# Whispers in Murky Waters: Bacterioplankton Interaction Networks Underpinning Ecosystem Health

**A Dissertation**
**Submitted to the Graduate School of**
**Information Science and Technology**
**Hokkaido University**

**Elroy Galbraith**

**June 16, 2022**

# Abstract

Accurate monitoring and prediction of aquatic ecosystems at multiple scales is crucial to maintaining the biosphere, especially given the role of rivers, estuaries, and oceans in species habitat provision and biogeochemical cycling in the land-atmosphere continuum. Communities of microscopic biota - the microbiome - perform many of these ecosystems regulating functions, emitting informative signals of environmental pressure via patterns in their collective structure and function. Still, the collective dynamics of the microbiome remains poorly understood. What might the characteristics of community structure and function be under optimal conditions? How does the environment destabilize community structure or function? Does an optimally structured microbiome imply a resilient microbiome? Are the many signals reducible to a habitat-specific portfolio that characterizes ecosystem health? In this dissertation I advance our understanding and inform ecohealth assessment and engineering by extracting phylogenetic, structural, and functional patterns from the collective dynamics of the microbiome, using a combination of information theory, network theory, and statistical physics, thereby simplifying ecosystem complexity. I also present the utility of envisioning the environment similarly: as an assembly of interacting elements whose destabilizing impact emerges from patterns within the network.

In Chapter 2, I demonstrate how it is possible to extract many signals from patterns in community phylogenetic relatedness, population abundance distributions, and species interactions within the Bacteriome network inferred from information fluxes, all informative of ecosystem state. Among these are the new info-theoretic Kleiber's Law between bacterioplankton phylum co-predictability (directed interactions) and population and community abundance uncertainty, with an average exponent $\phi \sim \frac{2}{3}$ in striking accordance with theoretical expectations; as well as the Phylogenetic Separation Rate ($\rho_{PS} = 0.01$) showing that communities accrue new functional groups much slower than new species ($\rho_{SR} = 0.4$). In Chapter 3, I

present the novel Eco-Evo Mandala, a multiscale map of the Bacteriome considering habitat-defined distributions, species interactions, and phylogeny, which signals community - and likely ecological - departures from relative theoretical optimality. The Mandala confirms that these departures are habitat-specific, mostly considering the structural and functional traits related to bacterioplankton abundance and interaction distributions (reflected by $\epsilon$ and $\lambda$ as power law and exponential distribution parameters, respectively), which are not linearly associated with each other. In Chapter 4, I describe how the Envirome - the collective assemblage of environmental drivers defined by environmental interactions - can pinpoint factors responsible for community disorganization, an idea which has been applied elsewhere but not ecological research. Disorganization within the Envirome meant higher environmental impacts causing larger disorganization within the Bacteriome toward random interaction topologies.

I end by envisioning a possible future for ecological research and practice that incorporates the perspective underlying this work. This work emphasizes how functional diversity regulating community optimality is different from taxonomic richness, and it must be combined with probabilistic bacterio-environmental covariations to create a fulsome picture of ecosystem state, particularly because full biological knowledge is presently too cumbersome to obtain. Consequently, eco-engineering policies and technologies should focus less on moderating the whispers from individual populations or environmental drivers, and more on orchestrating all components into a symphony of order and health.

# Acknowledgments

I would like to thank my esteemed supervisor Dr. Matteo Convertino for his invaluable supervision and support during my PhD program. My gratitude extends to the GI-CoRE Global Station for Big Data and Cybersecurity and the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan for the funding opportunity to undertake my studies at the Graduate School of Information Science and Technology, Hokkaido University. Additionally, I am also deeply grateful to Prof. Ohgane for his assistance and guidance during my final year, and Dr. P. R. Frade at the Natural History Museum, Vienna for his advice and insights. I also thank Dr. Louis Chan and Dr. Jie Li for valuable discussions on my research, and my wife, Dr. Sikopo Nyambe, for her support and encouragement. My appreciation also goes out to my family and friends for their encouragement and support throughout my studies.

# Publications

## Articles

1. Galbraith E., Convertino M., "The Eco-Evo Mandala: Simplifying Bacterioplankton Complexity into Ecohealth Signatures", *Entropy* 23.11 (2021):1471.

2. Galbraith, E. Li, J., and Convertino M., "In.To. COVID-19 Socio-epidemiological Co-causality", *Scientific Reports* 12:1 (2022):1-25.

3. Galbraith E., Frade, P.R., Convertino M., "Metabolic Shifts of Oceans: Summoning Bacterial Interactions", *Ecological Indicatiors* 138 (2022):108871.

## Conference Papers

1. Elroy Galbraith, Matteo Convertino, "On Structure, Function, and Services of the Ocean Bacterioplankton" Proceedings of 2018 Summer International Symposium on Big-Data, Cybersecurity and IoT, August 7-8, 2018.

2. Elroy Galbraith, Matteo Convertino, "The Ocean Environmental-Bacterioplankton Nexus" Complex System Symposium, 2020

## Patent

1. Elroy Galbraith, Matteo Convertino, Jie Li, Victor Del Rio-Vilas. "InTo (Infodemic Tomography) COVID-19: Social-epidemiological Co-causality", US Patent, Application Number: 63119650.

# Contents

# List of Figures

# Chapter 1

# Introduction

## 1.1 Ecosystem Monitoring During the Anthropocene

### 1.1.1 Global and Local Aquatic Ecosystem Stress

The Anthropocene, characterized by anthropogenic pollution and climate change, sees many ecosystems – at many scales – facing an existential crisis [1]. Maintaining the health of aquatic ecosystems, the ocean in particular, is critical to maintaining the health of the planet. The ocean and its tributaries cycle critical materials and nutrients around the planet [2]. It even plays a role in staving off biosphere collapse as it is the largest carbon-sink, thus moderating climate change due to carbon dioxide [3, 4]. Rivers metabolize much of the carbon originating from terrestrial sources, including anthropogenic activity, before it ever reaches the ocean [5], making rivers and near shore marine habitats critically important for biosphere maintenance.

My study explores the Great Barrier Reef, a marine habitat off the coast of Northern Australia. It is the world's largest coral reef system, providing invaluable environmental, social, and economic services to humans and other species within this region [6]. Though protected as a World Heritage Site, it is not spared the deleterious effects of anthropogenic perturbations such as catchment run-off (from coastal and riverine efflux), direct use, ecosystem (geomorphological) degradation and global climate change stress, like ocean temperature anomalies, which affect ecosystems at even the microbial scale [7–15]. These pressures negatively impact the biodiversity of the region, which underpins its geomorphology, cultural and economic values [6]. For example, coral bleaching events have been increasing in frequency and magnitude [10], decimating the breeding and feeding ground for several

species, including some endangered ones. To protect these habitats, and the planet at large, we need to find ways of monitoring that are robust, expedient, efficient, and informative for bioengineering solutions. However, this is difficult as these habitats are extremely complex, in that there are many interacting components, and many drivers whose impact is non-linear, time-lagged, and cumulative.

### 1.1.2 Marine Microbial Community Services

Global research on the Earth's oceans has revealed the large abundance, complexity, and ecological importance of microscopic organisms, which interact in communities called the microbiome [16–20]. These are large and highly diverse organizations of bacteria (bacterioplankton in the case of aquatic habitats), viruses and other microorganisms (e.g., eukaryotes), which can be free-living [16, 21, 22] or in host-associated communities [13, 23, 24]. The microbiome, particularly its bacterial component (the bacteriome), is the ideal candidate for ecohealth monitoring because, despite being dependent on water dynamics and quality, they are largely responsible for the planet's biogeochemical cycles [25–30] and are highly responsive to disturbances in these cycles and environment quality [23, 31, 32] presenting as variation in their population dynamics. Consequently, there is growing interest in decoding the dynamics of the bacteriome for indicating and predicting ecosystem health [33–36].

The Great Barrier Reef Marine Park Authority has acknowledged the need to include the microbiome in its monitoring and eco-engineering strategies [37], as there is value in understanding these dynamics within localities and regions [38], not merely at the global scale. Current recommendations include identifying and tracking the presence/absence or relative abundance of specific indicator taxa or microbial functions. The technology exists that can support this work at relatively low operational cost. However, research has stalled because it is infeasible to observe (sample and sequence) all these biological mechanisms at play [37]. One solution is to find a way to use less data to create equally accurate bioindicators, but this requires re-imagining what the microbiome is and the value of patterns as opposed to processes.

## 1.2 Community Complexity for EcoHealth Assessment

Microorganisms are like mini-computers executing complicated, distributed algorithms [39] that download, store, and transform materials, energy, and information within a network [40]. Consequently, in this work, I approach the bacteriome as a complex system. A complex system is a collection of numerous, autonomous yet hierarchically ordered units, whose functions and services emerge from the coordination of interactions among its units within a network [41]. These systems, beyond being scientifically intriguing, display characteristic features useful for assessments at multiple scales. For example, these systems display power law patterns [42] that were proven to coincide with a local or global energy minimum of the system's features. Consider river networks, whose power law distribution reflects the scale-free pattern of runoff and the energy minimum of water flow [43]. The stationary solution of the landscape evolution equation is a power law that is invariant across many orders of magnitudes; yet, it is linking patterns to processes clearly at stationarity. Previous studies on the neural tissue [44] have also shown how scale-invariance is not only occurring at a critical transition, but at a global stable state, and this is the case for rivers. Such invariance, which is a state of relative stability of a system's features, is a byproduct of self-organized criticality [45] but, more generally, of evolutionary dynamics leading to feasibly optimal states [46]. In this sense, some authors [47, 48] have equated invariance with robustness, although the latter does not correspond to invariant features necessarily: for example, random networks are also topologically robust (invariant but not scale-invariant) because any perturbation leads to another random state.

In general, ecosystems may not be locally stable and undergo critical transitions (as in [49]) but these fluctuations are invariant over time. For instance, runoff in rivers may exhibit local instability (permanent due to structural features such as sudden jumps, or temporary such as black-swan variability due to climate extremes) but, over large spatial and temporal scales, shows stability in distribution. Deviation from global stability or invariance is a worrisome signature of departure from optimality — for instance, driven by large habitat and climatic modifications. A recent analysis of dynamical stability and invariance (which is indubitably dependent on the scale of analysis) has been performed for fisheries by [50].Lastly, probabilistically speaking, the stable distribution family is sometimes referred to as the Lévy alpha-stable distribution to which power law distribution belongs. Thus, using this principle of relative optimality, one could evaluate the distribution of microbial

species' abundance and their interactions to assess the extent to which the bacteriome is optimally arranged (and stable) or its divergence from the theoretically optimal state.

Bacteriomes, such as that found in the human gut [51], exhibit signatures of evolutionary optimality (or self-organizing criticality) in which dynamical fluctuations are scale-free distributed, resulting in a relatively stable system. This self-organization manifests as indicative community diversity patterns, i.e., the spatio-temporal distribution of unique species or functions (underpinned by interactions), although the taxonomic resolution can mislead more than functional profiles because biological traits are differentially conserved along the phylogenetic tree [52, 53]. There are intra- and interspecies patterns defining community metabolism, which collapse to universal, scale invariant curves characteristic of environmental regimes [40, 54]. Some take the form of allometric (scaling) relationships, $Y \sim \beta X^{\alpha}$, where $Y$ is some dependent variable related to ecosystems' fluxes, like metabolic rate; $X$ is an allometric predictor related to structural ecosystem properties, like body mass or abundance; and $\alpha$ is the scaling exponent regulating energy expenditure. When these communities display a power law distribution in the frequency distribution of size classes among species, such as Taylor's law and the abundance–size spectrum [55, 56], their ecosystem is equally well-balanced ( [56] for plankton and [14] for oceanic biomass). Kleiber's Law [57, 58], $B \sim M^{\alpha}$, is the scaling law regulating how metabolic rate (B) scales with body mass (M) by a power coefficient ($\alpha$), that is typically in between $\frac{2}{3}$ [59, 60] or $\frac{3}{4}$ [61]. When applied to communities, Kleiber's Law can predict changes in community growth rate, diversity, and nutrient cycling rates [62, 63], because metabolism underpins multiple species dynamics during changes in resource supply [64], thereby indicating changes in ecosystems.

The community in each habitat can be seen as a meta-organism composed by species, much like a cell where populations of proteins – with their different structure and function – flow within organelles in the cell [65]; the human body where populations of cells flow to organs together with metabolites via the vascular network [66, 67]; or river basins where diverse populations of species (producers and consumers), or water and nutrients, flow among riverine communities via network streams [68]. In such systems, the flow of biotic and abiotic elements (where the latter can also be mapped as a functional network) implies energy expenditure over time, i.e., organism metabolism. Species and metabolites are eco-environmental flows, which in a discrete data realm are information flows [69]: species sense oth-

ers and the environment and their exchange of information is the primary trigger of changes requiring energy expenditure.

Biomass conversion in a system involves the interspecies abundance variation resulting from competitive (or cooperative) interactions, making information theoretic approaches, such as transfer entropy [70], suitable for signal detection. Transfer entropy (TE) is an assumption-free and probabilistic means of maximizing uncertainty reduction about species interactions [69, 71]. It is ideal because species interactions are usually non-normal, non-linear, bidirectional, highly unpredictable, and asynchronous: methods such as TE have been found to be superior to correlational approaches [72, 73], and other models for functional network inference [69]. [51] showed that the information flow (measured by TE) within the gut microbiome of healthy individuals is more scale-free than that within the gut microbiome of unhealthy individuals.

In order to quantify the systemic impact of the environment on potential community metabolism, I consider biogeochemical dynamics as a collective assembly of interacting factors called "envirome". Recent studies have highlighted how environmental factors – biogeochemical metabolic resources, habitat features, and cross-habitat dependencies – drive species and entire microbial communities in the ocean, including their metabolism, on global scales [16, 21, 22, 31, 74, 74–76]. However, the information about the environmental state contained in microbiome patterns, particularly at the habitat scale, is difficult to interpret given observations of strong, complex associations among those environmental factors [9, 21, 22]. Disentangling factors risks oversimplifying the relationship between the environment and microbiome effects, as well as neglecting bidirectional relationships such as biota-environment feedback. Human molecular biologists use this "enviromic" perspective, proposing that the entirety of an organism's environment, rather than any factor in isolation, impacts its genomic expression [77]. These factors represent a complex system of interacting variables, whose outcome is not the sum of their parts. As such, the ecological state of a habitat community can be disentangled from its other features – like static geomorphology and fluctuating biogeochemical variables – inferred as a network from their pair-wise interactions.

## 1.3 Eco-Environmental Nexus Probabilistic Mapping

If we agree that the collective dynamics of the bacteriome underpins ecosystem health, then informing eco-engineering solutions involves deciphering the orchestrated (habitat-specific) 'noise' of the entire community rather than the 'whispers' (small yet comparatively uninformative noise) of its individual members. I identify phylogenetic, structural, and functional patterns comprising the collective dynamics of these communities, then, I explore whether the collective dynamics of the marine bacteriome are reducible to a habitat-specific portfolio. I question how informative the genetic measurements for ecosystem health are compared to the statistical features, such as the distribution of abundance or information flow; and, whether an optimally structured bacteriome, so characterized by its collective dynamics, implies a resilient bacteriome.

In Chapter 2 I describe the materials and methods involved in this approach. Briefly, the data comprised ecological data - bacterioplankton phylogenetics and abundances, and environmental readings – from seven locations within the Great Barrier Reef grouped into four geomorphologically distinct marine habitats; the methods involved a combination of information theory, network theory, and statistical physics (see Fig. 1.1).

Chapter 3 presents the phylogenetic, structural and functional signals extracted from the bacterioplankton interaction networks. Biodiversity should be considered as variation in species functions that from the coordination of interactions [69, 78], rather than species richness because community ecological function results from the distribution of functions within; furthermore, richness (baselined by habitats as well as site- and time- specific) can misinform about ecological stability (or divergence from optimal coordination equivalently). Therefore, this chapter focuses on population traits and likely ecological functions, exploring such patterns as the rate at which new functional groups emerge over time, and the collective's metabolic activity – inferred from patterns of interdependencies as networks (i.e., the skeleton of biodiversity patterns in terms of species distribution and interaction) – and its organization (as entropy of their distribution) as determinants of an information-theoretic Kleiber's Law.

In Chapter 4 I disaggregate the bacteriome into statistical distributions of its features – its genetic dissimilarity, configurations, and coordination dynamics – and then model the relationship among them using the new Eco-Evo Mandala. This Mandala is a habitat-specific portfolio of metrics to consider during eco-engineering,

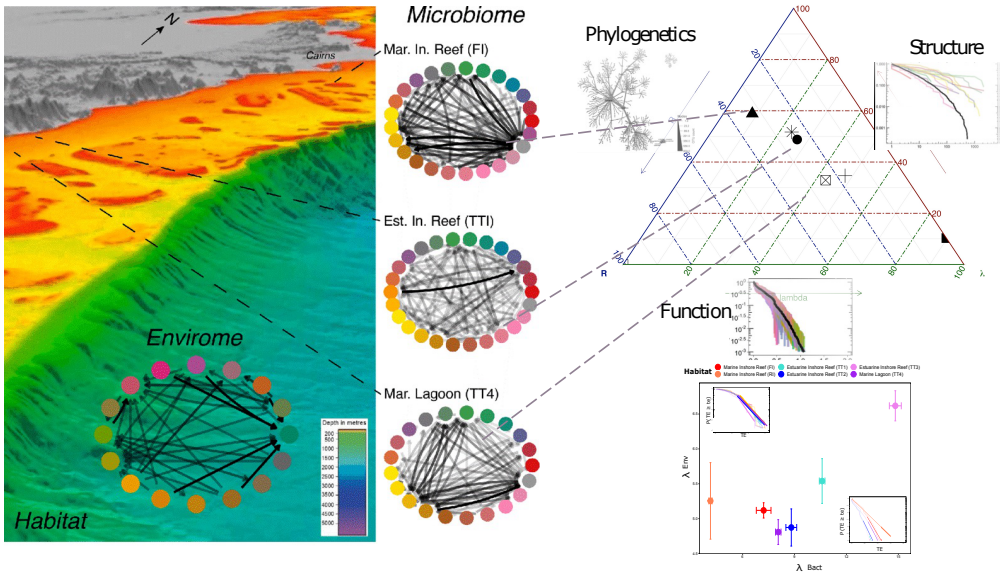


Figure 1.1: **Habitat, Envirome and Bacteriome.** The study investigates the systemic impact of networked environmental factors ("envirome") into the collective organization of bacterial entropy-based interactions, mediated by the habitat. **Left**: bathymetry of the Great Barrier Reef (source https://www.deepreef.org/) for the study area (Fig. 2.2A) and envirome from abiotic data. **Center**: bacterioplankton networks inferred using TE given relative abundance data, i.e., RSA (Fig. 2.2A). **Top Right**: Phase-space designed modelling the ecosystem health state given bacterioplankton phylogenetics, structure (configuration), and function (coordination). **Bottom Right**: Comparison of habitat-defined enviromes and bacteriomes characterizing environmental collective impact of bacterioplankton function.

that also enables a comparison between the value of genetic information compared to statistical information. It enables exploration of the association between the structural stability and dynamic stability of the system, and comparison of the between-state differences along separate axes. This method is applied at the community level as a measure of habitat state, and at the population level to explore species-specific signals.

Chapter 5 focuses on the environment's impact on the collective dynamics of the bacteriome. I present the results of investigating Kleiber's law specificity for different habitats considering similar bacteriome composition, as well as how bacteriome variability – in the form of Kleiber's Law exponent – is related to habitat structural factors and other systemic environmental fluctuations (i.e., hydrogeomorphic and biogeochemical factors and their distribution leading to eco-environmental disorganization). Afterward, I explore the impact of salient environmental pressure (where salience is interpreted in a network sense [79]) on these patterns.

In Chapter6 I discuss the implications of my findings for policy, practice, and future research. Because the principles being applied are theoretically universal, if this idea holds, then we would be able to monitor ecosystem states at multiple scales using community patterns with less need for precise biological knowledge. This greatly reduces the cost and time-lag involved in the current method of ecosystem monitoring. The big-data supporting this analysis could be used to design spatio-temporal forecasting technologies, that provides location specific, actionable insights within relatively short time-horizons. We could also model bacterio-environmental feedback loops, thereby informing more targeted interventions at different scales.

# Chapter 2

# Materials and Methods

## 2.1 Introduction

This study explores the community phylogenetic, structural, and functional dynamics of bacterioplankton underpinning ecosystem health, using a combination of information theory, network theory, and statistical physics. This chapter begins by describing the raw data before describing the methods used to explore the community features of interest. I describe the creation of the phylogenetic trees (phylogenies) used to support analysis of diversity and evolutionary rates. Next, I present how both the species and environmental driver interaction networks were inferred using transfer entropy. Then, I define the key network metrics: link weight which is proportional to interaction strength; node centrality, which means the relative importance of a population or environmental driver; and link salience, which measures the relative importance of the interaction between a pair of populations or environmental drivers. Next, the information-theoretic Kleiber's Law is defined, as well as the parameters used to characterize distributions in abundance and network link weight. How these parameters are combined into the Eco-Evo Mandala is then described, followed by the method used to quantify the sensitivity of community phylogenetic diversity to uncertainty in the environmental measures.

## 2.2 Marine Ecological Data

Bacterioplankton and environmental data, originally published by [9], were collected from 7 sites found along a 124KM long gradient of the Great Barrier Reef

between 2009 and 2013 (see Fig.2.2). The data were collected over different time lengths ($t$) for each location: $t$ = 15 in TT1; $t$ = 17 in TT2; $t$ = 17 in TT3; $t$ = 15 in TT4; $t$ = 10 in FI; $t$ = 9 in RI; and $t$ = 3 in TR. The sites were originally categorized into 3 classes, river (TR), plume (TT1, TT2, and TT3), and marine (FI, RI, and TT4), based on their proximity to the nearest influencing river. I categorized the sites into 4 habitat classes based on their proximity to the nearest influencing river and likely bathymetry: River (TR), Estuarine Inshore Reefs (TT1, TT2, and TT3), Marine Inshore Reef (FI and RI), and Marine Lagoon (TT4).

The bacterioplankton data from each site contained read counts (considered abundance) and taxonomic classifications of 1140 near-coastal, epipelagic, free-living, operational taxonomic units (OTUs) from the Bacteria and Archaea kingdoms, as well as unclassified OTUs (see Fig.2.2A). The OTUs – considered species from here onward – were identified following 16s rRNA gene amplicon sequencing of water samples taken from each location. The taxonomic classifications of OTUs were derived using the Hitman bioinformatics workflow (see [9] for a description of the workflow and parameters used). Rarefaction was not conducted for this analysis, and raw read counts based on sequence reads rather than relative read counts (relative abundance, used only to indicate species dominance) were used.

The environmental data were time series of measurements of biogeochemical variables taken simultaneously with water sampling for bacterioplankton readings (see Fig.2.2B). Environmental readings were also taken for temperature, salinity, silicate, suspended particulate matter, along with nutrients such as nitrogen, phosphorous, carbon, and chlorophyll a. Attempts were made to sample each site at least three times per year (June, October and March) but the time series lengths $T$ differed for each location (because of challenges with data collection): for TR, $t$ = 3; for TT1, $t$ = 15; for TT3, $t$ = 17; for TT2, $t$ = 17; for FI, $t$ = 10; for RI, $t$ = 9; and for TT4, $t$ = 15. Critical thresholds for each environmental variable were taken from [80].

## 2.3 Population Complexity Indicators

### 2.3.1 Probabilistic Characterization of Distributions

Community distributions of abundance (or interactions (TEs), see Sec.2.4) were characterized using either $\lambda$ or $\epsilon$ derived from an exceedance probability distribution

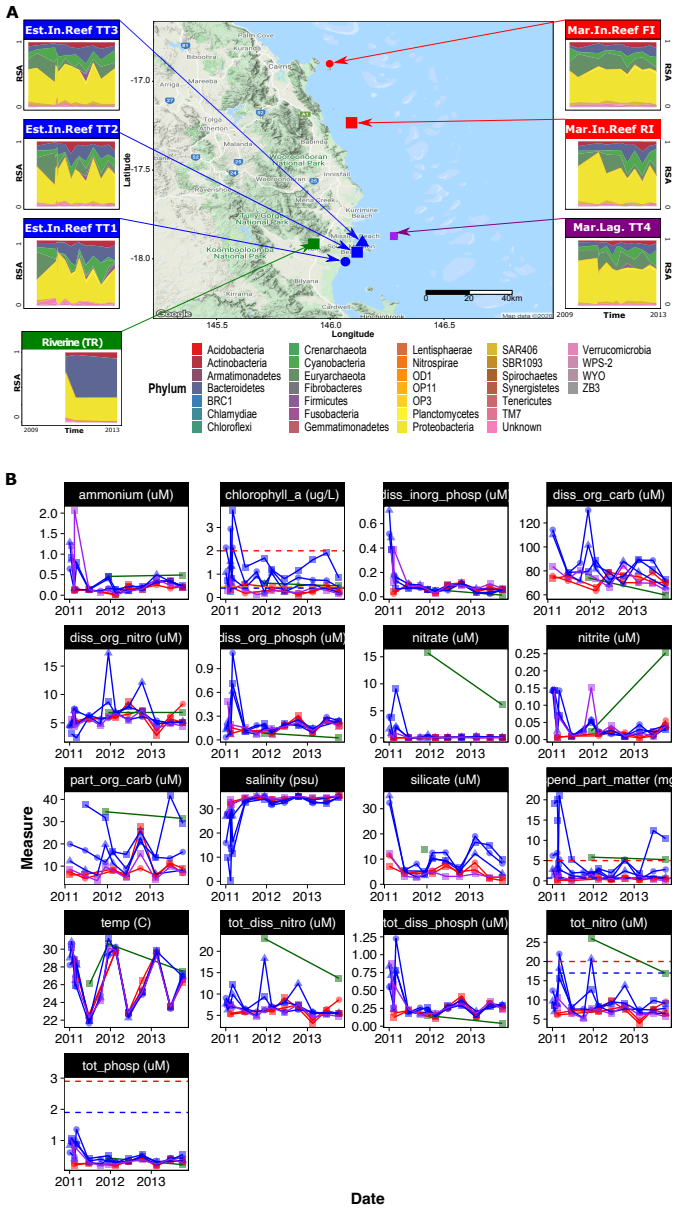


Figure 2.1: **Materials, Location, and Time Series.** **A** Google Map of the 7 sites in the Great Barrier Reef surrounded by time series of the relative abundance (0% to 100%) of phyla (color corresponding to the legend) observed in each site. **B** The color and shape of the points on the lines correspond to the location/habitat identified in **A**. The horizontal dashed red line indicates critical thresholds.

function (EPDF) as in [51]:

$$P(Y \geq y) \sim \begin{cases} e^{-\lambda_y} \\ y^{-\epsilon+1} \end{cases}, \tag{2.1}$$

where $Y$ was the abundances (or interactions from $G_{Bact}$ and $G_{Env}$ (see Section 2.4)) at the population (taxonomic phylum) or community level; $\lambda$ was used to characterize exponential distributions (see Fig.S3); and Zipfian (power-law) distributions were characterized with $\epsilon$ (see Fig. S1). The poweRlaw package [81] was used to estimate the parameters of all EPDFs using a maximum likelihood procedure to obtain the best fit for the distributions. Scale-freeness, and thus community or population optimality, increased with the value of $\epsilon$ or $\lambda$.

### 2.3.2 Taylor's Law

The variability of population abundances was characterized using the exponent $\nu$ of Taylor's Law [82]:

$$\langle x^2 \rangle \sim \langle x \rangle^{\nu}, \tag{2.2}$$

where $\langle x \rangle = E|X|$ is the mean abundance and $\langle x^2 \rangle = E|X^2| - E|X|^2$ is the variance of the abundance. This exponent was estimated for each population at the phylum taxonomic level within each community, as well as for the entire community at each location. The parameter $\nu$ was estimated by determining the slope of a linear regression between the base 2 logarithm of both the population mean and variance (see Fig.S2). Higher values of $\nu$ characterized communities or populations that experienced greater temporal instability in their abundance.

## 2.4 Multispecies Network Inference

Let $X = \{x_1, x_2, \ldots, x_t\}$ be the time series of bacterioplankton species abundance, or biogeochemical variables, from a location: where $x_i$ denotes the abundance or measurement for the species or biogeochemical variable $X$; and $t$ is the time series length in that location.

The information disorganization (uncertainty) of a species was quantified by the Shannon Entropy [83] of species abundance $H(Ab)$, defined as :

$$H(Ab) = \sum P(x_i) log_2 P(x_i), \tag{2.3}$$

where specifically in this equation $x_i$ was the abundance $Ab$ of species $X$.

Transfer entropy $TE$ was used to measure the paired interaction among species or biogeochemical variables [50, 51]:

$$TE_{X_i \to X_j} = \sum p(X_{j,t}, X_{j,t-1}, X_{i,t-1}) \cdot log_2 \left( \frac{p(X_{j,t}|X_{j,t-1}, X_{i,t-1})}{p(X_{j,t}|X_{j,t-1})}, \right), \tag{2.4}$$

where $X_{i,t}$ and $X_{j,t}$ are the values of pairs of species or drivers ($X_i$ and $X_j$) at time $t$ (the subscript $t-1$ denotes the value of $X_i$ or $X_j$ at $t-1$). The $TE$ was calculated using the box kernel estimator for continuous data from the JIDT Package [84], setting a history length of 1 and a kernel box size of 0.5 normalized units as the initialization parameters. TE quantifies the interdependence between OTUs as the information flux between them, considering the divergence, non-linearity, and asynchronicity in their abundance distribution, thereby reflecting predator–prey relationships.

An asymmetrical, weighted community matrix $W$ was then created based on TE:

$$w_{i \to j} = TE_{X_i \to X_j}, \tag{2.5}$$

where $i$ and $j$ in $w_{i \to j}$ were $X_i$ and $X_j$ in $TE_{X_i \to X_j}$, respectively. Two networks were created for each location: $G_{Bact} = (OTU, W_{Bact})$ for the bacteriome with its species $OTU$ and their interactions $W_{Bact}$ (see Fig.2.4); and $G_{Env} = (E, W_{Env})$ for the envirome (see Fig.S2) with its biogeochemical variables $E$ and their interactions $W_{Env}$.

### 2.4.1 Network Collective Metrics

#### Total Directed Interactions and Connections

Within these graphs, a node's total outgoing interaction $x_i \to x_j$ and its total incoming interaction $x_i \leftarrow x_j$ were measured as its $OTE$ and $ITE$, respectively:

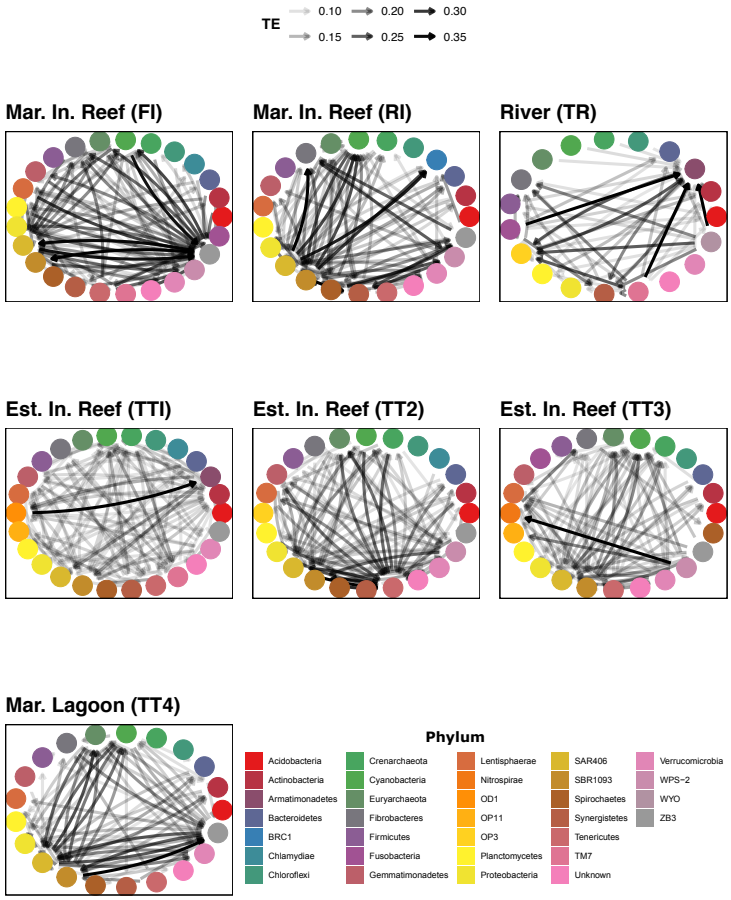


Figure 2.2: **Bacteriome Top Interaction Networks.** Nodes are phylum and the opacity of the links is proportional to TE: larger TEs are represented by more opaque links. The direction reflect the predictive influence. Networks show the top 20% of interactions that are power-law distributed, highlighting the tendency to scale-free core organization despite the overall organization is exponential.

$$OTE_i = \sum w_{x_i \to x_j}; \tag{2.6}$$

and

$$ITE_i = \sum w_{x_i \leftarrow x_j}. \tag{2.7}$$

The total number of connections (or phylum degree) $K$ for each node was calculated using

$$K_I = \sum (w_{x_i \leftarrow x_j})^0 + \sum (w_{x_i \to x_j})^0, \tag{2.8}$$

where $\sum (w_{x_i \leftarrow x_j})^0$ and $\sum (w_{x_i \to x_j})^0$ were the total number of incoming and outgoing edges, respectively.

### Weighted Node Centrality

The importance of a node was measured by its weighted node centrality [85]:

$$C_i = \frac{ote_i}{OTE_{max}} + \frac{ite_i}{ITE_{max}} + \frac{k_i}{K_{max}}, \tag{2.9}$$

where $ote$ was calculated using eq. 2.6, $ite$ was calculated using eq. 2.7 and $k$ was calculated using eq.2.8. Each term ($ote$, $ite$, and $k$) was normalized to the maximum value in each distribution of values ($OTE_{max}$, $ITE_{max}$, and $K_{max}$).

### Link Salience

A link-centric approach has been lacking to identify keystone species supporting network collective structure and function. This is particularly true when considering the salient impacts of the environment into the bacteriome. Equivalently, for enviromes, the most central factor (defining the most salient link) is the environmental factor that is coordinating the state of others, and whose removal would result is a substantially different environmental stress affecting the bacteriome, local habitat and climate dynamics at higher scales. Network analysis has been used to identify these keystone species by inferring species interaction matrices [86, 87] and then

using network metrics like node centrality [85] or link salience [88] to detect most important nodes and links. An interaction's importance to its network was measured by its salience $S$ [88] that determines the frequency a pair is involved into the collective organization of all ecosystem's interactions. All shortest path trees (SPTs) $T$ from the graphs $G$ ($G_{Bact}$ and $G_{Env}$) were extracted from effective distances by using the spTreeBellmanFord function from the optrees package [89]. The inverse of the edge weights defined by TE (i.e., $d_e = w_{i\to j}^{-1}$; see Eq.2.5) were used as the effective distances necessary for estimating the shortest path between nodes of the ecosystem interaction networks (i.e., species and environmental factors). Stronger interactions thus mean node-pairs that are more closely connected in the network, as their effective distance is smaller, and yet more likely for these pairs to be involved more frequently in a SPT. Then, the number of times a directed interaction $i \to j$ is a branch of an SPT is defined as:

$$T_I(i \to j) = \begin{cases} 1 & \text{if } i \to j \in T_i(w_{i\to j}^{-1}) \\ 0 & \text{otherwise} \end{cases}, \tag{2.10}$$

where $T_i$ was the SPT rooted at node $X_i$ and effective distances are used as the support on which pairs are counted.

The salience of an interaction $s_{i\to j}$ is the ratio between the times the interaction appears along an SPT and the total number $N$ of SPTs:

$$s_{i\to j} = \frac{1}{N}\sum_{x_i} T_i(i \to j). \tag{2.11}$$

## 2.5 Information-theoretic Kleiber's Law

The scaling relationship between total outgoing species interaction and entropy of abundance is defined as:

$$OTE \sim H(Ab)^{\Phi}, \tag{2.12}$$

where $OTE$ is the outgoing interaction of a species (eq. 2.6), $H(Ab)$ (see Eq. 2.3) was a species' information content given its abundance over time, and $\Phi$ is the scaling exponent. This was also applied to "populations" (here defined to be at the

taxonomic level of microbial phyla) in order to measure metabolic scaling at the population level (e.g., see Fig.S5).

## 2.6 Phylogenetic Analysis

### 2.6.1 Phylogenetic Tree Creation

Genetic data was unavailable, so the reference phylogenetic tree for this study was inferred from the taxonomic classification (rather than gene sequences) of all OTUs present in the dataset: kingdom, phylum, class, order, family, and species. Consequently, the internal nodes of the tree were the taxonomic class, and the tree-tips were the OTUs; all branch lengths were set to 1. When a taxonomic classification of an OTU was undefined, a unique label was supplied: for example, if the family of OTU $X$ was unidentified, its family was labelled as "family.otu.X". Trees were also created for each community (location), whereon the tips represented only those OTUs observed in that location (ie. abundance $> 0$). Phylogenies were created using the *ape* [90] and *phyloseq* [91] R packages.

### 2.6.2 Phylogenetic $\alpha$ Diversity

Using the reference phylogeny and the species abundances observed over time, I measured the phylogenetic $\alpha$-diversity of a sample in three ways: the number of tips (species richness) $S$; the number of nodes (phylogenetic diversity) $PD$; and the mean pair-wise distances between tips (phylogenetic separation) $PS$. Both $PD$ and $S$ for each sample from a location were measured using the *pd* function from the *picante* R package [92], which implements Faith's $PD$ [93] to calculate phylogenetic diversity. Higher $S$ meant a greater number of unique species, and since a taxonomic label summarizes a set of phylogenetic traits (features), higher $PD$ meant greater diversity of traits within a community.

### 2.6.3 Phylogenetic Distance

To determine $PS$ of a phylogeny, we used the mean pair-wise distances,

$$PR = S^{-1} \sum d_{ij}, \tag{2.13}$$

where $S$ is the number of species, and $d_{i,j}$ is the number of nodes along the path between the tips $OTU_i$ and $OTU_j$. It was calculated using the *ses.mpd* function of the picante R package [92]. Communities with higher $PS$ produced a more dispersed (less clumped) tree, suggesting greater phylogenetic dissimilarity among species within the community. It should be noted that, typically, the phylogenetic distance is calculated as the sum of all pairwise distances between the tips of the tree in contrast to our effective average above.

As each OTU was classified into a phylum $p$, the $PS$ for each phylum within the phylogeny $PS_p$ was calculated as the mean of all distances attributed to that phylum:

$$PS_p = \frac{1}{n} \sum_{p} d_{i_p,j}, \tag{2.14}$$

where $n$ is the number of pairs in the distance matrix and $d_{i_p,j}$ is the phylogenetic separation between OTU $i$ belonging to phylum $p$ and all other OTUs $j$ (see Fig.S4). Phyla with higher $PS_p$ are more genetically unrelated, or functionally dissimilar, to other phyla within their community.

### 2.6.4 Phylogenetic $\gamma$ Diversity

I measured the temporal phylogenetic accumulation within each community up to time $t$ by applying the principle of inclusion-exclusion,

$$\gamma_t = \sum_{1,t} |A_t| - \sum_{1 \le i < j \le t} |A_i \cap A_j| - \ldots + (-1)^{t+1} |A_1 \cap \ldots \cap A_t|, \tag{2.15}$$

to the finite set phylogeny tips (species) and associated branches, $A_1, \ldots, A_t$. This algorithm was executed to obtain $\gamma_{S,t}$, $\gamma_{PD,t}$, and $\gamma_{PR,t}$ by setting the cardinality of each set – i.e., $|A_t|$, $|A_i \cap A_j|$, and $|A_1 \cap \ldots \cap A_t|$ – to its $S$, $PD$, and $PR$ (see Sec.r2.6.2), respectively, depending on the variety of $\gamma$ desired. For example, to calculate $\gamma_{PD}$, the cardinality of each set was measured as the $PD$ of that set.

### 2.6.5 Community Evolutionary Rates

I estimated the rate at which phylogenies accumulate and expand over time using

$$\gamma = T^{\rho}, \tag{2.16}$$

where $\gamma$ is one of $\gamma_S$, $\gamma_{PD}$, or $\gamma_{PR}$, $T$ is time, and $w$ is the slope of the linearized relationship on log-transformed axes (i.e., $log(\gamma$ and $log(T)$).

## 2.7 Eco-Evo Mandala

Inspired by the work of [94] phytoplankton and [95] for microorganisms, we designed the Eco-Evo Mandala (see Figure 4.1A). Mandalas are visualizations of theoretical models or raw data of ecosystems along two or more dimensions and aim to (i) simplify observed complexities (in the physical or metaphysical planes) and (ii) detect ecosystem patterns [96]. Ref. [95] is a recent example of a "foraging Mandala" for aquatic microorganisms, where two axes were used to account for the local environment and individual biological adaptations (related to resource frequency and resource quality). Then, the Mandala allows us to reduce the vast biocomplexity of microbe–environment interactions into a pattern, where areas are distinct for a minimal number of fundamental parameters related to microbial strategies. This has also great relevance for the monitoring and management of aquatic environments by leveraging microbial information.

In this study, the proposed Eco-Evo Mandala is a ternary plot that illustrates how a bacterioplankton community departs from optimality - in a habitat-specific way - given the relationship among its structure ($\epsilon$ as abundance distribution parameter; see Equation (2.1)), function ($\lambda$ as interaction distribution parameter; see Equation (2.1)), and phylogenetic dissimilarity ($D = PS$ as effective phylogenetic distance; see Equation 2.13). The values on each axis are rescaled to the maximum in each series (using $x/x_{max}$, where $x$ is the value and $x_{max}$ is the maximum value observed). In the Mandala, $\epsilon$ and $\lambda$ reflect more ecological processes (such as local speciation, long-range dispersal, and environment-driven interactions), while $D$ reflects more long-term Evolutionary processes (such as local genetic make-up through adaptation and selection).

Communities departed from optimality the further they were plotted from the region reserved for high $D$, low $\lambda$, and low $\epsilon$—in this case, the lower-left third that is the theoretically optimal "healthy" state. Note that this healthy state is an "absolute" optimal (considering high organization $\epsilon$ and $\lambda$, and effective functional

diversity $D$); however, the habitat-specific optimal state may be located in another region of the Mandala, such as the ones we found for the observed habitats in this study.

## 2.8 Chapter Summary

This chapter detailed the materials and methods used to study bacterioplankton community complexity supporting ecosystem health. The diversity of a community was measured by the mean phylogenetic distance among community members; the structure was characterized by the distribution of population abundances; and the function was characterized by the distribution of community interactions. These qualities were combined into a phase-space describing how communities may have departed theoretical optimality. Finally, the state and impact of the environment on community dynamics was measured by the network metrics of the environmental drivers, the covariance of these metrics with those of the bacterioplankton network, and the sensitivity of select community features to uncertainty in the environment. In the following chapters, I present the results of applying these methods to ecological data from the Great Barrier Reef.

# Chapter 3

# The Bacteriome: Eco-evolutionary Signals from a Complex System

## 3.1 Introduction

Much of the planet, including marine ecosystems, are in need of ecohealth indicators to help in their management given the existential threat posed by climate change and other anthropogenic stressors. With improvements in DNA sequencing technology, more work can be done on utilizing the bacteriome as indicators of ecosystem health [31]. These communities of bacteria are foundational members of any ecosystem, providing essential nutrient cycling services contributing to the homeostasis of hosts and ecosystems at large. Furthermore, they are highly responsive to even slight anomalies in ecosystem conditions, such as abnormal temperatures, elevated acidity, or eutrophication. However, their potential as ecosystem indicators is yet to be fully exploited. This is largely because their communities are quite complex, comprised of numerous interacting populations, the biological function of whom little is known [97].

### 3.1.1 Community Trait Dynamics

Instead of focusing on individual species, I propose to analyze the community dynamics of the marine bacteriome, which contain many signals informative of environmental pressure. The color of this ”noise”, in a dynamical sense, reflects the network topology underlying the observed ecological phenomena over time [50, 69]. Traditional bioinformatics evaluates these systems by considering the dynamics of

phylogenetic traits such as taxonomy or species abundance. For example, the diversity and abundances of populations within these communities change with habitat geomorphology and physicochemistry [9, 12, 98–101]. However, the taxonomic resolution at which such analysis is conducted reveals (or hides) other signals about the ecosystem state, because biological traits are differentially conserved along the phylogenetic tree [52]. Others have demonstrated how functional profiles are more informative than taxonomic profiles [53], given that environmental pressure can select for species with competitively advantageous functional potential.

Characteristic of complex systems, the autonomous yet interacting members within the bacteriome can reconfigure and coordinate themselves into evolutionarily optimal configurations in response to environmental pressure. However, there is growing evidence that the range of possible configurations is constrained by ecological laws, such as Taylor's law, Kleiber's Law, and the abundance–size spectrum (Zipf's Law) [54–56], which are highly robust and seem to shift only in the presence of significant environmental pressure. Identifying such patterns in the bacteriome would greatly improve its ability to objectively signal ecosystem anomalies, at multiple scales, and even inform monitoring and management priorities.

### 3.1.2 Entropic Methods for Species Interaction Assessment

There are also entropy-based approaches to species interaction assessment, which can aid ecosystem monitoring. Community metabolic activity result from species interactions (cooperation or competition) which manifest as time-lagged covariations in population abundances. However, the nature of these covariations, particularly their direction, are difficult to identify using correlational approaches. Transfer entropy [70] considers that the two processes may asymmetrically co-predict each other at a certain time-lag, making it more appropriate for species interaction assessment than correlational methods [69, 72, 73]. Therefore, here, I explore the usefulness of TE for ecosystem evaluation given its ability to handle the complexities within biosystems.

Considering these ideas, in this chapter I explore the collective dynamics of the bacteriome for features which may be indicative of ecosystem state (health). Specifically, I search for entropic and power-law patterns that are less likely to be misinterpreted as they unlikely to deviate from theoretical expectations, except in the presence of environmental extremes. Thus I exploit the bacteriome as a complex eco-evolutionary information system.

## 3.2 Characterizing Community Configuration and Coordination

The configuration and coordination of the bacteriome was characterized using the distributions that manifested from its collective dynamics (see Sec. 2.3). Community structure was evaluated by the amount of scale-freeness in the distribution of species' abundances ($\epsilon$, see Sec.2.1): lower values characterized more optimal structural organizations. The river (TR) was considered the most structurally optimal ($\epsilon$ = 1.74), followed by the river mouth ($\epsilon$ = 1.82), with the northernmost marine inshore reef FI showing the poorest structure ($\epsilon$ = 2.85) (see Fig. S1A). Another metric for community and population structure characterization was $\nu$, which described how the variance in the abundance of populations scaled with their mean abundance (see Section 2.2A): higher values indicated more structurally stable communities. At the community level, congruent with Taylor's Law, $\nu$ ranged between 1.5 in the marine inshore reef FI and 1.7 in the river, approaching but not exceeding two (see Fig. S2A). The interaction between species was estimated as the transfer entropy between their respective abundance time series (see Eq.2.4). Given that ecological function is the outcome of species interaction, community function was characterized by the amount of scale-freeness in the distribution of species interactions: lower $\lambda$ signaled more optimal distributions of community interactions. Again, the river was measured to be the most optimal ($\lambda$ = 2.39), followed by the marine inshore reef RI ($\lambda$ = 4.14); the estuarine inshore reef TT3 had the poorest functional organization ($\lambda$ = 14.82) (see Fig.S3).

An inverse linear model provided a better fit between $\nu$ and $\epsilon$ than between $\nu$ and $\lambda$ (see Figure 3.1); linearity between $\epsilon$ and $\lambda$ was also weak (see Figure 4.1B). However, at the population level, the pattern of the relationship among the metrics was much less clear. The top phyla considering $\nu$ and $\epsilon$ were comparable, as similar populations were shared between them (see FigsS2B and S1B); contrarily, the top phyla considering $\lambda$ (see Fig.S3B) were much less comparable to either of the other two. For example, in the marine inshore reef (FI), the same six Phyla were among the top 10 Phyla for $\nu$ and $\epsilon$, but five were shared between $\lambda$ and $\epsilon$. Similarly, in the river (TR), more were shared between $\nu$ and $\epsilon$ than between $\lambda$ and $\nu$ or $\epsilon$. Consequently, from hereon, community structure was evaluated considering $\epsilon$: $\nu$ and $\epsilon$ appear to be encoding similar signals from the community; $\lambda$ measures a different axis of variation; and both $\epsilon$ and $\lambda$ are derived from PDFs.

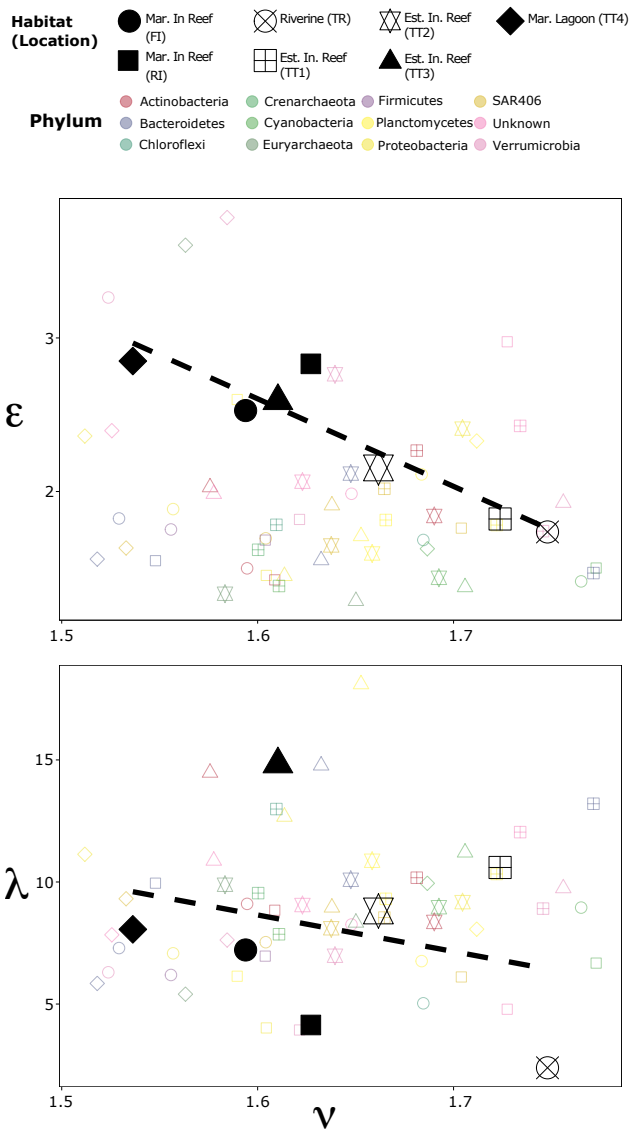


Figure 3.1: **Associations among $\nu$, $\epsilon$, and $\lambda$.** The shape of each point corresponds to the habitat (and location). The large black points are the community averages, while the smaller colored points indicate the values for specific phyla.

The phylum-level distribution of abundances best fit a Zipfian distribution. The mean phylum-level parameters ($|\epsilon|$) were highest in the marine inshore reef site RI ($|\epsilon|$ = 2.43), followed by the lagoon (TT4, $|\epsilon|$ = 2.35) and then the estuarine inshore reef site TT2 ($|\epsilon|$ = 2.06), least of all in the estuarine inshore reef site TT3 ($|\epsilon|$ = 1.72). In all estuarine inshore reefs and one marine inshore reef (FI), Euryarchaeota presented the lowest $\epsilon$ and was thus considered closest to optimal; in the other marine inshore reef (RI), Actinobacteria had the lowest; and for the marine lagoon, Bacteriodetes had the lowest. In the marine lagoon and one marine inshore reef (FI), the phylum with the highest $\epsilon$, and therefore furthest from optimal, was Verrumicrobia, while in the other marine inshore reef (RI), the highest was Euryarchaeota. In the estuarine inshore reef designated TT1, the phylum with the highest exponent was Gemmatimonadetes; in TT2, it was Acidobacteria, and in TT3, it was Firmicutes. (see Fig. S1)

Like the community scale distribution, the phylum-level distribution of TE also best fit an exponential distribution. The mean phylum-level parameters ($|\lambda|$) were highest in the estuarine inshore reef sites TT3, TT2, and TT1 ($|\lambda|$ = 12.8, 10.6, and 10, respectively), followed by the marine inshore reef site FI ($|\lambda|$ = 9.43), the marine lagoon (TT4, $|\lambda|$ = 9.16), and then the marine inshore reef site RI ($|\lambda|$ = 6.61). The phylum closest to interaction optimality (i.e., presenting the lowest $\lambda$) in the estuarine inshore reef TT1 was the Synergistes, while in TT2, it was Chlorofelxi, and in TT3, it was SBR1093; in the marine lagoon, Euryarchaeota had the lowest $\lambda$; for the marine inshore reef FI, the lowest $\lambda$ was for Crearchaeota, while in RI, the lowest was for Verrumicrobia. The phylum furthest from optimal (that is, with the highest $\lambda$) in the marine inshore reef TT1 was Fibrobacteres, while in TT2 and TT3, it was WPS-2; in the marine lagoon, Spirochaetes had the highest $\lambda$; for the marine inshore reef FI, the highest $\lambda$ was for Acidobacteria, while in RI, the highest was for Gemmatimonadetes. (see Fig.S3)

## 3.3 Habitat-specific Phylogenies

Diversity of these communities was measured as species richness $S$, phylogenetic diversity $PD$, and phylogenetic separation $PS$ (see Sec.2.6.2). Species richness, the number of unique species, was highest on average in the marine inshore reefs (FI and RI, $\bar{s}$ = 238, sd = 76.1), followed by the estuarine inshore reefs (TT1, TT2, TT3, $\bar{s}$ = 188, sd = 63.9), the marine lagoon (TT4, $\bar{s}$ = 180, sd = 65.6) and the river

(TR, $\bar{s}$ = 157, sd = 54). Similarly, $PD$, the total branch length (number of nodes considering all branch lengths equalled 1 unit) spanning the phylogenetic tree, was highest in the marine inshore reefs ($\bar{pd}$ = 353, sd = 99.6), followed by the estuarine inshore reef ($\bar{pd}$ = 292, sd = 83.2), the marine lagoon ($\bar{pd}$ = 275, sd = 88.3), and the river ($\bar{pd}$ = 262, sd = 79). The mean $PS$, the mean pair-wise distance among species on the phylogeny, was highest in the marine inshore reefs($PS$ = 7.68, $SES_{mpd}$ = 2.15), followed by the marine lagoon ($PS$ = 7.53, $SES_{mpd}$ = 1.88), the estuarine inshore reef ($PS$ = 7.43, $SES_{mpd}$ = 1.6), and the river ($PS$ = 6.82, $SES_{mpd}$ = 0.33). see (Fig. 3.2C)

### 3.3.1 Diversity-Time Relationship

Given all three metrics, diversity generally increased for all habitats, notwithstanding moments of decline (see Fig.3.2A). The community diversification rates (i.e., diversity-time relationships) were measured as the $S$-time, $PD$-time, and $PS$-time relationships (Fig.3.2A). The rate of species accumulation ($S$-time) and taxa accumulation ($PD$-time) were almost identical (0.457, std. err. = 0.038 and 0.408, std. err. = 0.032, respectively) while the phylogenetic separation rate ($PS$-time) was significantly lower (0.013, std. err. = 0.003). The $S$-time exponent ranged between 0.325 in the marine inshore reefs and 0.8 in the marine lagoon. The $PD$-time relationship exponent ranged between 0.292 in the marine inshore reef and 0.713 in the marine lagoon. The $PS$-time relationship exponent was negative in the river (-0.047), but ranged between 0.0115 in the estuarine inshore reef and 0.029 in the marine lagoon. By plotting the $PS$ against the $S$ of the community at each speciation event using log-transformed axes, I also explored the rate at which community phylogenies separated during speciation events. A speciation event occurs whenever a new species is added to the community. Overall, I observed that separation scales with speciation by an exponent of 0.04, ranging between 0.02 in the river and 0.05 in the marine habitats (both the inshore reef and lagoon). These observations imply that while habitats are diversifying taxonomically, by accumulating unique species, functionally, these communities are diversifying at a much slower rate as many of these new members are closely, phylogenetically related.

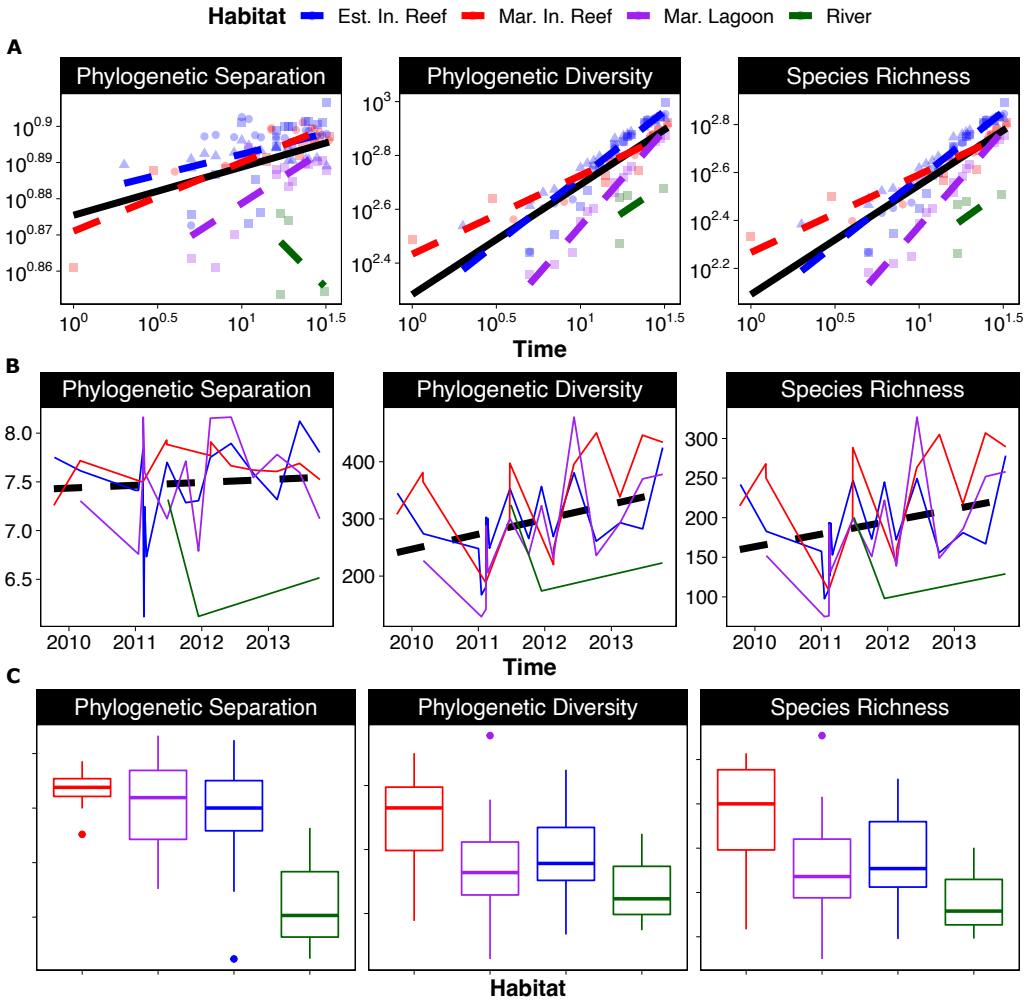


Figure 3.2: **Community Phylogenetic Expansion. A** Cumulative PS, PD, and PR (see Sec.2.6.4) over time in log-log space, where the slope is the growth rate of each value (see Eq.2.16). **B** PS, PD, and SR over time for each habitat. **C** The distribution of PS, PD, and SR values for each habitat.

### 3.3.2 Taxonomies versus Traits

My results not only emphasize that adding new species and adding new functional groups are different processes, but that the latter may be more important for ecohealth indication. Environmental selective pressure constrains the range of possible traits, i.e., phenotypes and functions, for adaptive complex systems like the bacteriome. Yet, I have shown that community phenotype distribution fluctuates more rapidly than community functional groups, suggesting that functional groupings are more robust to environmental variations. This may be because of the high rate of horizontal gene transfer characteristic of bacterial communities. Although genetic information is transferred vertically, from parent to child, bacterial communities have been known to transfer genes horizontally, even among distantly related populations, a feature which facilitates the high functional redundancy also characteristic of these communities. Consequently, others demonstrated how taxonomic diversification occurs within functional groups [53, 102, 103], indicating that they are complementary yet separate axes of variations for communities.

Although all communities displayed patterns approaching theoretical expectations, particularly for $\epsilon$ and $\nu$, the exponent values obtained varied between habitat classes. The biogeography of the global microbiome demonstrates that the geomorphological characteristics, like depth [22, 104], can restrict the presence of species: shallow waters with greater wave forcing tends to have less stable communities than deeper, more settled waters [105, 106]. Additionally, microbialization is known to occur during increased carbon sequestration [107], implicating nutrient gradients in microbiome reorganization. In comparing the results by habitat, the results in this chapter did not decouple geomorphological features from biogeochemical gradients to pinpoint the likely cause of this variation (as was done elsewhere in this study, see Chap.5). Nonetheless, my findings supports the idea that the habitat limits community organization. Altogether, this means that engineering solutions should consider the local habitat features and state instead of taking a more global approach.

## 3.4 Kleiber's Law of Interaction Organization

While further exploring the organization of community function, I discovered a scaling relationship between a species' information content and its total outgoing interactions. A species' information content was quantified as the Shannon entropy of its abundance ($H(Ab)$, see EQ.2.3): rare species with more unstable abundance

had higher information content than dominant species with stable abundance. The total outgoing interactions of a species was measured as its total outgoing transfer entropy ($OTE$, EQ.2.6). In an information-theoretic sense, higher $OTE$ indicates that a species disseminated information into (interacted with) their community at a higher rate than others. A non-linear relationship was found between $OTE$ and $H(Ab)$ across two orders of magnitude or more in a log-scale, indicating a scaling relationship (Fig.3.3). On average, the scaling exponent ($\Phi$, see Eq.2.12) was 0.63 ($\pm 0.182$): the marine lagoon (TT4) presented the steepest relationship ($\Phi$ = 0.814), followed by the marine inshore reefs (FI and RI: $\bar{\Phi}$ = 0.692), the estuarine inshore reefs (TT1, TT2, and TT3: $\bar{\Phi}$ = 0.655), and the river (TR, $\Phi$ = 0.245).

I observed that the Kleiber's exponents of marine sites (the inshore reefs FI and RI, and the lagoon TT4) and the estuarine sites (TT1, TT2, and TT3) differed more between distance-defined categories (i.e., when the bacterioplankton community was categorized solely by its distance from shore) than within categories: $\bar{\Phi}_{mar} = 0.732 \pm 0.071$ and $\bar{\Phi}_{est} = 0.656 \pm 0.063$, respectively. Thus, the uncertainty of a species, phylum or community abundance could predict its interaction rate; rarer species interacted more than dominant species and the potential for this relationship to vary by habitat.

When comparing a population's (i.e., all OTUs categorized within a phyla) relative abundance (see Fig.2.2) with its $\Phi$ (see Fig.3.4), phyla with the highest mean relative abundance (dominant, $> 10\%$) had lower $\Phi$ than low abundance (rare, $< 1\%$) phyla. Populations such as Proteobacteria, Bacteroidetes, and Cyanobacteria (highly abundant in all habitats) consistently had lower $\Phi$ than rare populations (likely habitat-specific) who had among the highest $\Phi$: like SBR1093 and ZB3 in the marine sites; Lentisphaerae in the estuarine sites; and Verrumicrobia in the lagoon. While $\Phi$ for most species remained sublinear, $\Phi$ for rare taxa exceeded 1; the marine habitats (FI and RI) were most likely to have taxa with superlinear exponents that go beyond the theoretical expectation.

I also investigated whether a population's $OTE$ scaled with its abundance to see whether interactions scaled with abundance as well. The $OTE$ of a population was determined by the mean $OTE$ of all species belonging to that population, while the abundance was the mean of the abundances of species belonging to that population. There was evidence of a scaling relationship between the abundance of a population and its $OTE$, although the range of population abundances was small ($< 1$ order of magnitude). The exponent was different than $\Phi$ estimated from the relationship between $OTE$ and $H(Ab)$, but the pattern among the habitats was comparable (Fig.

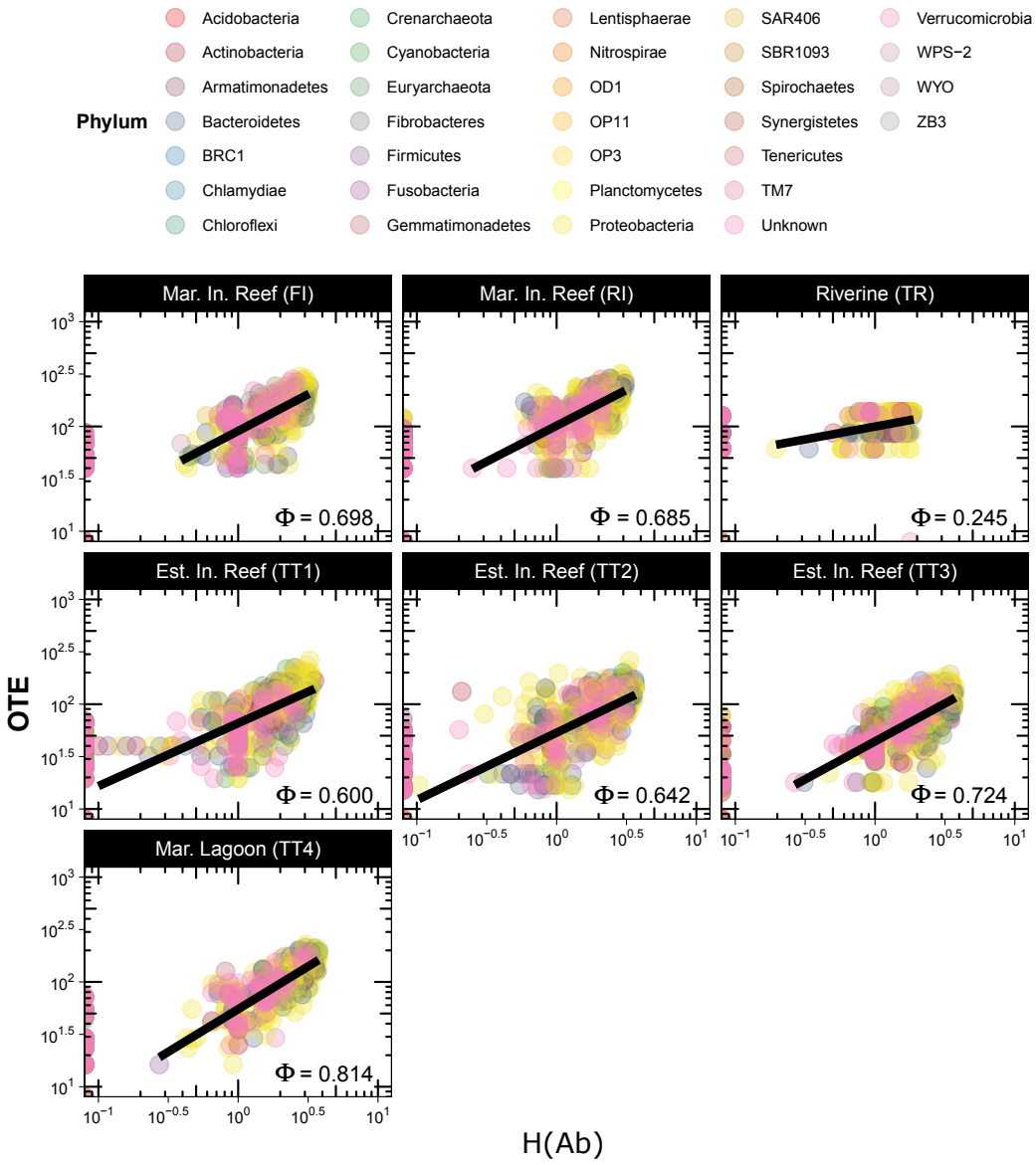


Figure 3.3: **Information theoretic Kleiber's Law.** Each panel illustrates the relationship between the Shannon Entropy of species abundance and their OTE across time within each habitat. Each point represents a species (OTU) and the color indicates its phylum. Black lines are the Kleiber's scaling law, $OTE \sim H(Ab)^{\Phi}$, where $\Phi$ is the scaling exponent characterizing each of the seven communities.

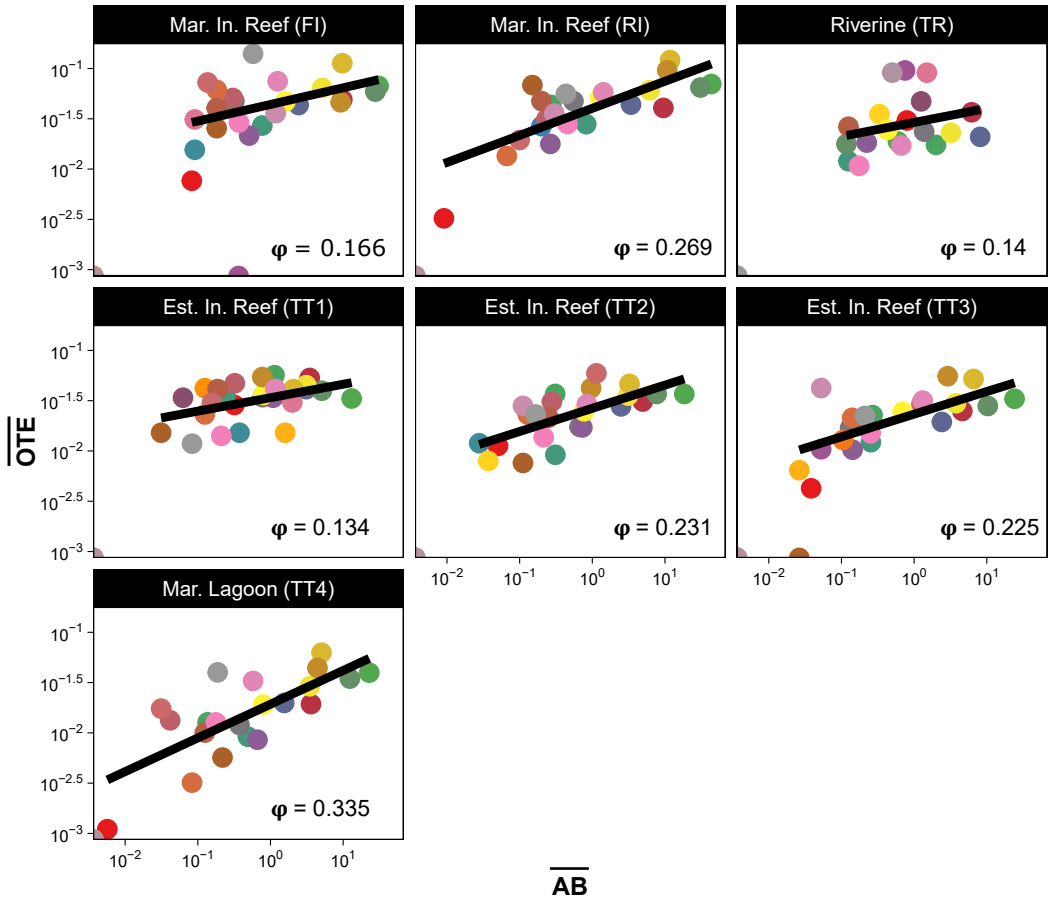


Figure 3.4: **Phylum-scale Kleiber's Law.** Each point represents a phylum and black lines are the best fit of Kleiber's scaling law $\overline{OTE} \sim \overline{Ab}^{\Phi}$ where the mean OTE and abundance is considered among species belonging to the same phylum.

S6): the marine lagoon again presented the steepest relationship ($\Phi$ = 0.335), followed by the marine inshore reefs ($\bar{\Phi}$ = 0.218), the estuarine inshore reefs ($\bar{\Phi}$ = 0.197), and the river ($\Phi$ = 0.14). Appropriate multipliers can transform between species and phylum abundance, and $H(Ab)$ is an assumption-free measurement of the probability uncertainty of species abundance. Therefore, both at the community and population scales, I confirmed that community interactions scale with abundance, and that the elevation of this relationship is habitat-specific.

### 3.4.1 Community Metabolic Scaling Inferred from Community Interactions

The Metabolic Theory of Ecology [62] posits that metabolic rates, governed by Kleiber's Law, predicts most ecological patterns. Metabolic activity is ultimately species interactions, manifested as dependencies between population dynamics, as all biological processes link species via interactions in a network [26, 40, 51, 69, 86]. After quantifying interaction as information flow, I observed that the interaction magnitude of the marine bacterioplankton scales with the unpredictability of its abundance raised to a power exponent $\sim \frac{2}{3}$ (as low as 0.6). The outlier was the river (TR), which could have presented a very small exponent due to data limitation constraining a proper inference of the community interaction network. Kleiber's Law purports that a species metabolic rate will scale proportionally with its mass by $\frac{2}{3}$ when considering the species' volume. [51] found a scaling relationship between species OTE and relative abundance with an exponent $\frac{1}{4}$ when analyzing the human gut microbiome; I too found approximately $\frac{1}{4}$ after exploring the relationship between phylum OTE and abundance. Species (i.e., OTU-level diversity) and phylum showed different interaction magnitudes, but these interactions can be transformed from one to another via proper scaling multipliers.

I propose that this information-theoretic Kleiber's Law (i.e., $OTE \sim H(AB)^{\phi}$) is a superior to biodiversity indicators and the prototypical Kleiber's Law (i.e., $B \sim M^{\alpha}$) as an indicator of eco-environmental condition. As to the former, competitive [108] and cooperative [109, 110] interactions regulates community productivity, explaining how biodiversity loss undermines ecosystem function. However, biodiversity metrics, like species richness, are poorly informative of ecosystem shifts because they do not characterize species collective organization, and changes in richness may not destabilize ecosystem function. As to the latter, the practicality of using the prototypical Kleiber's Law is limited by the need to choose an appropri-

ate metabolite to monitor, quickly yet accurately measure metabolite consumption rates, and measure body mass or volume. All this can be difficult, expensive, and slow *in-situ* at the microbiotic scale, for communities and ecosystems, especially when multiple trophic levels are considered. This requires much less knowledge to construct and is applicable to any trophic scale: all that is required is the spatio-temporal abundance of community members; the rest is derived mathematically. The interpretation would, nonetheless, be the same: variations in the exponent signal eco-environmental anomalies (see Sec.5.4.1). Altogether, these results underscore the notion that the bacteriome's interaction network topology is informative of ecohealth evaluation (as departure from optimal interaction topology) even in the absence of other biological data.

## 3.5 Keystone Species and Interactions

I identified the most important populations and population interactions within the bacteriomes by analyzing the inferred interaction networks. The importance of a population, identified as the taxonomic phyla, was measured by its node centrality (Fig. 3.5). ZB3 and/or SAR406 were among the top 2 most important species for all habitats, excluding the river; and dominant phyla, such as Cyanobacteria and Proteobacteria, were among the most important phyla for all habitats (though not the most important). Link salience was used to identify the most important population interactions within the bacteriomes. For all habitats, the most important interactions (links) were between a population with high node centrality and one with low node centrality, but never between two species of similarly high centrality (Fig. S12). For example, in one marine inshore reef (FI), the most salient links involved ZB3, with the highest node centrality, and Spirochaetes, whose node centrality did not feature among the top 10. In the lagoon the link between SAR406 and Firmictues was high on the list; SAR406 was most central but Firmicutes was not among the most central. Cyanobacteria were among the top 10 most central populations to one estuarine inshore reef (TT3), and formed a top pair with Fusboacteria, which was not among the top 10. Thus, I saw that these bacterioplankton interaction networks were centered around rare taxa that structured the networks via links with peripheral species.

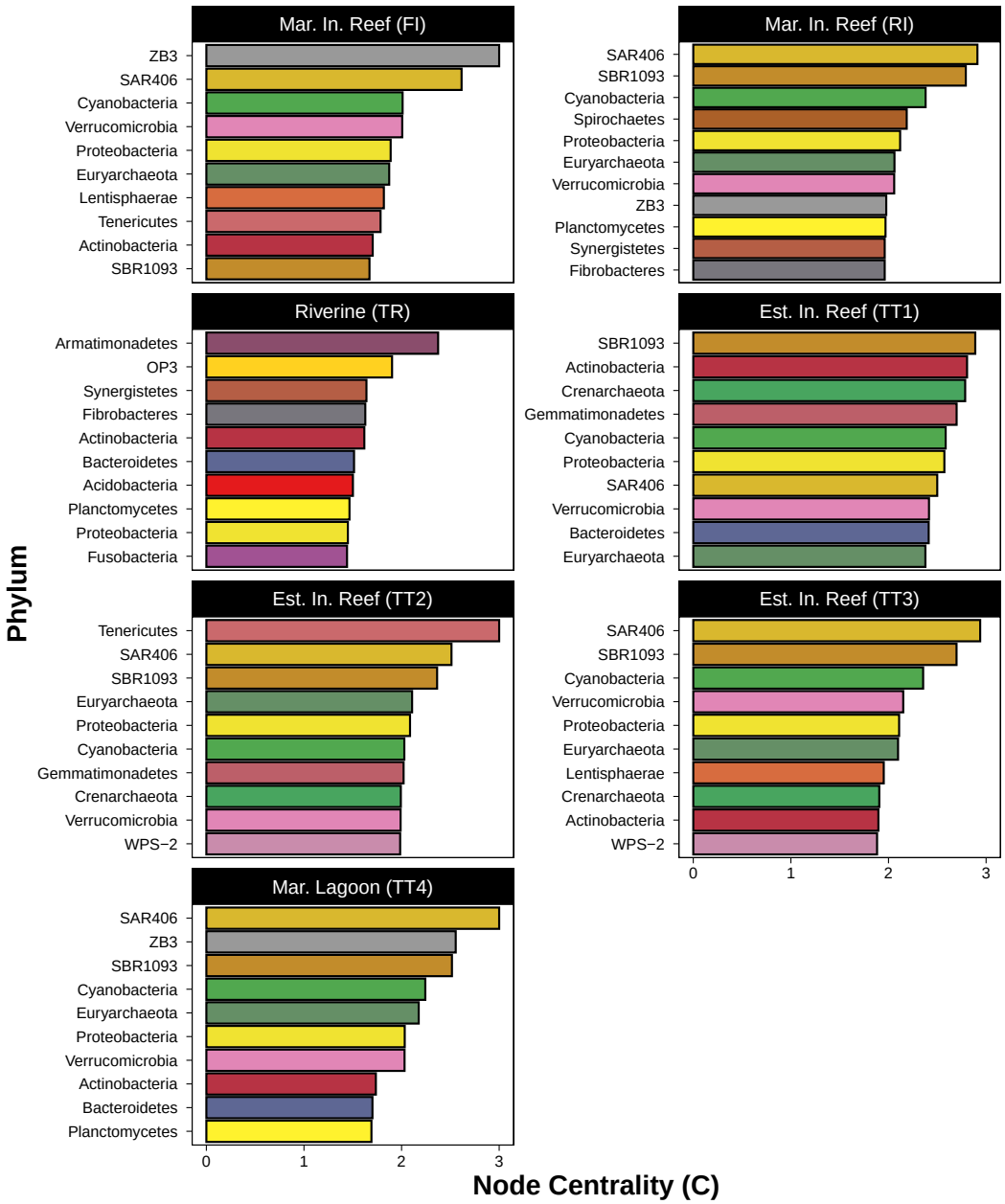


Figure 3.5: **Phylum Importance.** Top 10 important phyla ranked by node centrality for each habitat as in EQ.2.9, where the mean of TEs among phylum-specific species is considered. Consequently, $OTE$, $ITE$, and $k_{tot}$ of a phylum are, respectively, the mean of $OTE$, $ITE$, and $k_{tot}$ of species belonging to each phylum.

### 3.5.1 Habitat-specific Keystone Taxa Sculpt Community Response

Informing habitat-specific interventions requires the identification of keystone species and environmental factors participating in salient interactions that underpin ecosystem structure, function, and stability. By analyzing the topology of the bacteriomes and enviromes, I also identified keystone pairs likely responsible for maintaining ecosystem (theoretical) optimality. Biologically, keystone species [79] typically refers to individual species (nodes) but ignores functionally important interactions (node-pairs). However, these salient interactions underpin ecological functions and are thus equally (if not more) responsible for ecosystem state.

When measured by network link-salience, species involved in these keystone interactions are typically non-dominant (in terms of abundance) but spatio-temporally ubiquitous species (sometimes termed as "core species", e.g., see [111]), shaping community interaction patterns across multiple trophic levels. For example, Phyla such as Proteobacteria, Cyanobacteria, and Bacteroidetes constitute the cosmopolitan bacteriome [111, 112] as they are known to be dominant in marine ecosystems globally [16, 22]; particularly in coastal regions like the GBR studied here [9, 21, 113]. Here, I showed that these central taxa were also topologically central, as they maintained a high number of interactions within their communities. Their numerical dominance led to community (node) centrality because they participated in many broad processes, manifesting as high interconnectedness [79]. However, I observed that comparatively rare species, like ZB-3, WPS-2, and Gemmatimonadetes, were also among the most central species – in some cases more central than dominant species. The presence of rare taxa among the most central seemed habitat driven as, for example, ZB-3 was most central in the marine and lagoon sites but not the estuarine sites, while Gemmatimonadetes was most central in the estuarine sites and not the marine or lagoon sites. Furthermore, the top salient interactions were sometimes rooted at these rare species. Thus, true keystone species are functionally important regardless of their abundance [79, 114], underscoring the importance of rare species who perform niche-specific functions [113, 115]. I propose that dominant species form the structural and functional core of marine microbial communities, but the rare species drive the community's habitat-specific response to environmental pressure. This may be because, as this scaling relationship implies, the rare species have a higher interaction (metabolic) rate, and thus produce a disproportionate impact on community interactions.

## 3.6 Summary

The bacterial component of marine ecosystems – a keystone element into the organization of eukaryotic and viral components of the marine microbiome – has a likely impact into macro-diversity and environmental phenomena, such as algal blooms, primary processes, such as carbon sequestration (via biological and microbial loops), and multitrophic ecosystem health at large. Thus, unraveling its collective dynamics is important for ecosystem evaluation, although it remains a challenging feat. The collective dynamics of the bacteriome, specifically patterns of diversity, structure, and function, hold great potential for eco-health indication [11, 116]. However, this requires that the indicators extracted are robust against minor impacts, at multiple scales, given the typical (local) instability of nature. In this chapter I revealed habitat-specific, scale-free statistical patterns in the diversification, abundance distribution, and function distribution of microbial communities. The main findings are as follows:

- The collective dynamics of the bacteriome contain fractal patterns, typical of self-organizing complex systems. Importantly, these results showed that community structural and functional optimality, though related, represent separate axes; and that care must be taken when selecting optimality metrics as some, like Taylor's Law and Zipf's Law, seem to signal the closely related features of the community.

- Habitats, arguably through their geomorphic organization, determine a likely universal scale-free organization of bacterioplankton communities. This is achieved by constraining the ecological relevance of species much more than their taxonomic identity, leading to a comparatively slow rate of community adaptation, characteristic of resilient systems. Patterns in species diversity remain informative, but they can be improved by including information about species relatedness, thereby emphasizing the importance of phylogenetic trait distribution for ecosystem function.

- Kleiber's Law shows how collective interaction organization grows according to a certain power exponent (close to 2/3) when collective biomass evenness increases, making information flow a suitable proxy for general metabolic activity. These species interactions, underpinning and predicting biodiversity patterns, show a minimum average interaction value that becomes lower

when diversity increases due to the more power-law distribution of interactions. Kleiber's Law implies that community interactions are shaped by dominant species whose entropy is large on average; however, rare or unevenly distributed species with low entropy have the highest directed interactions affecting the collective and community scaling exponent. Rare species, with weak but sensitive interactions, are also identified by the highest node centrality and link salience. The observed variations in the Kleiber's Law exponent are habitat-specific: habitat features, like hydro-geomorphological features, modulate the effect of highly fluctuating environmental drivers such as temperature and nutrients.

- Bacterioplankton communities adapt to their habitat by accumulating new functions slower than new species, suggesting that ecological conservation should focus on community functional diversity much more than species richness.

In the next chapter, I demonstrate how all these ecological signals can be reduced into a multiscale map of ecosystem health.

# Chapter 4

# The Eco-Evo Mandala: Simplifying Bacterioplankton Complexity into Ecohealth Signatures

## 4.1 Introduction

### 4.1.1 Community Signals as Environment-Modulated Noise

In the preceding chapter, I revealed several probabilistic signals emanating from the collective dynamics of the bacteriome. With so many signals to choose from, at so many (taxonomic and spatio-temporal) scales, it is daunting to condense the complexity of the bacteriome into a simple yet useful portfolio of features that can inform ecosystem management. There is also a need to compare the usefulness of genetic and statistical features—such as information entropy – for ecohealth evaluation. Furthermore, these signals can themselves be misleading. For example, one can conclude that a community is optimally configured because there is scale-freeness in its abundance-size spectrum, but this may not necessarily mean that the dynamical stability, its resilience, is high. The adaptability of microbes could lead to a re-structuring that presents as critical, although the rate of interaction among species has irreversibly transitioned, indicating low resilience.

Thus, this chapter is guided by three questions. First, are the collective dynamics of the marine bacteriome reducible to a habitat-specific portfolio? Second, how informative are the genetic measurements for ecosystem health, compared to the statistical features, such as the distribution of abundance or information flow?

Third, does an optimally configured (structured) bacteriome, so characterized by its collective dynamics, imply a resilient bacteriome? To answer the first, the relationship among select statistical distributions of the bacteriome's features extracted previously (see Chapter 3)—its genetic dissimilarity, interaction dynamics, and functional dynamics— are modelled in the new Eco-Evo Mandala. This facilitates answering the second and third questions, because the Mandala enables exploration of the association between the structural stability and dynamic stability of the system, and comparison of the between-state differences along separate axes. This method is applied at the community level as a measure of habitat state, and at the population level to explore species-specific signals.

## 4.2 Signaling Ecosystem Optimality

Characterizations of community genetic relatedness ($D = PS$), configuration ($\epsilon$), and coordination ($\lambda$) were combined into the Eco-Evo Mandala (see Figure4.1) to evaluate ecosystem departure from optimal states. Recall that lower $D$ meant a more diverse community, lower $\epsilon$ meant more a more scale-free configuration, and lower $\lambda$ meant a more scale-free organization of interactions (coordination). Given the values measured from each community (see Chapter 3), the Mandala suggests that the river (TR) is the most optimal location within the region; the estuarine inshore reef TT3 was furthest from optimality for functional (coordination) reasons; and the marine inshore reef FI was furthest from optimal for structural reasons. Additionally, the Mandala illustrates the relative importance of each signal and habitat class to the optimality characterization. The communities varied least along the genetic dissimilarity axis compared to the structural or functional axes of the Mandala (see Figure 4.1B). Genetic dissimilarity ($D$) ranged between 6.5 and 6.9 (sd = 0.125), while $\epsilon$ ranged between 1.74 and 2.85 (sd = 0.459), and $\lambda$ ranged between 2.39 and 14.82 (sd = 4.102). Also, there were greater between-habitat differences than within-habitat differences. The river (TR) was plotted furthest from all other points, the estuarine inshore reefs grouped together, and the marine sites, both the lagoon (TT4) and the inshore reefs (FI and RI), were separated from the rest. Therefore, the signals chosen for ecosystem monitoring were habitat-specific, and that genetic dissimilarity was the least significant of the three when evaluating ecosystem optimality, as communities are more likely to depart from optimality for structural or functional reasons.

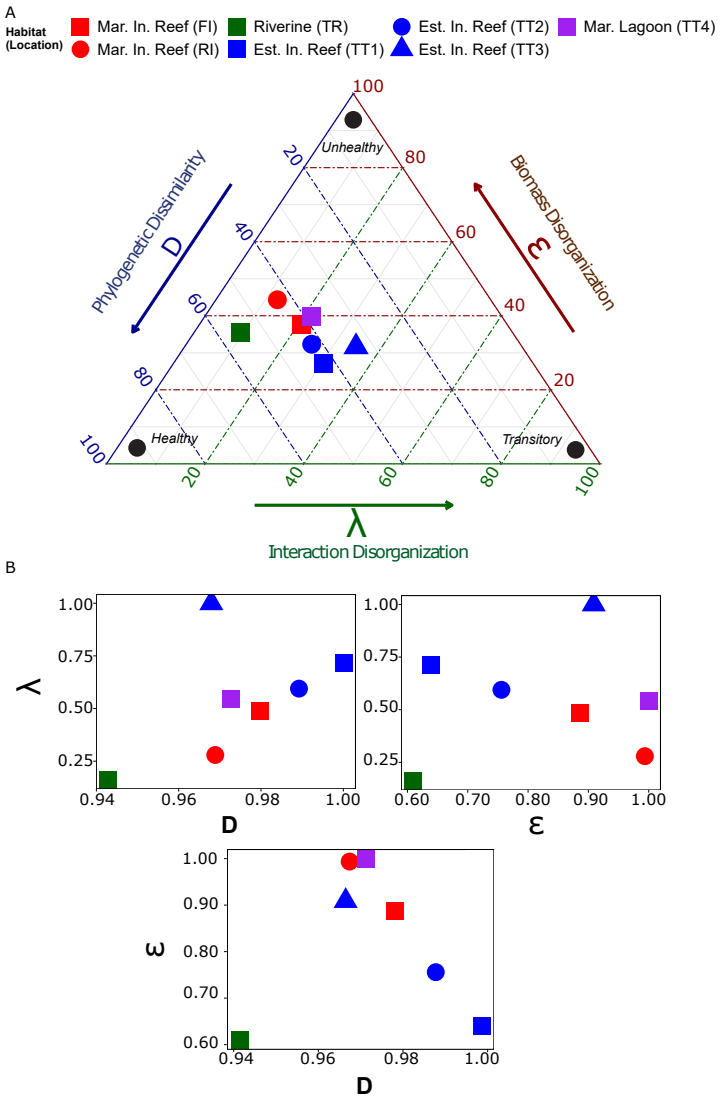


Figure 4.1: **The Eco-Evo Mandala.** (**A**). The community state on the Mandala along the three fundamental traits conveys information about the bacteriome's state (low, moderately, and highly stressed at the vertices as extreme states, and any other state) as divergence from the expected relative optimality. It is read counter-clockwise, the colored arrows exterior to the plot serving as guides: the blue axis $D$ measures the amount of genetic dissimilarity within the network; the green axis $\lambda$ measures the amount of organization in interactions among populations, and the red axis $\epsilon$ measures the amount of structural organization. (**B**) Pair-wise plots of the Mandala axes: $D - \lambda$, $D - \epsilon$, and $\epsilon - \lambda$.

Previous work on the microbiome (bacteriomes, as well as other components) has concluded that community diversity, the distribution of abundances, and the distribution of functions represent separate axes of variation, each capable of signaling the ecosystem state.[1] High $D$ (high diversity) is ideal because it facilitates functional redundancy and creates pools of dormant genes that can be used for atypical environmental conditions [53]. When highly diverse communities encounter stress, they can respond quickly by recruiting previously dormant populations who specialize in the stressed state. It also maintains community resilience despite population loss: functional redundancy ensures that necessary functions are maintained by allowing for population replacements. Low $\epsilon$ and low $\lambda$ are considered optimal because they indicate that the distributions of abundances (configuration) and interactions (coordination underpinning function) are scale-free. Scale-free distributions, most notably in the abundance of biological species, have been associated with healthy biological systems [14, 51]. Biological systems tend to self-organize to optimal states, a phenomenon known as self-organizing criticality [45], which manifest scale-freeness in the distribution of features [117]. Therefore, habitats plotted in the portion of the Mandala reserved for high $D$, low $\lambda$, and low $\epsilon$ were considered closest to optimality because an optimal habitat should present a diverse bacteriome, with a scale-free distribution of functions and abundances.

This notwithstanding, these are theoretical expectations and the distribution or distribution parameter indicative of true optimality for the ocean bacterioplankton in relation to a habitat type remains unknown. A distribution type and its parameter reflect a network; therefore, what is examined here is variation in community assembly organizations that are likely evolutionarily optimal or departing from an optimal state. However, it is possible for the habitat bacteriome to be reorganized into topologies different to scale-free networks. For instance, shallow lagoons without any major structural habitat forcing may lead to feasible optimality of the bacteriome as distributed exponentially; these theoretically suboptimal conditions may also be related to recurrent biogeochemical loads. Consequently, these observations are likely about relative habitat optimality, so future research is needed to identify the network topology of the bacterioplankton (and its distributions) associated with the optimal habitat baseline; this will enable objective quantification of departure from optimality due to diffuse and point-source stress such as climate oscillations (e.g., heatwaves) and nutrient loads.

[1] Note that here "signaling" is purposely used in an information-theoretic perspective to underline the communication between species and environment manifested by time-series data.

### 4.2.1 Habitat Inference from Bacterioplankton Organization

The ubiquity and high motility of many marine bacteria, coupled with the high porosity of habitats could mean that communities are roughly the same, rendering the marine bacteriomes uninformative of habitat type or state, and thus only useful on a planetary scale. However, note greater between-class differences than within-class differences, supporting the observation of others that microbial communities collectively respond to the state and nature of their habitat in characteristic patterns [21, 31, 55, 118, 119].

The geomorphology, hydrodynamics, and water quality (temperature, salinity, nutrient loads, etc.) of these habitats may be responsible for the differentiation of microbial communities, which enables the identification of habitat class from community dynamics. Hydrodynamics coupled with eutrophication can increase microbial biomass in habitats [120], resulting in the relatively high abundance of particular species whose motility is positively impacted [121]. The river and estuarine sites present a shallower, more two-dimensional habitat constantly flushed by wave dynamics, which expedites the entry and exit of species and nutrients [122]. This may explain the comparatively low but exponentially distributed interactions of the estuarine sites, particularly the river mouth TT1, and the high yet less exponentially distributed interactions (indicative of a slower decay in the EPDF of interactions) of the marine inshore reef and marine lagoon. The low residence time of the water in the river and estuarine sites does not allow a stable environment for the build-up of interactions; on the other hand, the marine sites are less hydrodynamically turbulent and accessible from more angles, creating a calmer, more settled environment in which populations may grow and interact. The estuarine environment is more greatly impacted by local perturbation, such as industrial runoff from agricultural lands, which greatly increases nutrient loads, promoting growth—especially that of potentially toxic species – which may explain their less structured interactions and relatively stable abundances [9]. The marine sites are more impacted by global stressors such as temperature and salinity, with limited impact on nutrient availability [9], which would keep its interactions more stable. As such, fluctuations in structural stability and functional dynamics could be able to be indicative of the habitat geomorphology and environmental perturbation. This pattern would need to be explored further in other domains to confirm its validity.

Other ecological processes also impact community organization, not merely habitat geomorphology and environmental stress. Bacteriome dynamics are also

shaped by top-down factors, such as viral dynamics [123]. Recent studies focused on a complete characterization of ocean microbiome dynamics, i.e., the interaction of eukaryotes, prokaryotes, and viruses observed in several habitats around the world, to capture universal dynamical patterns. For instance, recently, [124] considered sampled microbiomes in the Pearl River Delta and found associations with algal blooms and salinity gradients. [125] recently observed during diverse eukaryotic species blooms the Megavirus occurrence in the Uranouchi Inlet (JP), confirming the importance of viruses in the bacterioplankton dynamics. Debates still exist about the causal pathways between eukaryotes, prokaryotes, and viruses, and so further research is needed.

Simply by considering abundance and interaction distribution at the community level, habitat identification, and their potential relative stability or divergence from the theoretically optimal distribution, is possible. Additionally, phylogenetic dissimilarity did not help much with characterizing habitat patterns, although it may aid in understanding factors such as organization processes and effective diversity versus taxonomic diversity. By zooming out from the phylum to the OTU scale—where viral dynamics may appear evident—the uncertainty around habitat statistics, related to coupling abundance and interaction distribution parameters ($\epsilon$ and $\lambda$, respectively), increases because micro-determinants of populations become more important. Overall, the Mandala is capable of signaling habitat-specific dynamics of bacterioplankton communities at the macro-level by using eco-evolutionary traits, and this can also shed light on the importance of micro- versus macro-scale processes with relevance to ecological monitoring (i.e., what, where, and how much to sample bacterioplankton in relation to environmental dynamics).

## 4.3 Intra-Community Health Characterization

Considering the population (phylum) scale characterizations (see Chapter 3, interesting patterns emerged, which were consistent across ecological scales. First, the average behavior of populations (phyla) approximated the behavior of the entire community, resulting in the clustering of phyla around their habitats in the phase space (see Fig.4.2). Additionally, I observed novel inter-population relationships, which could be useful for ecosystem health evaluation. For example, the estuarine inshore reef TT3 presented Cyanobacteria, which were closer to optimality than Proteobacteria, and this community overall was farthest of the seven sites from op-

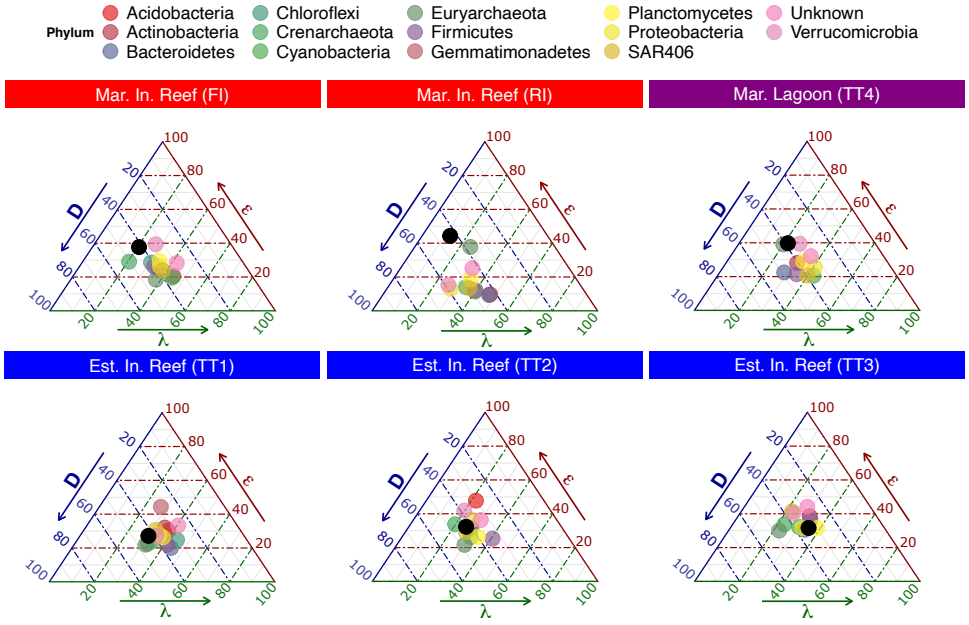


Figure 4.2: **Phylum-level Mandala.** Each panel is an Eco-Evo Mandala for the community sampled at each site. The colored points are the phyla while the black point is the community average. Only those phyla for which all three dimensions were calculable are displayed; the river (TR) is excluded, as its $\lambda$ was not calculable for any of the populations (though it was for the entire population).

timality. Conversely, the marine inshore reef RI presented Proteobacteria closer to optimality than Cyanobacteria, and this community overall was not too far from optimality. Cyanobacteria are widely implicated as problematic for marine ecosystems when they dominate microbial assemblages [126], while Proteobacteria are usually the dominant species characteristic of stable marine environments. Thus, it may be possible to discern the optimality of a habitat by comparing the structural and functional stability of select species.

### 4.3.1 Phylogenetic and Probabilistic Ecohealth Characterization

It has been argued that the inclusion of biological knowledge can improve models of biosystem dynamics [127]. Phenotypic distributions hold information about the microbiome's habitat [52], but care must be taken in selecting which traits to monitor: some, such as genetic dissimilarity, which contributes to diversity (rich-

ness), can be misleading. Additionally, the scale at which this analysis occurs, whether the population or community level, is important. These results show that a community's distributions of abundance and information flow are potentially more informative about the habitat than its diversity, similar to others [128] who demonstrated how structural and dynamical indicators were sufficient for measuring system health. Although some studies have equated diversity with ecosystem health, some have found that diseased habitats can display higher microbial diversity than healthy habitats, particularly because of an increase in viruses [129–131]. These findings also demonstrated that analysis of the information entropy dynamics at the community scale—rather than at the population scale—could inform about the habitat state. This is important because there are many microbes that are as yet uncultured [132, 133], so much about their phylogenetics and ecological function remains unknown. Even if they were cultured, sometimes, in vitro behavior is different from in situ behavior [134]. Therefore, ecosystem health monitoring requires knowledge about the statistical dynamics of the entire community, rather than the taxonomic identity and ecological function of any particular species.

The Eco-Evo Mandala considers the distribution of traits within the entire community to evaluate the ecosystem state. When evaluating ecosystem health, one can focus on identifying keystone species [79], core to their ecological networks and whose exceedance of predefined thresholds destabilizes ecosystems. This justifies focusing on individual populations, because altering one can alter the entire system. However, while this implies the importance of the community interactions, it does not explicitly consider community dynamics; furthermore, it ignores the complexity of these community dynamics. The ecological function of the bacteriome is the result of the interactions of all species within the meta-population, not any one in particular. This complexity means that the effects of one population can be offset by the collective dynamics of the entire system, rendering the information provided by a single population potentially misleading. Consequently, it is better to monitor entire communities when trying to evaluate an ecosystem, rather than specific populations. Here, I emphasize that community dynamics are not only informative of ecosystem health, but sometimes more so than population dynamics.

## 4.4 Summary

Traditional approaches to ecosystem diagnosis can miss habitat-specific features by focusing on abnormal readings in single species (or environmental drivers) rather than community patterns. Community patterns are, for instance, distribution functions of the collective organization of species (abundance, interaction, and phylogenetic dissimilarity from a distance/dissimilarity perspective, where the distance is associated with the autocorrelation function of these features), which may be independent of the spatial, temporal, and biological scale of organization. This invariance across scales is associated with stable or scale-free distribution, manifesting the relative optimality of a community, and the exponent of the distribution is likely habitat-specific (modulated by environmental dynamics) rather than blueprinted by biology. Thus, it is much more appropriate to focus on the collective distribution of abundance and interactions (as information flow due to the data-driven approach of ecosystem monitoring [69]) revealing ecosystem organization, rather than treating any specific bacterium and its abundance in isolation. By decoupling and decoding the genetic, structural, and functional signals emanating from aquatic bacteriomes (here, focusing on the bacterioplankton but extendable to eukaryotic and viral components), I presented the new Eco-Evo Mandala as a ecosystem monitoring tool which utilizes the proximity of the collective dynamics of the bacteriome to characterize ecosystem state (as community organization state). The Mandala is derived from the organization of population abundance, population interactions, and population phylogenetic relatedness. Applied to this data, results suggested that the river was closest to theoretical optimality, but the communities departed optimality in habitat-specific ways. In fact, there were habitat-specific differences in the identity of populations closest to theoretical optimality; and these tended to be largely responsible for the state of the entire community.

The results: (i) confirmed the relatively low importance of community genetic dissimilarity for ecosystem organization (and response, potentially); and (ii) illustrated the complementarity but distinctness between structural and dynamic stability (biomass and interaction distributions, where the latter is much more sensitive to environmental fluctuations); and (iii) explored the habitat specificity of the community response to environmental stress by highlighting the important protective role of structural complexity, such as in marine reef and riverine habitats. In conclusion, it is possible—and preferable—to derive robust ecosystem optimality indicators from the information (probabilistic distributions) extracted from the biocomplexity of

bacterioplankton communities.

In the following chapter, I explore the environment's role in orchestrating the noise emanating from these communities. I introduce the novel concept of the environe, conceptualizing a habitat's biogeochemical features as an assembly of drivers which collectively destabilize the organization with the bacteriome.

# Chapter 5

# The Envirome: Habitat-specific Destabilizers of Bacterioplankton Interaction

## 5.1 Introduction

So far, I have focused on the dynamics of the bacterioplankton from these habitats as indicators of ecosystem state. However, habitats are a complex system of feedbacks between (micro-)biota and their environment: the geophysical features and biogeochemical drivers, such as temperature, salinity, and nutrient gradients (notwithstanding other biota), of their habitat. Especially given the shifts due to climate change, there is a need to understand how the interactions and organization of the biogeochemical drivers, in particular, impact the collective dynamics of the bacteriome. This will elucidate the extent and rate at which instability among this large number of synergetic drivers disorganizes the bacteriome. However, this highly connected nature of environmental drivers (see [9, 22] for example) makes it difficult to evaluate the importance a driver: the impact of any driver will be likely be non-linear, moderated or mediated by another. Furthermore, the habitat geomorphology may exert some constraints on these assemblages, such that the importance of any driver is relative to the habitat in question. Proper examination of the environment requires exploring the collective impact of these drivers on the disorganization of biotic configuration and coordination. In this chapter, I present the results of applying an "enviromic" approach to studying the collective impact of

environmental features on the bacteriome in the Great Barrier Reef. The drivers are treated as interacting nodes in a complex system, and their network characteristics are compared to those of the bacteriome (interaction network of bacterioplankton see Chapter 3) to discuss how the environment is likely disrupting the bacteriome's structure (configuration) and function (coordination). The most important drivers and interactions between drivers are determined given the driver's centrality and the interaction's link salience, respectively, within each habitat's network.

## 5.2 Environmental Impact

The measure of most of the 17 environmental variables were lower in the marine (inshore reef and lagoon) sites than in the estuarine and river sites, except temperature and salinity (see Fig.2.2B). The values for nutrients (nitrogen, phosphorus, and carbon) in the estuarine and river sites experienced a large increase between 2011 and 2012 before retreating to pre-increase levels; this was not observed in the marine sites. The estuarine inshore reefs and river exceeded the highest established threshold values for total nitrogen, chlorophyll A, and suspended particulate matter, but not for total phosphorus; the marine sites (inshore reefs and lagoon) did not exceed these thresholds. Linear regression suggested that most environmental variables were non-linearly correlated with phylogenetic $\alpha$-diversity, excepted for temperature which presented a linear fit, though weak (see Fig.5.1). Notwithstanding this, generally there was an inverse relationship between the environmental variables and diversity; but in the marine inshore reefs there as a positive relationship between the nutrients and phylogenetic $\alpha$-diversity.

## 5.3 Habitat Modulation of the Envirome Impact

The habitat's modulation of the impact of the environment on bacterioplankton metabolism was explored by comparing the topologies of the enviromes and bacteriomes inferred from locations in each habitat class. Enviromes were networks (Fig. S5), inferred using the same process as for the bacteriomes, wherein nodes represented biogeochemical drivers; and edges, their interaction, quantified by TE. Each location's envirome and bacteriome was characterized by the distribution of its interactions (EQ. 2.1). The epdf of interactions in both networks were exponential, (see Figs.5.2B and C) leading us to use $\lambda$ to describe them: lower $\lambda$ meant a more

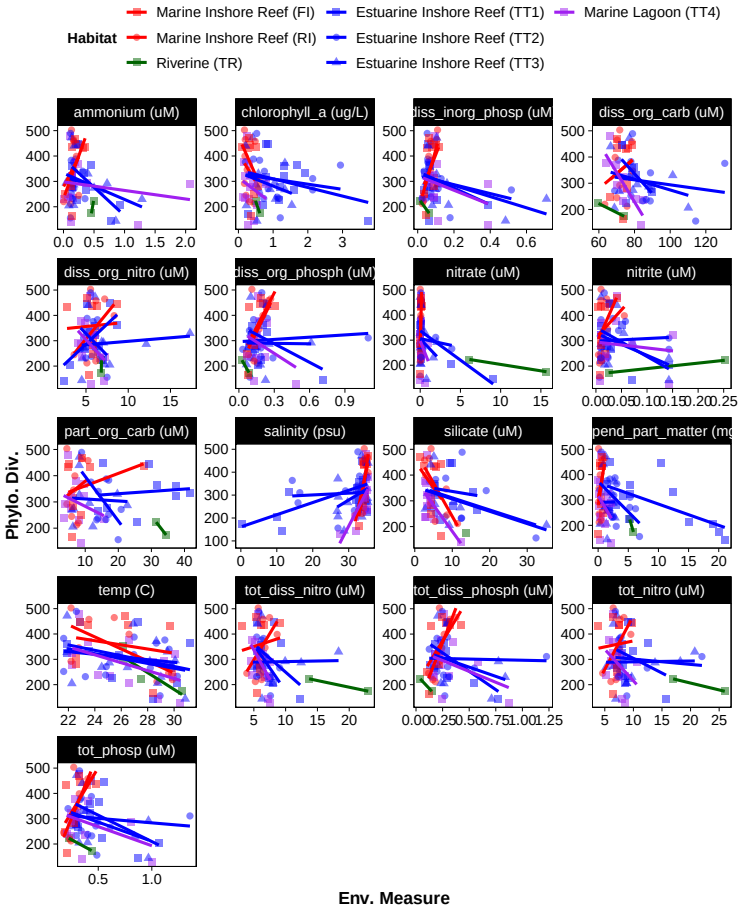


Figure 5.1: **Biogeochemistry and Diversity** The correlation between biogeochemical environmental drivers and community diversity is mostly non-linear.

non-randomly configured network that had one node acting as a hub at the center of, and coordinating, all other interactions. The envirome of the lagoon (TT4) had the lowest exponent ($\lambda$ = 4.81) while the estuarine inshore reefs had the highest ($\bar{\lambda}$ = 5.68, $\lambda_{TT3}$ = 6.62, $\lambda_{TT1}$ = 5.52, and $\lambda_{TT2}$ = 4.87); median values were achieved by the marine inshore reefs ($\bar{\lambda}$ = 5.19, $\lambda_{FI}$=5.12 and $\lambda_{RI}$ = 5.54). I also saw that the bacteriomes from the marine habitats ($\lambda_{RI}$ = 4.1, $\lambda_{FI}$ = 7.2, and $\lambda_{TT4}$ = 8.1) had lower $\lambda$ than the bacteriomes from the estuarine sites ($\lambda_{TT2}$ = 8.8, $\lambda_{TT1}$ = 10.6, and $\lambda_{TT3}$ = 14.8). Thus, I observed that habitats with more non-randomly configured enviromes, like the lagoon and marine inshore reefs, had more non-randomly configured bacteriomes (Fig. 5.2A). To illustrate, notice in Fig. S4 how the bacteriomes from the marine inshore reefs (FI and RI) have few nodes involved in most of the interactions, while the estuarine inshore reefs had interactions spread out among many more nodes. Additionally, I observed that $\lambda$ varied less among the envirome values than among the bacteriomes', and there was habitat clustering on the bacteriome axis: the marine sites (FI, RI, and TT4) were separated from the estuarine sites (Fig 5.2). So, although a randomly configured envirome can destabilize the bacteriome, interactions of the bacteriome were partitioned more by its habitat than by its envirome.

I also explored the enviromes impact on species interactions by comparing a community's $\Phi$ (see 3) with its envirome $\lambda$ (Fig. 5.3B). Here, the envirome demonstrated little explanatory power for species interactions, leaving $\Phi$ to vary by habitat class. In contrast, the distribution of species abundances within each community seemed to relate strongly to $\Phi$. I applied EQ.2.1 to the species abundance and found the distributions were power-law and ranged between 1.75 in the river (TR) to 2.85 in the lagoon (TT4), with the northernmost inshore reef, FI, possessing the median value ($\epsilon$ = 2.5). A lower parameter ($\epsilon$) meant the abundances were more non-randomly configured, and closer to a Zipf's law distribution. Figure 5.3A showed linearity between $\Phi$ and $\epsilon$, therefore I concluded that $\Phi$ was explained well by the configuration of the bacteriomes.

## 5.4 Keystone Eco-environmental Drivers and Interactions

The most important biogeochemical drivers to each habitat were identified as the envirome nodes with the largest centrality (EQ.2.9), which considered the sum of all

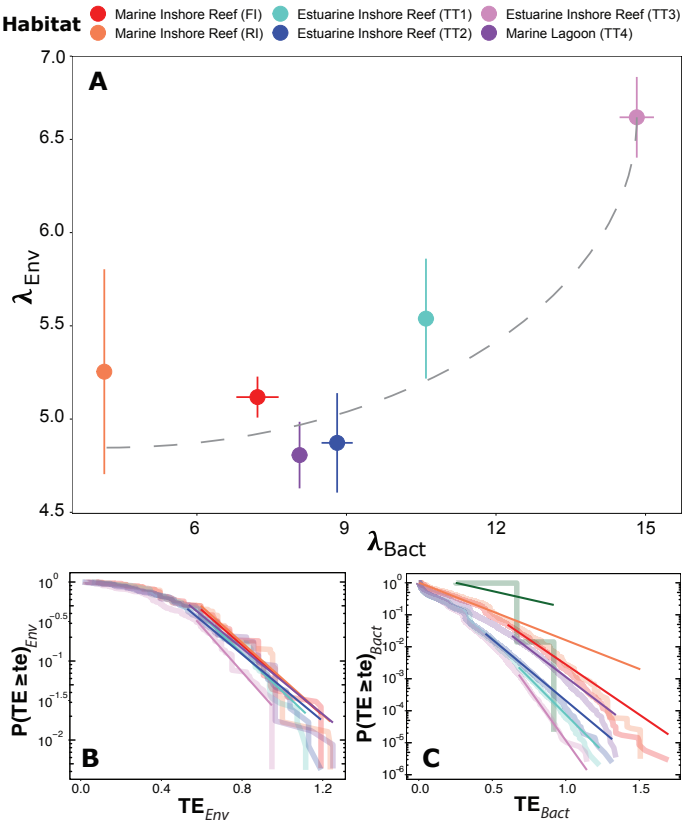


Figure 5.2: **Systemic Environment Impact on Bacteriome Interaction Organization.** (A) relationship between scale parameters of bacteriome and envirome interaction exponential distributions (TE); lower $\lambda$ meant higher organization of interactions (leading to a log-normal or power-law distribution). Thus, the higher the $\lambda$ the larger the systemic impact due to disorganized environmental forcing. Error bars represent the standard error of community-scale TEs. (B) epdf of envirome TE. (C) epdf of bacteriome TE.

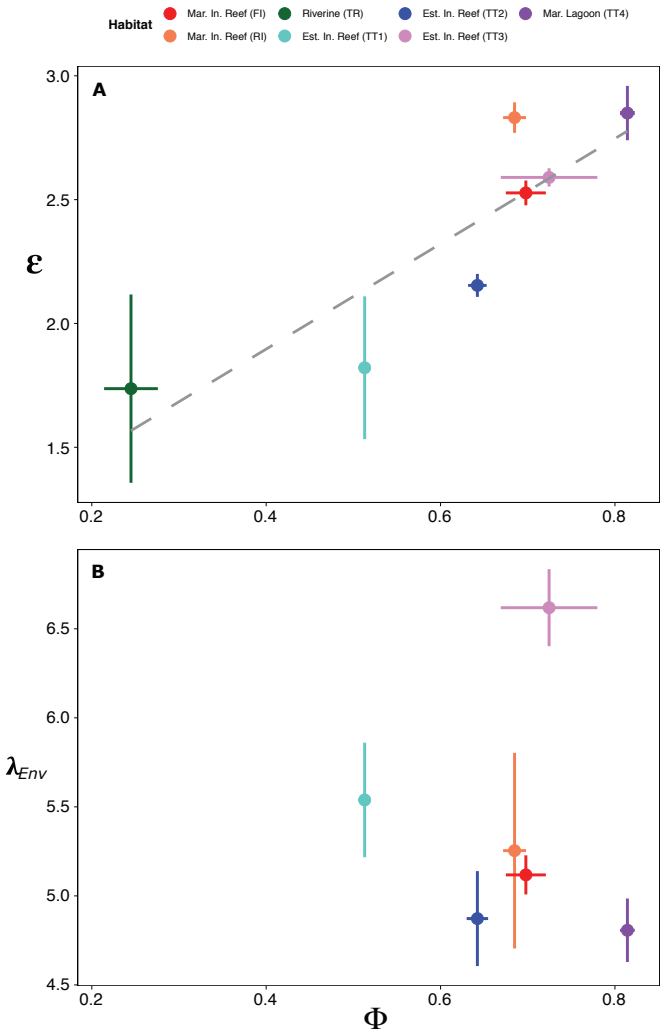


Figure 5.3: **Relationship between Abundance and Envirome Interaction Distribution Parameters and Kleiber's $\Phi$.** (A) $\epsilon$ is the power-law exponent of the abundance EPDF (EQ.2.1) and $\Phi$ is the Kleiber's law scaling exponent (Eq. 2.12). The lower $\epsilon$ the more power-law the distribution, and yet the more stable the abundance dynamics, arguably related to habitat structural features. (B) scale parameter of envirome interaction exponential distribution and $\Phi$. Higher $\lambda$ meant larger systemic impacts due to disorganized environmental forcing. The river (TR) is not shown due to insufficient data to determine $\lambda_{Env}$.

incoming and outgoing interactions, and the number of connections. Link salience (EQ.2.11) was used to identify the most important biogeochemical interactions to each habitat. Chlorophyll A and dissolved organic carbon were among the top 10 drivers of all habitats, and featured among the most important interactions for all habitats except the marine lagoon. Temperature featured among the top 10 drivers of the estuarine inshore reefs, but was not found among the most important drivers within the marine inshore reefs or lagoon (Fig. S13). Figure5.4 shows that the most important interactions in the estuarine inshore reef TT2 were between temperature and salinity. Besides temperature, no other physical driver (e.g., silicate, salinity or suspended particulate matter) was among the top 10 drivers, only nutrients. In particular, nitrite and phosphorus (total dissolved and dissolved organic phosphorus) had the highest centrality, and were the roots of the most important interactions as well. Ammonium was the most central driver and root of the most important interactions in one marine inshore reef (FI); in the other (RI) salinity was the most important driver and formed the root of the most important interactions. In the lagoon, the most important driver was suspended particulate matter, having the highest centrality and rooting the most important interactions. The results suggest that the habitat has some impact on the importance of a biogeochemical driver to the envirome.

### 5.4.1 Habitat-specific Shifts in Bacterioplankton Scaling

Interestingly, although scale-invariant and robust properties like Kleiber's Law ($\phi$) and Zipf's Law ($\epsilon$) differed between habitat classes, the organization among the biogeochemical drivers ($\lambda_{Env}$) did not explain well the distribution of exponents. This agrees with global [76, 135] and regional [9, 136, 137] analyses of marine plankton interactomes that have shown that the habitat filters the impact of biogeochemical drivers on community membership and interactions.

While the assemblage of biogeochemicals (envirome) provide the conditions and materials for metabolism, the habitat geomorphology constrains the availability of those materials and the configuration of the community metabolizing them. For example, river courses provide more of a two-dimensional tract of space for hosts and free-living species to navigate, while pelagic sites – like the marine lagoon – provide a much broader range of options for movement [122]. Habitats that allow more dispersal (because they are open and physically connected to others) minimize the selective effect of environmental pressure on local community configuration

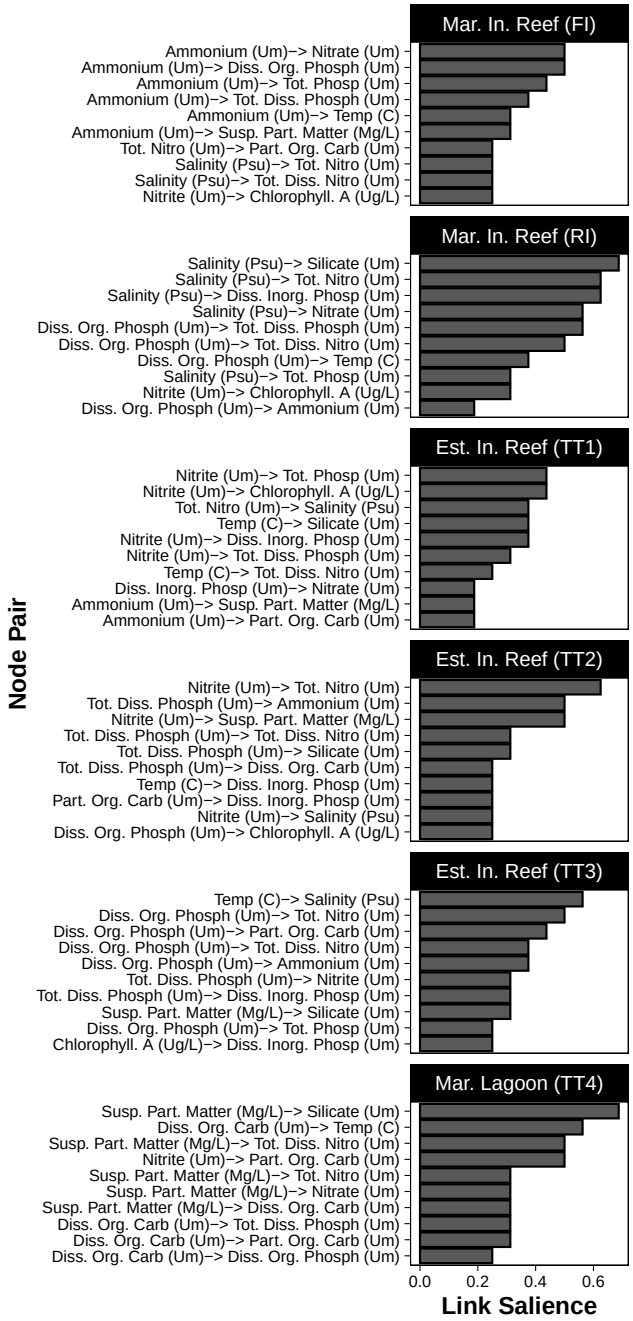


Figure 5.4: **Interaction Importance.** Top 10 phyla pairs ranked by the link salience (see Eq. 2.11) for each habitat type

[138, 139]. Other geophysical features, such as water flow [16] and wind convection [137], impact the movement of hosts and food sources within a site. This may be why the topologies of the bacteriomes, characterized by the distribution of their interactions ($\lambda_{Bact}$), appeared grouped along a habitat gradient: the marine sites were clustered separately from the estuarine sites. When the function of either a highly connected unit (targeted at hubs or keystone species) or a substantially large number of units (system-wide) is destabilized the entire system is in jeopardy of collapse; but random (untargeted), localized disturbances would have minimal impact on the overall system function [140]. Consequently, if bacterioplankton metabolism were only the result of habitat biogeochemistry, then contrary to this observation, the marine lagoon, with the greatest scale-freeness in its envirome and thus capable of system-wide impact, would have had more scale-freeness in its bacteriome than the marine inshore reefs'.

Like the bacteriome, the optimal value of the envirome exponents is not known so, care must be taken when declaring any ecosystem "optimal" or "suboptimal", leaving aside divergence from theoretical scale-free optimality. More research, across a wider spatio-temporal domain, would be needed to determine baseline habitat-specific optimality and deviations due to historical environmental change. This notwithstanding, I suggest that ecosystem managers use this approach to signal the departure of an ecosystem from a theoretically optimal state.

### 5.4.2 Key-stone Drivers

Informing habitat-specific interventions requires the identification of keystone species and environmental factors participating in salient interactions that underpin ecosystem configuration, coordination, and stability. By analyzing the topology of the bacteriomes and enviromes, I also identified keystone pairs likely responsible for maintaining ecosystem (theoretical) optimality. Biologically, keystone species [79] typically refers to individual species (nodes) but ignores functionally important interactions (node-pairs). However, these salient interactions underpin ecological functions and are thus equally (if not more) responsible for ecosystem state. Similarly, important environmental drivers must be identified, considering their individual and joint impacts on ecosystems. This represents a novel approach necessary for more effective habitat-specific ecosystem management.

These results also validate the suitability of the envirome concept for analyzing the relative importance of environmental features. The central environmental

factors differed by habitat proximity to the coast and were suggestive of the environmental history of the habitat. Estuarine sites, highly impacted by agricultural runoff especially during high rainfall periods (like 2011's Cyclone Yasi) [9] presented nutrients as very central to their enviromes, while habitats further from shore, less impacted by agricultural runoff, presented more hydrogeomorphic drivers (e.g., salinity, silicate, and suspended particle matter) as central to their network. Other abiotic factors like salinity and suspended particular matter were more central the further from the river mouth; nutrient concentrations, like total dissolved phosphorus and dissolved organic phosphorus, were more central closer to the river mouth. This may be reflecting the reduced concentration of organic and inorganic nutrients further from shore [21, 113]. Specifically, POC was higher for estuarine inshore reefs, likely related to suspended particle matters as well as DOC (Fig. S2) due to river carbon exports or other carbon upwelling from BC habitats. This can be a signature of higher carbon sequestration for estuarine areas, as expected. Chlorophyll A was also higher for estuarine sites due to the larger presence of submerged vegetation and microalgae. DON is also larger for estuarine sites due to the larger concentration of nutrients from coastal effluxes. Dissolved Inorganic and Organic Nitrogen, and Salinity were the only factors likely determining an increase in local phylum diversity $\alpha$. All other factors were likely associated to a decrease in $\alpha$ (Fig. S3). Future research would be needed to confirm this pattern, but enviromes could be used to indicate habitat type - in the same way indicator species are used to infer habitat type.

The abiotic variables represent an established set of factors that are strongly linked to water quality and impacting biotic community coordination to different degrees. Future research could investigate the optimum portfolio of environmental factors necessary to facilitate accurate ecosystem characterization [141] in short- and long-term, that is useful for ecosystem managers. While this may challenge the network associations observed between envirome and bacteriome topologies, likely it would not impact the metabolic scaling conditioned by habitat type.

## 5.5 Summary

The geomorphology and biogeochemical components of any habitat create a collective pressure on the functioning of biota at all scales. Additionally, much like the biota they impact, these abiotic components interact in a complex system of direct

and indirect impacts. Therefore, a complex system approach is most appropriate for evaluating the organization among these drivers and features, and for identifying critical (keystone) components and interactions responsible for biota disorganization. The enviromic approach is able to provide such information.

- Global and local environmental stress – related to ocean temperature anomalies and river basin/coastal biogeochemical loads – act similarly on any habitats, but the collective environmental impact is related to microbial disorganization dictated by habitat-specific features and historical stress. Abiotic factors have a systemic impact into species interactions mediated by habitats. Yet, habitats sculpt the collective organization of environmental factors (envirome), where the most important ones in estuarine habitats are nutrients, while hydrogeomorphic factors (e.g., salinity, silicate, and suspended particle matter) are more central for marine habitats. The primary impact of microbial disorganization is likely dissolved organic carbon that is the main metabolic flux (in input and output) of microbial interactions; therefore, variability in microbial interaction organization is a signature of ecosystem dysbiosis and potentially informing about variability in microbial pump and climatic impacts.

In the next chapter, I discuss the implications of the Eco-Evo Mandala and the Envirome for ecosystem monitoring and management of the Great Barrier Reef and other regions.

# Chapter 6

# Discussion and Conclusion

## 6.1 Summary of Findings

The aim of this study was to inform the design of an ecosystem monitoring protocol that exploited the collective dynamics of the microbiome. The dynamics of individual species are informative of ecosystem values, given their relationship with specific drivers and functions. However, when examined collectively, patterns in the microbiome can provide even more robust, accurate, and even inexpensive readings of ecosystem parameters. In Chapter 1, I presented multiple patterns of community organization and coordination derived from a statistical analysis of abundance profiles that facilitated ecosystem categorization, as possibly healthy, transitory, or unhealthy, at multiple scales. These patterns signal the departure of community features from theoretical optima, namely, scale-free distributions typical of self-organizing complex systems. Focusing on the information entropy within these dynamics revealed a novel scaling relationship which could be used to pinpoint populations undermining community function. In Chapter 2, I displayed the Eco-Evo Mandala that serves as a multiscale, probabilistic map of habitat states. In order to understand how environmental drivers collectively destabilize these biotic patterns, I introduced the idea of the Envirome in Chapter 3, a network of interacting biogeochemical drivers whose distributions describe the environment's stability, and whose node and link centrality define driver importance.

In this chapter, considering the theory and findings of this work, I present some recommendations for policy, research, and technology aimed at safeguarding marine ecosystems. I discuss what I see as some of the shortcomings of current policy,

and how the findings of my research facilitate a more affordable and sustainable approach to ecosystem monitoring and management, that can be applied both regionally and globally. I outline areas of future research that remain unresolved despite the evidence herein provided, and new questions which emerged. Finally, I describe how my work supports eco-enginnering solutions.

## 6.2 Enhancing Monitoring Protocol

The Great Barrier Reef Marine Authority (GBRMA) recently has seen the need to include the marine microbiome in its monitoring and management toolkit [37]. Among the recommendations are to combine microbial community and environmental parameters in the search for microbial indicators of ecohealth; establish and test diagnostic protocols based on these indicators; and integrate these protocols into the reef monitoring program. This approach requires robust microbial baselines established after a rigorous sampling exercise utilizing cutting edge DNA sequencing technology.

This work represents an important step forward in the search for accurate yet parsimonious indicators of ecosystem states, focused on marine ecosystems. Traditional approaches to ecohealth assessment fail to decouple site-specific and universal habitat-related ecological features; and they may incorrectly estimate fluctuations of ecological communities over time, particularly when comparing different habitats. Limitations of traditional approaches are also related to the consideration of abnormal temporal values of species abundance or environmental factors one-at-a-time, rather than probabilistic community patterns and their temporal persistence. For example, scale-free stability of ecosystem functions, related to the optimal collective organization of species interactions, indicate ecosystem states meaningful of optimal eco-environmental organization over space and time [50]. Therefore, it is more appropriate to focus on the collective distribution of abundance and interactions, rather than labeling any specific bacterium or its abundance as ”safe” and ”toxic”. My work does not ignore the dynamics of individuals or populations, rather it evaluates the parts with respect to the whole. This not only ensures greater accuracy when trying to characterize the state of a community and its environment, but it also informs how the dynamics of the parts relate to the whole, filling out a multiscale map of systems dynamics.

Similarly, individual environmental drivers should not be the focus of analyses,

given that their organization (or disorganization) determines the collective environmental stress to which biological communities respond. In fact, the GBRMA implements a Cumulative Impact Policy [142] which considers the direct, indirect and consequential impacts and the incremental and compounding effects of these impacts over time, including past, present and reasonably foreseeable future pressures. The Envirome approach is well suited to support the execution of this policy. In this work, I emphasize the greater importance that the collective organization – of the coupled microbiome and envirome – plays in defining ecosystem health as divergence from optimal ecosystem function, in contrast to analyses of a single bacterium or environmental factor in isolation.

The GBRMA monitors select ecosystem values to inform assessments of the area. My work presents new values worth inclusions in the current management approach because they implicitly measure the others. For example, phylogenetic separation rate ($\rho$) and the abundance-size spectrum ($\epsilon$) impact community biodiversity, function, and ultimately ecosystem services. As metabolism underpins functions and constrains many other community dynamics, ecologists can characterize the state of an ecosystem and alterations of its functioning using the info-theoretic Kleiber's Law ($\phi$) herein described. From the ranking of outward information flow, ecologists and environmental scientists can identify problematic taxa and environmental drivers that require close monitoring and control to preserve and enhance ecosystem function. Finally, my work establishes objective baselines derived from universal patterns, albeit theoretical, which can be applied at multiple spatial and temporal scales.

## 6.3 Future Research

Fraser and Hill [143] rightly point out that limiting the microbiome to only data from these Bacteria kingdom would be an error, because the microbiome also includes microscopic viruses, eukaryotes, and fungi. Unfortunately, this work only uses data from the Bacteria and Archaea kingdoms, notwithstanding some as-yet unclassified taxa, and lacks data on viral dynamics—and other biotic and abiotic factors, potentially—and disallowing the quantification of these other important features. Characterization of bacterioplankton patterns alone remains informative due to the centrality of bacteria in many ecosystem functions. This notwithstanding, future research needs to explore these findings among more kingdoms, either

independently or jointly. For example, it would be quite interesting to see how microeukaryotes compare to bacteria in terms of the distribution of their interactions and abundances, or whether the slow phylogenetic separation rate observed among bacteria applies to viruses as well. Importantly, this work explores patterns and not processes, based on the data available, along three main phenotypic axes reflecting abundance distribution, abundance-based interactions, and phylogenetics (ecological and evolutionary community traits). Thus, the model and the Mandala can be extended to other domains easily: the whole microbiome (eukaryotic and viral communities too) rather than only being used for prokaryotes.

Building on this more inclusive data, a future research direction could also expand to multiple trophic levels. Environmental DNA technologies now enables the tracking of all life-forms that ever occupied a habitat. The analytical approaches described here could be used to explore the big data obtainable from n=environmental DNA (eDNA) sequencing. In so doing, we would be one step closer to truly confirming the existing of ecological patterns which represent universal laws. Doing so is critical to our development of accurate eco-engineering solutions.

## 6.4 Eco-engineering Solution

Greater efficiency and more data enable unprecedented spatio-temporal granularity, which can be used to facilitate eco-monitoring cyber-technologies. For example, during the Covid-19 pandemic, researchers were able to use the big-data available from social media to perform spatio-temporal forecasts of disease spread and health care pressure. A similar signalling system could be designed to provide now- and forecasts of broad ecosystem states.

There are several microbiome related remediation protocols currently in development and showing success. Microbiome transplantation has shown promise as a means of alleviating human illness [144]. There are also designer microbiomes, specially crafted to metabolize pollutants, correct nutrient fluxes thereby improving the resistance and resilience of ecosystems, or provide novel services, such as new biofuels, foods, or antibiotics [145]. Recent work outlined the simple community assembly rules necessary to design communities capable of metabolizing a single type of organic compound [146], but this is likely to become more complicated when many multiple compounds need to be degraded. The main contribution of my work is a modelling approach which can be applied at multiple scales because

it uses scale-free features of the microbiome network. This will make it easier to forecast how the introduction of new species, or how the increase in population size of existing species, could alter ecosystem function.

## 6.5 Conclusion

The present state of many ecosystems is difficult to measure objectively, and the future states remain unpredictable, despite our increasing knowledge about the supporting patterns and processes. However, we must continue to search for signatures of healthy systems, and signals of undesirable transitions. Keystone members and drivers are critical, by definition, however the collective behavior of ecologically important communities provides the best opportunity for ecohealth indication. In this work, I extracted multiple, robust signals of ecological state that manifest because communities trend toward evolutionarily optimal configurations and coordination. These collective dynamics, such as the abundance-size spectrum and information-abundance scaling, constrain the response of communities to environmental stress and reflect other salient habitat-specific features. By using these metrics alone, we are able to identify likely habitat geomorphology and biogeochemical organization. This collective, systems approach was also shown to be useful in discerning and explaining how the environment disturbs ecosystem service provision. Understanding processes remains important for more precise control over interactions, but this doing so will become more infeasible as the volume of data to extract from nature increases. This big eco-data will require probabilistic, info-theoretic methods, as described here, if we truly wish to convert the noises from nature into meaningful symphonies.

# Bibliography


[1] W. Steffen, K. Richardson, J. Rockström, S. E. Cornell, I. Fetzer, E. M. Bennett, R. Biggs, S. R. Carpenter, W. De Vries, C. A. De Wit *et al.*, "Planetary boundaries: Guiding human development on a changing planet," *Science*, vol. 347, no. 6223, p. 1259855, 2015.

[2] M. J. Fasham, *Ocean biogeochemistry: the role of the ocean carbon cycle in global change*. Springer Science & Business Media, 2003.

[3] G. A. McKinley, A. R. Fay, N. S. Lovenduski, and D. J. Pilcher, "Natural variability and anthropogenic trends in the ocean carbon sink," *Annual review of marine science*, vol. 9, no. 1, pp. 125–150, 2017.

[4] P. J. Webster, "The role of hydrological processes in ocean-atmosphere interactions," *Reviews of Geophysics*, vol. 32, no. 4, pp. 427–476, 1994.

[5] A. K. Aufdenkampe, E. Mayorga, P. A. Raymond, J. M. Melack, S. C. Doney, S. R. Alin, R. E. Aalto, and K. Yoo, "Riverine coupling of biogeochemical cycles between land, oceans, and atmosphere," *Frontiers in Ecology and the Environment*, vol. 9, no. 1, pp. 53–60, 2011.

[6] G. B. R. M. P. Authority, *Great Barrier Reef Region Strategic Assessment: Strategic assessment report*. GBRMPA, Townsville, 2014.

[7] G. De'ath, K. E. Fabricius, H. Sweatman, and M. Puotinen, "The 27–year decline of coral cover on the great barrier reef and its causes," *Proceedings of the National Academy of Sciences*, vol. 109, no. 44, pp. 17 995–17 999, 2012.

[8] J. R. Zaneveld, D. E. Burkepile, A. A. Shantz, C. E. Pritchard, R. McMinds, J. P. Payet, R. Welsh, A. M. Correa, N. P. Lemoine, S. Rosales *et al.*,

"Overfishing and nutrient pollution interact with temperature to disrupt coral reefs down to microbial scales," *Nature communications*, vol. 7, no. 1, pp. 1–12, 2016.

[9] F. E. Angly, C. Heath, T. C. Morgan, H. Tonin, V. Rich, B. Schaffelke, D. G. Bourne, and G. W. Tyson, "Marine microbial communities of the great barrier reef lagoon are influenced by riverine floodwaters and seasonal weather events," *PeerJ*, vol. 4, p. e1511, 2016.

[10] T. P. Hughes, J. T. Kerry, M. Álvarez-Noriega, J. G. Álvarez-Romero, K. D. Anderson, A. H. Baird, R. C. Babcock, M. Beger, D. R. Bellwood, R. Berkelmans *et al.*, "Global warming and recurrent mass bleaching of corals," *Nature*, vol. 543, no. 7645, pp. 373–377, 2017.

[11] A. R. Harborne, A. Rogers, Y.-M. Bozec, and P. J. Mumby, "Multiple stressors and the functioning of coral reefs," *Annual Review of Marine Science*, 2017.

[12] D. Varkey, S. Mazard, T. C. Jeffries, D. J. Hughes, J. Seymour, I. T. Paulsen, and M. Ostrowski, "Stormwater influences phytoplankton assemblages within the diverse, but impacted sydney harbour estuary," *PloS one*, vol. 13, no. 12, 2018.

[13] L. G. Wilkins, M. Leray, A. O'Dea, B. Yuen, R. S. Peixoto, T. J. Pereira, H. M. Bik, D. A. Coil, J. E. Duffy, E. A. Herre *et al.*, "Host-associated microbiomes drive structure and function of marine ecosystems," *PLoS biology*, vol. 17, no. 11, p. e3000533, 2019.

[14] R. F. Heneghan, I. A. Hatton, and E. D. Galbraith, "Climate change impacts on marine ecosystems through the lens of the size spectrum," *Emerging Topics in Life Sciences*, vol. 3, no. 2, pp. 233–243, 2019.

[15] E. J. Howells, A. G. Bauman, G. O. Vaughan, B. C. Hume, C. R. Voolstra, and J. A. Burt, "Corals in the hottest reefs in the world exhibit symbiont fidelity not flexibility," *Molecular ecology*, vol. 29, no. 5, pp. 899–911, 2020.

[16] G. Salazar, F. M. Cornejo-Castillo, V. Benítez-Barrios, E. Fraile-Nuez, X. A. Álvarez-Salgado, C. M. Duarte, J. M. Gasol, and S. G. Acinas, "Global diversity and biogeography of deep-sea pelagic prokaryotes," *The ISME journal*, vol. 10, no. 3, pp. 596–608, 2016.

[17] J. Kleinteich, F. Hildebrand, M. Bahram, A. Y. Voigt, S. A. Wood, A. D. Jungblut, F. C. Küpper, A. Quesada, A. Camacho, D. A. Pearce *et al.*, "Pole-to-pole connections: similarities between arctic and antarctic microbiomes and their vulnerability to environmental change," *Frontiers in Ecology and Evolution*, vol. 5, p. 137, 2017.

[18] S. Sunagawa, S. G. Acinas, P. Bork, C. Bowler, D. Eveillard, G. Gorsky, L. Guidi, D. Iudicone, E. Karsenti, F. Lombard *et al.*, "Tara oceans: towards global ocean ecosystems biology," *Nature Reviews Microbiology*, vol. 18, no. 8, pp. 428–445, 2020.

[19] S. Cao, W. Zhang, W. Ding, M. Wang, S. Fan, B. Yang, A. Mcminn, M. Wang, B.-b. Xie, Q.-L. Qin *et al.*, "Structure and function of the arctic and antarctic marine microbiota as revealed by metagenomics," *Microbiome*, vol. 8, no. 1, pp. 1–12, 2020.

[20] M. G. Cardozo-Mino, E. Fadeev, V. Salman-Carvalho, and A. Boetius, "Spatial distribution of arctic bacterioplankton abundance is linked to distinct water masses and summertime phytoplankton bloom dynamics (fram strait, 79° n)," *Frontiers in microbiology*, vol. 12, 2021.

[21] P. R. Frade, B. Glasl, S. A. Matthews, C. Mellin, E. A. Serrão, K. Wolfe, P. J. Mumby, N. S. Webster, and D. G. Bourne, "Spatial patterns of microbial communities across surface waters of the great barrier reef," *Communications biology*, vol. 3, no. 1, pp. 1–14, 2020.

[22] S. Sunagawa, L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti *et al.*, "Structure and function of the global ocean microbiome," *Science*, vol. 348, no. 6237, 2015.

[23] D. G. Bourne, K. M. Morrow, and N. S. Webster, "Insights into the coral microbiome: underpinning the health and resilience of reef ecosystems," *Annual Review of Microbiology*, vol. 70, 2016.

[24] B. Glasl, G. J. Herndl, and P. R. Frade, "The microbiome of coral surface mucus has a key role in mediating holobiont health and survival upon disturbance," *The ISME journal*, vol. 10, no. 9, pp. 2280–2292, 2016.

[25] B. Norrman, U. L. Zwelfel, C. S. Hopkinson Jr, and F. Brian, "Production and utilization of dissolved organic carbon during an experimental diatom bloom," *Limnology and Oceanography*, vol. 40, no. 5, pp. 898–907, 1995.

[26] A. Z. Worden, M. J. Follows, S. J. Giovannoni, S. Wilken, A. E. Zimmerman, and P. J. Keeling, "Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes," *Science*, vol. 347, no. 6223, 2015.

[27] M. A. Moran, "The global ocean microbiome," *Science*, vol. 350, no. 6266, 2015.

[28] P. G. Falkowski, T. Fenchel, and E. F. Delong, "The microbial engines that drive earth's biogeochemical cycles," *science*, vol. 320, no. 5879, pp. 1034–1039, 2008.

[29] J. A. Fuhrman, "Microbial community structure and its functional implications," *Nature*, vol. 459, no. 7244, pp. 193–199, 2009.

[30] J. A. Fuhrman, J. A. Cram, and D. M. Needham, "Marine microbial community dynamics and their ecological interpretation," *Nature Reviews Microbiology*, vol. 13, no. 3, pp. 133–146, 2015.

[31] B. Glasl, D. G. Bourne, P. R. Frade, T. Thomas, B. Schaffelke, and N. S. Webster, "Microbial indicators of environmental perturbations in coral reef ecosystems," *Microbiome*, vol. 7, no. 1, pp. 1–13, 2019.

[32] J. A. Fuhrman, I. Hewson, M. S. Schwalbach, J. A. Steele, M. V. Brown, and S. Naeem, "Annually reoccurring bacterial communities are predictable from ocean conditions," *Proceedings of the National Academy of Sciences*, vol. 103, no. 35, pp. 13 104–13 109, 2006.

[33] B. Glasl, N. S. Webster, and D. G. Bourne, "Microbial indicators as a diagnostic tool for assessing water quality and climate stress in coral reef ecosystems," *Marine Biology*, vol. 164, no. 4, p. 91, 2017.

[34] B. Glasl, D. G. Bourne, P. R. Frade, and N. S. Webster, "Establishing microbial baselines to identify indicators of coral reef health," *Microbiology Australia*, vol. 39, no. 1, pp. 42–46, 2018.

[35] S. Roitman, F. J. Pollock, and M. Medina, "Coral microbiomes as bioindicators of reef health," in *Population Genomics: Marine Organisms*. Springer, 2018, pp. 39–57.

[36] C. Astudillo-García, S. M. Hermans, B. Stevenson, H. L. Buckley, and G. Lear, "Microbial assemblages and bioindicators as proxies for ecosystem health status: potential and limitations," *Applied microbiology and biotechnology*, vol. 103, no. 16, pp. 6407–6421, 2019.

[37] N. Webster and H. Gorsuch, *Monitoring additional values within the Reef 2050 Integrated Monitoring and Reporting Program: final report of the microbes expert group*. Great Barrier Reef Marine Park Authority, 2020.

[38] C. J. Patrick, K. E. McCluney, A. Ruhi, A. Gregory, J. Sabo, and J. H. Thorp, "Multi-scale biodiversity drives temporal variability in macrosystems," *Frontiers in Ecology and the Environment*, vol. 19, no. 1, pp. 47–56, 2021.

[39] A. S. Hahn, K. M. Konwar, S. Louca, N. W. Hanson, and S. J. Hallam, "The information science of microbial ecology," *Current opinion in microbiology*, vol. 31, pp. 209–216, 2016.

[40] P. A. Marquet, R. A. Quiñones, S. Abades, F. Labra, M. Tognelli, M. Arim, and M. Rivadeneira, "Scaling and power-laws in ecological systems," *Journal of Experimental Biology*, vol. 208, no. 9, pp. 1749–1769, 2005.

[41] J. Ladyman, J. Lambert, and K. Wiesner, "What is a complex system?" *European Journal for Philosophy of Science*, vol. 3, no. 1, pp. 33–67, 2013.

[42] A.-L. Barabási and E. Bonabeau, "Scale-free networks," *Scientific american*, vol. 288, no. 5, pp. 60–69, 2003.

[43] I. Rodriguez-Iturbe and A. Rinaldo, *Fractal river basins: chance and self-organization*. Cambridge University Press, 2001.

[44] M. Martinello, J. Hidalgo, A. Maritan, S. Di Santo, D. Plenz, and M. A. Munoz, "Neutral theory and scale-free neural dynamics," *Physical Review X*, vol. 7, no. 4, p. 041071, 2017.

[45] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality," *Physical review A*, vol. 38, no. 1, p. 364, 1988.

[46] J. Hidalgo, J. Grilli, S. Suweis, M. A. Munoz, J. R. Banavar, and A. Maritan, "Information-based fitness and the emergence of criticality in living systems," *Proceedings of the National Academy of Sciences*, vol. 111, no. 28, pp. 10 095–10 100, 2014.

[47] G. Zimatore, M. Tsuchiya, M. Hashimoto, A. Kasperski, and A. Giuliani, "Self-organization of whole-gene expression through coordinated chromatin structural transition," *Biophysics Reviews*, vol. 2, no. 3, p. 031303, 2021.

[48] S. Nikolov, E. Yankulova, O. Wolkenhauer, and V. Petrov, "Principal difference between stability and structural stability (robustness) as used in systems biology," *Nonlinear Dynamics, Psychology, and Life Sciences*, vol. 11, no. 4, pp. 413–33, 2007.

[49] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk, and G. Sugihara, "Early-warning signals for critical transitions," *Nature*, vol. 461, no. 7260, pp. 53–59, 2009.

[50] J. Li and M. Convertino, "Temperature increase drives critical slowing down of fish ecosystems," *PloS one*, vol. 16, no. 10, p. e0246222, 2021.

[51] ——, "Optimal microbiome networks: macroecology and criticality," *Entropy*, vol. 21, no. 5, p. 506, 2019.

[52] J. B. Martiny, S. E. Jones, J. T. Lennon, and A. C. Martiny, "Microbiomes in light of traits: a phylogenetic perspective," *Science*, vol. 350, no. 6261, 2015.

[53] S. Louca, L. W. Parfrey, and M. Doebeli, "Decoupling function and taxonomy in the global ocean microbiome," *Science*, vol. 353, no. 6305, pp. 1272–1277, 2016.

[54] S. Zaoli, A. Giometto, E. Marañón, S. Escrig, A. Meibom, A. Ahluwalia, R. Stocker, A. Maritan, and A. Rinaldo, "Generalized size scaling of metabolic rates based on single-cell measurements with freshwater phytoplankton," *Proceedings of the National Academy of Sciences*, vol. 116, no. 35, pp. 17 323–17 329, 2019.

[55] Y. Wang, J. Pan, J. Yang, Z. Zhou, Y. Pan, and M. Li, "Patterns and processes of free-living and particle-associated bacterioplankton and archaeaplankton

communities in a subtropical river-bay system in south china," *Limnology and Oceanography*, vol. 65, pp. S161–S179, 2020.

[56] K. K. Cavender-Bares, A. Rinaldo, and S. W. Chisholm, "Microbial size spectra from natural and nutrient enriched ecosystems," *Limnology and Oceanography*, vol. 46, no. 4, pp. 778–789, 2001.

[57] M. Kleiber *et al.*, "Body size and metabolism," *Hilgardia*, vol. 6, no. 11, pp. 315–353, 1932.

[58] J. K. L. da Silva, G. J. Garcia, and L. A. Barbosa, "Allometric scaling laws of metabolism," *Physics of Life Reviews*, vol. 3, no. 4, pp. 229–261, 2006.

[59] L. Von Bertalanffy, "Quantitative laws in metabolism and growth," *The quarterly review of biology*, vol. 32, no. 3, pp. 217–231, 1957.

[60] P. S. Dodds, D. H. Rothman, and J. S. Weitz, "Re-examination of the "3/4-law" of metabolism," *Journal of Theoretical Biology*, vol. 209, no. 1, pp. 9–27, 2001.

[61] V. M. Savage, J. F. Gillooly, W. H. Woodruff, G. B. West, A. P. Allen, B. J. Enquist, and J. H. Brown, "The predominance of quarter-power scaling in biology," *Functional Ecology*, vol. 18, no. 2, pp. 257–282, 2004.

[62] J. H. Brown, J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West, "Toward a metabolic theory of ecology," *Ecology*, vol. 85, no. 7, pp. 1771–1789, 2004.

[63] S. Zaoli, A. Giometto, A. Maritan, and A. Rinaldo, "Covariations in ecological scaling laws fostered by community dynamics," *Proceedings of the National Academy of Sciences*, vol. 114, no. 40, pp. 10 672–10 677, 2017.

[64] A. M. Segura and G. Perera, "The metabolic basis of fat tail distributions in populations and community fluctuations," *Frontiers in Ecology and Evolution*, vol. 7, p. 148, 2019.

[65] E. Hacisuleyman, L. A. Goff, C. Trapnell, A. Williams, J. Henao-Mejia, L. Sun, P. McClanahan, D. G. Hendrickson, M. Sauvageau, D. R. Kelley *et al.*, "Topological organization of multichromosomal regions by the long intergenic noncoding rna firre," *Nature structural & molecular biology*, vol. 21, no. 2, pp. 198–206, 2014.

[66] J. R. Banavar, A. Maritan, and A. Rinaldo, "Size and form in efficient transportation networks," *Nature*, vol. 399, no. 6732, pp. 130–132, 1999.

[67] J. R. Banavar, M. E. Moses, J. H. Brown, J. Damuth, A. Rinaldo, R. M. Sibly, and A. Maritan, "A general basis for quarter-power scaling in animals," *Proceedings of the National Academy of Sciences*, vol. 107, no. 36, pp. 15 816–15 820, 2010.

[68] A. Maritan, R. Rigon, J. Banavar, and A. Rinaldo, "Network allometry," *Geophysical Research Letters*, vol. 29, no. 11, pp. 3–1, 2002.

[69] J. Li and M. Convertino, "Inferring ecosystem networks as information flows," *Scientific reports*, vol. 11, no. 1, pp. 1–22, 2021.

[70] T. Schreiber, "Measuring information transfer," *Physical review letters*, vol. 85, no. 2, p. 461, 2000.

[71] C. M. Cullen, K. K. Aneja, S. Beyhan, C. E. Cho, S. Woloszynek, M. Convertino, S. J. McCoy, Y. Zhang, M. Z. Anderson, D. Alvarez-Ponce *et al.*, "Emerging priorities for microbiome research," *Frontiers in microbiology*, vol. 11, p. 136, 2020.

[72] A. Carr, C. Diener, N. S. Baliga, and S. M. Gibbons, "Use and abuse of correlation analyses in microbial ecology," *The ISME journal*, vol. 13, no. 11, pp. 2647–2655, 2019.

[73] A. R. Coenen and J. S. Weitz, "Limitations of correlation-based inference in complex virus-microbe communities," *MSystems*, vol. 3, no. 4, pp. e00 084–18, 2018.

[74] R. Logares, I. M. Deutschmann, P. C. Junger, C. R. Giner, A. K. Krabberød, T. S. Schmidt, L. Rubinat-Ripoll, M. Mestre, G. Salazar, C. Ruiz-González *et al.*, "Disentangling the mechanisms shaping the surface ocean microbiota," *Microbiome*, vol. 8, no. 1, pp. 1–17, 2020.

[75] F. Polazzo, S. K. Roth, M. Hermann, A. Mangold-Döring, A. Rico, A. Sobek, P. J. Van den Brink, and M. C. Jackson, "Combined effects of heatwaves and micropollutants on freshwater ecosystems: Towards an integrated assessment of extreme events in multiple stressors research," *Global change biology*, vol. 28, no. 4, pp. 1248–1267, 2022.

[76] G. Lima-Mendez, K. Faust, N. Henry, J. Decelle, S. Colin, F. Carcillo, S. Chaffron, J. C. Ignacio-Espinosa, S. Roux, F. Vincent *et al.*, "Determinants of community structure in the global plankton interactome," *Science*, vol. 348, no. 6237, 2015.

[77] A. P. Teixeira, J. M. Dias, N. Carinhas, M. Sousa, J. J. Clemente, A. E. Cunha, M. von Stosch, P. M. Alves, M. J. Carrondo, and R. Oliveira, "Cell functional enviromics: unravelling the function of environmental factors," *BMC systems biology*, vol. 5, no. 1, pp. 1–16, 2011.

[78] M. Ushio, "Interaction capacity underpins community diversity," *BioRxiv*, 2020.

[79] S. Banerjee, K. Schlaeppi, and M. G. van der Heijden, "Keystone taxa as drivers of microbiome structure and functioning," *Nature Reviews Microbiology*, vol. 16, no. 9, pp. 567–576, 2018.

[80] A. Thompson, C. Lønborg, P. Costello, J. Davidson, M. Logan, M. Furnas, K. Gunn, M. Liddy, M. Skuza, S. Uthicke *et al.*, "Marine monitoring program: Annual report of aims activities 2013-2014. inshore water quality and coral reef monitoring," 2014.

[81] C. S. Gillespie, "Fitting heavy tailed distributions: the powerlaw package," *arXiv preprint arXiv:1407.3492*, 2014.

[82] L. Taylor and I. Woiwod, "Temporal stability as a density-dependent species characteristic," *The Journal of Animal Ecology*, pp. 209–224, 1980.

[83] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[84] J. T. Lizier, "Jidt: an information-theoretic toolkit for studying the dynamics of complex systems," *Frontiers in Robotics and AI*, vol. 1, p. 11, 2014.

[85] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social networks*, vol. 32, no. 3, pp. 245–251, 2010.

[86] M. Novak, J. D. Yeakel, A. E. Noble, D. F. Doak, M. Emmerson, J. A. Estes, U. Jacob, M. T. Tinker, and J. T. Wootton, "Characterizing species

interactions to understand press perturbations: what is the community matrix?" *Annual Review of Ecology, Evolution, and Systematics*, vol. 47, pp. 409–432, 2016.

[87] E. Delmas, M. Besson, M.-H. Brice, L. A. Burkle, G. V. Dalla Riva, M.-J. Fortin, D. Gravel, P. R. Guimarães Jr, D. H. Hembry, E. A. Newman *et al.*, "Analysing ecological networks of species interactions," *Biological Reviews*, vol. 94, no. 1, pp. 16–36, 2019.

[88] D. Grady, C. Thiemann, and D. Brockmann, "Robust classification of salient links in complex networks," *Nature communications*, vol. 3, no. 1, pp. 1–10, 2012.

[89] M. Fontenla, *optrees: Optimal Trees in Weighted Graphs*, 2014, r package version 1.0. [Online]. Available: https://CRAN.R-project.org/package=optrees

[90] E. Paradis and K. Schliep, "ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R," *Bioinformatics*, vol. 35, pp. 526–528, 2019.

[91] P. J. McMurdie and S. Holmes, "phyloseq: an r package for reproducible interactive analysis and graphics of microbiome census data," *PloS one*, vol. 8, no. 4, p. e61217, 2013.

[92] S. Kembel, P. Cowan, M. Helmus, W. Cornwell, H. Morlon, D. Ackerly, S. Blomberg, and C. Webb, "Picante: R tools for integrating phylogenies and ecology," *Bioinformatics*, vol. 26, pp. 1463–1464, 2010.

[93] D. P. Faith, "Conservation evaluation and phylogenetic diversity," *Biological conservation*, vol. 61, no. 1, pp. 1–10, 1992.

[94] R. Margalef, "Life-forms of phytoplankton as survival alternatives in an unstable environment," *Oceanologica acta*, vol. 1, no. 4, pp. 493–509, 1978.

[95] V. I. Fernandez, Y. Yawata, and R. Stocker, "A foraging mandala for aquatic microorganisms," *The ISME Journal*, vol. 13, no. 3, pp. 563–575, 2019.

[96] P. M. Glibert, "Margalef revisited: a new phytoplankton mandala incorporating twelve dimensions, including nutritional physiology," *Harmful Algae*, vol. 55, pp. 25–30, 2016.

[97] A. M. Thomas and N. Segata, "Multiple levels of the unknown in microbiome research," *BMC biology*, vol. 17, no. 1, pp. 1–4, 2019.

[98] N. Li, H. Zhao, G. Jiang, Q. Xu, J. Tang, X. Li, J. Wen, H. Liu, C. Tang, K. Dong *et al.*, "Phylogenetic responses of marine free-living bacterial community to phaeocystis globosa bloom in beibu gulf, china," *Frontiers in microbiology*, vol. 11, p. 1624, 2020.

[99] L. Weber, P. González-Díaz, M. Armenteros, V. M. Ferrer, F. Bretos, E. Bartels, A. E. Santoro, and A. Apprill, "Microbial signatures of protected and impacted northern caribbean reefs: changes from cuba to the florida keys," *Environmental microbiology*, vol. 22, no. 1, pp. 499–519, 2020.

[100] J. Li, X. Jiang, Z. Jing, G. Li, Z. Chen, L. Zhou, C. Zhao, J. Liu, and Y. Tan, "Spatial and seasonal distributions of bacterioplankton in the pearl river estuary: The combined effects of riverine inputs, temperature, and phytoplankton," *Marine pollution bulletin*, vol. 125, no. 1-2, pp. 199–207, 2017.

[101] T. C. Jeffries, M. L. Schmitz Fontes, D. P. Harrison, V. Van-Dongen-Vogels, B. D. Eyre, P. J. Ralph, and J. R. Seymour, "Bacterioplankton dynamics within a large anthropogenically impacted urban estuary," *Frontiers in microbiology*, vol. 6, p. 1438, 2016.

[102] S. Louca, S. Jacques, A. P. Pires, J. S. Leal, D. S. Srivastava, L. W. Parfrey, V. F. Farjalla, and M. Doebeli, "High taxonomic variability despite stable functional structure across microbial communities," *Nature ecology & evolution*, vol. 1, no. 1, pp. 1–12, 2016.

[103] S. Louca, M. F. Polz, F. Mazel, M. B. Albright, J. A. Huber, M. I. O'Connor, M. Ackermann, A. S. Hahn, D. S. Srivastava, S. A. Crowe *et al.*, "Function and functional redundancy in microbial systems," *Nature ecology & evolution*, vol. 2, no. 6, pp. 936–943, 2018.

[104] C. Bienhold, L. Zinger, A. Boetius, and A. Ramette, "Diversity and biogeography of bathyal and abyssal seafloor bacteria," *PloS one*, vol. 11, no. 1, p. e0148016, 2016.

[105] L. J. Hamdan, R. B. Coffin, M. Sikaroodi, J. Greinert, T. Treude, and P. M. Gillevet, "Ocean currents shape the microbiome of arctic marine sediments," *The ISME journal*, vol. 7, no. 4, pp. 685–696, 2013.

[106] P. E. Galand, M. Potvin, E. O. Casamayor, and C. Lovejoy, "Hydrography shapes bacterial biogeography of the deep arctic ocean," *The ISME journal*, vol. 4, no. 4, pp. 564–576, 2010.

[107] A. F. Haas, M. F. Fairoz, L. W. Kelly, C. E. Nelson, E. A. Dinsdale, R. A. Edwards, S. Giles, M. Hatay, N. Hisakawa, B. Knowles *et al.*, "Global microbialization of coral reefs," *Nature microbiology*, vol. 1, no. 6, pp. 1–7, 2016.

[108] D. S. Maynard, T. W. Crowther, and M. A. Bradford, "Competitive network determines the direction of the diversity–function relationship," *Proceedings of the National Academy of Sciences*, vol. 114, no. 43, pp. 11 464–11 469, 2017.

[109] M. J. McInerney, J. R. Sieber, and R. P. Gunsalus, "Syntrophy in anaerobic global carbon cycles," *Current opinion in biotechnology*, vol. 20, no. 6, pp. 623–632, 2009.

[110] B. Schink, "Energetics of syntrophic cooperation in methanogenic degradation," *Microbiology and molecular biology reviews*, vol. 61, no. 2, pp. 262–280, 1997.

[111] A. Shade and J. Handelsman, "Beyond the venn diagram: the hunt for a core microbiome," *Environmental microbiology*, vol. 14, no. 1, pp. 4–12, 2012.

[112] P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon, "The human microbiome project," *Nature*, vol. 449, no. 7164, pp. 804–810, 2007.

[113] K. Richa, C. Balestra, R. Piredda, V. Benes, M. Borra, A. Passarelli, F. Margiotta, M. Saggiomo, E. Biffali, R. Sanges *et al.*, "Distribution, community composition, and potential metabolic activity of bacterioplankton in an urbanized mediterranean sea coastal zone," *Applied and environmental microbiology*, vol. 83, no. 17, pp. e00 494–17, 2017.

[114] M. E. Power, D. Tilman, J. A. Estes, B. A. Menge, W. J. Bond, L. S. Mills, G. Daily, J. C. Castilla, J. Lubchenco, and R. T. Paine, "Challenges in the quest for keystones: identifying keystone species is difficult—but essential to understanding how loss of species will affect ecosystems," *BioScience*, vol. 46, no. 8, pp. 609–620, 1996.

[115] A. Jousset, C. Bienhold, A. Chatzinotas, L. Gallien, A. Gobet, V. Kurm, K. Küsel, M. C. Rillig, D. W. Rivett, J. F. Salles *et al.*, "Where less may be more: how the rare biosphere pulls ecosystems strings," *The ISME journal*, vol. 11, no. 4, pp. 853–862, 2017.

[116] C. E. Lawson, W. R. Harcombe, R. Hatzenpichler, S. R. Lindemann, F. E. Löffler, M. A. O'Malley, H. García Martín, B. F. Pfleger, L. Raskin, O. S. Venturelli *et al.*, "Common principles and best practices for engineering microbiomes," *Nature Reviews Microbiology*, vol. 17, no. 12, pp. 725–741, 2019.

[117] F. Bauchinger, "Self-organized criticality in the gut microbiome," Ph.D. dissertation, uniwien, 2015.

[118] M. J. van Oppen, P. Bongaerts, P. Frade, L. M. Peplow, S. E. Boyd, H. T. Nim, and L. K. Bay, "Adaptation to reef habitats through selection on the coral animal and its associated microbiome," *Molecular ecology*, vol. 27, no. 14, pp. 2956–2971, 2018.

[119] A. Apprill, L. G. Weber, and A. E. Santoro, "Distinguishing between microbial habitats unravels ecological complexity in coral microbiomes," *Msystems*, vol. 1, no. 5, 2016.

[120] P. M. Meirelles, A. C. Soares, L. Oliveira, L. Leomil, L. R. Appolinario, R. B. Francini-Filho, R. L. de Moura, R. T. de Barros Almeida, P. S. Salomon, G. M. Amado-Filho *et al.*, "Metagenomics of coral reefs under phase shift and high hydrodynamics," *Frontiers in microbiology*, vol. 9, p. 2203, 2018.

[121] B. S. Lambert, V. I. Fernandez, and R. Stocker, "Motility drives bacterial encounter with particles responsible for carbon export throughout the ocean," *Limnology and Oceanography Letters*, vol. 4, no. 5, pp. 113–118, 2019.

[122] J. L. Falter, R. J. Lowe, Z. Zhang, and M. McCulloch, "Physical and biological controls on the carbonate chemistry of coral reef waters: effects of metabolism, wave forcing, sea level, and geomorphology," *PloS one*, vol. 8, no. 1, 2013.

[123] C.-E. T. Chow, D. Y. Kim, R. Sachdeva, D. A. Caron, and J. A. Fuhrman, "Top-down controls on bacterial community structure: microbial network analysis of bacteria, t4-like viruses and protists," *The ISME journal*, vol. 8, no. 4, pp. 816–829, 2014.

[124] T. Zhang and P. Wang, "The response of bacterial communities to organic matter in the surface sediment of the pearl river estuary," *Plankton and Benthos Research*, vol. 17, no. 1, pp. 45–56, 2022.

[125] F. Prodinger, H. Endo, Y. Takano, Y. Li, K. Tominaga, T. Isozaki, R. Blanc-Mathieu, Y. Gotoh, T. Hayashi, E. Taniguchi *et al.*, "Year-round dynamics of amplicon sequence variant communities differ among eukaryotes, imitervirales and prokaryotes in a coastal ecosystem," *FEMS microbiology ecology*, vol. 97, no. 12, p. fiab167, 2021.

[126] R. P. Rastogi, D. Madamwar, and A. Incharoensakdi, "Bloom dynamics of cyanobacteria and their toxins: environmental health impacts and mitigation strategies," *Frontiers in microbiology*, vol. 6, p. 1254, 2015.

[127] M. Yousef, A. Kumar, and B. Bakir-Gungor, "Application of biological domain knowledge based feature selection on gene expression data," *Entropy*, vol. 23, no. 1, p. 2, 2021.

[128] L. P. Medeiros, C. Song, and S. Saavedra, "Merging dynamical and structural indicators to measure resilience in multispecies systems," *Journal of Animal Ecology*, vol. 90, no. 9, pp. 2027–2040, 2021.

[129] M. Ziegler, A. Roik, A. Porter, K. Zubier, M. S. Mudarris, R. Ormond, and C. R. Voolstra, "Coral microbial community dynamics in response to anthropogenic impacts near a major city in the central red sea," *Marine pollution bulletin*, vol. 105, no. 2, pp. 629–640, 2016.

[130] T. Bruce, P. M. Meirelles, G. Garcia, R. Paranhos, C. E. Rezende, R. L. de Moura, R.-F. Filho, E. O. Coni, A. T. Vasconcelos, G. Amado Filho *et al.*,

"Abrolhos bank reef health evaluated by means of water quality, microbial diversity, benthic cover, and fish biomass data," *PloS one*, vol. 7, no. 6, p. e36687, 2012.

[131] E. A. Dinsdale, O. Pantos, S. Smriga, R. A. Edwards, F. Angly, L. Wegley, M. Hatay, D. Hall, E. Brown, M. Haynes *et al.*, "Microbial ecology of four coral atolls in the northern line islands," *PloS one*, vol. 3, no. 2, p. e1584, 2008.

[132] J. Overmann and C. Lepleux, "Marine bacteria and archaea: Diversity, adaptations, and culturability," in *The Marine Microbiome*. Springer, 2016, pp. 21–55.

[133] C. Lok, "Mining the microbial dark matter," *Nature News*, vol. 522, no. 7556, p. 270, 2015.

[134] D. Jung, L. Liu, and S. He, "Application of in situ cultivation in marine microbial resource mining," *Marine Life Science & Technology*, pp. 1–14, 2020.

[135] S. Chaffron, E. Delage, M. Budinich, D. Vintache, N. Henry, C. Nef, M. Ardyna, A. Zayed, P. Junger, P. Galand *et al.*, "Environmental vulnerability of the global ocean plankton community interactome," *bioRxiv*, 2020.

[136] A. Eiler, D. H. Hayakawa, and M. S. Rappé, "Non-random assembly of bacterioplankton communities in the subtropical north pacific ocean," *Frontiers in microbiology*, vol. 2, p. 140, 2011.

[137] S. K. Yeo, M. J. Huggett, A. Eiler, and M. S. Rappé, "Coastal bacterioplankton community dynamics in response to a natural disturbance," *PLoS One*, vol. 8, no. 2, p. e56207, 2013.

[138] L. E. Bertassello, A. F. Aubeneau, G. Botter, J. W. Jawitz, and P. Rao, "Emergent dispersal networks in dynamic wetlandscapes," *Scientific reports*, vol. 10, no. 1, pp. 1–10, 2020.

[139] Y. Suzuki and E. P. Economo, "From species sorting to mass effects: spatial network structure mediates the shift between metacommunity archetypes," *Ecography*, vol. 44, no. 5, pp. 715–726, 2021.

[140] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, "Breakdown of the internet under intentional attack," *Physical review letters*, vol. 86, no. 16, p. 3682, 2001.

[141] C. Mellin, S. Matthews, K. R. Anthony, S. C. Brown, M. J. Caley, K. A. Johns, K. Osborne, M. Puotinen, A. Thompson, N. H. Wolff *et al.*, "Spatial resilience of the great barrier reef under cumulative disturbance impacts," *Global change biology*, vol. 25, no. 7, pp. 2431–2445, 2019.

[142] G. B. R. M. P. Authority, *Cumulative impact management policy*. Great Barrier Reef Marine Park Authority, 2018.

[143] C. J. Fraser and E. C. Hill, "The problem with 'microbiome'," *Diversity*, vol. 13, no. 4, p. 138, 2021.

[144] C. R. Kelly, C. Ihunnah, M. Fischer, A. Khoruts, C. Surawicz, A. Afzali, O. Aroniadis, A. Barto, T. Borody, A. Giovanelli *et al.*, "Fecal microbiota transplant for treatment of clostridium difficile infection in immunocompromised patients," *The American journal of gastroenterology*, vol. 109, no. 7, p. 1065, 2014.

[145] M. Strous and C. Sharp, "Designer microbiomes for environmental, energy and health biotechnology," *Current opinion in microbiology*, vol. 43, pp. 117–123, 2018.

[146] T. N. Enke, M. S. Datta, J. Schwartzman, N. Cermak, D. Schmitz, J. Barrere, A. Pascual-García, and O. X. Cordero, "Modular assembly of polysaccharide-degrading marine microbial communities," *Current Biology*, vol. 29, no. 9, pp. 1528–1535, 2019.

# Chapter 7

# Appendix

# Appendix A

# Supplement for Chapter 3

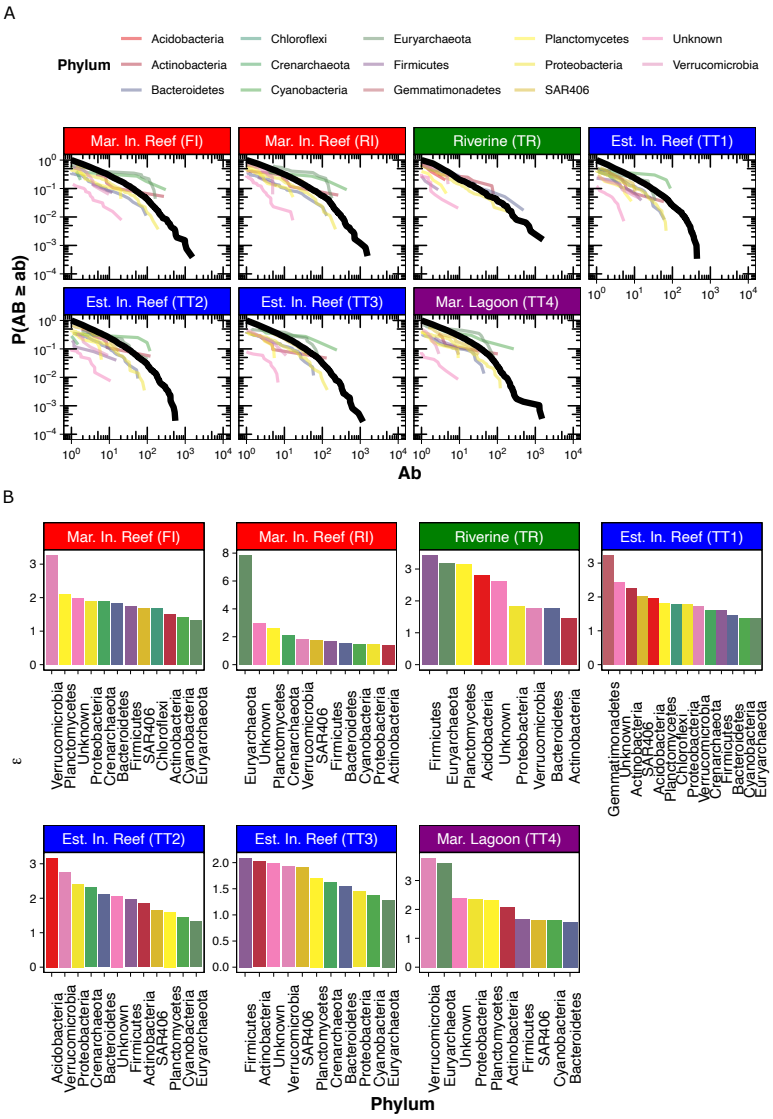


Figure S1: **Power-law Distribution of Abundance.** (A) The epdf of abundance for populations (colored lines) and the communities (black lines). (B) The top 10 phyla given the exponent of the distribution of their abundance.

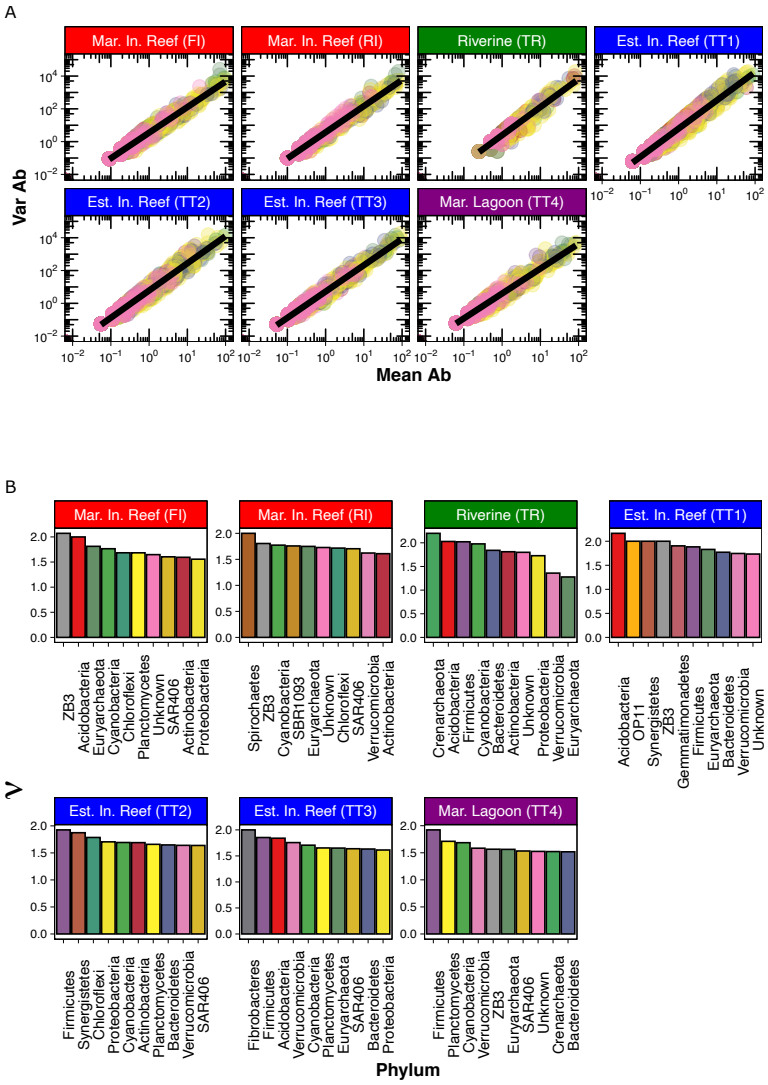


Figure S2: **Taylor's Law of Abundance.** (A) The relation between the variance of OTU abundance and the mean of OTU abundance for populations (colored points) and the communities (black lines). (B) The top 10 phyla given the Taylor's law exponent of their abundance.

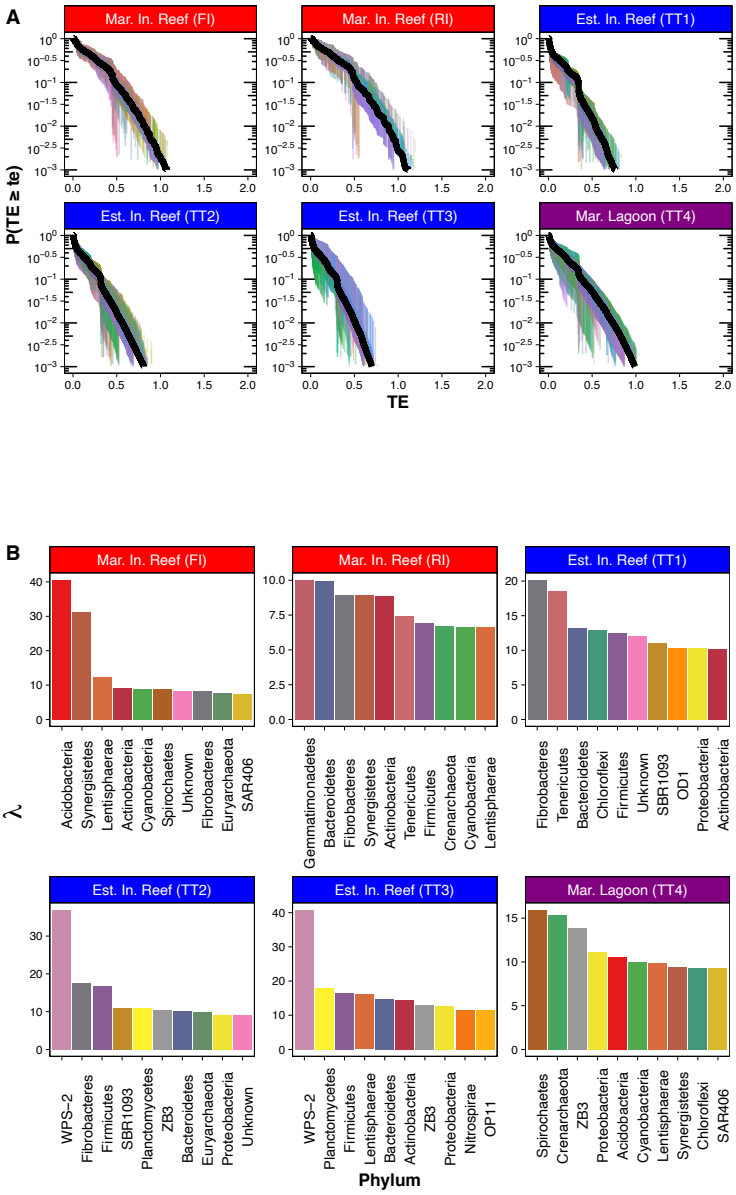


Figure S3: **Exponential Distribution of Interaction** (A) The epdf of interactions (TE) for populations (colored lines) and the communities (black lines). (B) The top 10 phyla given the exponent of the distribution of their interactions.

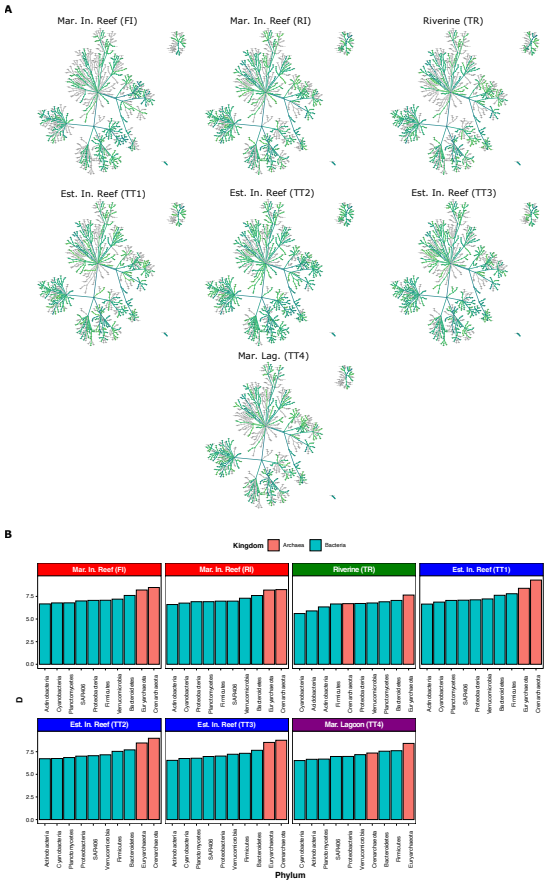


Figure S4: **Phylogenetic Dissimilarity** (A) The phylogenetic trees inferred for each site, given the taxonomic classification of all OTUs observed. A colored node and edge indicates that the taxon was observed in that site; a gray node or edge was not observed. (B) The top 10 phyla given their phylogenetic distances.

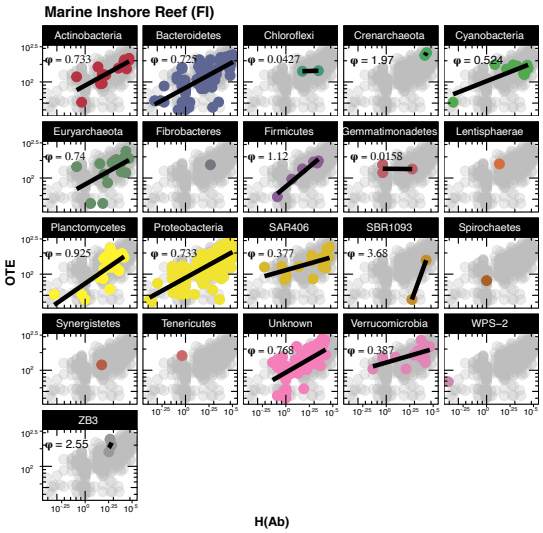


Figure S5: **Species-scale Kleiber's Law for Marine Reefs.** Background gray points are all species (OTUs) for all phyla found within the marine reef habitat FI, while colored points identify OTUs that belong to each phyla indicated in the label above each panel. Black lines are the Kleiber's scaling law $OTE \sim H(Ab)^{\Phi}$; the line is shown only when the fit was possible.

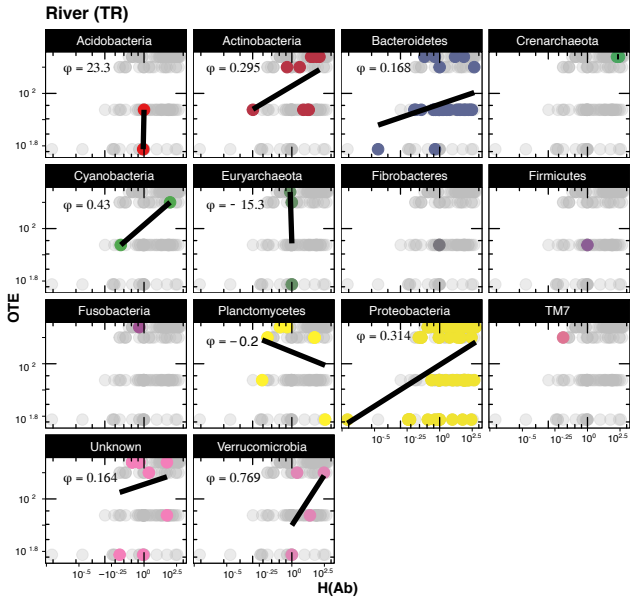


Figure S6: **Species-scale Kleiber's Law for Estuarine Reefs.** Background gray points are all species (OTUs) for all phyla found within the riverine habitat TR, while colored points identify OTUs that belong to each phyla indicated in the label above each panel. Black lines are the Kleiber's scaling law $OTE \sim H(Ab)^{\Phi}$; the line is shown only when the fit was possible.

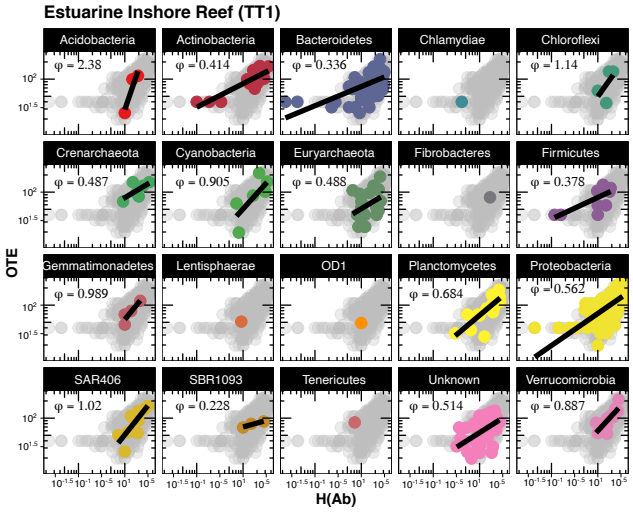


Figure S7: **Species-scale Kleiber's Law for Estuarine Reefs.** Background gray points are all species (OTUs) for all phyla found within the estuarine reef habitat TT1, while colored points identify OTUs that belong to each phyla indicated in the label above each panel. Black lines are the Kleiber's scaling law $OTE \sim H(Ab)^{\Phi}$; the line is shown only when the fit was possible.

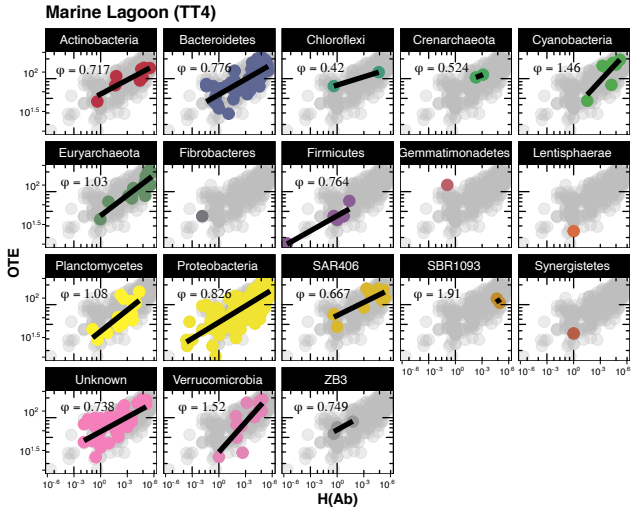


Figure S8: **Species-scale Kleiber's Law for Lagoons.** Background gray points are all species (OTUs) for all phyla found within the lagoon habitat TT4, while colored points identify OTUs that belong to each phyla indicated in the label above each panel. Black lines are the Kleiber's scaling law $OTE \sim H(Ab)^{\Phi}$; the line is shown only when the fit was possible.

# Appendix B

# Supplement for Chapter 4

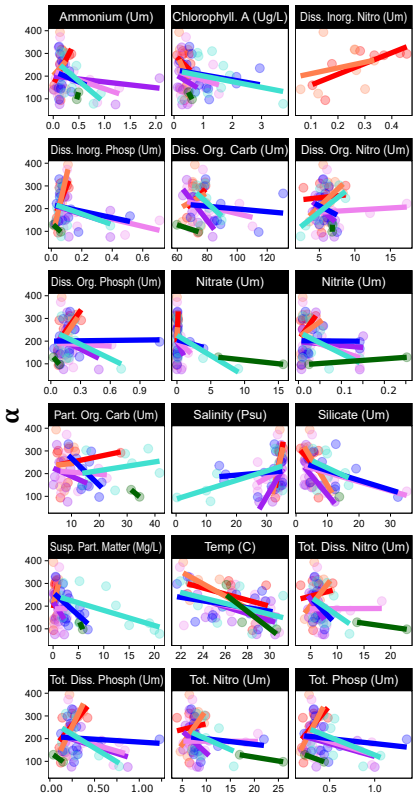


Figure S1: **$\alpha$-diversity by Environmental Driver.** Relationships between community $\alpha$-diversity (the number of distinct OTUs that is not necessarily proportional to the number of distinct phyla for each habitat emphasizing: (i) the representativeness of phylum-diversity for community patterns; and (ii) the dependence of macro ecological indicators on biological scales) and environmental variables are shown fitted by a linear relationship for each habitat type identified by the color of the line as in Fig.2.2B.

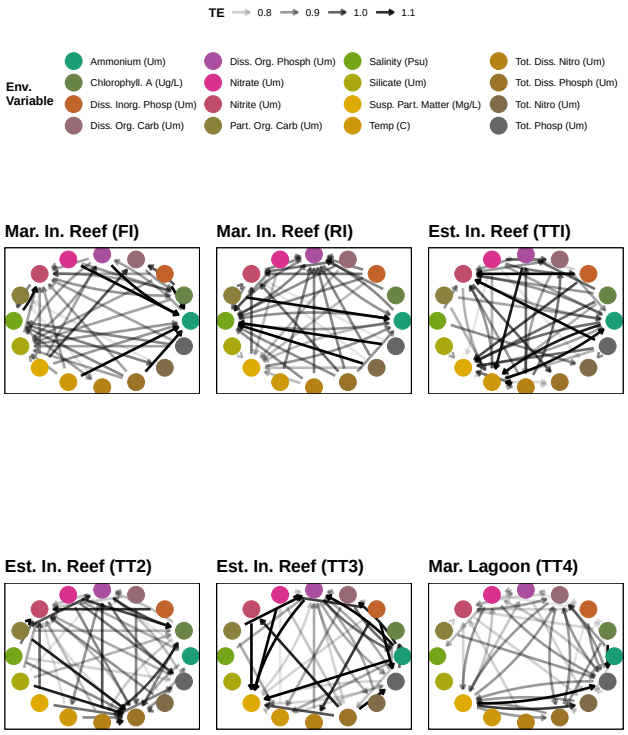


Figure S2: **Envirome Top Interaction Networks.** Nodes are environmental variables, and the opacity of the links is proportional to TE: larger TEs are represented by more opaque links. The direction reflect the predictive influence. Networks show the top 20% of interactions that are power-law distributed, highlighting the tendency to scale-free core organization despite the overall organization is exponential. For the river (TR) the network cannot be inferred due to insufficient environmental data to fit probability distribution functions.

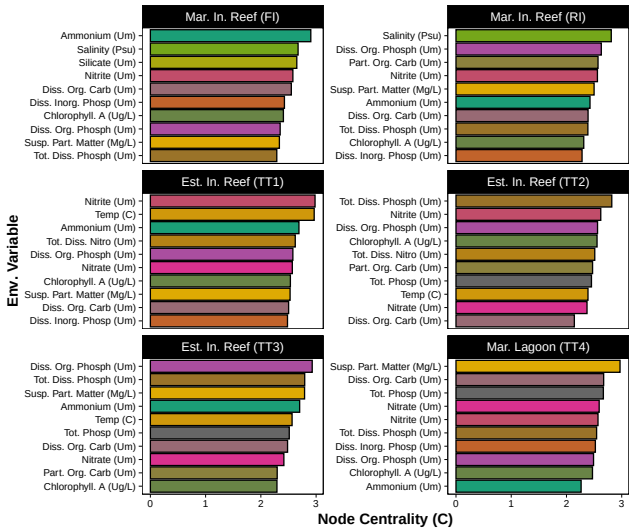


Figure S3: **Envirome Driver Importance.** Top 10 important environmental drivers ranked by node centrality for each habitat as in Eq. 2.9, where the mean of TEs among drivers is considered.