.9621	**1.00E-04**	0.28269	**1.00E-04**	8.9327	**1.00E-04**	0.19361	**1.00E-04**	3.2364	**0.0184**	0.20319	**1.00E-04**	5.9358	**5.00E-04**
	Site	0.2467	**1.00E-04**	2.4032	0.0018	0.59949	**1.00E-04**	1.904	**0.0463**	0.48998	**1.00E-04**	3.2152	**5.00E-04**	0.51621	**1.00E-04**	1.3225	0.231
	Size fraction	0.09587	**1.00E-04**	23.483	**1.00E-04**												

B. 18S rRNA dataset

Indices	Factors	Adonis		Betadispers		100 µm Fraction				500 µm Fraction				Sessile			
						Adonis		Betadisper		Adonis		Betadisper		Adonis		Betadisper	
		R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*	R2	*p-value*	F	*p-value*
Jaccard	Region	0.11718	**1.00E-04**	18.895	**1.00E-04**	0.2572	**1.00E-04**	10.812	**2.00E-04**	0.17995	**1.00E-04**	11.93	**1.00E-04**	0.185	**1.00E-04**	2.3568	0.0615
	Site	0.24863	**1.00E-04**	5.1346	**1.00E-04**	0.5499	**1.00E-04**	2.1661	**0.0216**	0.47319	**1.00E-04**	4.7099	**1.00E-04**	0.49185	**1.00E-04**	2.4244	**0.0129**
	Size fraction	0.08208	**1.00E-04**	26.368	**1.00E-04**												
Bray-Curtis	Region	0.10711	**1.00E-04**	2.6927	**0.035**	0.33303	**1.00E-04**	4.265	**0.004**	0.21136	**1.00E-04**	0.6791	0.6182	0.24209	**1.00E-04**	3.0351	**0.0246**
	Site	0.23116	**1.00E-04**	0.7612	0.738	0.64575	**1.00E-04**	3.1199	**0.001**	0.51127	**1.00E-04**	2.0382	**0.0251**	0.5864	**1.00E-04**	0.8912	0.5992
	Size fraction	0.19603	**1.00E-04**	1.2511	0.2905												
UniFrac	Region	0.06103	**1.00E-04**	14.426	**1.00E-04**	0.15273	**1.00E-04**	7.6599	**2.00E-04**	0.11342	**1.00E-04**	13.297	**1.00E-04**	0.10783	**1.00E-04**	2.907	**0.0291**
	Site	0.1646	**1.00E-04**	2.9341	**1.00E-04**	0.42998	**1.00E-04**	2.668	**0.0051**	0.38515	**1.00E-04**	3.4336	**6.00E-04**	0.38236	**1.00E-04**	1.4384	0.1527
	Size fraction	0.0393	**1.00E-04**	10.931	**2.00E-04**												

Supplemental Table S3-6. Beta diversity indices (PERMANOVA) of eukaryote diversity across five regions (Aceh, Pulau Seribu, Bali, Raja Ampat and Teluk Cenderawasih) in Indonesia. Analysis were calculated from total diversity per ARMS (e.g. summing diversity of all 3-set of fraction per ARMS.

Indices	**Factors**	**COI**				**18S rRNA**			
		Adonis		**Betadisper**		**Adonis**		**Betadisper**	
		R2	***p-value***	**F**	***p-value***	**R2**	***p-value***	**F**	***p-value***
Jaccard	**Region**	0.25639	**1.00E-04**	15.115	**1.00E-04**	0.24919	**1.00E-04**	7.6072	**1.00E-04**
	Site	0.57836	**1.00E-04**	1.4549	0.148	0.54774	**1.00E-04**	3.1812	**0.0014**
Bray-Curtis	**Region**	0.28044	**1.00E-04**	5.055	**0.002**	0.27728	**1.00E-04**	1.9632	0.1129
	Site	0.62667	**1.00E-04**	1.3239	0.2282	0.60414	**1.00E-04**	1.7766	0.0635

Supplemental Table S3-7. Operational Taxonomic Units (OTU) richness across from five regions of Indonesia from A) COI and B) 18S rRNA metabarcoding data, including total OTU diversity, diversity of the 100 µm fraction, 500 µm fraction, and sessile fractions.

A. COI metabarcoding dataset

Region	Total OTU	100 µm OTUs	100 µm % Total	500 µm OTUs	500 µm % Total	Sessile OTUs	Sessile % Total
Aceh	2063	1325	64.20%	1021	49.50%	639	31.00%
Pulau Seribu	2090	1244	59.50%	939	44.90%	788	37.70%
Bali	2236	1428	63.90%	873	39.00%	729	32.60%
Raja Ampat	4198	3036	72.30%	1528	36.40%	931	22.20%
Teluk Cendrawasih	4235	3215	75.90%	1473	34.80%	1033	24.40%
Average			67.20%		40.90%		29.60%

B. 18S rRNA metabarcoding dataset

Region	Total OTU	100 µm OTUs	100 µm % Total	500 µm OTUs	500 µm % Total	Sessile OTUs	Sessile % Total
Aceh	4309	3005	69.70%	2107	48.90%	1753	40.70%
Pulau Seribu	4577	2835	61.90%	2002	43.70%	2449	53.50%
Bali	4567	2959	64.80%	1628	35.60%	2127	46.60%
Raja Ampat	5801	3613	62.30%	1829	31.50%	2867	49.40%
Teluk Cendrawasih	5751	4151	72.20%	2184	38.00%	2447	42.50%
Average			66.20%		39.60%		46.50%

Supplemental Table S3-8. Unique Operational Taxonomic Units (OTU) richness across from five regions of Indonesia from A) COI and B) 18S rRNA metabarcoding data, including total unique OTU diversity, unique diversity of the 100 µm fraction, 500 µm fraction, and sessile fractions.

A. COI metabarcoding dataset

Region	Total unique OTU	Unique OTUs % Total	100 µm OTUs	100 µm % Total	500 µm OTUs	500 µm % Total	Sessile OTUs	Sessile % Total
Aceh	1268	61.50%	852	67.20%	700	55.20%	442	34.90%
Pulau Seribu	1329	63.60%	808	60.80%	680	51.20%	549	41.30%
Bali	1432	64.00%	955	66.70%	625	43.60%	729	50.90%
Raja Ampat	3276	78.00%	2460	75.10%	1277	39.00%	713	21.80%
Teluk Cendrawasih	3144	74.20%	2523	80.20%	1135	36.10%	692	22.00%
Average		**68.30%**		**70.00%**		**45.00%**		**34.20%**

B. 18S rRNA metabarcoding dataset

Region	Total unique OTU	Unique OTUs % Total	100 µm OTUs	100 µm % Total	500 µm OTUs	500 µm % Total	Sessile OTUs	Sessile % Total
Aceh	2358	54.70%	1579	67.00%	1213	51.40%	1108	47.00%
Pulau Seribu	2548	55.70%	1514	59.40%	1244	48.80%	1529	60.00%
Bali	2676	58.60%	1634	61.10%	1148	42.90%	2127	79.50%
Raja Ampat	3651	62.90%	2178	59.70%	1310	35.90%	2029	55.60%
Teluk Cendrawasih	3228	56.10%	2260	70.00%	1421	44.00%	1547	47.90%
Average		**57.60%**		**63.40%**		**44.60%**		**58.00%**

Supplemental Table S3-9. Operational Taxonomic Units (OTU) diversity captured in a set of three ARMS based on A) COI and B) 18S rRNA metabarcoding.

A. COI metabarcoding dataset

Region	Maximum OTUs per 3-ARMS Set	Minimum OTUs per 3-ARMS Set	Average OTUs per 3-ARMS Set	S.D.	% Regional OTU Diversity
Aceh	920	721	945.6	70.7	45.80%
Pulau Serbu	854	760	826.3	40.2	39.50%
Bali	987	836	887	64.2	39.70%
Raja Ampat	1485	1228	1372.5	103.1	32.70%
Teluk Cendrawasih	1853	1286	1494.5	227.1	35.30%

B. 18S rRNA metabarcoding dataset

Region	Maximum OTUs per 3-ARMS Set	Minimum OTUs per 3-ARMS Set	Average OTUs per 3-ARMS Set	S.D.	% Regional OTU Diversity
Aceh	2212	1555	2062.3	252	47.90%
Pulau Seribu	2265	2001	2084.8	111.2	45.50%
Bali	2350	1789	2037	209.9	44.60%
Raja Ampat	2592	2180	2409.3	176.1	41.50%
Teluk Cendrawasih	2831	2356	2653.5	186.5	46.10%

Supplemental Table S3-10. Operational Taxonomic Units (OTU) diversity captured in a single ARMS based on A) COI and B) 18S rRNA metabarcoding.

A. COI metabarcoding dataset

Region	Total OTUs per location	Average total OTUs per ARMS	S.D.	Average local diversity in 1 ARMS	% Regional OTU Diversity
Aceh	2063	438.8	77.8	52.4%	21.3%
Pulau Seribu	2090	410.3	102.9	49.7%	19.6%
Bali	2236	422.3	103.1	47.5%	18.9%
Raja Ampat	4198	639.2	154.6	46.5%	15.2%
Teluk Cendrawasih	4235	732.3	78.4	49.4%	17.3%
Average	2964.4	528.6	103.4	49.1%	18.5%

B. 18S rRNA metabarcoding dataset

Region	Total OTUs per location	Average total OTUs per ARMS	S.D.	Average local diversity in 1 ARMS	% Regional OTU Diversity
Aceh	4309	1192	175.7	58.3%	27.7%
Pulau Seribu	4577	1156.8	191	55.5%	25.3%
Bali	4567	1066.3	313.1	52.1%	23.3%
Raja Ampat	5801	1234.4	205.3	51.3%	21.3%
Teluk Cendrawasih	5751	144.5	150.3	54.5%	25.1%
Average	5001.0	958.8	207.1	54.3%	24.5%

References

Ackiss, A. S., Pardede, S., Crandall, E. D., Ablan-Lagman, M. C. a, Ambariyanto, Romena, N., Barber, P. & Carpenter, K. E. (2013). Pronounced genetic structure in a highly mobile coral reef fish, Caesio cuning, in the Coral Triangle. *Marine Ecology Progress Series*, *480*, 185–197. https://doi.org/10.3354/meps10199

Aguilar, C., & Reimer, J. D. (2010). Molecular phylogenetic hypotheses of zoanthus species (Anthozoa:Hexacorallia) using RNA secondary structure of the internal transcribed spacer 2 (ITS2). *Marine Biodiversity*, *40*(3), 195–204. https://doi.org/10.1007/s12526-010-0043-2

Ainsworth, T. D., Heron, S. F., Ortiz, J. C., Mumby, P. J., Grech, A., Ogawa, D., Eakin, C. M. & Leggat, W. (2016). Climate change disables coral bleaching protection on the Great Barrier Reef. *Science*, *352*(6283), 338–342. https://doi.org/10.1126/science.aac7125

Al-Rshaidat, M. M. D., Snider, A., Rosebraugh, S., Devine, A. M., Devine, T. D., Plaisance, L., Knowlton, N. & Leray, M. (2016). Deep COI sequencing of standardized benthic samples unveils overlooked diversity of Jordanian coral reefs in the northern Red Sea. *Genome*, 59, 724–737.

Allen, G. R. (2008). Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conserv: Mar. Freshw. Ecosyst.* 18: 541–556. https://doi.org/10.1002/aqc

Allen, G. R., & Werner, T. B. (2002). Coral reef fish assessment in the ‘ coral triangle ’ of southeastern Asia. *Environmental Biology of Fishes*, *65*, 209–214.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*. https://doi.org/10.1016/S0022-2836(05)80360-2

Andersen, K. S., Kirkegaard, R. H., Karst, S. M.& Albertsen, M (2018). “ampvis2: an R package to analyse and visualise 16S rRNA amplicon data.” *bioRxiv*. http://dx.plos.org/10.1371/journal.pone.0132783

Barber, P. H., Ablan-lagman, M. C. A., Cahyani, D., Crandall, E. D., Ravago-gotanco, R., Juinio-meñez, M. A., Mahardika, I.G.N., Shanker, K., Starger, C.J., Toha, A.H.A., Anggoro, A.W. and Willette, D.A. (2014). Advancing biodiversity research in developing countries : the need for changing paradigms. *Bulletin of Marine Science*. *90*(1). 2014 Rosenstiel School of Marine & Atmospheric Science of the University of Miami

Barber, P. H., & Erdmann, M. V. (2006). Comparative Phylogeography of three Codistributed Stomatopods: Origins and Timing of Regional Lineage Diversification in the Coral Triangle, *60*(9), 1825–1839.

Barber, P. H., Cheng, S. H., Erdmann, M. V., Tengardjaja, K. & Ambariyanto (2011) Evolution and conservation of marine biodiversity in the Coral Triangle: insights from stomatopod Crustacea. In: Crustacean Issues 19 Phylogeography and Population Genetics in Crustacea (eds Held C, Koenemann S and Schubart CD), pp. 129–156; CRC Press, Boca Raton.

Bellwood, D. R. (2001). Regional-Scale Assembly Rules and Biodiversity of Coral Reefs. *Science*, *292*(5521), 1532–1535. https://doi.org/10.1126/science.1058635

Bellwood, David R., & Meyer, C. P. (2009). Searching for heat in a marine biodiversity hotspot. *Journal of Biogeography*, *36*(4), 569–576. https://doi.org/10.1111/j.1365-2699.2008.02029.x

Bellwood, David R, Hughes, T. P., Connolly, S. R., & Tanner, J. (2005). Environmental and geometric constraints on Indo-Pacific coral reef biodiversity. *Ecology Letters*, *8*, 643–651. https://doi.org/10.1111/j.1461-0248.2005.00763.x

Bouchet, P., Lozouet, P., & Maestrati, P. (2002). Assessing the magnitude of species richness in tropical marine environments : exceptionally high numbers of molluscs at a New Caledonia site. *Biological Journal of the Linnean Society*, 75, 421–436.

Brandl, S. J., Tornabene, L., Goatley, C. H. R., Casey, J. M., Morais, R. A., Côté, I. M., Baldwin, C. C., Parravicini, V., Schiettekatte, N. M. D. & Bellwood, D. R. (2019). Demographic dynamics of the smallest marine vertebrates fuel coral reef ecosystem functioning. *Science* 364:1189–1192. https://doi.org/10.1126/science.aav3384

Brandt, M. I., Trouche, B., Quintric, L., Wincker, P., Poulain, J., & Arnaud-Haond, S. (2020). A flexible pipeline combining clustering and correction tools for prokaryotic and eukaryotic metabarcoding. *Peer Community In Ecology*, 100043. https://doi.org/10.24072/pci.ecology.100043

Bucklin, A., Steinke, D., & Blanco-bercial, L. (2011). DNA Barcoding of Marine Metazoa. *Annu. Rev. Mar. Sci*, *3*, 471–508. https://doi.org/10.1146/annurev-marine-120308-080950

Burke, L., Reytar, K., Spalding, M., & Perry, A. (2012). *Reefs at risk, Revisited In the Coral Triangle*. *National Geographic*. https://doi.org/10.1016/0022-0981(79)90136-9

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*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fiere, N. & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, *108 Suppl*(Supplement_1), 4516–4522. https://doi.org/10.1073/pnas.1000080107

Carpenter, K. E., Barber, P. H., Crandall, E. D., Ablan-Lagman, M. C. A., Ambariyanto, Mahardika, G. N., Manjaji-Matsumoto, B. M., Juinio-Meñez, M. A., Santos, M. D., Starger, C.J. & Toha, A. H. A. (2011). Comparative Phylogeography of the Coral Triangle and Implications for Marine Management. *Journal of Marine Biology*, *2011*, 1–14. https://doi.org/10.1155/2011/396982

Carugati, L., Corinaldesi, C., Dell'Anno, A., & Danovaro, R. (2015). Metagenetic tools for the census of marine meiofaunal biodiversity: An overview. *Marine Genomics*, *24*, 11–20. https://doi.org/10.1016/j.margen.2015.04.010

Chain, F. J. J., Brown, E. A., Macisaac, H. J., & Cristescu, M. E. (2016). Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. *Diversity and Distributions*, *22*(5), 493–504. https://doi.org/10.1111/ddi.12427

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*. https://doi.org/10.1890/13-0133.1

Chen, H., & Boutros, P. C. (2011). VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-12-35

Cheng, S. H., Anderson, F. E., Bergman, A., Mahardika, G. N., Muchlisin, Z. A., Dang, B. T., Calumpong, H. P., Mohamed, K. S., Sasikumar, G., Venkatesan, V. & Barber, P. H. (2013). Molecular evidence for co-occurring cryptic lineages within the Sepioteuthis cf . lessoniana species complex in the Indian and Indo-West Pacific Oceans. https://doi.org/10.1007/s10750-013-1778-0

Cleary, D. F. R., Polonia, A. R. M., Renema, W., Hoeksema, B. W., Wolstenholme, J., Tuti, Y., & De Voogd, N. J. (2014). Coral reefs next to a major conurbation: A study of temporal change (1985-2011) in Coral cover and composition in the reefs of Jakarta, Indonesia. *Marine Ecology Progress Series*, *501*, 89–98. https://doi.org/10.3354/meps10678

Clouse, R., Janies, D., & Kerr, A. M. (2005). Resurrection of Bohadschia bivittata from B. marmorata (Holothuroidea: Holothuriidae) based on behavioral, morphological, and mitochondrial DNA evidence. *Zoology*. https://doi.org/10.1016/j.zool.2004.07.007

Counsell, C. W. W., Donahue, M. J., Edwards, K. F., Franklin, E. C., & Hixon, M. A. (2018). Variation in coral-associated cryptofaunal communities across spatial scales and environmental gradients. *Coral Reefs*. https://doi.org/10.1007/s00338-018-1709-7

Cowart, D. A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., & Arnaud-Haond, S. (2015). Metabarcoding is powerful yet still blind: A comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS ONE*.

Cruz-Trinidad, A., Aliño, P. M., Geronimo, R. C., & Cabral, R. B. (2014). Linking Food Security with Coral Reefs and Fisheries in the Coral Triangle. *Coastal Management, 42:2,* 160-182. https://doi.org/10.1080/08920753.2014.877761

Curd, E. E., Gold, Z., Kandlikar, G. S., Gomer, J., Ogden, M., O'Connell, T., Pipes, L., Schweizer, T., Rabichow, L., Lin, M., Shi, B., Barber, P., Kraft, N., Wayne, R. & Meyer, R. S. (2018). Anacapa Toolkit: an environmental DNA toolkit for processing multilocus metabarcode datasets. *BioRxiv*, 488627. https://doi.org/10.1101/488627

De Boer, T. S., Naguit, M. R. A., Erdmann, M. V., Ablan-Lagman, M. C. A., Ambariyanto, A., Carpenter, K. E., Toha, A. H. A. & Barber, P. H. (2014). Concordance between phylogeographic and biogeographic boundaries in the Coral Triangle: Conservation implications based on comparative analyses of multiple giant clam species. *Bulletin of Marine Science*, *90*(1). https://doi.org/10.5343/bms.2013.1003

Enochs, I. C., & Manzello, D. P. (2012). Responses of cryptofaunal species richness and trophic potential to coral reef habitat degradation. *Diversity*. https://doi.org/10.3390/d4010094

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C. M., Boyer, F., Pompanon, F., Rayé, G. & Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.12338

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7881515

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01312-x

Gardner, S. G., Camp, E. F., Smith, D. J., Kahlke, T., Osman, E. O., Gendron, G., Hume, B. C. C., Pogoreutz, C., Voolstra, C. R. & Suggett, D. J. (2019). Coral microbiome diversity reflects mass coral bleaching susceptibility during the 2016 El Niño heat wave. *Ecology and Evolution*, *9*(3), 938–956. https://doi.org/10.1002/ece3.4662

Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, *13*(5), 851–861. https://doi.org/10.1111/1755-0998.12138

Giebner, H., Langen, K., Bourlat, S. J., Kukowka, S., Mayer, C., Astrin, J. J., Misof, B. & Fonseca, V. G. (2020). Comparing diversity levels in environmental samples: DNA sequence capture and metabarcoding approaches using 18S and COI genes. *Molecular*

Ecology Resources. https://doi.org/10.1111/1755-0998.13201

Gold, Z., Sprague, J., Kushner, D. J., Zerecero Marin, E. & Barber P. H. (*in press*) eDNA metabarcoding as a biomonitoring tool for marine protected areas. *PLOS One.*

Grajales, A., Aguilar, C., & Sánchez, J. A. (2007). Phylogenetic reconstruction using secondary structures of Internal Transcribed Spacer 2 (ITS2, rDNA): Finding the molecular and morphological gap in Caribbean gorgonian corals. *BMC Evolutionary Biology*, *7*, 1–9. https://doi.org/10.1186/1471-2148-7-90

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., Del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A. L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P. & Christen, R. (2013). The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*(D1). https://doi.org/10.1093/nar/gks1160

Guo, L., Sui, Z., Zhang, S., Ren, Y., & Liu, Y. (2015). Comparison of potential diatom 'barcode' genes (The 18S rRNA gene and ITS, COI, rbcL) and their effectiveness in discriminating and determining species taxonomy in the Bacillariophyta. *International Journal of Systematic and Evolutionary Microbiology*. https://doi.org/10.1099/ijs.0.000076

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., & Troedsson, C. (2014). Characterization of the 18s rRNA gene for designing universal eukaryote specific primers. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0087624

Hebert, P. D. N., Ratnasingham, S., Waard, J. R. De, B, P. R. S. L., & Jeremy, R. (2003). Barcoding animal life : cytochrome c oxidase subunit 1 divergences among closely related species Barcoding animal life : cytochrome c oxidase subunit 1 divergences among closely related species. https://doi.org/10.1098/rsbl.2003.0025

Hoegh-Guldberg, O. (2011). The Impact of Climate Change on Coral Reef Ecosystems. In Z. Dubinsky & N. Stambler (Eds.), *Coral Reefs: An Ecosystem in Transition*. Springer.

Hoegh-Guldberg, O., Hoegh-Guldber, H., Veron, J. E. (Charlie), Green, A., Gomez, E. D., Lough, J., King, M., Ambariyanto, Hansen, L., Cinner, J., Dews, G., Russ, G., Schuttenberg, H. Z., Peñaflor, E. L., Eakin, C.M., Christensen, T. R. L., Abbey, M., Areki, F., Kosaka, R. A., Tewfik, A. & Oliver, J. (2009). *The Coral Triangle and Climate Change: Ecosystems, People and Societies at Risk*.

Hoeksema, B. W. (2007). Delineation of the Indo-Malayan centre of maximum marine biodiversity: the Coral Triangle. In: Renema W (ed) Biogeography, time, and place: distribu- tions, barriers, and islands. Springer, Netherlands, pp 117- 178

Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.12613

Hughes, T. P., Anderson, K. D., Connolly, S. R., Heron, S. F., Kerry, J. T., Lough, J. M., Baird, A. H., Baum, J. K., Berumen, M. L., Bridge, T. C., Claar, D. C., Eakin, C. M., Gilmour, J. P., Graham, N. A. J., Harrison, H., Hobbs, J. P. A., Hoey, A. S., Hoogenboom, M., Lowe, R. J., McCulloch, M. T., Pandolfi, J. M., Pratchett, M., Schoepf, V., Torda, G. & Wilson, S. K. (2018). Spatial and temporal patterns of mass bleaching of corals in the Anthropocene. *Science* 359(6372):80-83. https://doi.org/10.1126/science.aan8048

Ivanova, N. V, Zemlak, T. S., Hanner, R. H., & Hebert, P. D. N. (2007). Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, (January 2006). https://doi.org/10.1111/j.1471-8286.2007.01748.x

Juhel, J. B., Utama, R. S., Marques, V., Vimono, I. B., Sugeha, H. Y., Kadarusman, Pouyaud, L., Dejean, T., Mouillot, D. & Hocdé, R. (2020). Accumulation curves of environmental DNA sequences predict coastal fish diversity in the coral triangle: EDNA predict fish diversity. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1930). https://doi.org/10.1098/rspb.2020.0248rspb20200248

Karcher, D. B., Roth, F., Carvalho, S., El-Khaled, Y. C., Tilstra, A., Kürten, B., Struck, U., Jones, B. H. & Wild, C. (2020). Nitrogen eutrophication particularly promotes turf algae in coral reefs of the central Red Sea. *PeerJ*, *2020*(4), 1–25. https://doi.org/10.7717/peerj.8737

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4). https://doi.org/10.1093/molbev/mst010

Kelly, R. P., Closek, C. J., O'Donnell, J. L., Kralj, J. E., Shelton, A. O., & Samhouri, J. F. (2017). Genetic and Manual Survey Methods Yield Different and Complementary Views of an Ecosystem. *Frontiers in Marine Science*, *3*(January), 1–11. https://doi.org/10.3389/fmars.2016.00283

Kelly, R. P., O'Donnell, J. L., Lowell, N. C., Shelton, A. O., Samhouri, J. F., Hennessey, S. M., Feist, B. E. & Williams, G. D. (2016). Genetic signatures of ecological diversity along an urbanization gradient. *PeerJ*, *4*, e2444. https://doi.org/10.7717/peerj.2444

Kirkendale, L. A, & Meyer, C. P. (2004). Phylogeography of the Patelloida profunda group (Gastropoda: Lottidae): diversification in a dispersal-driven marine system. *Molecular Ecology*, *13*(9), 2749–2762. https://doi.org/10.1111/j.1365-294X.2004.02284.x

Knowlton, N., Brainard, R. E., Fisher, R., Moews, M., Plaisance, L., & Caley, M. J. (2010). Coral Reef Biodiversity. In A. D. McIntyre (Ed.), *Life in the World's Oceans* (pp. 65–77). Blackwell Publishing Ltd.

Kochzius, M., & Nuryanto, A. (2008). Strong genetic population structure in the boring giant clam, Tridacna crocea, across the Indo-Malay Archipelago: Implications related to evolutionary processes and connectivity. *Molecular Ecology*, *17*, 3775–3787. https://doi.org/10.1111/j.1365-294X.2008.03803.x

Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, *112*(7), 2076–2081. https://doi.org/10.1073/pnas.1424997112

Leray, M., & Knowlton, N. (2016a). Censusing marine eukaryotic diversity in the twenty-first century. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1702), 20150331. https://doi.org/10.1098/rstb.2015.0331

Leray, M., & Knowlton, N. (2016b). Visualizing Patterns of Marine Eukaryotic Diversity from Metabarcoding Data Using QIIME, *1452*, 219–235. https://doi.org/10.1007/978-1-4939-3774-5

Leray, M., Meyer, C. P., & Mills, S. C. (2015). Metabarcoding dietary analysis of coral dwelling predatory fish demonstrates the minor contribution of coral mutualists to their highly partitioned, generalist diet. *PeerJ*, *3*. https://doi.org/10.7717/peerj.1047

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T. & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*(1). https://doi.org/10.1186/1742-9994-10-34

Lovejoy, C., Massana, R., & Pedros-Alio, C. (2007). Diversity and Distribution of Marine Microbial Eukaryotes in the Arctic Ocean and Adjacent Seas. *Applied and Environmental Microbiologyl*, *72*(5), 3085–3095. https://doi.org/10.1128/AEM.72.5.3085

Luddington, I. A., Kaczmarska, I., & Lovejoy, C. (2012). Distance and Character-Based Evaluation of the V4 Region of the 18S rRNA Gene for the Identification of Diatoms (Bacillariophyceae). *PLoS ONE*, *7*(9). https://doi.org/10.1371/journal.pone.0045664

Manzari, C., Lionetti, C., Erchia, A. M. D. & Pesole, G., (2015). LifeWatch – MoBiLab Report. Institute of Biomembranes and Bioenergetics, Consiglio Nazionale delle Ricerche, Bari, Italy. (February), 1–8.

Marwayana, O.N. (2018). Biodiversity and Distribution of Marine Fishes in Indonesia inferred by Environmental DNA (Master's thesis, University of California, Los Angeles (UCLA), Los Angeles, California, USA). Retrieved from https://escholarship.org/uc/item/1sx4k83d

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061217

Meyer, C. P., & Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, *3*(12), e422. https://doi.org/10.1371/journal.pbio.0030422

Mustika, P. L. K., Gunawan, T., & Erdmann, M. V. (2013). *A Marine Rapid Assessment (MRAP) of the Anambas Islands Marine Tourism Park, 3-31 May 2012*. Denpasar.

Mustika, P. L. K., Ratha, I. M. J., & Purwanto, S. (2012). *Bali Marine Rapid Assessment Program 2011*. Denpasar.

Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., Green, L. E. & Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.12895

Ogle, D. H., P. Wheeler. & A. Dinno. 2018. FSA: Fisheries Stock Analysis. R package version 0.8.22, https://github.com/droglenc/FSA).

Oksanen, J. (2017). Vegan: ecological diversity. *R Package Version 2.4-4*. https://doi.org/10.1029/2006JF000545

Pavan-Kumar, A., Gireesh-Babu, P., & Lakra, W. S. (2015). DNA Metabacoding: a new approach for rapid biodiversity assessment. *Journal of Cell Science and Molecular Biology*. Volume 2. Issues 1.

Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., & Esling, P. (2016). Protist metabarcoding and environmental biomonitoring: Time for change. *European Journal of Protistology*, *55*, 12–25. https://doi.org/10.1016/j.ejop.2016.02.003

Pearman, J. K., Leray, M., Villalobos, R., Machida, R. J., Berumen, M. L., Knowlton, N., & Carvalho, S. (2018). Cross-shelf investigation of coral reef cryptic benthic organisms reveals diversity patterns of the hidden majority. *Scientific Reports*, *8*(1), 1–17. https://doi.org/10.1038/s41598-018-26332-5

Pearman, J. K., Anlauf, H., Irigoien, X., & Carvalho, S. (2016). Please mind the gap - Visual census and cryptic biodiversity assessment at central Red Sea coral reefs. *Marine Environmental Research*, *118*, 20–30. https://doi.org/10.1016/j.marenvres.2016.04.011

Peñaflor, E. L., Skirving, W. J., Strong, A. E., Heron, S. F., & David, L. T. (2009). Sea-surface temperature and thermal stress in the Coral Triangle over the past two decades. *Coral Reefs*, *28*(4), 841–850. https://doi.org/10.1007/s00338-009-0522-8

Plaisance, L., Brainard, R., Caley, M. J., & Knowlton, N. (2011a). Using DNA Barcoding and Standardized Sampling to Compare Geographic and Habitat Differentiation of Crustaceans: A Hawaiian Islands Example. *Diversity*, (4), 581–591. https://doi.org/10.3390/d3040581

Plaisance, L., Caley, M. J., Brainard, R. E., & Knowlton, N. (2011b). The Diversity of Coral Reefs : What Are We Missing ?, *6*(10). https://doi.org/10.1371/journal.pone.0025026

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0009490

Puillandre, N., Strong, E. E., Bouchet, P., Boisselier, M. C., Couloux, A., & Samadi, S. (2009). Identifying gastropod spawn from DNA barcodes: Possible but not yet practicable. *Molecular Ecology Resources*. https://doi.org/10.1111/j.1755-0998.2009.02576.x

Radulovici, A. E., Archambault, P., & Dufresne, F. (2010). DNA Barcodes for Marine Biodiversity: Moving Fast Forward?, (February), 450–472. https://doi.org/10.3390/d2040450

Ransome, E., Geller, J. B., Timmers, M., Leray, M., Mahardini, A., Sembiring, A., Collins, A. G. & Meyer, C. P. (2017). The importance of standardization for biodiversity comparisons: A case study using autonomous reef monitoring structures (ARMS) and metabarcoding to measure cryptic diversity on Mo'orea coral reefs, French Polynesia. *PLoS ONE*, *12*(4), 1–19. https://doi.org/10.1371/journal.pone.0175066

Reimer, J. D., Ono, S., Takishita, K., Tsukahara, J., & Maruyama, T. (2006). Molecular Evidence Suggesting Species in the Zoanthid Genera Palythoa and Protopalythoa (Anthozoa: Hexacorallia) Are Congeneric. *Zoological Science*, *23*(1), 87–94. https://doi.org/10.2108/zsj.23.87

Roberts, C. M., McClean, C. J., Veron, J. E., Hawkins, J. P., Allen, G. R., McAllister, D. E., Mittermeier, C. G., Schueler, F. W., Spalding, M., Wells, F., Vynne, C. & Werner, T. B. Marine biodiversity hotspots and conservation priorities for tropical reefs. Science. 2002 Feb 15;295(5558):1280-4. doi: 10.1126/science.1067728. PMID: 11847338.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*. https://doi.org/10.7717/peerj.2584

Seabird McKeon, C., & Moore, J. M. (2014). Species and size diversity in protective services offered by coral guard-crabs. *PeerJ*. https://doi.org/10.7717/peerj.574

Slaveykova, V., Sonntag, B., & Gutiérrez, J. C. (2016). Stress and Protists: No life without stress. *European Journal of Protistology*, *55*, 39–49. https://doi.org/10.1016/j.ejop.2016.06.001

Spalding, M. D., Fox, H. E., Allen, G. R., Davidson, N., Ferdaña, Z. A., Finlayson, M., Halpern, B. S., Jorge, M. A., Lombana, A., Lourie, S. A., Martin, K. D., McManus, E., Molnar, J., Recchia, C. A., Robertson, J. & Robertson, J. (2007). Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. *BioScience*, *57*(7), 573–583. https://doi.org/10.1641/B570707

Stewart, H. L., Holbrook, S. J., Schmitt, R. J., & Brooks, A. J. (2006). Symbiotic crabs maintain coral health by clearing sediments. *Coral Reefs* 25:609-615. https://doi.org/10.1007/s00338-006-0132-7

Stier, A. C., Gil, M. A., McKeon, C. S., Lemer, S., Leray, M., Mills, S. C., & Osenberg, C. W. (2012). Housekeeping mutualisms: Do more symbionts facilitate host performance? *PLoS ONE*, *7*(4), 2–7. https://doi.org/10.1371/journal.pone.0032079

Tornabene, L., Valdez, S., Erdmann, M., & Pezold, F. (2015). Support for a "Center of Origin" in the Coral Triangle: Cryptic diversity, recent speciation, and local endemism in a diverse lineage of reef fishes (Gobiidae: Eviota). *Molecular Phylogenetics and Evolution*, *82*(PA), 200–210. https://doi.org/10.1016/j.ympev.2014.09.012

Turak E, DeVantier L (in press) Biodiversity and conservation priorities of reef-building corals in the Papuan Bird's Head Seascape. In: Katz LS, Firman A, Erdmann MV (eds) A Rapid Marine Biodiversity Assessment of Teluk Cenderawasih and the FakFak-Kaimana Coastline of the Papuan Bird's Head Seascape, Indonesia. RAP Bulletin of Biological Assessment. Conservation International, Washington, D.C.

Turak E, Souhoka J (2003) Coral diversity and the status of coral reefs in the Raja Ampat Islands. In: Donnelly R, Neville D, Mous P (eds) Report on a rapid ecological assessment of the Raja Ampat Islands, Papua, Eastern Indonesia, held October 30 – November 22, 2002. The Nature Conservancy Southeast Asia Center for Marine Protected Areas, Sanur, Bali Indonesia.

van der Meij, S. E. T., Suharsono, & Hoeksema, B. W. (2010). Long-term changes in coral assemblages under natural and anthropogenic stress in Jakarta Bay (1920-2005). *Marine Pollution Bulletin*, *60*(9), 1442–1454. https://doi.org/10.1016/j.marpolbul.2010.05.011

Veron, J. E. N., Devantier, L. M., Turak, E., & Green, A. L. (2009). Delineating the Coral Triangle. *Galaxea, Journal of Coral Reef Studies*, *11*, 91–100.

Vogler, C., Benzie, J., Barber, P. H., Erdmann, M. V., Ambariyanto, Sheppard, C., Tenggardjaja, K., Gérard, K. & Wörheide, G. (2012). Phylogeography of the crown-of-thorns starfish in the Indian ocean. *PLoS ONE*, *7*(8), 1–10. https://doi.org/10.1371/journal.pone.0043499

Voris, H. K. (2000). Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems and time durations. *Journal of Biogeography*, *27*(5), 1153–1167. https://doi.org/10.1046/j.1365-2699.2000.00489.x

Wangensteen, O. S., & Turon, X. (2016). Metabarcoding Techniques for Assessing Biodiversity of Marine Animal Forests. In *Marine Animal Forests* (pp. 1–34). https://doi.org/10.1007/978-3-319-17001-5

Ward, R. D., Holmes, B. H., & O'Hara, T. D. (2008). DNA barcoding discriminates echinoderm species. *Molecular Ecology Resources*. https://doi.org/10.1111/j.1755-0998.2008.02332.x

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York. Media*. https://doi.org/10.1007/978-0-387-98141-3

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593