University of Essex

THE DIVERSITY AND FUNCTIONING OF COASTAL MICROBIAL COMMUNITIES

Andrew Davis

A thesis submitted for the degree of MSD

School of Life Sciences

March 2020

ACKNOWLEDGEMENTS

The author would like to thank Professor Alex Dumbrell for his supervision throughout this dissertation and both Doctor David Clark and Professor Graham for intellectual input. All bioinformatics was carried out using the high-power computing cluster [ceres] at the University of Essex.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

BEF – Biodiversity-ecosystem functioning

CBESS - Coastal Biodiversity and Ecosystem Service Sustainability

NMDS - Non-metric multidimensional scaling

OTU - Operational taxonomic unit

OOB – Out-of-bag

GHG – Greenhouse Gas

APL - Average path length

PCA – Principal component analysis

PC – Principal component

ANOVA - Analysis of variance

GPP - gross primary productivity

BCR - Benthic community respiration

ABSTRACT

Microbial communities are notably complex and underpin many important global biogeochemical processes. However, there is still much that we do not know regarding the structuring dynamics of microbial communities and their biodiversity-ecosystem functioning (BEF) relationships. Therefore, there is great importance in researching these aspects in greater detail. In order to do this kind of analysis, large datasets are needed to properly capture the nuances of microbial communities. The CBESS (Coastal Biodiversity and Ecosystem Service Sustainability) dataset used here is one such dataset. Random forest analysis, a type of machine learning, was applied here to an expansive dataset of microbial metabarcode reads and environmental measures across multiple spatial scales. Two different coastal habitats were used to create further distinctions. Random forest models were created with and without environmental measures and it was found that notable differences across domain, spatial scale, and habitat were only observed when environmental measures were included. The relative importance of environmental factors increased both when scale increased and when specific habitat models were constructed, emphasizing the role that scale and context play in the interpretation of this type of analysis. Across domains, the relative importance of taxonomic factors was much higher in bacteria, indicating a possible increased role of dispersal limitations for that domain. Co-occurrence networks were then constructed using the same dataset to investigate if there were any relationships between microbial network structure and process profiles across spatial scales. Significant relationships were only found at the smallest spatial scale with bacteria and archaea exhibiting nearly all the relationships. In those two domains, clear ecological patterns were linked with specific ecological processes for both network size and network interconnectedness. Across taxonomic levels, bacteria had stronger links at higher taxonomic levels and archaea had the strongest links at

the genus level. Overall, this study reveals both the complex spatial dynamics of microbial community structuring and BEF relationships that exist between microbial communities and globally important processes.

CHAPTER 1: GENERAL INTRODUCTION

Belowground microbial communities are defined as any microbial community that exist at or below the sediment surface and the community ecology of these systems differ significantly from aboveground terrestrial systems (Fierer and Jackson, 2006; Bardgett and van der Putten, 2014). Despite this difference less effort is put into studying belowground diversity. What is understood about these systems reveals stark differences in diversity patterns and the drivers behind them (Hendershot et al., 2017). While knowing the composition of a microbial community does not automatically reveal its functional ability (Baldrian, 2019), there is still plenty of value in community diversity studies. These values range from uncovering the underlying dynamics that structure microbial communities to better understanding any potential links between microbial diversity patterns and important ecosystem functions.

Studying these dynamics also has value in other fields of ecology. Experiments have demonstrated that microbes show notably complex degrees of spatiotemporal behaviour (Gore, 2018) with their abundances fluctuating noticeably across scales (Ji et al., 2019). Studying spatiotemporal effects on microbial communities is still an underdeveloped field and there is much that is still not known (Dumbrell et al., 2011). Research thus far indicates that the scale of observation plays an important role in the interpretation of microbial communities and their relationships. Different factors affect communities at different spatial scales so unique patterns naturally emerge at different levels (Bardgett and van der Putten, 2014). Therefore, using larger scale gradients often does not accurately represent the heterogeneity of microbial systems properly since smaller-scale spatial variations would not be accounted for (Hendershot et al., 2017). The short-lived and dynamic nature of microbes means that they quickly adapt to new conditions. These smaller-scale spatial variations are often only observed at scales smaller than what is needed to capture similar dynamics in larger organisms

(Lauber et al. 2013). Just looking at broadscale patterns would not properly capture the smaller scale variations seen in microbes (Barberán et al., 2014).

The microbial communities that underpin coastal systems are just some of the many belowground community types that are still poorly understood. Coastal ecosystems have a lot to offer regarding investigating BEF relationships in microbes. As important transition zones between terrestrial and marine habitats (Levin et al., 2001), coastal ecosystems offer the opportunity for researchers to study life in highly dynamic systems. These natural variations allow for in-depth research into the functional relationships of microbial communities (Baumann et al., 2015). Because of their status as transition zones, coastal ecosystems are important buffers zones that absorb and process nutrients such as nitrogen that can be harmful to marine habitats in higher concentrations than normal (Bowen et al., 2011; Vivanco et al., 2015). However, complications arise when studying these systems since previous research into both soil and sediment communities must be considered at different degrees depending on the environmental characteristics of a given habitat. For example, tidal changes affect bacterial abundances in estuarine waters, creating another potential dispersal mechanism that must be considered (Chen et al., 2019a). Coastal systems also serve vital roles as nurseries to commercially important fishes and as buffers for both erosion and nutrient run-off so better understanding the microbial communities present will help in researching these aspects as well (Beck et al., 2001; Gedan et al., 2011; Vivanco et al., 2015).

One of the biggest hurdles in microbial ecology is the amount of data and processing power that is needed disentangle causative mechanisms. Addressing these issues, recent advances in next-generation sequencing (NGS) have created new avenues for bioinformatic analysis in many fields (Goodwin et al., 2016). In the case of metabarcoding, the targeted amplification and sequencing of phylogenetic marker genes, increasing the depth and size of sequence libraries while lowering costs allows for more detailed analysis (Sinclair et al., 2015). Once sequencing is done the sequences can then be clustered based on similarity and assigned taxonomy. This process creates OTUs (operational taxonomic units) that can then be used in subsequent analysis of microbial communities in their natural habitat (Ward et al., 1990; Kim and Isaacson, 2015; Gore et al., 2018). By using OTU data a researcher can look at what is in a system via the occurrence of OTUs and how much is there via the relative abundance of OTUs (Clark et al., 2018).

The data used are just as important as the methods used. This statement is especially true when investigating microbial communities. Larger datasets provide unique opportunities to understand communities at depths and scales that normal studies cannot. For instance, more intensive sampling is required to capture the full diversity of microbial communities (Hermans et al., 2019). Scaling up the size of datasets is typically done by conducting a meta-analysis, but accounting for methodological differences across studies typically limits the ability to do more in-depth analysis. These differences are seen in a recent meta-analysis focused on microbial community structuring using random forest analysis. This study found that technical factors play a major role in influencing relative abundance measures (Ramirez et al., 2018b). By removing technical variation from sampling and analysis, the factors which are important in structuring microbial communities can be better elucidated. The CBESS (Coastal Biodiversity and Ecosystem Service Sustainability) dataset used here combines the analytical advantages of being a larger dataset with the technical consistency of a single study. With a focus on better understanding the biodiversity, ecosystem processes, and ecosystem services of UK coastal habitats, this dataset opens the door to many novel insights (Raffaelli et al, 2014).

The overarching aim of this thesis is to investigate microbial community structuring, BEF relationships in microbial communities, and how these two aspects change across contexts. There is still much that is not known about each of these aims so research of this kind has great importance. To cover each aim metabarcode reads and environmental measures from the CBESS dataset were used. Random forest models were used to identify factors important in structuring communities and co-occurrence networks were used to identify links between community structure and important processes. Spatial scales, multiple habitats, and multiple domains were all used to see how results differed in each context.

- In Chapter 2 the general materials and methods used throughout are described.
- In Chapter 3 the structuring of microbial communities is investigated using random forest models.
- In Chapter 4 BEF relationships are analysed using co-occurrence networks.
- In Chapter 5 the broader implications and takeaways from this work are discussed.

CHAPTER 2: GENERAL MATERIALS AND METHODS

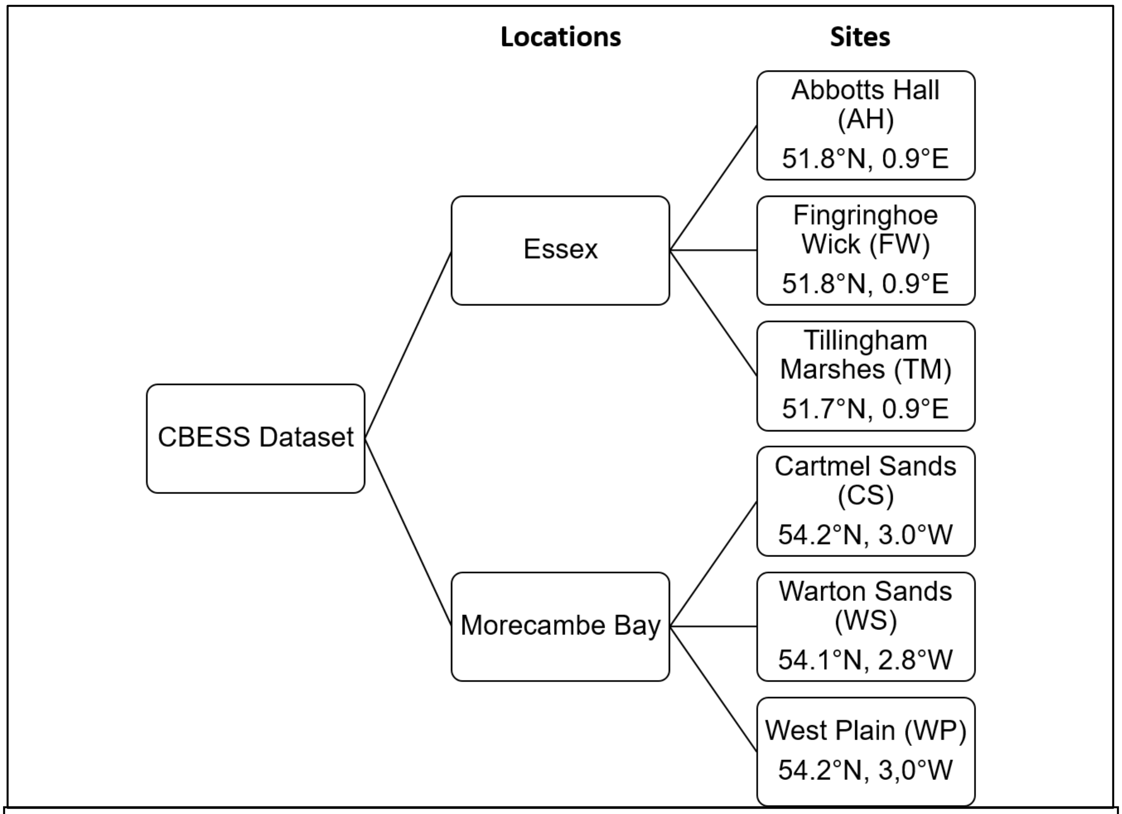


Figure 1. Flowchart describing the sampling hierarchy of the CBESS dataset. There were 528 total samples taken, 264 at each location, and 88 at each site. 24 salt marsh samples and 24 mud flat samples were taken at each site, equally divided seasonally between summer and winter.

Study Site

In this study the CBESS dataset was used. CBESS was a sub-project of the NERC (Natural Environment Research Council) funded BESS (Biodiversity and Ecosystem Service Sustainability) program that focused on coastal areas (Raffaelli et al, 2014). There were two main geographic locations sampled, Essex and Morecambe Bay (Figure 1). Each location had three sites and each site contained a salt marsh and a mud flat, providing 12 unique areas for sampling. For a map level view of the sampling sites please refer to Alzarhani et al. (2019).

At each site-specific habitat, 22 soil cores (5-cm diameter, 15-cm depth) were taken for microbial community metabarcoding in both summer and in winter, yielding

Table 1. Brief description of the environmental factors measured. All the factors aside from nutrient and GHG fluxes are used for analysis in chapter 3 and nutrient and GHG fluxes are used in chapter 4. Some measures were only recorded in salt marshes (*) or mud flats (**) and are identified accordingly. For the bioturbation measures, the species in question were any animals found in 10cm cylindrical cores. (B) Biotic measures and (A) abiotic factors are tagged respectively.

Category	Measurement	Notes
Sediment Chlorophyll (B)	Chlorophyll a, b and c concentration	-
Sediment Colloidal Carbohydrates (B)	Colloidal carbohydrate concentration	-
NDVI (Normalised Difference Vegetation Index) (B)	Spectral reflectance	-
NPP (Net primary Productivity) (B)	kg / m^2 yr	Mass of the dry vegetation
Particle Size (A)	% of each size sediment type	-
Sediment Stability by CSM (Cohesive Strength Meter) (A)	stagnation pressure in N / m^2	Resistance of surface to erosive force via critical threshold for surface erosion
Sediment Water Content (A)	% Water content	-
TOC (Total Organic Carbon) (B)	Loss on Ignition	Loss after ignition (g) and % composition
Light and dark nutrient and oxygen fluxes	μmol / m^2 hr	Oxygen, NPOC (non-purgeable organic carbon) Nitrate, NOx, Ammonia, Nitrate, Phosphate and Silicate
Greenhouse Gas Fluxes (GHG)	μmol$_{CO2}$ / m^2 h (NPP, BCR, GPP) nmol$_{CH4}$ / m^2 s (CH_4) nmol$_{N2O}$ / m^2 s (N_2O)	NPP - (Net primary Productivity) BCR - (Benthic Community Respiration) GPP - (Gross primary Productivity) CH_4 - (Methane) N_2O - (Nitrous Oxide)
Bulk Density* (B)	dry weight (kg)	taken from bulk density rings

Electrical Conductivity* (A)	Electrical Conductivity	proxy for salinity
pH* (A)	pH	-
Plant Cover* (B)	Plant % cover	Percentage cover of each plant species
Plant Height* (B)	height (cm)	Direct measurements of plant height
Root Biomass* (B)	dry weight (kg) / m^2	Root biomass of 3 soil depths
Sediment Erosion Rate* (A)	% mass loss / hr	Sediment cores placed in flume tank flow
Shear Stress* (A)	N / cm^2	Measured at three soil depths
Soil Moisture Content* (A)	Soil moisture content (%)	Field soil moisture content measured from bulk density samples
Soil Organic Matter* (B)	Soil organic matter (%)	Soil organic matter assessed by LOI (Loss-on-ignition) using bulk density samples, taken adjacent to flume core
Standing Crop* (B)	dry weight (kg)/m^2	50cm x 25cm area of above ground vegetation cut to ground level and dried to give indication of standing crop biomass
Vegetation Projected Area* (B)	Veg area (mm^2) / Horizontal extent (mm)	Area of vegetation normalised by the horizontal extent of the section analysed

Individual Bioturbation**	BPi per m^2	BPi = √B x Mi x Ri Bi = individual biomass Mi = individual species mobility Ri = individual species reworking
Population Bioturbation**	BPp per m^2	BPp = BPi x Ai Ai = Individual species abundance per m^2

44 total samples for each site-specific habitat. Overall, there were 528 total samples taken. For each soil sample an extensive list of plant, soil, and nutrient measures were taken from the 1x1m quadrat associated with each core in the summer and winter of 2013 (Table 1). Note that certain measures were not taken at every site due to logistical constraints. There was still enough coverage of measurements across sites and habitats to conduct in-depth analysis though. To limit the effect of space and time on sampling data, field sampling was coordinated across sampling teams and taken at similar spaces and times. To minimize the influence of sampling on measurements, a hierarchy of observations was created to attain accurate readings. For a more detailed explanation of the hierarchical sampling please refer to section 4.2 of chapter 2 of Raffaelli et al. (2014). The CBESS environmental data can be found at https://catalogue.ceh.ac.uk/eidc/documents#term=cbess&page=1.

Molecular analysis

Following field sampling DNA was extracted from 0.05 g of homogenised dry roots from each soil core using the MoBio PowerPlant DNA isolation kit. DNA samples were then run through PCR cycles using Illumina's recommended amplicon-sequencing protocols (Dumbrell et al., 2016). Libraries were multiplexed by adding Nextera XT indices with an 8-cycle PCR to differentiate between samples and bead purification of PCR products was done using Agencourt AMPure XP PCR Purification beads. The 16S rRNA Klindworth primers, 341f (5'-CCTACGGGNGGCWGCAG-3') and 785r (5'-GACTACHVGGGTATCTAATCC-3'), (Klindworth et al., 2013) were used to target bacteria. 16S 344f (5'-ACGGGGYGCAGCAGGCGCGA-3') (Raskin et al., 1994) and 915r (5'-GTGCTCCCCCGCCAATTCCT-3') (Stahl and Amann, 1991) primers were used for archaea and 18S 574*f (5'-CGGTAAYTCCAGCTCYV-3') and 1132r (5'-CCGTCAATTHCTTYAART-3') (Hugerth et al., 2014) were used for eukaryotes. Each PCR product was then sequenced in an Illumina HiSeq machine 2500 in rapid run mode run at NBAF (NERC Biomolecular Analysis Facility) in Liverpool yielding 300+300bp paired-end sequences.

Bioinformatics analyses

After sequencing, data clean up and filtering was run following recommendations in Dumbrell et al. (2016). Sickle v1.33 (Joshi and Fass, 2011) was run for quality trimming on all reads with a minimum length filter of 275bp. SPAdes v3.10.1 (Nurk et al., 2013) was then run for error correction with the BayesHammer denoising algorithm (Nikolenko et al., 2013). Paired-end assembly was then run on the remaining reads. PANDAseq v2.11 (Masella et al., 2012) with the PEAR algorithm (Zhang et al., 2014) was used for paired-end assembly. PEAR was used over the PANDAseq algorithm because PEAR has been found to align more paired-end sequences compared to PANDAseq (Schirmer et al., 2015). Any sequences smaller than 300bp were then filtered out before OTU clustering. Only the forward reads were used for archaea community analysis because the length of the region targeted by the archaea primers does not allow for paired-end assembly. Using this approach has been shown to have marginal effects on results (Clark et al., 2017).

OTU tables were created for each respective domain using the open-source VSEARCH v2.4.3 pipeline (Rognes et al., 2016). Reads were first dereplicated, removing singleton sequences to omit likely artefacts (Flynn et al., 2015). The dereplicated reads were then sorted by abundance and used to pick centroid sequences. UCHIME was run to filter chimeras from the centroids using the RDP database (release 11. The remaining centroids were used to create the OTU table. OTUs were selected at a 97% similarity level, corresponding to the intragenus similarity level between species (Yarza et al., 2014). Taxonomy was assigned to each OTU table using the RDP classifier v2.11 and the RDP database (Wang et al., 2007). For the eukaryotic reads the RDP classifier was trained using the PR^2 (Protist Ribosomal Reference) database (Guillou et al., 2013). The PR^2 database primarily targets protists but metazoans, land plants, fungi, and eukaryotic organelles are included in comparatively less detail.

All subsequent analysis was conducted using R v3.5.0 (R Core Team, 2018). Prior to running analysis each OTU table was rarefied to account for variable read numbers across samples. To prevent rarefying to too low of a read number certain samples were filtered out based on rarefaction curves. Samples with read counts below 100,000 for the bacteria OTU table, 5,000 for the archaea OTU table, and 4,600 for the eukaryotic OTU table were filtered out for each domain. Using the vegan package (Oksanen et al., 2018) NMDS plots were then created for each domain to visualize how microbial community composition differed across sites, location, and habitat type. Plots were made using both a presence/absence algorithm (Jaccard) and abundance-based algorithm (Morisita-horn).

Each table was then reformatted to prepare for subsequent analysis. To do this the OTU data was replicated for each sample, creating an expanded table with samples as rows and each row containing the abundance and taxonomic identification for a given OTU. Rows with an abundance of 0 were filtered out and the remaining table was one-hot encoded based on the taxonomic information. One-hot encoding was the process by which the variables signifying taxonomic categories at each level were removed and new binary variables were added for each unique category. This table was then summarised, resulting in a table with each row representing a sample and columns representing the relative abundance for each taxonomic level found in the data. Columns for sample-specific information were included as well. These tables were created for each domain using the full taxonomic data, just the salt marsh data, and just the mud flat data. For each table created there was a version made that included environmental measures as well. These tables were used for all subsequent analysis.

CHAPTER 3: MICROBIAL COMMUNITY STRUCTURING DYNAMICS

Introduction

Two prevailing theories exist for explaining ecological structuring: niche theory (Tokeshi, 1990) and neutral theory (Chave, 2004). Under niche theory coexistence is possible because species respond to environmental factors in different ways, ultimately creating a community structure where each species fits into a unique niche (Leibold and McPeek, 2006). In contrast, the neutral theory assumes all species have the same demographic traits (Chave, 2004) with community structure being dictated by stochastic processes and dispersal limitations (Rosindell et al., 2011). Disentangling the effects of deterministic versus stochastic processes in microbial community structuring is a current field of interest and work done thus far has revealed interesting insights (Ning et al., 2019). While communities are influenced by stochastic community-level processes, previous research in specific study systems have demonstrated that microbial community composition is largely determined by environmental factors (Dumbrell et al., 2010; Stegen et al., 2012). For example, the evolution and retention of new genes in bacterial populations seem to be largely environmentally driven (Noda-García et al., 2019) and environmental effects have been shown on specific microbial genera (Hu et al., 2019). However, the exact effects of environmental drivers on microbial communities are still very much unknown (Lladó et al., 2018) and there is evidence that stochasticity plays a role in structuring as well (Albright et al., 2019).

Different structuring patterns occur at different scales and the patterns that emerge may be indicative of overall structuring dynamics. From certain perspectives these changes in dynamics could indicate that the nested patterns of microbial communities are more due to environmental selection than dispersal limitation

(Menegotto et al., 2019). However, this shift to environmental selection could also be due to there naturally being more environmental heterogeneity seen at larger scales, confounding the interpretation of structuring dynamics (Zinger et al., 2019). Certain environmental factors may also be influencing the prevalence of stochasticity in community structuring which is important to consider (Albright et al., 2019).

Microbial communities being the immensely diverse communities that they are makes them rather difficult to analyse using traditional methods. However, recent applications of machine learning methods have demonstrated that the value it holds in analysing microbial communities (Qu et al., 2019). Machine learning has been useful for predicting certain environmental associations in microbial species (Martínez-García et al., 2016) and random forest analysis specifically has been used to investigate microbial structuring patterns, revealing interesting insights (Ramirez et al., 2018b). Random forests are a form of decision tree-based machine learning that creates multiple iterations of trees using a different subset of predictors in each iteration. In this analysis predictors were the variables used to build a given random forest model. The advantages of random forests lie in their ability to work with data that has many more predictors than it does samples to identify variables that are important in differentiating samples. As machine learning methods continue to be developed and improved the potential applications of them continue to increase so it is well worth applying them more in the field of microbiology.

Random forest analysis was used to ensure that the full breadth of environmental factors included in the dataset was accounted for. Random forest models have the advantage of being able to work with many predictors, so it made sense to utilize them in this study. Analysis of the important taxonomic and environmental variables focused on answering three main questions: (1) how does the relative importance of factors change across spatial scales, (2) what role does habitat type play in structuring importance, and (3) how do structuring dynamics differ across microbial domains? It was hypothesized that environmental variables would increase in importance as scale increased, differences would be observed between habitats, and similar overarching patterns would be shown across microbial domains.

Methods

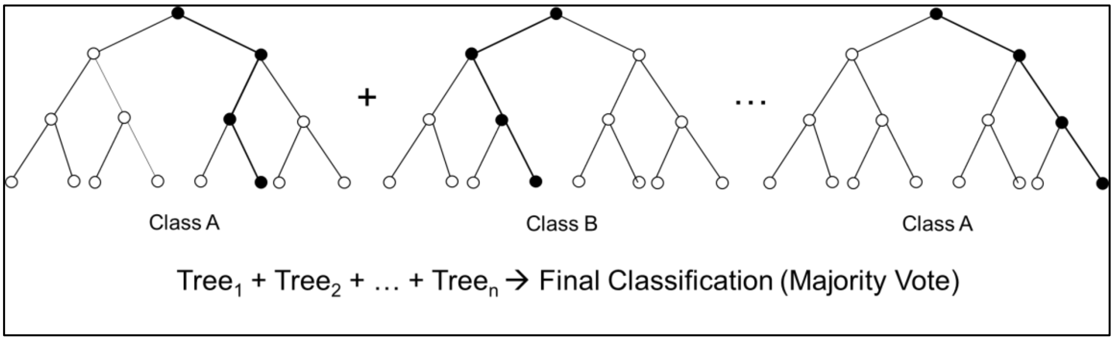


Figure 2. Overview of the classification framework for a random forest model. Each sample has n number of decisions trees created and each has a subset of the predictors used to build the model. Each tree is then evaluated based on the sample values and the final classification is determined by the classification that occurs the most frequently. This process is repeated for each sample.

All random forest analysis was run using the randomForest package in R (Liaw and Wiener, 2002). Random forests analysis is a form of decision tree-based machine learning that creates multiple iterations of trees using a different subset of predictors in each iteration (figure 1). The advantages of random forests lie in their ability to work with many predictors to identify which variables are important in classifying samples. Random forests can either be run unsupervised or supervised. Unsupervised runs use only predictors with no classification or regression to predict. Instead, they compare the original data to synthetically generated data using the original data as a reference. In the case of this analysis, unsupervised runs were useful in that they yielded the importance of each predictor in differentiating between samples. In comparison, supervised models were run against a dependent variable and the importance of a given predictor was how important it was in classifying the dependent variable. Variable-importance (the mean decrease in accuracy, or how much the accuracy of the model would decrease if a given predictor was removed from the model) was used throughout this analysis to signify importance. To account for scale differences across models, unscaled permutation variable-importance was used for all models. For each model, 10,000 trees were created, and the number of predictors sampled per tree was set to default (square root of the number of variables).

To analyze relative importance in structuring communities, unsupervised random forests were run for each taxonomy only group. Supervised models were run against site and location to investigate differences in variable-importance across different scales. MDS plots were created as well to compare the clustering of samples compared to the NMDS plots. To visualize the relative variable-importance for each model, the sum of all variable-importance values greater than zero for each model was taken and used to get the relative variable-importance of each predictor on a scale of zero to one. These were then plotted as stacked bar plots to compare how the relative variable-importance of different taxonomic levels in structuring communities differs depending on domain, habitat, and spatial scale. Correlations were taken across each model grouping as well to investigate how the relative variable-importance of predictors changed across scales.

The same process was done using the data containing environmental measures to look at the relative variable-importance measures for classifying samples, sites, and locations. For plotting purposes there were too many environmental variable categories to visualize them all in a meaningful manner. Instead, the environmental variables were grouped as either abiotic or biotic to make visual interpretation possible. Significant variable categories and their potential implications were still discussed though.

Another useful feature of random forest models is that supervised models can be used to predict new data. In the case of this analysis the taxonomic structure models built only using one habitat type were used to predict the data from the other habitat type. Predicting taxonomic data from one habitat using a model trained on the other habitat is useful because it exposes potential biases in extrapolating models to a context they have not been exposed to. It can also help identify taxa that are important in structuring communities in both habitats. Predicting was not done using the models made using the data from both habitats because predicting data that was used to train the model being applied to it is counterintuitive. Predictions were also not done on the models that included environmental data because some measures were only taken at only one of the habitat types and environmental measures would have to have been assumed to run cross-habitat predictions.

Results

OTU Analysis

After quality filtering approximately 220 million bacteria sequences were clustered into 140,983 OTUs, 272 million archaea sequences were clustered into 45,726 OTUs, and 82 million eukaryotes sequences were clustered into 14,058 OTUs. As expected, the average unique OTU richness for each sample followed roughly the same scale pattern to the OTU counts (bacteria: 6300 +/- 2455; archaea: 688.8 +/- 674.3; eukaryotes: 174.4 +/- 90.6). Richness patterns across sites for each domain were consistent aside from noticeably higher bacteria and archaea richness measures in the Essex mud flat sites. Community structure analysis via NMDS plots revealed multiple groupings based on location and on habitat (figure 2). All of the domains showed the same overall pattern as seen in the figure but the clearest distinctions were observed with the bacterial data. In comparison, both the archaeal and

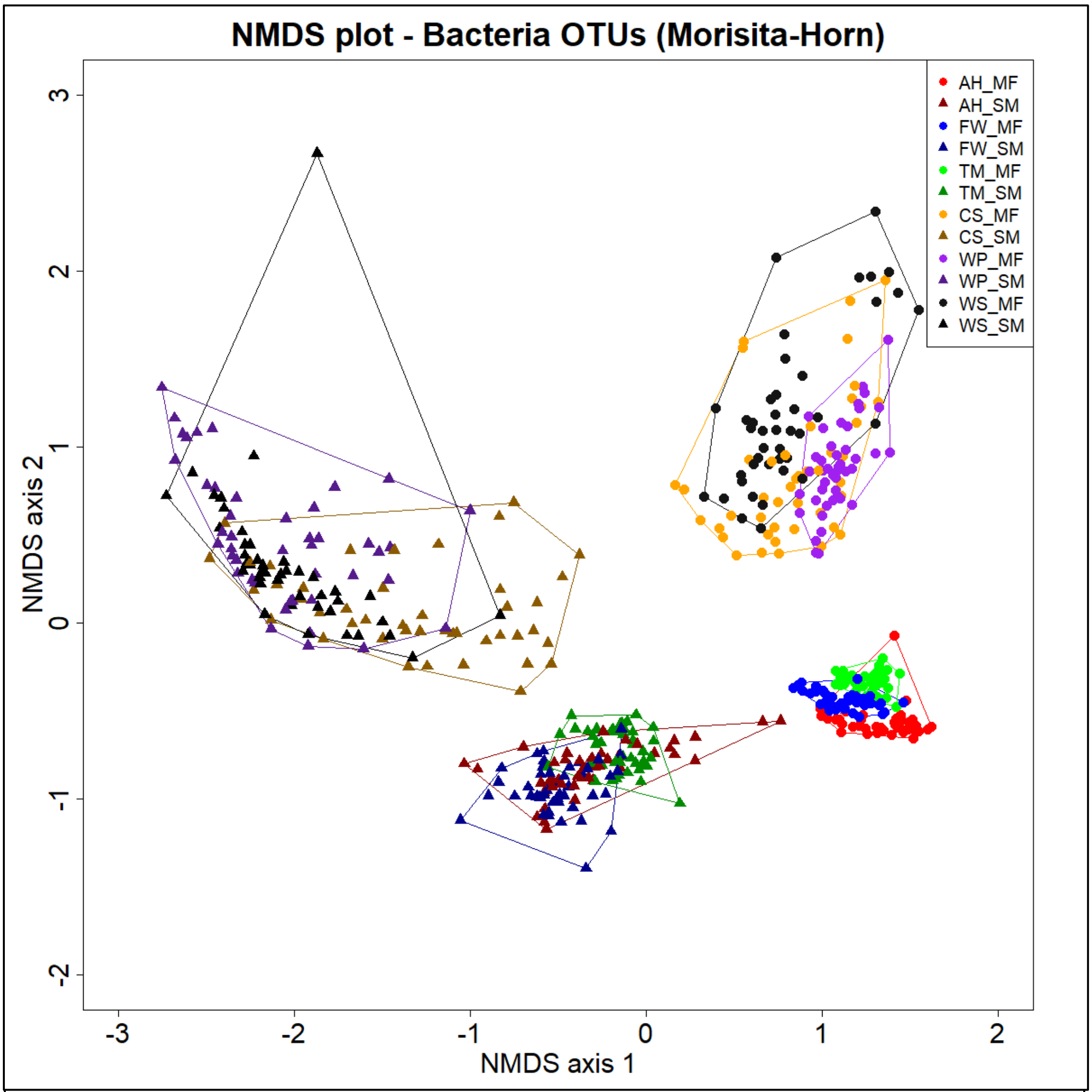


Figure 2. NMDS (Non-metric multidimensional scaling) plot of bacterial OTUs. Points represent samples, shapes represent habitat type, and color represents the specific site. OTUs showed distinct groupings based on both location since AH, FW, and TM represented were the Essex sites and CS, WP, and WS were at Morecambe Bay. Habitat groupings between salt marsh samples and mud flat samples are seen as well.

eukaryotic data had more overlap in their groupings. However, there was more overlap observed with these domains, eukarya in particular. After assigning taxonomy and filtering the data 2,452, 1,070, and 2,870 unique taxonomic categories were present eukaryotic data respectively. There were some bacterial taxa identified using the

eukaryotic data respectively. There were some bacterial taxa identified using the

archaeal primers but those were not removed in the interest of accurately representing what was captured by the archaea primers.

Relative Importance in Structuring Communities

The output of the random forest models showed similar distinctions as the NMDS plots, confirming that the random forest models grouped samples correctly. When data from only one habitat was modelled the MDS plots showed the location split more clearly than when the full dataset was used, highlighting how the number of factors differentiating sampling sites affects modelling. There did not seem to be any consistent patterns across spatial scales regarding the relative variable-importance of taxonomic levels across the different conditions (figure 3). However, there was a clear difference seen across domains with bacteria showing the highest relative variable-importance at the genus level. The eukaryotes were similar to the bacteria at the lower taxonomic levels but that is at least partly due to the eukaryotes only having taxonomic identification at the class level and below. Meanwhile, the archaea showed closer to even distribution of variable-importance across taxonomic levels. The archaea models also showed variable-importance in the domain which was a by-product of the archaea primers capturing some bacterial 16S rRNA.

For these taxonomic models a general pattern of the variable-importance of sample-level predictors being better correlated with site than location was observed in all models except for the mud flat models where the relationship was not clear (figure 4). The same pattern was seen with location level variable-importances being better correlated with site-level variable-importances than sample-level variable-

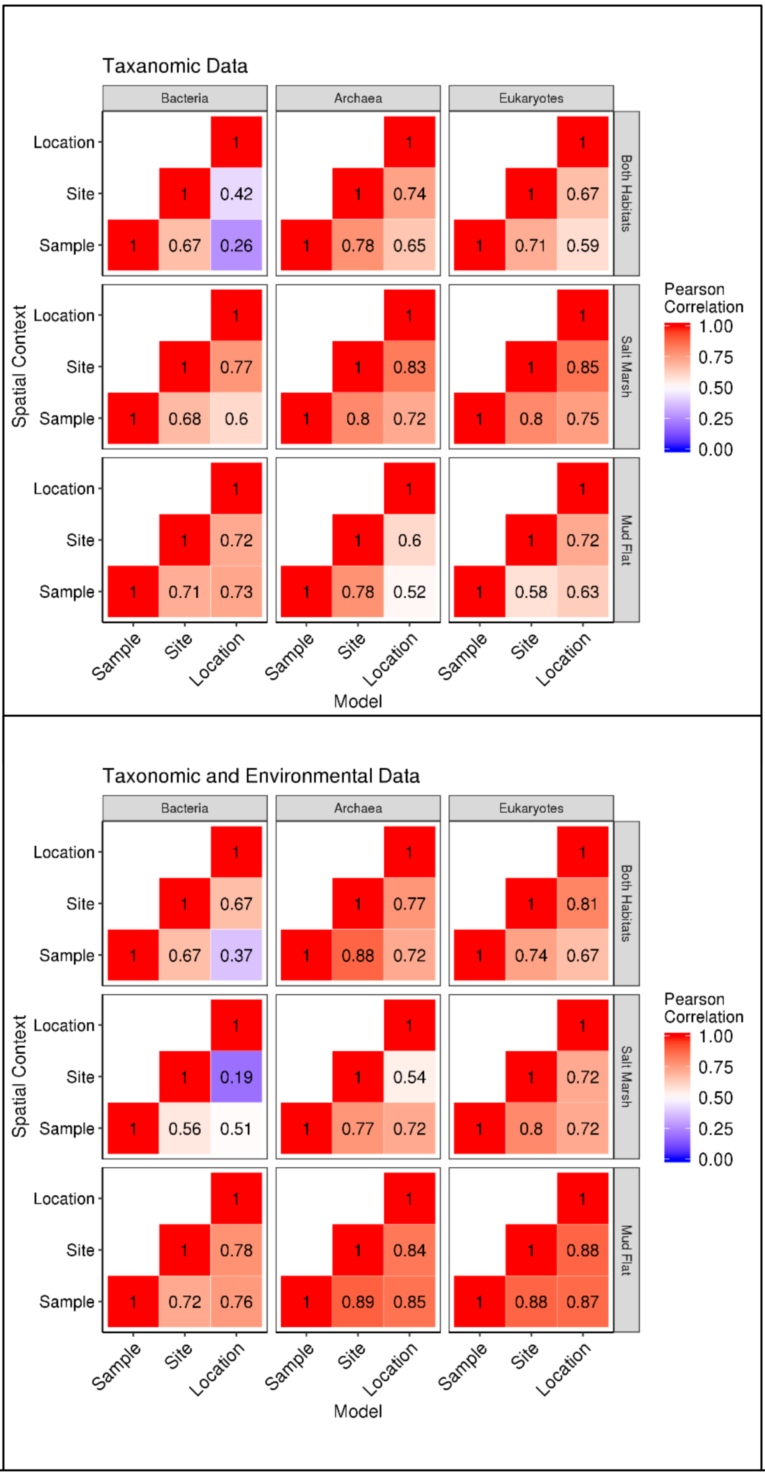


Figure 3. Correlations between the relative importance variables used to build random forest models across scales and domains. Models were created using datasets of (top) relative abundance values of taxonomic identifiers at each taxonomic level (bottom) and of both the same relative abundance values and environmental factors.

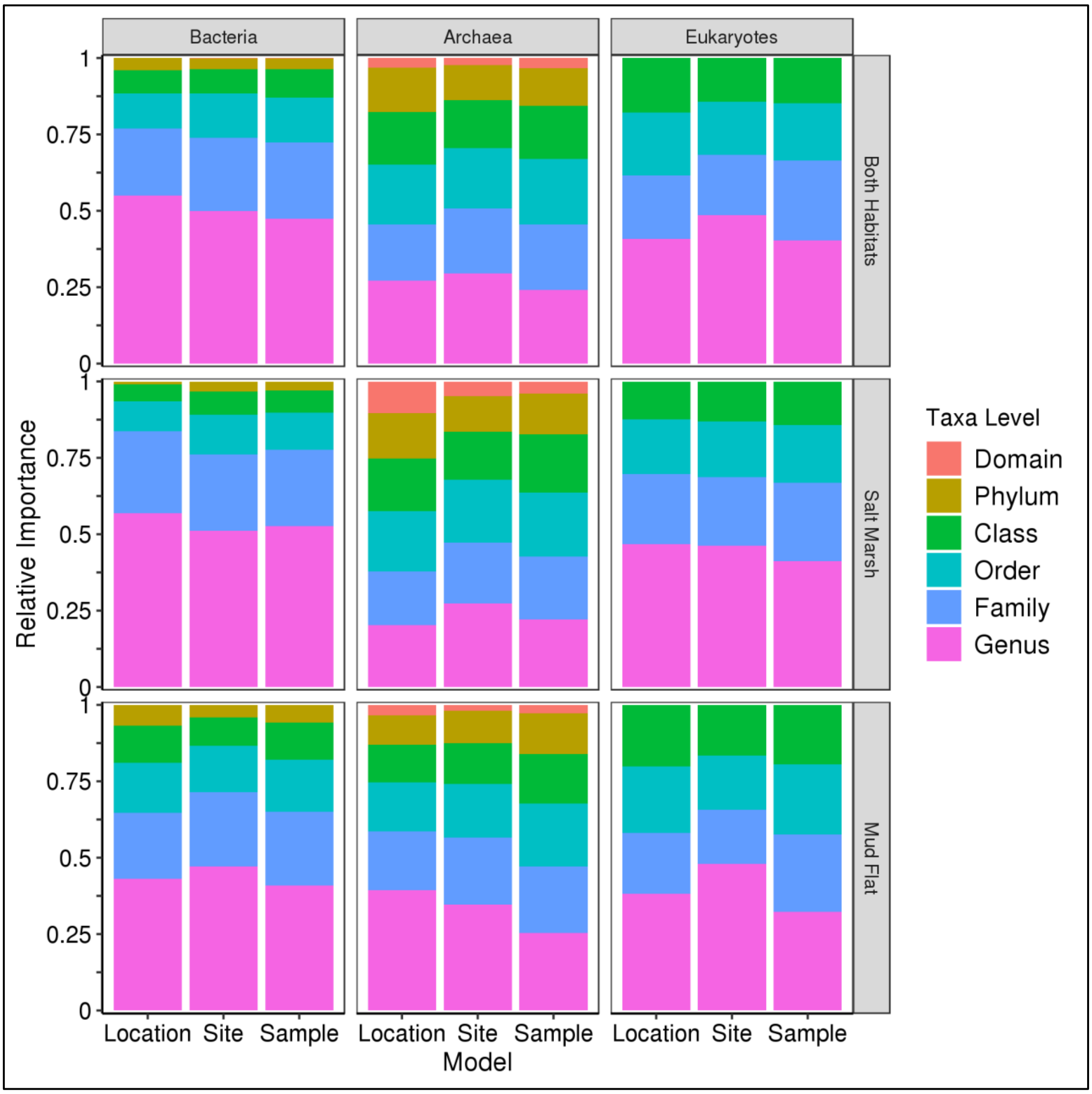


Figure 4. Relative importance values of random forest models built at various habitat, spatial scale, and microbial domain contexts. Models were created using datasets of relative abundance values of taxonomic identifiers at each taxonomic level. Aside from heightened importances of lower taxonomic levels, no clear trends were observed that cannot be explained by domain specific contexts.

importances. For the most part the salt marsh models had the highest correlations, followed by the full dataset models, and then the mud flat models. The correlations between models were generally between 0.6 to 0.7 with the differences being approximately between 0.1 and 0.2. Only the bacteria model using the full data showed any drastic changes in correlation between scales.

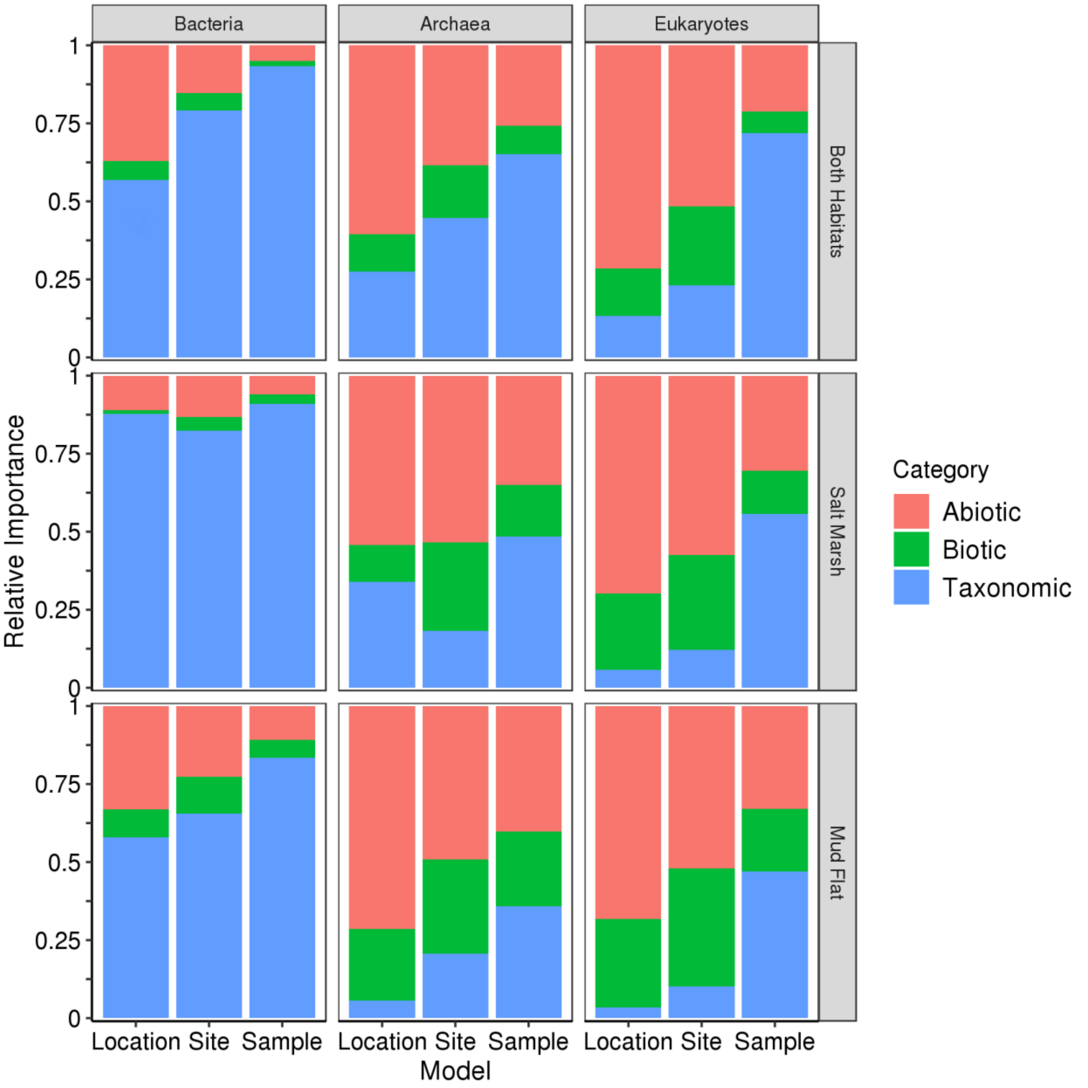


Figure 5. Relative importance values of random forest models built at various habitat, spatial scale, and microbial domain contexts. Models included both taxonomic and environmental data. Clear trends are observed across all scales, especially so for spatial scale and for domain.

Adding environmental measures to the models increased the clarity of the spatial distinctions for all models. There was a clear spatial scale pattern seen regarding the relative variable-importance of abiotic, biotic, and taxonomic factors (figure 5). Aside from the salt marsh models for bacteria and archaea taxonomic factors were consistently the most important at the sample level. The category then showed reduced variable-importance as scale increased. In other words, the relative variable-importance of environmental factors decreased as the scale of observation decreased. Taxonomic factors were the most important when data from both habitats were modelled, followed by the salt marsh models, and then the mud flat data. There were also clear differences observed across domains as well. The bacteria models had by far the highest variable-importance of taxonomic factors. In comparison, the archaea and eukaryotic models both had much higher variable-importance of environmental factors. A noticeable portion of that gain went to biotic factors at the site level.

Looking at the specific environmental factors for the bacteria models, soil geometry measures made up almost all the total environmental variable-importances. Total organic carbon and sediment chlorophyll measures were the only factors of any significance in the small biotic category slices. In the archaea and eukaryote models soil geometry measures still took up approximately all the relative variable-importance for the abiotic factors across the board but these models also had significant contributions from the biotic category that varied between habitats. The full data models for these domains still only had a showing of mainly total organic carbon measures followed by a smaller contribution of sediment chlorophyll measures. The mud flats had the same basic pattern as well, but variable-importance was clear for a mud flat specific category, bioturbation potential. In contrast, sediment chlorophyll had little variable-importance in the salt marsh models and the relative contribution of total organic carbon measures was reduced. Instead, the biotic categories for the two salt marsh models had a mix of salt marsh specific factors: bulk density, soil organic matter, and plant coverage. Of these factors plant coverage did show increased variable-importance as spatial scale increased. Soil moisture content and pH were also only taken in the salt marsh samples, but their variable-importance made up a marginal percentage of the abiotic factors.

Adding environmental variables improved the correlations between factor significance levels for the full dataset models (figure 3). The correlations for the salt marsh only eukaryotic models and all the mud flat models were increased and homogenized across scales as well. Meanwhile, the correlations between scales were lowered for the salt marsh bacteria and archaea models. Correlations dropped especially between site and location. These differing responses also switched what models had the highest and lowest correlations. The mud flat models now had the highest and the salt marsh models had the lowest in this context.

Predicting Across Habitats

The OOB (out-of-bag) estimate of the error rate for each random forest model, or the expected classification error rate for each model, differed by habitat, scale, and domain (figure 6). For reference, a lower OOB estimate indicates that the model is expected to have more accurate classifications. The models that incorporated environmental predictors were expected to be much more accurate and this hypothesis was reflected in their OOB estimates. Habitat differences were also observed with the mud flat only models having a lower OOB estimate and the salt marsh models having a higher OOB estimate compared to when the full dataset was used. Scale differences were obvious with the OOB estimates for models that classified based on location being much lower than the site-based models. Domain-specific differences were observed as well. Eukaryotic models displayed the highest OOB estimates for the site-based models, with the bacteria having the smallest estimates and the archaea models having estimates that were slightly larger.

Eukaryotic models were also the only ones with OOB estimates of any significance for the location-based models.

Contrary to the pattern seen with expected classification error rates, classifying the taxonomic data from one habitat with the model created using the data from the other habitat had the Eukaryotic models as the most accurate (figure 7). This pattern held true for classifying both location and site. The situation became more nuanced

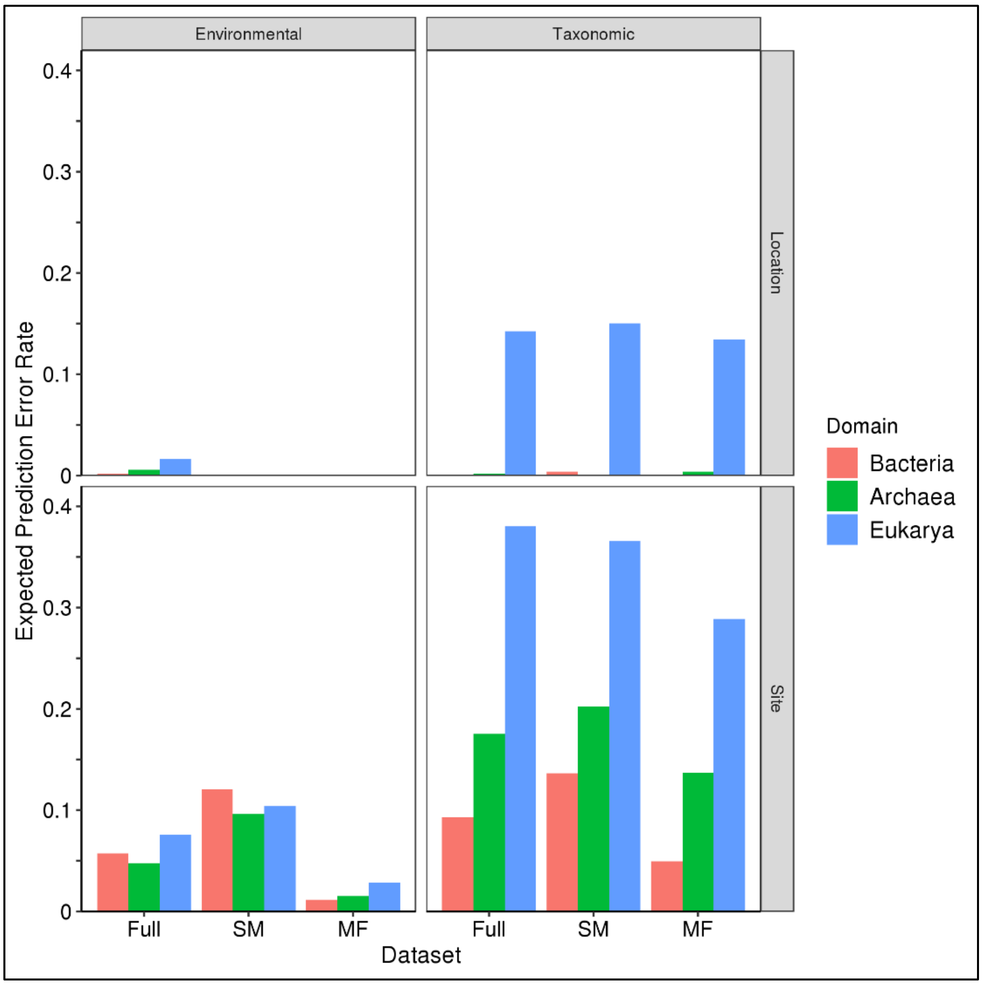


Figure 6. Expected prediction error rate of supervised random forest models built across multiple scale. Including environmental data in models decreased error rates in most cases. Location level models had higher error rates than site, the smaller spatial scale. Eukaryote models had much higher error rates in most cases as well. Little differences were seen between salt marsh models (SM), mud flat models (MF), and models using data from both habitats.

when the correct classification rate of specific factors was looked at. For the bacteria and archaea models the salt marsh models were heavily biased towards Essex and the mud flat models were biased towards Morecambe Bay. Both models seemed to be biased towards specific sites as well, which in turn influenced their location predictions as well. Meanwhile, the eukaryotic salt marsh models showed a smaller

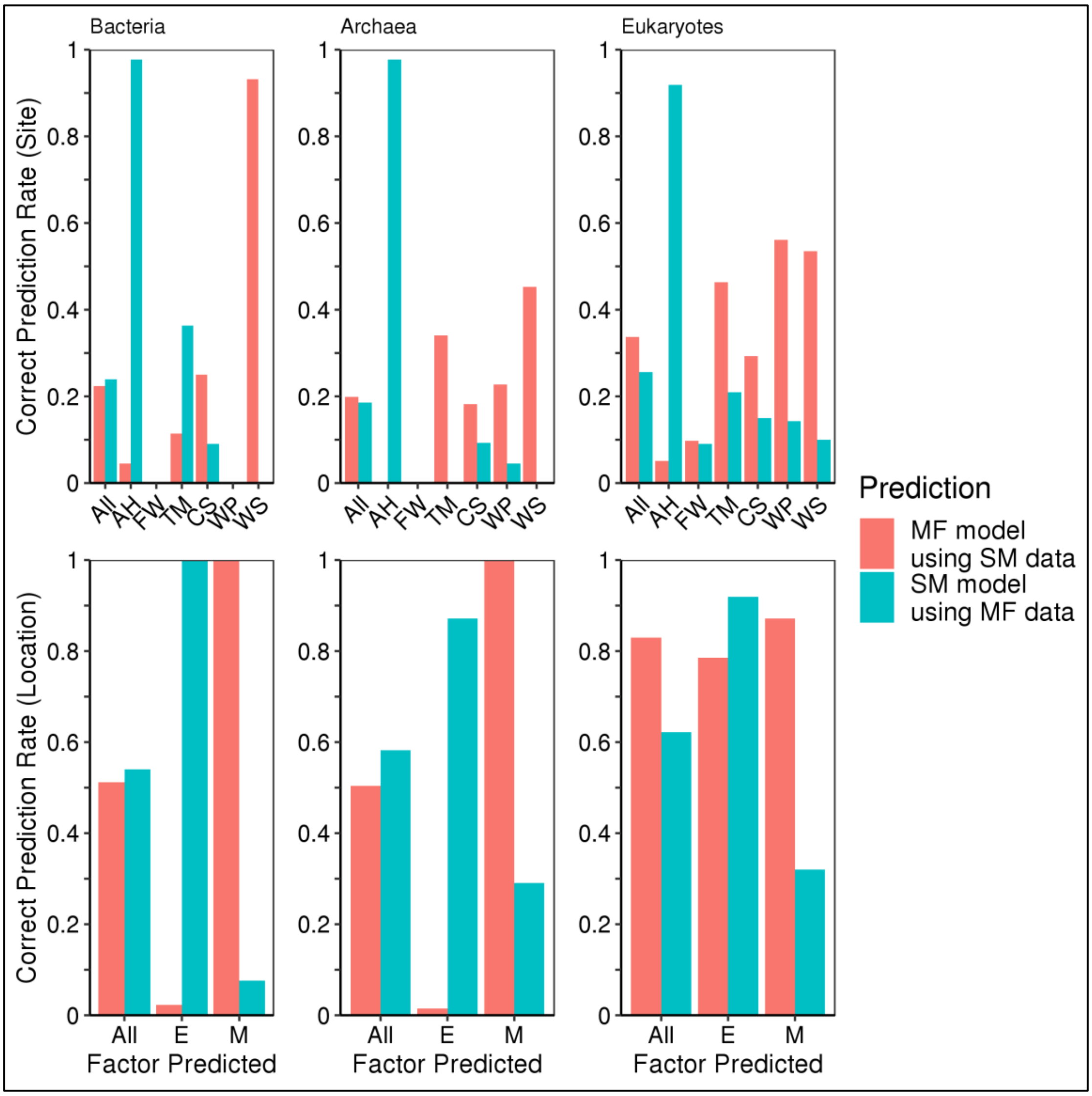


Figure 7. Correct prediction rate when classifying samples from one habitat using the model created from the other habitat. Predictions were done at the site and the location scale. Bacteria and archaea predictions both showed significant bias towards specific spatial contexts while microbial eukaryote prediction accuracies were more consistent.

bias towards Essex and the mud flat models showed very little bias. When analysing the validity of these biases, it was confirmed that they were not a by-product of a lack of overlapping taxonomic groups between habitats. Instead, the biases were likely a by-product of differing relative abundances of taxa between mud flats and salt marshes.

Discussion

Scale Effects on Structuring

Results showed that spatial scale plays an important role in interpreting what factors are important in structuring and classifying microbial communities (Bardgett and van der Putten, 2014). The gradient used should correspond to the study organism and, as seen here, the heterogeneity of microbial communities is only properly captured at smaller-scales (Hendershot et al., 2017). The factors being considered does play a role in this observation though. When only taxonomic factors were considered individual factor variable-importances did change but the overall relative variable-importance of categories did not. In contrast, clear categorical patterns were observed when the environmental factors surrounding the microbial communities were considered as well. Microbial communities do not exist in isolation, so it makes sense that both environmental interactions and interactions between microbes need to be considered to best understand microbial ecology (Gore, 2018).

Environmental factors were consistent within random forest model groupings so changes in relative variable-importance values were dependent on what taxonomic domain was used, which habitat was considered, and what spatial scale the model classified. Therefore, if environmental factors had a higher relative variable-importance in a random forest model then variation in environmental factors was of more use in classifying samples. This higher variable-importance also implies that environmental factors are more useful in explaining variation in taxonomic abundances, potentially indicating an increased role of environmental selection. Greater heterogeneity in environmental factors naturally exists at larger scales, so this study’s observation of environmental selection being more prevalent as scale increases makes sense (Zinger et al., 2019).

Conversely, a higher variable-importance of taxonomic factors in classifying spatial categories can be interpreted as dispersal limitations and stochastic effects having greater effects in structuring communities. Dispersal limitations can mask relationships between environmental factors and community compositions in a variety of ways, so it is important to not understate their effects (Evans et al., 2017). The concept of metacommunities helps explain the likely role of dispersal limitations in the microbial communities studied here. Most of the species in microbial communities likely draw from a regional metacommunity with specific local compositions depending on conditions. The size of the metacommunities in this study are likely smaller than the site level category but their presence still plays a role in structuring. Not all microbes have the capacity for global distributions so the makeup of a given metacommunity should in theory define the dispersal limitations of different microbes (Lindström and Langenheder, 2012, Barberán et al., 2014). Combine these metacommunities with the fact that initial species abundances are important in future community composition and the role of dispersal limitations in structuring communities becomes clear (Fukami, 2015; Dann et al., 2019). Extinctions in belowground systems are poorly studied so regional extinctions may have played a role in defining the realized dispersal ranges of various microbes as well (Veresoglou et al., 2015). Environmental factors usually play more significant roles at larger scales so it was expected that dispersal limitations would show a decreased relative variable-importance. Alternatively, as scale increases metacommunities can be viewed as dispersal networks so the elevated level of taxonomic variable-importances in certain models can be explained as well.

Habitat Effects on Structuring Microbial Communities

The increased variable-importance of taxonomic factors in the models that used data from both habitats signifies that more than just spatial scales affect the interpretation of structuring dynamics. Modelling the data from only one habitat type resulted in environmental factors being more important than in the both habitat models. Using data from multiple habitats added more variation that the model had to explain, ultimately at the expense of the variable-importance of environmental factors. One such factor category was soil particle metrics.

Soil particle metrics were the most important environmental factor for differentiating samples in all the models which is consistent with other studies (Vera-Gargallo et al, 2019). This importance stems from the physical structure of belowground systems creating microhabitat complexity, allowing for considerable amounts of niche partitioning (Decaëns, 2010). However, this category had reduced relative influences in structuring when both habitats were considered (Stegen et al., 2012). As seen by the NMDS done in this analysis, the different habitats had different microbial community compositions (Lindström and Langenheder, 2012), but not all particle size measures had distinct habitat groupings. In the context of random forest analysis, having overlap in particle size measures across habitats created uncertainty in classification, reducing the relative variable-importance of the category. Instead, the more distinct differences in taxonomic abundances had higher variable-importance in structuring. Looking at single habitats did not have this issue, increasing the variable-importance of this category. Keep in mind that the relative variable-importance of soil particle metrics was still high despite the reduction when both habitats were considered. Not all environmental factors were reduced in the cross-habitat models though. Total organic carbon was an influential category in all models which was expected because of the effect of nutrient availability on community composition (Yin et al., 2019).

The presence of vegetation played numerous roles in the salt marsh samples. The roots of plants create unique microhabitats that harbour microbial populations (Whipps, 2001). These roots also produce root exudates which play an important role in both a plant's nutrient uptake and in shaping microbial community composition (Haichar et al., 2012; Guyonnet, et al., 2018). This array of plant-related feedbacks helped to create the much more nuanced consortium of biotic factors that were important in structuring and classifying the salt marsh samples. Additionally, plant evenness has been shown to be important in determining microbial community structure and it is likely that this evenness and its associated feedbacks has variable effects depending on both scale and microbial domain (Sun et al., 2019). Correlation analysis revealed that these variable effects are most apparent when comparing factors important to structuring salt marsh samples at the site level compared to the location level. This variable scaling effect was not seen in the eukaryotes, implying that there are different factors at play in structuring that domain that are more consistent across spatial scales. Traditionally an important factor in structuring communities (Hendershot et al., 2017; Rath et al., 2019), pH was another salt marsh specific measure. However, it was not a very important factor in the models created here. Only having one variable associated with the category when looked at in the breadth of all the other factors played a role in reducing the overall variable-importance of pH.

The mud flat models exhibited similar trends as the salt marsh models with environmental structuring being more prevalent compared to the full models. However, the contextual reasons behind these trends are different. In other soil systems without plant life, also known as incipient soil systems, microbial communities have been found to be selected primarily through environmental drivers (Sengupta et al., 2019). For instance, bioturbation by invertebrates plays an important role in structuring by creating different oxygen regimes for microbial communities (Li et al., 2019) which helps explain the increased relative variable-importance of the biotic category observed here. However, when soil systems first form microbes play a larger role in shaping their environment in incipient systems (Sengupta et al., 2019). Therefore, the mud flat communities observed here could still be affected by higher dispersal limitations than what was seen since the microbial communities would have played a larger role in creating the current environmental conditions.

Another important temporal factor is tidal fluxes. These fluxes can be a dispersal mechanism for microbes and mud flats are subjected to greater tidal pressures than salt marshes due to their lack of vegetation (Pennings et al., 2005; Coulon et al., 2012; Yao et al., 2019). These increased tidal effects may help shuffle microbes throughout each site, decreasing the effect of dispersal limitations and lowering the relative variable-importance of taxonomic factors (Stegen et al., 2013). In this scenario soil geometry and other local conditions still play a role in determining where microbes settle, maintaining the role of environmental factors (Li et al., 2019). Note that the possible role of tidal dispersion could not be assessed directly in this study due to their temporal nature. Temporal dynamics, such as tides, were observed across multiple scales but this study only contained seasonal variations (Lauber et al. 2013).

Domain Effects on Structuring Microbial Communities

Trends were observed in relative variable-importances across domains for the taxonomic only models and for the full environmental models. The most meaningful interpretation of the taxonomic only models stems from the higher relative variable-importance of lower taxonomic groups in structuring which was due to multiple dynamics. More related taxa are more likely to be ecologically similar (Stegen et al., 2013) and positive interactions between functional groups show positive covariance in occurrences (Lee et al., 2019). Therefore, lower taxonomic groups have stronger groupings both within their taxa and with functional groups they interact positively with, increasing their relative variable-importance in structuring communities. In the case of the full environmental models, microbial interactions with environmental factors and other microbes both played important roles in structuring. Each microbial domain has a different set of interactions, so it makes sense that variable-importance in structuring samples differed by domain in this analysis (Gore, 2018).

Why bacteria showed a much higher degree of taxonomic variable-importance was due to multiple factors. While recent research suggests that bacteria do not have strong functional redundancy from a diversity perspective (Peter et al., 2011; Delgado-Baquerizo et al., 2016), bacteria still show high degrees of functional redundancy across taxa (Burke et al., 2011; Nielsen et al., 2011; Banerjee et al., 2016). Therefore, past observations of the domain’s lack of any strong links between taxonomic diversity and functional ability (Wang et al., 2019) may be indicative of more stochastic structuring mechanics. This neutral theory-based approach implies that dispersal limitations and other stochastic mechanisms like ecological drift would cause bacterial communities in different locations to be noticeably different taxonomically, without directly impacting the functional ability of the communities. These community patterns were observed here in the bacteria models and, to a lesser extent the other domains, but it does not automatically support neutral theory. Neutral models have predicted bacterial abundances better in certain systems (Moroenyane et al., 2019), but biological interactions within taxonomic groups played a significant role in creating the observed dispersal limitations and cannot be discounted.

Dispersal limitations can occur via restriction of movement or the restriction of successful establishment (Hanson et al., 2012), so species interactions can play an important role in shaping communities as well (Friedman et al., 2017; Zhao et al., 2019a). These interactions can vary from competition shaping the composition of ammonia-oxidizing bacteria (Veresoglou et al., 2018) to mutualism playing a role in shaping range expansions (Momeni et al., 2013). Bacteria have also been shown to interact with their environment by altering pH to create more favourable growth conditions. This pH change results in inhibiting the growth of other bacteria that are not well adapted to the dominant pH (Dohi and Mougi, 2018). In this analysis, this inhibition likely translated to an increased variable-importance for taxa that have this ability. Network analysis has identified keystone taxa as being important in structuring bacterial communities (Banerjee et al., 2018) and this random forest analysis has expanded this relationship further by identifying taxonomic factors as being important in structuring communities even in the context of environmental factors. Additionally, extinction events in belowground systems are poorly studied but they may have influenced the distributions observed in this analysis as well (Veresoglou et al., 2015). Neutral theory-based structuring did play a role in bacteria samples analysed here, but biological interactions still seem to be the driving force structuring communities.

In contrast, the archaea and eukaryote samples in this analysis were largely influenced by environmental factors at larger scales and it is difficult to ascribe any functional reason as to why this occurred. It is important to keep in mind that environmental variables were the same for each model. Considering that consistency and the reasoning used to explain the behaviour of the bacteria models, it seems that the environment played a much greater role in structuring the archaea and eukaryote communities. Sample level variable-importances still showed a strong taxonomic presence so it seems that scale plays a critical role in structuring dynamics even when domains show noticeable differences in structuring dynamics. It is also worth noting that the eukaryotic models had much higher classification error rates. Larger organisms are more stochastically driven (Zinger et al., 2019) and stochastic processes have been observed more in eukaryotic communities (Li et al., 2017). Therefore, the higher classification error rates of the eukaryotic models indicate that the reduced relative variable-importance of factors in classifying samples is reflective of the more random distribution of microbial eukaryotes instead of a heightened variable-importance of environmental factors. In this case, eukaryotic communities followed a more neutral structuring patterns, albeit for different reasons than the bacterial communities.

Regardless, the environmental factors that showed high relative variable-importance in structuring made sense for both domains. Soil geometry and total organic carbon, the top two environmental factors, have both been shown to play an important role in determining the structure of microbial communities, with soil particle metrics also showing high variable-importance in differentiating samples (Zhang et al., 2015; Waldrop et al., 2017; Vera-Gargallo et al, 2019). Environmental filtering (Zhao et al., 2019; Liu et al., 2019) having a larger effect on archaea and eukaryotic communities may be a big contributor to the patterns observed in this study but further research needs to be done to verify this claim.

Noticeable patterns were found when predicting spatial classifications across habitats as well. Most soil diversity was found in both habitats across domains but there were significant differences in diversity patterns based on habitat so high predictive accuracy was not expected (Ramirez et al., 2014). However, bacteria and archaea models showing almost identical predictions while the eukaryotic models differed was not expected. Instead of structuring trends being the delimiter between groups it seems that expected model accuracy was the delimiter. While the bacteria and archaea models may have had more accurate location-specific taxonomic indicators, that did not transfer well when predicting across habitats. Instead, it was the eukaryotes that were more predictive despite having higher classification error rates. The more equal model classification rates of the eukaryotes provide more evidence for the more stochastic nature of the domain.

Modelling prokaryotic communities was more likely to be skewed by sites that show unique characteristics. Therefore, how prone a model is to outliers may be based on the degree of stochasticity seen in the study organisms. The differences in correct prediction rates between habitats seen in this analysis further highlight the strong heterogeneity of microbial communities and how it limits the ability to extrapolate microbial biodiversity patterns (Lladó et al., 2018). This extrapolation ability would be even more limited in habitats that have higher amounts of unclassified microbes due to the higher degree of uncertainty in community compositions (Delgado-Baquerizo, 2019).

Conclusion

Based on this random forest analysis it is clear that the relative importance of factors in structuring microbial communities and the environment around them differs across spatial scales, habitat types, and domain. Taxonomic based dispersal limitations only being visible at the sample level in most cases points to the significance of spatial scale in structuring dynamics. These dynamics also differed across habitats since different habitats showed different scaling trends and different variable-importances. Domain arguably had the biggest effect on interpretation though. Bacteria seemed to be structured almost entirely by dispersal limitations while archaea and eukaryotic communities were much more environmentally driven, prompting questions regarding the underlying mechanics that led to this divide. Domain also had an effect when using single habitat models for prediction of data new to the models, with the prokaryotic domains showing high bias to specific sites and locations and the eukaryotes being more balanced. However, the differences shown in the eukaryotic models may have been due to their high classification error rates, creating alternate interpretations of their data depending on how much the classification error rate was considered. Overall this study revealed many novel insights regarding the structuring of microbial communities and the applications of machine learning in microbial ecology.

CHAPTER 4: BIODIVERSITY-ECOSYSTEM FUNCTIONING RELAIONSHIPS IN MICROBIAL COMMUNITIES

Introduction

The relationships between biodiversity and ecosystem functioning (BEF) has been a long-running subject of interest in ecology, but there is still much to learn (Reiss et al., 2009; Gamfeldt et al., 2013). As biodiversity is linked to ecosystem functions (Hooper et al., 2005; Duffy et al., 2017), there is tremendous value in conserving biodiversity to maintain functionality (Cardinale et al. 2012; Isbell et al., 2015; Cavicchioli et al., 2019). There are many ways to study BEF relationships with taxonomy, functional ability, and the relationships between these measures all being used. While there has been an increased push for the use of functional measures in BEF research, taxonomic measures do still hold merit (Flynn et al., 2011; Gagic et al., 2015; Laureto et al, 2015; van der Plas, 2019).

In the case of microbial communities, both microbial abundance and the composition of a microbial community have significant effects on functioning (Strickland et al., 2009; García-Palacios et al., 2014; Lipson and Xu, 2019). Furthermore, many ecosystem functions are underpinned by microbial biodiversity (Wall et al., 2012). In addition to nutrient cycling, soil microbial communities play important roles in crop production (Jeanne et al., 2019) and in maintaining the health and stability of the soil (Dubey et al., 2019). While BEF relationships have been established for belowground systems, the limited amount of research on them limits the depth of understanding of these relationships (Bardgett and van der Putten, 2014; Hendershot et al., 2017). In comparison to soil systems, coastal sediment communities are naturally different due to tidal and other aquatic effects but much of the same still applies to them. Sediment communities are also crucial in ecosystem process delivery via nutrient cycling and there are still many questions surrounding sediment microbial diversity dynamics (Bowen et al., 2001; Levin et al., 2001; Moulton et al., 2016).

Different microbial domains associate with process rates in different ways so it is natural to expect that BEF relationships vary across microbial domains (Hale et al., 2019). Bacteria play key roles in the cycling of carbon, nitrogen, and phosphorous (Lladó et al., 2017) and are important in pollutant degradation and nitrogen removal. They can also be important in promoting plant growth by absorbing nutrients and promoting nitrogen fixation (Chen et al., 2018). Archaea are also widespread in soils and other moderate environments despite being better known for their more extremophilic roles (Bates et al., 2011). Archaea are important in ammonium oxidation amongst many other nutrient cycles (Tourna et al., 2008). For instance, they are the taxon that carries out most of the methanogenesis that occurs in nature (Kietäväinen and Purkamo, 2015), an important step in the decomposition process (Peng et al., 2008). Meanwhile, microbial eukaryotes mainly operate as decomposers, predators, producers, and parasites (Bik et al., 2012). The sheer amount of diversity in each domain does limit the ability to directly ascribe a measured function as being directly due to a specific taxonomic group in most cases. However, there are still associations that emerge between taxonomic levels (Barberán et al, 2012).

Proper scaling is essential for the capture of potential BEF relationships. Differences have been observed in the biodiversity requirements to maintain BEF relationships across spatiotemporal scales indicating the importance of scaling in the interpretation of BEF relationships. The processes investigated in a given study also have a large bearing on results since studying individual processes underestimates the importance of biodiversity in maintaining multifunctional ecosystems (Hector and Bagchi, 2007; Isbell et al., 2011). Accounting for all these factors in BEF research while still answering meaningful questions can be difficult since doing so requires combining concepts from multiple fields of ecology (Wang and Brose et al., 2018). In the case of microbes, the complex and still relatively unknown dynamics of their communities have limited the ability of researchers to uncover relationships using traditional methods. This limitation is especially apparent when attempting to extrapolate findings outside of the specific context in which they were gathered (Lladó et al., 2018).

Studying communities via network analysis has potential to reveal novel insights by analysing the interaction dynamics of microbial communities (Faust and Raes, 2012; Ramirez et al., 2018a). Microbial species interactions play an important role in community structuring and process delivery (Brophy et al., 2017; Friedman et al., 2017). Therefore, network analysis is useful for capturing the underlying structuring dynamics and functional ability of communities. Functional ability relates more to the number of connections in a community and not the number of species so network analysis has been especially useful in that regard. Previous work using network methods have already revealed non-random structuring of microbial communities, associations across taxonomic levels, and unexpected taxonomic and functional relationships (Barberán et al, 2012). Having the ability to capture these interactions goes a long way in enabling researchers to study these complex communities with more resolution.

Recent advances in network analysis methods have revealed new insights into the existence of keystone taxonomic groups and guilds that are important in structuring microbial communities (Banerjee et al., 2018). Interactions amongst microbes within a community play an important role in influencing many other microbial interactions so it makes sense that certain nodes would emerge as being structurally important (Mamet et al., 2019). These nodes could potentially play roles in BEF relationships as well since some functions such as cellulose degradation rely more on specific species composition than overall biodiversity (Peter et al., 2011). Furthermore, these important taxa are not always the most abundant taxa in a given system since rarer taxa have been identified as being important for structuring soil communities as well (Ramirez et al., 2018b). To what extent these rare taxa play a role is still up for debate as the presence and role of the “rare microbial biosphere” is still contested (Louca et al., 2019).

There is a debate surrounding the usefulness of network analysis in studying microbial communities as well. Some argue that the ability of co-occurrence networks to capture relationships may be overstated. The complex nonlinear dynamics, the compositionality of sequencing data, and the natural variability of microbial communities all limit the use of correlation-based analysis in inferencing interactions (Carr et al., 2019). Additionally, some associations are captured better than others across methods with competitive communities being easier to capture than predator-prey and parasitic communities (Hirano and Takemoto, 2019). Even if networks perfectly captured all the interactions in a community, some processes are impacted independent of biodiversity changes (Li et al., 2018). Therefore, not all process patterns can be explained via microbial diversity anyways. Nevertheless, the applications of network analysis in microbial BEF relationship research are huge and well worth pursuing.

The analysis in this chapter used the CBESS dataset to investigate potential BEF relationships amongst microbial communities. Microbial community co-occurrence networks were constructed and used to test for potential BEF relationships. Three main questions guided this analysis: (1) are there any relationships between microbial network dynamics and ecosystem processes, (2) what role does spatial scale play in capturing these relationships, and (3) how do these relationships change across domains? It was hypothesized that relationships would exist for network size and network interconnectedness and relationship capture was expected to be better at the site level. Differences across domains were expected as well.

Methods

Co-occurrence networks were made using a modified version of the function used by Ju et al. (2014). In this function a correlation matrix was constructed using the relative abundance table. Spearman's correlation was used for the correlation matrix because of the nonlinearity of microbial distributions across samples. The correlation values and their associated p-values were then extracted. Benjamini-Hochberg standard false discovery rate correction was used to adjust p-values. The correlation matrix was then filtered to only include the correlation values that passed both the correlation threshold (0.6) and the p-value threshold ($p < 0.01$). This filtering included both positive and negative correlations. The filtered matrix was then transformed into a node list and an edge list. A node list represents the unique identifiers for all of the nodes used to construct the network and an edge list represents all of the connections between nodes. A checkerboard-score (C-score) test run was on the inputted data as well to evaluate the randomness of the community and to see if that randomness was significantly different from that of a null model. A higher C-score indicates that a network had a lower degree of randomness. Overall this function yielded a list with C-score metrics, the full correlation and p-value matrices, the filtered correlation matrices, node lists, and edge lists. All networks were retained for further analysis regardless if they passed the C-score test or not.

Network topology metrics were then extracted from networks after they were constructed using the igraph package (Csardi and Nepusz, 2006). The metrics collected all worked to capture the size and interconnectedness of the networks in some manner:

Nodes (also known as vertices) – the number of taxa present in the network

Edges - the number of connections between nodes.

Average degree - the average number of edges across all nodes.

Density - the number of edges divided by the number of possible edges.

Diameter - the longest path between any two nodes

Average Path Length (APL) - the average of the shortest paths between all pairs of nodes.

Modularity - the degree of distinctness between different node groups. A modularity value of 1 indicates that a network contains lots of sub-clusters with little interconnectedness.

Clustering Coefficient - the probability that the adjacent vertices of a vertex are connected.

Bootstrapped null models (n =1,000) were generated for each network to compare the observed topology metrics to the topology metrics of the randomly generated networks.

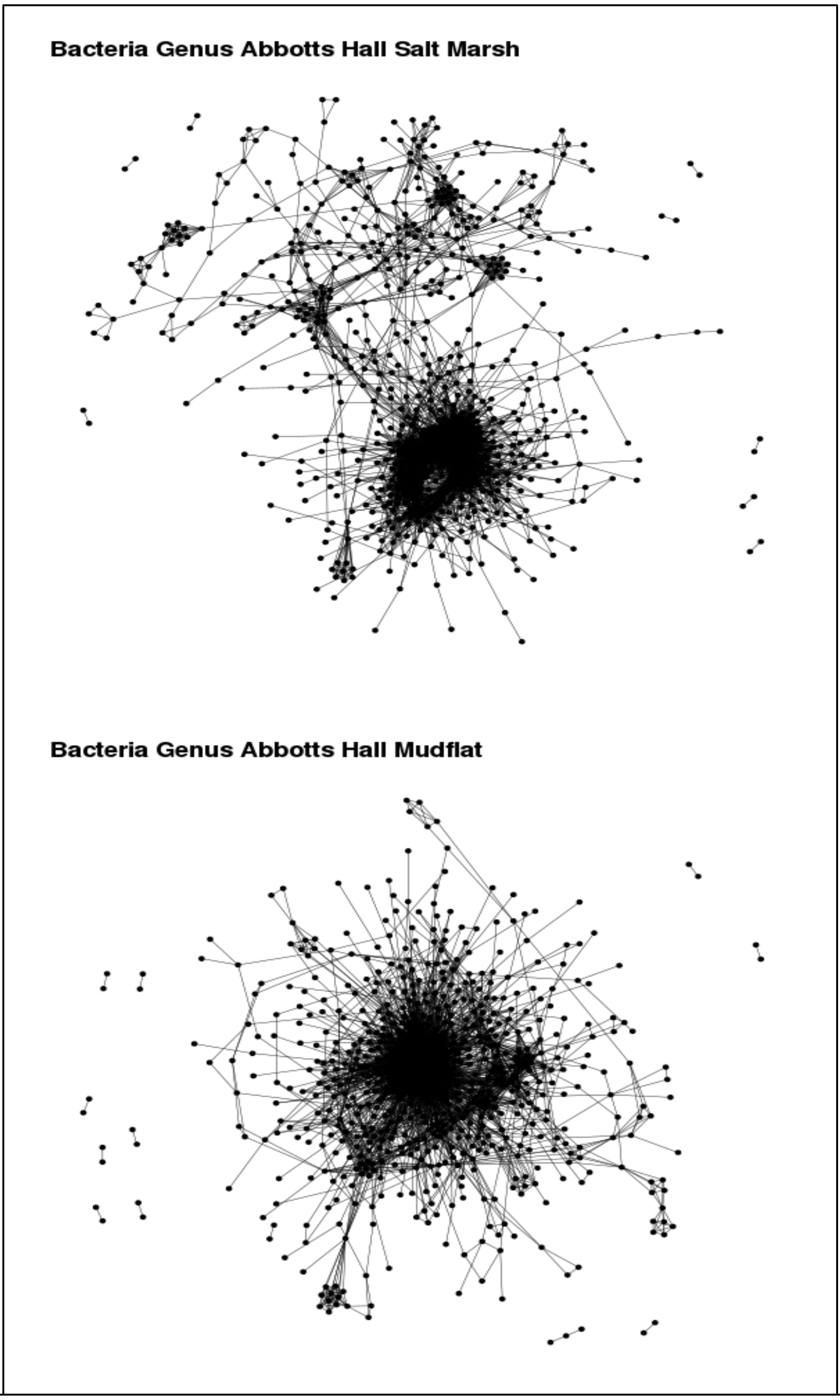
Co-occurrence networks were constructed for each taxonomic level within each domain except for the phylum level for the Eukaryotes. Networks were made for each factor at each spatial scale and for each habitat type as well. Networks were also made using the full dataset and using data from both habitats. Co-occurrence networks were visualized using the tidygraph package (Pedersen, 2019).

To investigate how co-occurrence network metrics influenced ecosystem processes, principal component analysis (PCA) plots were created to visualize all the ecosystem process measures in a 2D ordination space. Scaled PCAs were created using the *prcomp* function in R. A PCA differentiating between habitat types was constructed for processes at the site, location, and full dataset scale. The first three principal components (PC) were retained for each PCA in order to capture enough of the explained variation for all the PCAs. The PCA scores of each PC were then analysed to see if there were any significant relationships with microbial community co-occurrence network metrics. To analyse these relationships linear models were created between a given PC and a given network metric, with the PC as the response variable. An ANOVA (Analysis of variance) test was then run to evaluate if the relationship was statistically significant or not. The p-values were extracted from each value and the nature of each relationship was visualized to enable comparison of relationships across spatial and taxonomic scales. This analysis was also run using the average richness at each taxonomic level for each domain.

Results

Co-Occurrence Networks

Co-occurrence networks were successfully constructed across all domains, taxonomic levels, spatial scales, and habitats (figure 8). Overall the distribution of the network topology metrics followed logical trends. Metrics for site and location level were largely consistent aside from site-level networks having higher values for measures that reflected network size. Across both scales, differences were observed between domains, taxonomic levels, and habitats (table 2). Trends across taxonomic levels were observed in node counts, edge counts, modularity, density, APL, diameter, and average degree. Most patterns were straightforward but modularity and average degree did show more nuanced trends. Modularity only dropped in certain archaea contexts and average degree only decreased in bacteria and eukarya. Noticeable domain trends only emerged in node counts, edge counts, average degree, and modularity. The same was true for all but average degree for habitat trends.



Bacteria Genus Abbotts Hall Salt Marsh

Bacteria Genus Abbotts Hall Mudflat

Figure 8. Examples of co-occurrence networks made using different microbial communities. Co-occurrence networks were made at multiple spatial scales across two different habitats for each taxonomic level within each microbial domain. Each co-occurrence network displayed its own unique patterns specific to the contexts in which it was created.

Some metric specific behaviour was observed as well. Observed clustering coefficient measures varied from 0.9 to 0.35. Note that not all the mud flat archaea models had clustering coefficient measures due to networks being too fragmented. Modularity measures were near zero for almost all bacteria networks while both archaea and eukaryotes had modular structures (values > 0.4) at lower taxonomic levels (Ju et al., 2014).

It was difficult to identify any consistent interconnectedness trends for each domain. For bacteria, observed APL and diameter values being greater than expected from a random network indicate longer travel paths than expected. Modularity levels being near 0 indicate more overall interconnectedness than expected though. Archaea seem to be more interconnected than expected with all those metrics aside from diameter showing the favourable patterns. In comparison, the eukaryotes followed the same trend as the bacteria model aside from their modularity measures being inconsistent indicating lower interconnectedness than expected. Overall network size, complexity, and interconnectedness differed to varying degrees based on domain, habitat, and spatial scale.

Table 2. Behavior of each of network metric in microbial co-occurrence networks. Networks were made across taxonomic levels for bacteria, archaea, and eukarya. Two habitats, salt marshes and mud flats, were sampled and two different spatial scales were used for network construction, with site being the smaller scale.

Network Metric	**Trend as Taxonomic Level Increased**	**Spatial Trend**	**Domain Trend**	**Habitat Trend**
Nodes	Decreased	Higher at site level	Salt Marsh: bacteria > eukarya > archaea Mud Flat: eukarya >= bacteria > archaea	Slightly higher in salt marshes
Edges	Decreased	Higher at site level	Salt Marsh: bacteria > eukarya > archaea Mud Flat: eukarya >= bacteria > archaea	Slightly higher in salt marshes
Average Degree	Decreased only in bacteria and archaea	-	bacteria > eukarya > archaea	-
Density	Near 0 at lower levels and then increased	-	-	-
Diameter	Generally decreased	Higher at site level	-	-
Average Path Length (APL)	Generally decreased	Higher at site level	-	-
Modularity	Dropped in certain archaea contexts	-	bacteria: consistently near 0 Scattered positive values in other two	More variable in single habitat networks
Clustering Coefficient	-	-	-	-

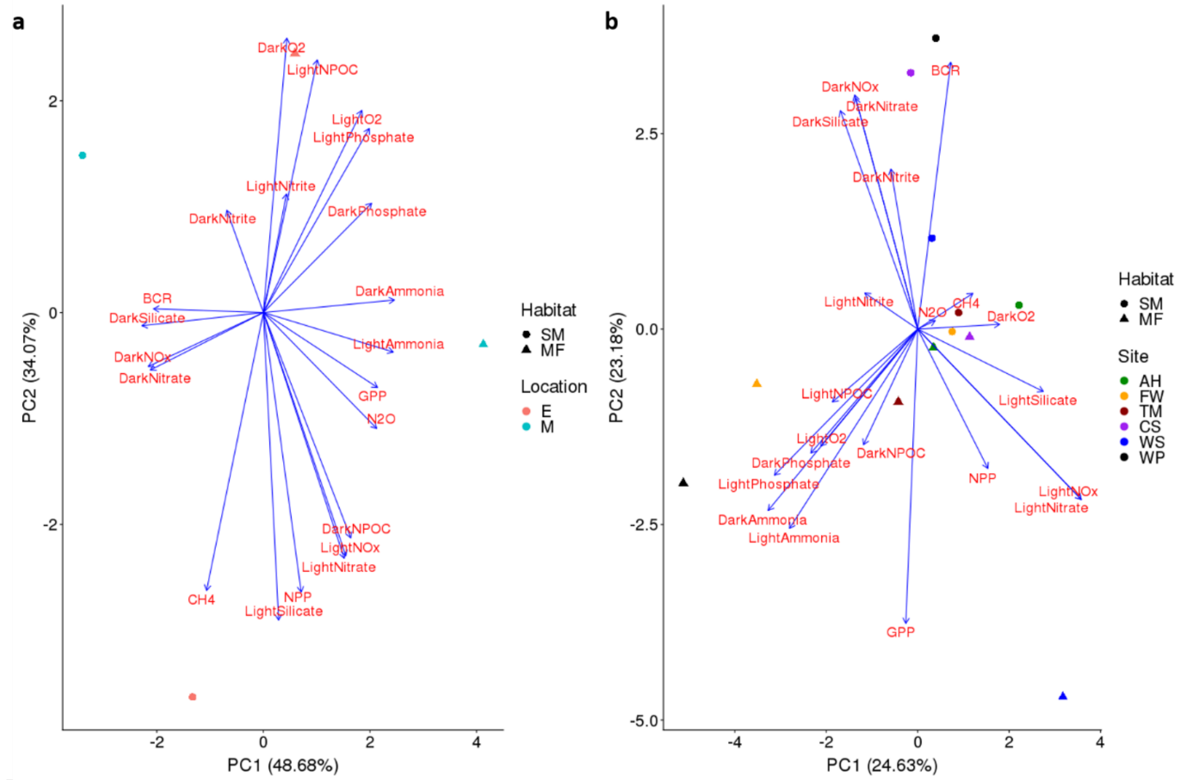


Figure 9. PCA biplots of the ecosystem processes measured in the CBESS dataset at the (a) location level and (b) site level, with site level being the smaller spatial scale. Each axis represents a principal component and how much variation it explains. Points represent the spatial context and vectors represent the loadings. Each loading is labeled based on the process rate it represents. Light measures were taken in sunlight and and dark measures were not (NPOC - Non-Purgeable Organic Carbon). Habitat differences were better observed and loadings displayed much tighter groupings at the site level.

Relationships between Network Structure and Ecosystem Processes

The PCA plots of process profiles each had three main components (figure 9). The proportion of variation explained by each PC dropped as spatial scale decreased. For the PCA using the full dataset, approximately 100% of the variation was captured in PC1. In comparison, both the location and site ordinations had many more PCs generated. At the location level, PC1, PC2, and PC3 explained 49%, 34%, and 18% of the variation respectively. Meanwhile, PC1, PC2, and PC3 only explained 25%,

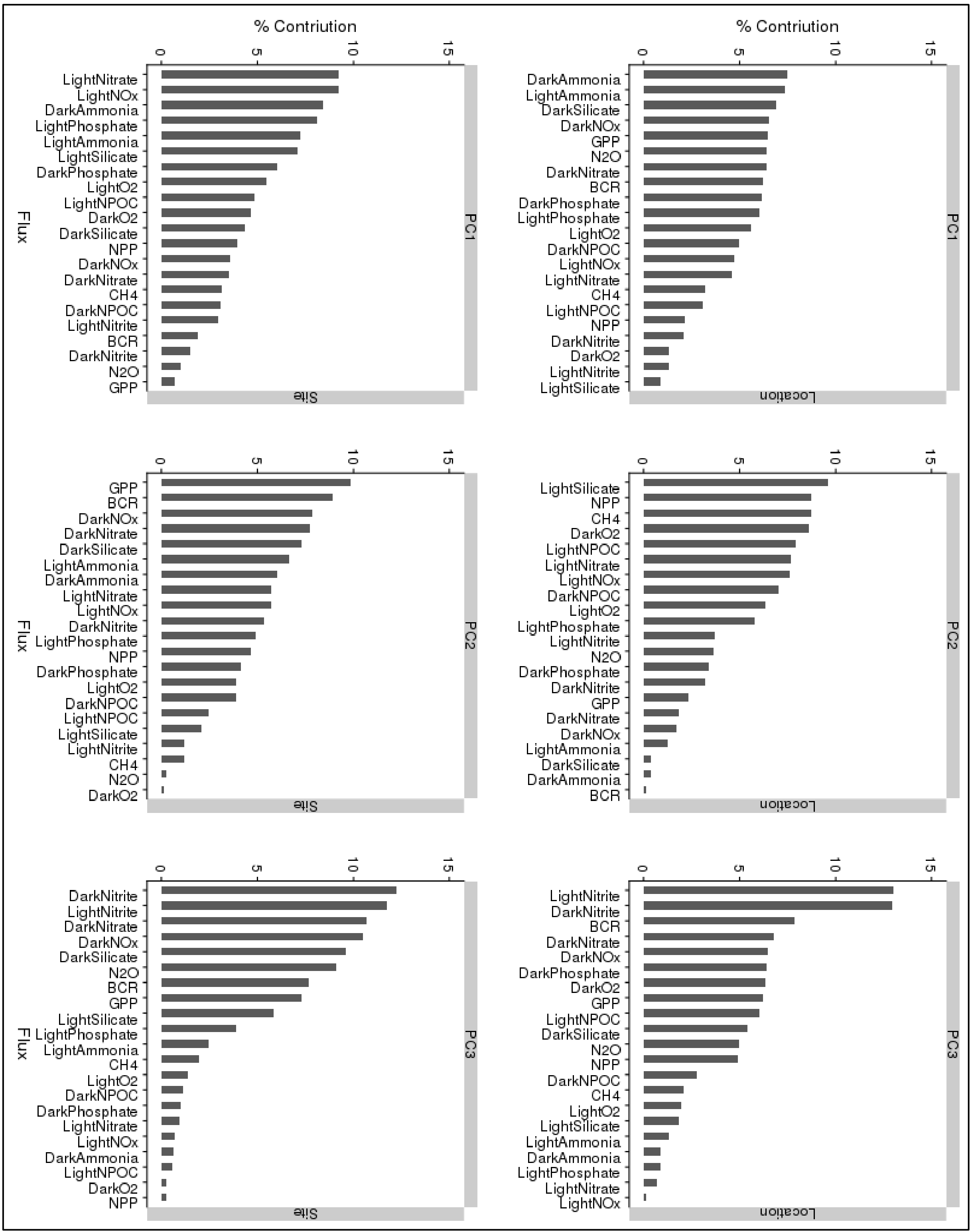


Figure 10. Contributions of process rates to the first three PCs for the location and site level PCAs sorted by highest contribution. While each PC displayed different contribution patterns, PC1 at the location level and PC2 at the site level were similar.

explained by the first two PCs in the site model was the primary reason why the first 23%, and 15% of the variation at the site level. The limited degree of variation three PCs were retained for further analysis.

Looking at the loadings of each process (figure 10), each process explained PC1 of the ordination at the full dataset level equally. In contrast, the contributions were variable at the location and site level and an interesting trend emerged across spatial scales. The PC contributions for PC1 at the location level and PC2 at the site level follow almost the same trend. Looking at the fluxes based on highest contribution, site-level PC1 looks to be majority driven by nutrient fluxes in the light and site-level PC2 looks to be majority driven by GPP, BCR, and nutrient fluxes in the dark fluxes. Location level PC1 looks to semi-mirror site-level PC2 from this perspective as well.

Spatially, the PCA plots showed similar patterns to the NMDS of OTU relative abundances regarding habitat patterns. There was little to no overlap in flux profiles across habitats, but the scores were closer together at the site level. This closeness indicated that there was more variability in flux profiles which led to less distinction between spatial categories. Additionally, there was more distance in PCA scores for the Morecambe sites when compared to the Essex sites. As expected, the scores for the process profiles that included both habitats were approximately at the midpoint between the scores for the habitat-specific groupings.

The site-level also had more distinct groupings of loadings and its light/dark fluxes were often closer together as well. Additionally, certain process couplings were retained across scales. However, the directionality in relation to geographic context and habitat of these couplings was not consistent. GHG fluxes were scattered in both but GPP (gross primary productivity) and BCR (benthic community respiration) had apparent negative correlations in both contexts.

Testing the relationships between PCs and network metrics at the genus level revealed significant relationships mainly at the site level spatial scale (figure 11). Limited relationships were present at the location level and only for archaea. At the site level, PC2 was the primary PC involved and the presence of relationships

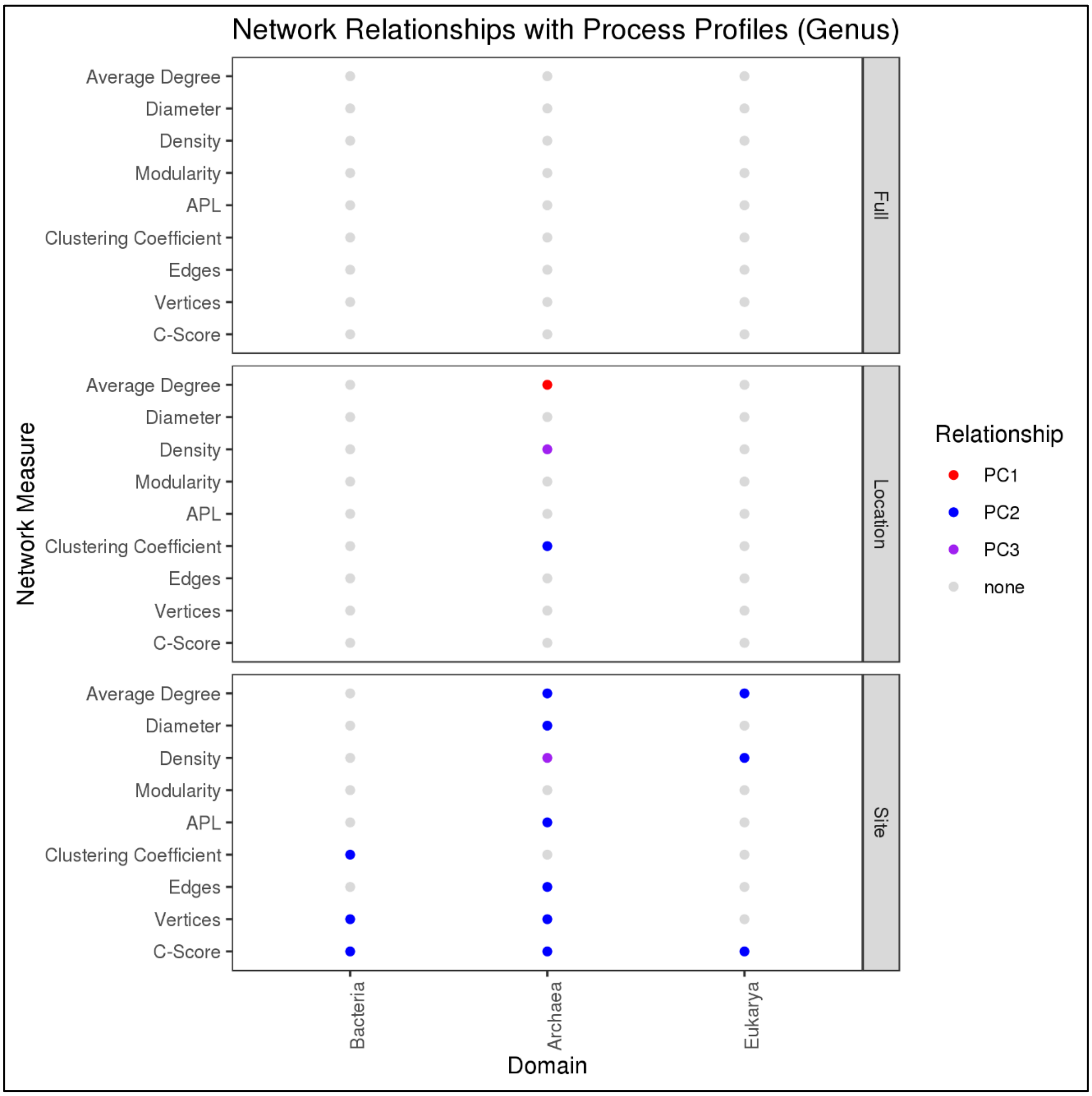


Figure 11. Significant relationships (as indicated via a p-value < 0.05 on an ANOVA test evaluating linear relationships between metrics and process profile) observed between genus level microbial co-occurrence networks and process profiles. Colored dots represent the presence of a relationship with each color signifying a specific principal component (PC). Both domain and spatial seemed to play a role in the observation of relationships.

seemed to be domain-dependent. Archaea had significant relationships for a bulk of its network metrics while bacteria only had relationships present for vertices and clustering coefficient and eukarya only had relationships for average degree and density. Not seeing much of anything at the location and full dataset levels suggests there are either limitations in obtaining observations of this kind at larger spatial delimiters or that the lack of significant relationships was a by-product of the smaller vector sizes used in relationship testing at the larger scales in this analysis.

Looking at these same relationships at higher taxonomic levels (figure 12) revealed that PC relationships are much different at higher taxonomic levels. At the site level, archaea had the most relationships at the genus level but loses nearly all its relationships aside from those with vertices at higher taxonomic levels. Meanwhile, bacteria picked up many relationships at the order and class levels. These relationships were mainly with PC1 instead of PC2 though. In contrast to both bacteria and archaea, eukarya stayed largely consistent with its lack of relationships across taxonomic levels. Another wrinkle that was added at higher taxonomic levels was that more relationships were seen at the location level. Most of these relationships were with location level PC1 which, as previously mentioned, had very similar patterns in what fluxes were contributing to it as PC2 at the site level. No relationships were seen at the full spatial scale even in this context.

When the direction of all the significant network relationships was investigated interesting relationships emerged. There were clear trends at the site level for bacteria where PC1 had a strong negative relationship with network interconnectedness and PC2 having a positive relationship with network size. Note that the significance found in bacteria was mainly found at the order and class levels which should be put into perspective. PC2 had a strong positive relationship with network size for archaea as well. The eukaryotes and their limited number of relationships only showed a slight

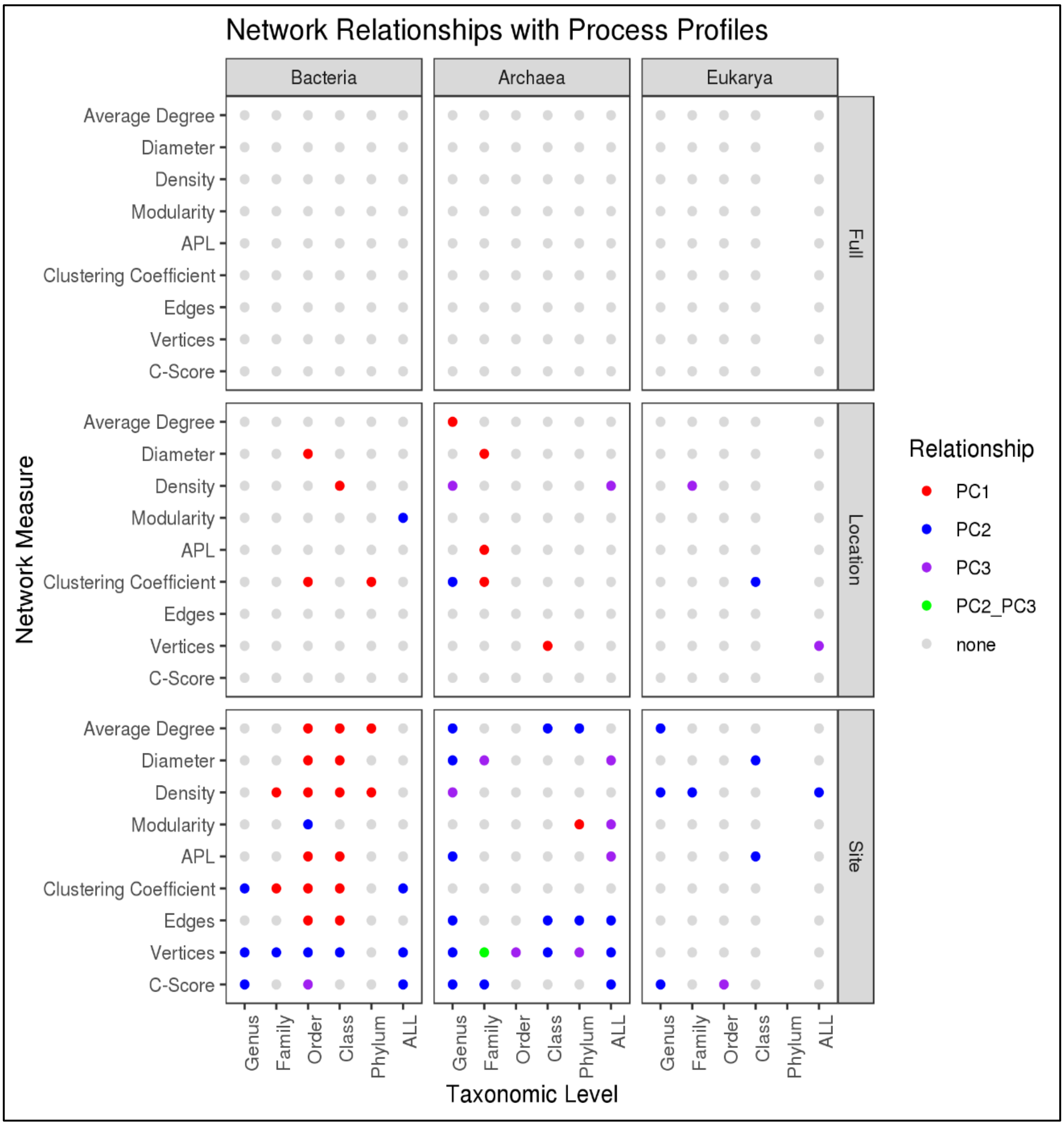


Figure 12. Significant relationships (as indicated via a p-value < 0.05 on an ANOVA test evaluating linear relationships between metrics and process profile) between microbial co-occurrence networks across all taxonomic levels and process profiles. Colored dots represent the presence of a relationship with each color signifying a specific principal component (PC). Just like the genus level, domain and spatial scale both played a role in the presence of relationships.

relation between interconnectedness and PC2. At the location level PC-network relationships were much weaker. The only semblance of a relationship at this level was PC1 having a positive relationship with interconnectedness for archaea and for the order and class level of bacteria.

When the same analysis was run using average taxonomic richness at each taxonomic level the results differed greatly (figure 13). Aside from bacteria's relation to site-level PC1 none of the network metric trends were present for richness. The only other trend seen was the consistent relationship between eukaryotic richness and PC3 at the location level which was associated with nitrogen cycling and community respiration.

Combining the directionality of process relationships with the loadings indicates that the archaea and higher taxonomic level bacteria communities in the CBESS dataset are more likely to have higher effluxes of silicates, NOx, and nitrate and will have higher influxes of ammonia as network size increases. These communities will also have more BCR and they will pull in more CO_2. Higher taxonomic level bacteria communities will likely also have higher effluxes of almost all nutrients in the light as well as higher effluxes of ammonia and phosphate in both the light and the dark as interconnectedness increases. In the location context the interconnectedness pattern is barely present for archaea and higher taxonomic level bacteria communities. At this context more interconnectedness is mainly associated with a higher efflux of ammonia and phosphate in the light and the dark and a higher efflux of two GHGs, GPP and N_2O. Meanwhile, less interconnectedness is associated with a higher efflux of silicate, NOx, and nitrate in the dark and a higher efflux of BCR.

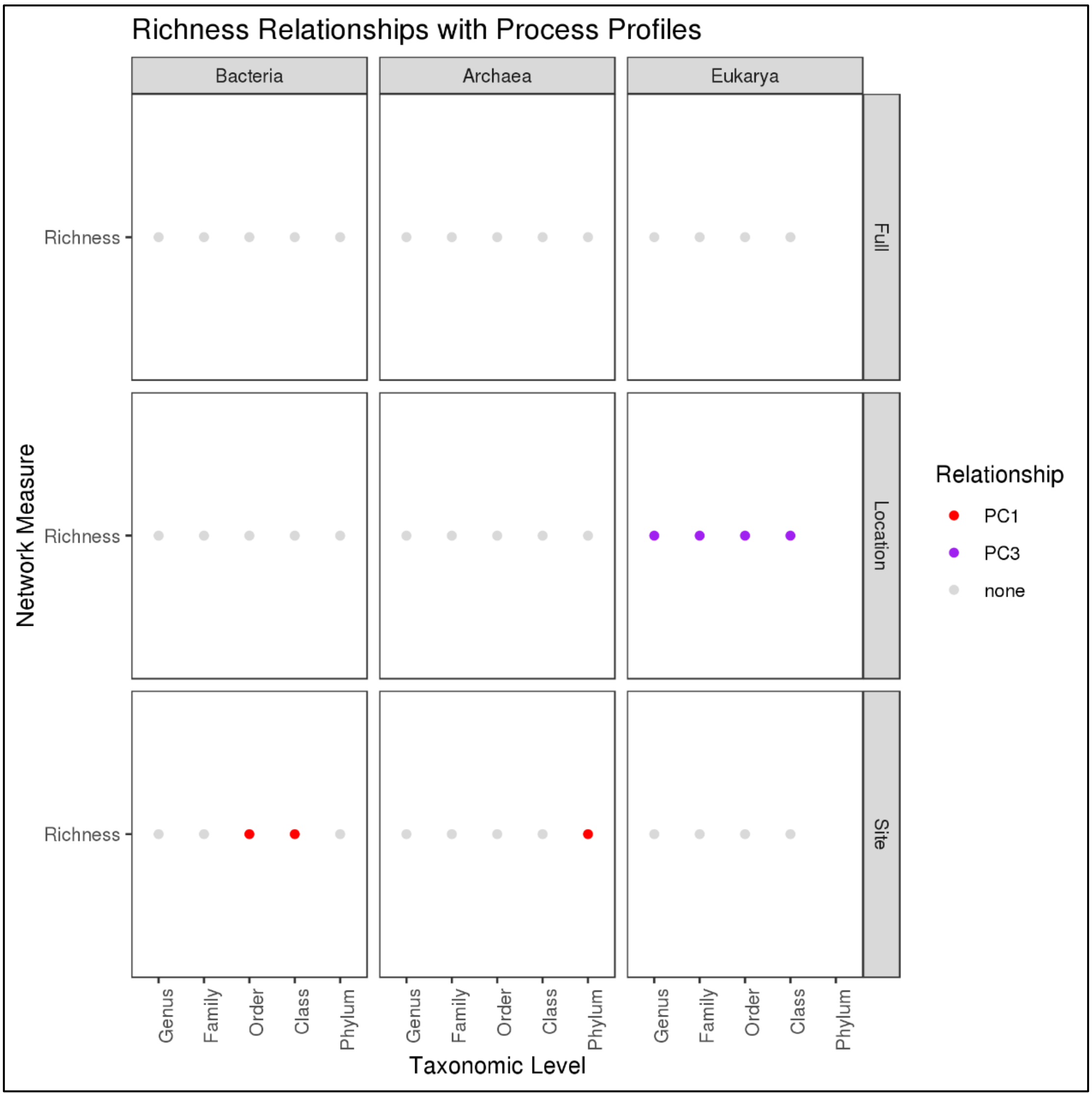


Figure 13. Significant relationships (as indicated via a p-value < 0.05 on an ANOVA test evaluating linear relationships between metrics and process profile) observed between average taxonomic richness and process profiles. Colored dots represent the presence of a relationship with each color signifying a specific principal component (PC). Compared to network relationships with process

Discussion

Overall it is evident that network size and network interconnectedness both play important roles in defining ecosystem processes within a system. It also seems that both network size and network interconnectedness have relationships with specific processes with an increase in a given factor leading to greater efflux rates. For instance, microbial species interactions drive nitrogen fluctuations (Penn et al., 2019) and the overall size and interconnectedness of the prokaryotic networks here do as well. Increasing network size also had a positive relationship with respiration. This relationship was interesting since microbial community respiration is related to activity levels so there was a connection between network size and total microbial activity (Cheng et al., 2013).

This analysis supports the positive effect of microbial diversity and the role it plays in the multifunctionality of ecosystems (Jiao et al., 2019; van der Plas, 2019). Furthermore, process delivery in microbes involves lots of interactions so process rates would be better maintained in larger and more interconnected networks based on the theory of functional redundancy (Feit et al., 2019). While finding diversity-multifunctionality relationships is great it does make elucidating the exact dynamics of specific processes at the whole system perspective understandably difficult (Wagg et al., 2014). However, getting information of that resolution from this analysis would always have been difficult since inferring ecological influences from complex networks is still a challenge. Regardless of network complexity, using many environmental factors in the process profile ordination may have limited the depth of variation that was captured as well (Tang et al., 2019).

Nevertheless, network analysis revealed much more than what was shown from the richness perspective. Network analysis has revealed many unexpected relationships in microbial communities already (Barberán et al, 2012) so it was not surprising that as this analysis revealed the BEF relationships that it did. Furthermore, the amount of BEF relationships present may be larger than what was captured here. Any intraspecific interactions present in the microbial communities were not captured in this analysis (Giri et al., 2019). There were also plenty of unidentified microbial taxa so it is possible that expanded sequence classification libraries with increased depth and range of identification would help further solidify the relationships observed here (Borrel et al., 2019).

Spatial Effects on Relationships

Spatial context had some of the most prominent effects on the presence of network-process relationships in this analysis. The effects on spatial scaling were clear even when just looking at the network metrics. There was much higher variability in many metrics related to network size and network interconnectedness. As shown in the random forest analysis, taxonomic factors were most important at the site level and then quickly fell off in variable-importance as scale increased. In the case of this analysis that fall in taxonomic variable-importance translates to not as much network complexity being captured at larger scales. While the effects of biodiversity on processes have been found to increase with scale in the past (Cardinale et al., 2012), increasing scale in this analysis had the opposite effect instead.

Microbial communities display incredibly complex heterogeneity patterns so maintaining a certain degree of spatial resolution is important to properly capture the network dynamics of these communities. From a macroecology perspective larger-scales are good for generalizing but are less detailed (Barberán et al., 2014). Therefore, it makes sense that increasing spatial scale caused a decrease in network complexity (Hendershot et al., 2017). There are also many ecological reasons behind this drop in detail. Dispersal limitations that come about via metacommunity dynamics and founder effects would cause co-occurrence relationships to be lost at larger scales when dispersal limitations are smaller than the scale used in a given analysis (Fukami, 2015; Dann et al., 2019). Potential dispersal limitations have already been discussed in chapter 3 and many of those limitations are relevant here as well. Microbes have also demonstrated that local coevolutionary interactions can emerge between species which are likely lost at larger scales (Essarioui et al., 2019).

Microbial community structure is also influenced by different factors at different scales (Bardgett and van der Putten, 2014). Therefore, the dynamics of microbial co-occurrence networks may be decoupled from the ecosystem processes evaluated in this analysis at larger scales. However, local-scale effects are likely more important than larger-scale regional ones in shaping microbial communities so if this decoupling were to exist then it would not be too drastic (Lladó et al., 2018). Furthermore, both the variability of microbial communities and the variability between process profiles naturally seen between sites is masked as scale increases (Jing et al., 2015). From an analytical perspective this translates to fewer data points being considered when evaluating potential relationships. This drop may then have influenced the number of BEF relationships observed at the larger scale.

Domain Differences

Substantial differences were observed across the three domains looked at in this study. Bacteria’s high diversity of OTUs and clear distinctions across locations and habitats led to expectations that its co-occurrence networks would be the most complex and that it would have noticeable BEF relationships. However, despite the bacterial networks generally being the largest most of the expected network relationships did not emerge. Instead, most of the significant network relationships with process profiles were not seen until the order and class taxonomic levels. However, there is a potential explanation for these observations. Random forest analysis from chapter 3 indicated that bacterial communities seem to be much more heterogeneous, with an array of dispersal limitations, founder effects (Dann et al., 2019), and other factors creating more distinct community structures. Dispersal limitations can mask relationships between environmental factors and microbial community composition (Evans et al., 2017) so network relationships likely exist at lower taxonomic levels for bacteria despite them not being observed here.

While the complexity and interconnectedness dynamics of the archaeal networks were not as clear the bacterial networks, the domain did show the genus level relationships that were expected from the bacterial networks. Both the higher variability of the domain’s network metric measures and the drop in relationships after the genus level can be explained by the nature of the domain and its branching and wide-ranging phylogeny (Brochier-Armanet et al., 2011). Interactions and connections between the more differentiated microorganisms in this domain would likely be more context-dependent. Relationships would also be expected to fall off since interactions between more distant taxonomic groupings would not be expected. Meanwhile, the eukaryotic microbes only had one weak interconnectedness relationship at best across all taxonomic levels. The lack of BEF relationships seen in this domain was not unexpected since larger organisms display a higher degree of stochasticity in their dispersal (Zinger et al., 2019). Evidently this stochasticity translated to the co-occurrence networks of eukaryotic microbial communities not having a relationship with process profiles.

Environmental Considerations

It is important to acknowledge the role that environmental factors may have played in this analysis. After all ecosystem processes are essentially abiotic properties interacting with the functional traits of organisms (Lavorel and Garnier 2002). The possible effect of environmental factors should not be understated since they have been the strongest predictors of process rates in other studies (Graham et al., 2016). For instance, changes in nutrient levels have been found to alter a wide range of processes in soil microbial systems (Kuske et al., 2019; Yan et al., 2019; Yin et al., 2019). These additions can even affect the microbial communities present as well (Geyer and Barrett, 2019). However, microbial communities have been found to exhibit notable stoichiometric plasticity in response to nutrient imbalances so there is a degree of persistence that communities can show to changing nutrient levels (Fanin et al., 2017).

The possible role of macrobiota in process delivery is also worth considering (Moulton et al., 2016). Aboveground-belowground interactions have significant effects on microbial communities that can alter process rates. For instance, microbial co-occurrence patterns have been shown to change across vegetation conditions (Chen et al., 2019b) and salt marshes plants have been shown to have a strong influence on diazotrophs and ammonia oxidizers (Wang et al., 2018). Plants also influence microbial communities via root dynamics affecting microbial community composition, the active microbiota community, and nutrient cycling rates (Haichar et al., 2012; Iversen et al., 2017; Guyonnet, et al., 2018; Schmid et al., 2018). Plant soil feedbacks act both ways in soils though, with soil microbial communities altering plant growth dynamics (van der Putten et al., 2016). Plants played much more of a role in the salt marshes compared to mud flats due to the differences in habitat characteristics, but aboveground flora still have a presence in mud flats. For example, different algal covers impact both community composition and nutrient fluxes in mud flat sediments (Bishop and Kelaher et al., 2013; Gonzalez et al., 2013).

On the note of algae, it is important to also cover the influence that the tides have on these coastal habitats. Tidal pressures do influence process rates, but microbial groups do differ in their responses to the redox transitions that occur during tidal changes (Frindte et al., 2016). Gene expression patterns also change less during the transition from oxic to anoxic (high-tide) than from anoxic to oxic conditions (low-tide) and different processes occur at different rates depending on the depth and mixing conditions (Gantzer and Stephen, 2003; Hewson et al., 2014). All told temporal dynamics play a role in BEF relationships in coastal systems through the effect of tides on process rates. However, it is unlikely that these temporal dynamics were captured in this analysis since sampling was only done in areas that were not submerged.

Aspects of the microbial communities themselves must be considered as well. It is important not to discount the effects of stochastic assembly processes, phenotypic plasticity, and other factors that create patterns in microbial communities not linked to process rates (Graham et al., 2016). Physical characteristics of microbial communities also play a role in functioning, especially in the case of processes such as nitrification and respiration (Suarez et al., 2019).

While factors not covered in this analysis likely play a role in regulating microbial BEF relationships, the role played by the microbial communities in dictating these relationships is still apparent. Microbes exhibit multiple ecotypes within single species, allowing them to persist throughout fluctuating environmental conditions so the surrounding environment is far from the ultimate determiner of ecosystem processes (García-García et al., 2019). Also, the exact effects of environmental drivers on microbial BEF relationships may still be unknown (Llado et al., 2018), but microbial communities do exhibit strong feedback loops with their environment. Therefore, the high predictive value of environmental factors regarding ecosystem processes is likely partly due to the effects of microbial communities on environmental variables.

Conclusion

While analytical and theoretical limitations were present in this analysis, the presence of relationships between microbial community networks and the functioning of ecosystems is evident. Furthermore, the role of spatial scale in observing these relationships is clear as well since network-related BEF relationships only showed up at the more detailed site level. These relationships differed across domains as well. Bacteria and archaea networks both exhibited similar relationships with nitrogen cycling and other fluxes. The bacterial relationships were more nuanced though since relationships were only observed at higher taxonomic levels. Dispersal limitations potentially masked the presence of relationships at lower taxonomic levels but further research is required to confirm that theory. However, if that is true then BEF links may be shared across prokaryotic communities to a certain extent. In comparison, eukaryotes did not exhibit any strong functional relationships and this was likely due to the higher stochasticity of the domain. Environmental contexts must always be considered in microbial community analysis like this, but even when that is done the strength of the relationships between important functions and the microbial communities observed here is still evident.

CHAPTER 5: GENERAL DISCUSSION

The value of this work lies just as much in its insights into the structuring dynamics and the BEF relationships of microbial communities as it does the research questions it poses. There is a lot of potential in investigating the phylogenetic and functional relationships of the taxa that were important in structuring. This potential is further exemplified in taxa that were important in structuring co-occurrence networks (Banerjee et al., 2018). How structuring and BEF drivers change across scales is also of great interest. The questions posited throughout this dissertation regarding spatial scale offer support for multiple research paths. The degree to which temporal dynamics impacted the patterns observed here are well worth investigating as well.

Creating and sharing more datasets like the CBESS dataset has great value because of the insights that can only be attained by using large datasets. While the focus of this work is on understanding ecological relationships, the implications of this work extend into biomonitoring efforts as well. Random forest analysis illustrated the limitations of extrapolating microbial data across habitats, but it is still worth considering the possibility of developing models for monitoring habitats using microbial communities (Cordier et al., 2019). Maintaining the diversity of habitats is important not only for BEF relationships but also for human health so BEF research has considerable potential in helping shed light on important biodiversity-ecosystem service relationships as well (Cardinale et al. 2012; Duncan et al., 2015; Wall et al., 2015).

The future use of machine learning in the field of microbial ecology is worth touching upon as well. Machine learning has proven to be invaluable in nearly every field, especially as Big Data has risen in popularity, and microbial ecology is no exception. The sheer size of data that microbial ecologists work with makes integrating machine learning seem like a logical choice in most cases. Numerous studies, some of which highlighted in this analysis, have already demonstrated the value that machine learning brings. Looking forward, further integrating machine learning into the analysis of metagenomics data seems to be one of the areas that has the greatest potential due to the comprehensive picture that metagenomic data provides already. Machine learning is not the only useful tool at a researcher's disposal though and care should be taken to ensure there are not methods that are better equipped to solve a given problem (Tarca et al., 2007).

Despite the analytical and theoretical limitations present in this analysis, it is still clear that microbial community structuring dynamics and BEF relationships differ across spatial scales, habitat types, and domains. The clear trends observed across the spatial scales indicate the importance of perspective in spatial analysis. The trends also point to the importance of perspective in studies that do not use multiple spatial scales since the patterns present at a given scale may mask expected relationships. The heterogeneity of microbial communities and the many dynamics behind it should not be underestimated (Hendershot et al., 2017).

Spatial scale seemed to be especially important in both the interpretation of structuring mechanics and in the presence of BEF relationships. For instance, dispersal limitations were only visible at the sample level in most cases and network-related BEF relationships only showed up at the smaller-scale site level. Some of the largest differences were seen across domains though. Bacteria seemed to be structured almost entirely by dispersal limitations while archaea and eukaryotic communities were much more environmentally driven. In comparison, bacteria and archaea networks both exhibited similar BEF relationships with nitrogen cycling and other fluxes, with eukaryotes not exhibiting any strong functional relationships. It is interesting that structuring dynamics were more domain-specific while BEF patterns more closely followed the divide between prokaryotes and eukaryotes.

The exact reasons was to why the domain-specific structuring pattern emerged are largely discussed in chapter 2 but it is still worth revisiting the wider implications of this observation. The exact contributions of different structuring effects were not revealed in this analysis but it is still clear that each domain interacts with the environment differently. Therefore care should be taken to ensure domain specific structuring effects are taken into account when analysing data and referencing other work.

Further research is recommended for both the community structuring and the BEF relationships front specifically as well for the interplay between them. Overall, by approaching the questions relating to microbial community structuring and BEF relationships from alternative angles, this study revealed many novel insights and opens the door to more research questions.

REFERENCES

Albright, M.B.N., Chase, A.B., Martiny, J.B.H., 2019. Experimental Evidence that Stochasticity Contributes to Bacterial Composition and Functioning in a Decomposer Community. mBio 10, e00568-19. https://doi.org/10.1128/mBio.00568-19

Alzarhani, A.K., Clark, D.R., Underwood, G.J., Ford, H., Cotton, T.A. and Dumbrell, A.J., 2019. Are drivers of root-associated fungal community structure context specific?. The ISME journal, 13, 330-1344.

Baldrian, P., 2019. The known and the unknown in soil microbial ecology. FEMS Microbiol. Ecol. https://doi.org/10.1093/femsec/fiz005

Banerjee, S., Kirkby, C.A., Schmutter, D., Bissett, A., Kirkegaard, J.A., Richardson, A.E., 2016. Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil. Soil Biology and Biochemistry 97, 188–198. https://doi.org/10.1016/j.soilbio.2016.03.017

Banerjee, S., Schlaeppi, K., van der Heijden, M.G.A., 2018. Keystone taxa as drivers of microbiome structure and functioning. Nature Reviews Microbiology 16, 567–576. https://doi.org/10.1038/s41579-018-0024-1

Barberán, A., Bates, S.T., Casamayor, E.O., Fierer, N., 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. The ISME Journal 6, 343–351. https://doi.org/10.1038/ismej.2011.119

Barberán, A., Casamayor, E.O., Fierer, N., 2014. The microbial contribution to macroecology. Front. Microbiol. 5. https://doi.org/10.3389/fmicb.2014.00203

Bardgett, R.D., van der Putten, W.H., 2014. Belowground biodiversity and ecosystem functioning. Nature 515, 505–511. https://doi.org/10.1038/nature13855

Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R., Fierer, N., 2011. Examining the global distribution of dominant archaeal populations in soil. ISME J 5, 908–917. https://doi.org/10.1038/ismej.2010.171

Baumann, H., Wallace, R.B., Tagliaferri, T., Gobler, C.J., 2015. Large Natural pH, CO2 and O2 Fluctuations in a Temperate Tidal Salt Marsh on Diel, Seasonal, and Interannual Time Scales. Estuaries and Coasts 38, 220–231. https://doi.org/10.1007/s12237-014-9800-y

Beck, M.W., Heck, K.L., Able, K.W., Childers, D.L., Eggleston, D.B., Gillanders, B.M., Halpern, B., Hays, C.G., Hoshino, K., Minello, T.J., Orth, R.J., Sheridan, P.F., Weinstein, M.P., 2001. The Identification, Conservation, and Management of Estuarine and Marine Nurseries for Fish and Invertebrates. BioScience 51, 633–641. https://doi.org/10.1641/0006-3568(2001)051[0633:TICAMO]2.0.CO;2

Bik, H.M., Porazinska, D.L., Creer, S., Caporaso, J.G., Knight, R., Thomas, W.K., 2012. Sequencing our way towards understanding global eukaryotic biodiversity. Trends in Ecology & Evolution 27, 233–243. https://doi.org/10.1016/j.tree.2011.11.010

Bishop, M.J., Kelaher, B.P., 2013. Replacement of native seagrass with invasive algal detritus: impacts to estuarine sediment communities. Biol Invasions 15, 45–59. https://doi.org/10.1007/s10530-012-0267-0

Borrel, G., Adam, P.S., McKay, L.J., Chen, L.-X., Sierra-García, I.N., Sieber, C.M.K., Letourneur, Q., Ghozlane, A., Andersen, G.L., Li, W.-J., Hallam, S.J., Muyzer, G., de Oliveira, V.M., Inskeep, W.P., Banfield, J.F., Gribaldo, S., 2019. Wide diversity of methane and short-chain alkane metabolisms in uncultured archaea. Nat Microbiol 4, 603–613. https://doi.org/10.1038/s41564-019-0363-3

Bowen, J.L., Ward, B.B., Morrison, H.G., Hobbie, J.E., Valiela, I., Deegan, L.A., Sogin, M.L., 2011. Microbial community composition in sediments resists perturbation by nutrient enrichment. ISME J 5, 1540–1548. https://doi.org/10.1038/ismej.2011.22

Brochier-Armanet, C., Forterre, P., Gribaldo, S., 2011. Phylogeny and evolution of the Archaea: one hundred genomes later. Current Opinion in Microbiology, Ecology and industrial microbiology / Special section: Archaea 14, 274–281. https://doi.org/10.1016/j.mib.2011.04.015

Brophy, C., Dooley, Á., Kirwan, L., Finn, J.A., McDonnell, J., Bell, T., Cadotte, M.W., Connolly, J., 2017. Biodiversity and ecosystem function: making sense of numerous species interactions in multi-species communities. Ecology 98, 1771–1778. https://doi.org/10.1002/ecy.1872

Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., Thomas, T., 2011. Bacterial community assembly based on functional genes rather than species. Proceedings of the National Academy of Sciences 108, 14288–14293. https://doi.org/10.1073/pnas.1101591108

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., Naeem, S., 2012. Biodiversity loss and its impact on humanity. Nature 486, 59–67. https://doi.org/10.1038/nature11148

Carr, A., Diener, C., Baliga, N.S., Gibbons, S.M., 2019. Use and abuse of correlation analyses in microbial ecology. The ISME Journal 1. https://doi.org/10.1038/s41396-019-0459-z

Cavicchioli, R., Ripple, W.J., Timmis, K.N., Azam, F., Bakken, L.R., Baylis, M., Behrenfeld, M.J., Boetius, A., Boyd, P.W., Classen, A.T., Crowther, T.W., Danovaro, R., Foreman, C.M., Huisman, J., Hutchins, D.A., Jansson, J.K., Karl, D.M., Koskella, B., Mark Welch, D.B., Martiny, J.B.H., Moran, M.A., Orphan, V.J., Reay, D.S., Remais, J.V., Rich, V.I., Singh, B.K., Stein, L.Y., Stewart, F.J., Sullivan, M.B., van Oppen, M.J.H., Weaver, S.C., Webb, E.A., Webster, N.S., 2019. Scientists' warning to humanity: microorganisms and climate change. Nat. Rev. Microbiol. https://doi.org/10.1038/s41579-019-0222-5

Chave, J., 2004. Neutral theory and community ecology: Neutral theory and community ecology. Ecology Letters 7, 241–253. https://doi.org/10.1111/j.1461-0248.2003.00566.x

Chen, G., Huang, J., Fang, Y., Zhao, Y., Tian, X., Jin, Y., Zhao, H., 2018. Microbial community succession and pollutants removal of a novel carriers enhanced duckweed treatment system for rural wastewater in Dianchi Lake basin. Bioresour. Technol. 276, 8–17. https://doi.org/10.1016/j.biortech.2018.12.102

Chen, X., Wei, W., Wang, J., Li, H., Sun, J., Ma, R., Jiao, N., Zhang, R., 2019. Tide driven microbial dynamics through virus-host interactions in the estuarine ecosystem. Water Res. 160, 118–129. https://doi.org/10.1016/j.watres.2019.05.051

Chen, M., He, S., Li, J., Hu, W., Ma, Y., Wu, L., Ge, G., 2019. Co-occurrence patterns between bacterial and fungal communities in response to a vegetation gradient in a freshwater wetland. Can. J. Microbiol. https://doi.org/10.1139/cjm-2019-0147

Cheng, F., Peng, X., Zhao, P., Yuan, J., Zhong, C., Cheng, Y., Cui, C., Zhang, S., 2013. Soil Microbial Biomass, Basal Respiration and Enzyme Activity of Main Forest Types in the Qinling Mountains. PLOS ONE 8, e67353. https://doi.org/10.1371/journal.pone.0067353

Clark, D.R., Ferguson, R.M.W., Harris, D.N., Nicholass, K.J.M., Prentice, H.J., Randall, K.C., Randell, L., Warren, S.L., Dumbrell, A.J., 2018. Streams of data from drops of water: 21st century molecular microbial ecology. Wiley Interdisciplinary Reviews: Water 5, e1280. https://doi.org/10.1002/wat2.1280

Clark, D.R., Mathieu, M., Mourot, L., Dufossé, L., Underwood, G.J.C., Dumbrell, A.J., McGenity, T.J., 2017. Biogeography at the limits of life: Do extremophilic microbial communities show biogeographical regionalization? Global Ecology and Biogeography 26, 1435–1446. https://doi.org/10.1111/geb.12670

Coulon, F., Chronopoulou, P.-M., Fahy, A., Païssé, S., Goñi-Urriza, M., Peperzak, L., Acuña Alvarez, L., McKew, B.A., Brussaard, C.P.D., Underwood, G.J.C., Timmis, K.N., Duran, R., McGenity, T.J., 2012. Central Role of Dynamic Tidal Biofilms Dominated by Aerobic Hydrocarbonoclastic Bacteria and Diatoms in the Biodegradation of Hydrocarbons in Coastal Mudflats. Applied and Environmental Microbiology 78, 3638–3648. https://doi.org/10.1128/AEM.00072-12

Cordier, T., Lanzén, A., Apothéloz-Perret-Gentil, L., Stoeck, T. and Pawlowski, J., 2019. Embracing environmental genomics and machine learning for routine biomonitoring. Trends in microbiology, 27, 387-397.

Csardi G, Nepusz T: The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. http://igraph.org

Dann, L.M., Clanahan, M., Paterson, J.S., Mitchell, J.G., 2019. Distinct niche partitioning of marine and freshwater microbes during colonisation. FEMS Microbiol. Ecol. 95. https://doi.org/10.1093/femsec/fiz098

Decaëns, T., 2010. Macroecological patterns in soil communities. Global Ecology and Biogeography 19, 287–302. https://doi.org/10.1111/j.1466-8238.2009.00517.x

Delgado-Baquerizo, M., 2019. Obscure soil microbes and where to find them. The ISME Journal 1. https://doi.org/10.1038/s41396-019-0405-0

Delgado-Baquerizo, M., Giaramida, L., Reich, P.B., Khachane, A.N., Hamonts, K., Edwards, C., Lawton, L.A., Singh, B.K., 2016. Lack of functional redundancy in the relationship between microbial diversity and ecosystem functioning. Journal of Ecology 104, 936–946. https://doi.org/10.1111/1365-2745.12585

Dohi, M., Mougi, A., 2018. A coexistence theory in microbial communities. R Soc Open Sci 5, 180476. https://doi.org/10.1098/rsos.180476

Dubey, A., Malla, M.A., Khan, F., Chowdhary, K., Yadav, S., Kumar, A., Sharma, S., Khare, P.K., Khan, M.L., 2019. Soil microbiome: a key player for conservation of soil health under changing climate. Biodivers Conserv. https://doi.org/10.1007/s10531-019-01760-5

Duffy, J.E., Godwin, C.M., Cardinale, B.J., 2017. Biodiversity effects in the wild are common and as strong as key drivers of productivity. Nature 549, 261–264. https://doi.org/10.1038/nature23886

Dumbrell, A.J., Ashton, P.D., Aziz, N., Feng, G., Nelson, M., Dytham, C., Fitter, A.H., Helgason, T., 2011. Distinct seasonal assemblages of arbuscular mycorrhizal fungi revealed by massively parallel pyrosequencing. New Phytologist 190, 794–804. https://doi.org/10.1111/j.1469-8137.2010.03636.x

Dumbrell, A.J., Ferguson, R.M.W., Clark, D.R., 2016. Microbial Community Analysis by Single-Amplicon High-Throughput Next Generation Sequencing: Data Analysis – From Raw Output to Ecology, in:

McGenity, T.J., Timmis, K.N., Nogales, B. (Eds.), Hydrocarbon and Lipid Microbiology Protocols. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 155–206. https://doi.org/10.1007/8623_2016_228

Dumbrell, A.J., Nelson, M., Helgason, T., Dytham, C., Fitter, A.H., 2010. Relative roles of niche and neutral processes in structuring a soil microbial community. ISME J 4, 337–345. https://doi.org/10.1038/ismej.2009.122

Duncan, C., Thompson, J.R., Pettorelli, N., 2015. The quest for a mechanistic understanding of biodiversity–ecosystem services relationships. Proc Biol Sci 282, undefined-undefined. https://doi.org/10.1098/rspb.2015.1348

Essarioui, A., LeBlanc, N., Otto-Hanson, L., Schlatter, D.C., Kistler, H.C., Kinkel, L.L., 2019. Inhibitory and nutrient use phenotypes among co-existing Fusarium and Streptomyces populations suggest local coevolutionary interactions in soil. Environmental Microbiology 0. https://doi.org/10.1111/1462-2920.14782

Evans, S., Martiny, J.B.H., Allison, S.D., 2017. Effects of dispersal and selection on stochastic assembly in microbial communities. ISME J 11, 176–185. https://doi.org/10.1038/ismej.2016.96

Fanin, N., Fromin, N., Barantal, S., Hättenschwiler, S., 2017. Stoichiometric plasticity of microbial communities is similar between litter and soil in a tropical rainforest. Sci Rep 7, 12498. https://doi.org/10.1038/s41598-017-12609-8

Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. Nature Reviews Microbiology 10, 538–550. https://doi.org/10.1038/nrmicro2832

Feit, B., Blüthgen, N., Traugott, M., Jonsson, M., 2019. Resilience of ecosystem processes: a new approach shows that functional redundancy of biological control services is reduced by landscape simplification. Ecology Letters 0. https://doi.org/10.1111/ele.13347

Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. PNAS 103, 626–631. https://doi.org/10.1073/pnas.0507535103

Flynn, D.F.B., Mirotchnick, N., Jain, M., Palmer, M.I., Naeem, S., 2011. Functional and phylogenetic diversity as predictors of biodiversity–ecosystem-function relationships. Ecology 92, 1573–1581. https://doi.org/10.1890/10-1245.1

Flynn, J.M., Brown, E.A., Chain, F.J.J., MacIsaac, H.J., Cristescu, M.E., 2015. Toward accurate molecular identification of species in complex environmental samples: testing the performance of sequence filtering and clustering methods. Ecology and Evolution 5, 2252–2266. https://doi.org/10.1002/ece3.1497

Friedman, J., Higgins, L.M., Gore, J., 2017. Community structure follows simple assembly rules in microbial microcosms. Nature Ecology & Evolution 1, 0109. https://doi.org/10.1038/s41559-017-0109

Frindte, K., Allgaier, M., Grossart, H.-P., Eckert, W., 2016. Redox stability regulates community structure of active microbes at the sediment-water interface. Environ Microbiol Rep 8, 798–804. https://doi.org/10.1111/1758-2229.12441

Fukami, T., 2015. Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. Annu. Rev. Ecol. Evol. Syst. 46, 1–23. https://doi.org/10.1146/annurev-ecolsys-110411-160340

Gagic, V., Bartomeus, I., Jonsson, T., Taylor, A., Winqvist, C., Fischer, C., Slade, E.M., Steffan-Dewenter, I., Emmerson, M., Potts, S.G., Tscharntke, T., Weisser, W., Bommarco, R., 2015. Functional identity and diversity of animals predict ecosystem functioning better than species-based indices. Proc Biol Sci 282. https://doi.org/10.1098/rspb.2014.2620

Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., Ruiz-Jaen, M.C., Fröberg, M., Stendahl, J., Philipson, C.D., Mikusiński, G., Andersson, E., Westerlund, B., Andrén, H., Moberg, F., Moen, J., Bengtsson, J., 2013. Higher levels of multiple ecosystem services are found in forests with more tree species. Nature Communications 4, 1340. https://doi.org/10.1038/ncomms2328

Gantzer, C.J., Stefan, H.G., 2003. A model of microbial activity in lake sediments in response to periodic water-column mixing. Water Res. 37, 2833–2846. https://doi.org/10.1016/S0043-1354(03)00110-6

García-García, N., Tamames, J., Linz, A.M., Pedrós-Alió, C., Puente-Sánchez, F., 2019. Microdiversity ensures the maintenance of functional microbial communities under changing environmental conditions. ISME J 1–15. https://doi.org/10.1038/s41396-019-0487-8

García-Palacios, P., Vandegehuchte, M.L., Shaw, E.A., Dam, M., Post, K.H., Ramirez, K.S., Sylvain, Z.A., Tomasel, C.M. de, Wall, D.H., 2015. Are there links between responses of soil microbes and ecosystem functioning to elevated CO2, N deposition and warming? A global perspective. Global Change Biology 21, 1590–1600. https://doi.org/10.1111/gcb.12788

Gedan, K.B., Kirwan, M.L., Wolanski, E., Barbier, E.B., Silliman, B.R., 2011. The present and future role of coastal wetland vegetation in protecting shorelines: answering recent challenges to the paradigm. Climatic Change 106, 7–29. https://doi.org/10.1007/s10584-010-0003-7

Geyer, K.M., Barrett, J.E., 2019. Unimodal productivity - diversity relationships among bacterial communities in a simple polar soil ecosystem. Environ. Microbiol. 0. https://doi.org/10.1111/1462-2920.14639

Giri, S., Waschina, S., Kaleta, C., Kost, C., 2019. Defining Division of Labour in Microbial Communities. J. Mol. Biol. https://doi.org/10.1016/j.jmb.2019.06.023

Gonzalez, D.J., Smyth, A.R., Piehler, M.F., McGlathery, K.J., 2013. Mats of the nonnative macroalga, Gracilaria vermiculophylla, alter net denitrification rates and nutrient fluxes on intertidal mudflats. Limnology and Oceanography 58, 2101–2108. https://doi.org/10.4319/lo.2013.58.6.2101

Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. 17, 333–351. https://doi.org/10.1038/nrg.2016.49

Gore, J., 2018. Simple organizing principles in microbial communities. Curr. Opin. Microbiol. 45, 195–202. https://doi.org/10.1016/j.mib.2018.11.007

Graham, E.B., Knelman, J.E., Schindlbacher, A., Siciliano, S., Breulmann, M., Yannarell, A., Beman, J.M., Abell, G., Philippot, L., Prosser, J., Foulquier, A., Yuste, J.C., Glanville, H.C., Jones, D.L., Angel, R., Salminen, J., Newton, R.J., Bürgmann, H., Ingram, L.J., Hamer, U., Siljanen, H.M.P., Peltoniemi, K., Potthast, K., Bañeras, L., Hartmann, M., Banerjee, S., Yu, R.-Q., Nogaro, G., Richter, A., Koranda, M., Castle, S.C., Goberna, M., Song, B., Chatterjee, A., Nunes, O.C., Lopes, A.R., Cao, Y., Kaisermann, A., Hallin, S., Strickland, M.S., Garcia-Pausas, J., Barba, J., Kang, H., Isobe, K., Papaspyrou, S., Pastorelli, R., Lagomarsino, A., Lindström, E.S., Basiliko, N., Nemergut, D.R., 2016. Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes? Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.00214

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W.H.C.F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.-L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P., Christen, R., 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res 41, D597–D604. https://doi.org/10.1093/nar/gks1160

Guyonnet, J.P., Guillemet, M., Dubost, A., Simon, L., Ortet, P., Barakat, M., Heulin, T., Achouak, W., Haichar, F. el Z., 2018. Plant Nutrient Resource Use Strategies Shape Active Rhizosphere Microbiota Through Root Exudation. Front Plant Sci 9. https://doi.org/10.3389/fpls.2018.01662

Haichar, F. el Z., Roncato, M.-A., Achouak, W., 2012. Stable isotope probing of bacterial community structure and gene expression in the rhizosphere of Arabidopsis thaliana. FEMS Microbiol Ecol 81, 291–302. https://doi.org/10.1111/j.1574-6941.2012.01345.x

Hale, L., Feng, W., Yin, H., Guo, X., Zhou, X., Bracho, R., Pegoraro, E., Penton, C.R., Wu, L., Cole, J., Konstantinidis, K.T., Luo, Y., Tiedje, J.M., Schuur, E.A.G., Zhou, J., 2019. Tundra microbial community taxa and traits predict decomposition parameters of stable, old soil organic carbon. ISME J 1–15. https://doi.org/10.1038/s41396-019-0485-x

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., Martiny, J.B.H., 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol 10, 497–506. https://doi.org/10.1038/nrmicro2795

Hector, A., Bagchi, R., 2007. Biodiversity and ecosystem multifunctionality. Nature 448, 188–190. https://doi.org/10.1038/nature05947

Hendershot, J.N., Read, Q.D., Henning, J.A., Sanders, N.J., Classen, A.T., 2017. Consistently inconsistent drivers of microbial diversity and abundance at macroecological scales. Ecology 98, 1757–1763. https://doi.org/10.1002/ecy.1829

Hermans, S.M., Buckley, H.L., Lear, G., 2019. Perspectives on the Impact of Sampling Design and Intensity on Soil Microbial Diversity Estimates. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.01820

Hewson, I., Eggleston, E.M., Doherty, M., Lee, D.Y., Owens, M., Shapleigh, J.P., Cornwell, J.C., Crump, B.C., 2014. Metatranscriptomic analyses of plankton communities inhabiting surface and subpycnocline waters of the Chesapeake Bay during oxic-anoxic-oxic transitions. Appl. Environ. Microbiol. 80, 328–338. https://doi.org/10.1128/AEM.02680-13

Hirano, H., Takemoto, K., 2019. Difficulty in inferring microbial community structure based on co-occurrence network approaches. BMC Bioinformatics 20, 329. https://doi.org/10.1186/s12859-019-2915-1

Hooper, D.U., Chapin, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J.H., Lodge, D.M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A.J., Vandermeer, J., Wardle, D.A., 2005. Effects of Biodiversity On Ecosystem Functioning: A Consensus of Current Knowledge. Ecological Monographs 75, 3–35. https://doi.org/10.1890/04-0922

Hu, J., Liu, S., Yang, W., He, Z., Wang, J., Liu, H., Zheng, P., Xi, C., Ma, F., Hu, B., 2019. Ecological Success of the Nitrosopumilus and Nitrosospira Clusters in the Intertidal Zone. Microb Ecol. https://doi.org/10.1007/s00248-019-01359-x

Hugerth, L.W., Muller, E.E.L., Hu, Y.O.O., Lebrun, L.A.M., Roume, H., Lundin, D., Wilmes, P., Andersson, A.F., 2014. Systematic Design of 18S rRNA Gene Primers for Determining Eukaryotic Diversity in Microbial Consortia. PLOS ONE 9, e95567. https://doi.org/10.1371/journal.pone.0095567

Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W.S., Reich, P.B., Scherer-Lorenzen, M., Schmid, B., Tilman, D., van Ruijven, J., Weigelt, A., Wilsey, B.J., Zavaleta, E.S., Loreau, M., 2011. High plant diversity is needed to maintain ecosystem services. Nature 477, 199–202. https://doi.org/10.1038/nature10282

Isbell, F., Tilman, D., Polasky, S., Loreau, M., 2015. The biodiversity-dependent ecosystem service debt. Ecology Letters 18, 119–134. https://doi.org/10.1111/ele.12393

Iversen, C.M., McCormack, M.L., Powell, A.S., Blackwood, C.B., Freschet, G.T., Kattge, J., Roumet, C., Stover, D.B., Soudzilovskaia, N.A., Valverde-Barrantes, O.J., van Bodegom, P.M., Violle, C., 2017. A global Fine-Root Ecology Database to address below-ground challenges in plant ecology. New Phytologist 215, 15–26. https://doi.org/10.1111/nph.14486

Jeanne, T., Parent, S.-É., Hogue, R., 2019. Using a soil bacterial species balance index to estimate potato crop productivity. PLoS ONE 14, e0214089. https://doi.org/10.1371/journal.pone.0214089

Ji, B.W., Sheth, R.U., Dixit, P.D., Huang, Y., Kaufman, A., Wang, H.H., Vitkup, D., 2019. Quantifying spatiotemporal variability and noise in absolute microbiota abundances using replicate sampling. Nature Methods 1. https://doi.org/10.1038/s41592-019-0467-y

Jiao, S., Xu, Y., Zhang, J., Hao, X., Lu, Y., 2019. Core Microbiota in Agricultural Soils and Their Potential Associations with Nutrient Cycling. mSystems 4, e00313-18. https://doi.org/10.1128/mSystems.00313-18

Jing, X., Sanders, N.J., Shi, Yu, Chu, H., Classen, A.T., Zhao, K., Chen, L., Shi, Yue, Jiang, Y., He, J.-S., 2015. The links between ecosystem multifunctionality and above- and belowground biodiversity are mediated by climate. Nature Communications 6, 8159. https://doi.org/10.1038/ncomms9159

Joshi, N.A. and Fass, J.N., 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files.

Ju, F., Xia, Y., Guo, F., Wang, Z., Zhang, T., 2014. Taxonomic relatedness shapes bacterial assembly in activated sludge of globally distributed wastewater treatment plants. Environ. Microbiol. 16, 2421–2432. https://doi.org/10.1111/1462-2920.12355

Kietäväinen, R., Purkamo, L., 2015. The origin, source, and cycling of methane in deep crystalline rock biosphere. Front Microbiol 6. https://doi.org/10.3389/fmicb.2015.00725

Kim, H.B., Isaacson, R.E., 2015. The pig gut microbial diversity: Understanding the pig gut microbial ecology through the next generation high throughput sequencing. Vet. Microbiol. 177, 242–251. https://doi.org/10.1016/j.vetmic.2015.03.014

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41, e1. https://doi.org/10.1093/nar/gks808

Kuske, C.R., Sinsabaugh, R.L., Gallegos-Graves, L.V., Albright, M.B.N., Mueller, R., Dunbar, J., 2019. Simple measurements in a complex system: soil community responses to nitrogen amendment in a *Pinus taeda* forest. Ecosphere 10, e02687. https://doi.org/10.1002/ecs2.2687

Lamb, P.D., Hunter, E., Pinnegar, J.K., Creer, S., Davies, R.G., Taylor, M.I., 2019. How quantitative is metabarcoding: A meta-analytical approach. Mol. Ecol. 28, 420–430. https://doi.org/10.1111/mec.14920

Lauber, C.L., Ramirez, K.S., Aanderud, Z., Lennon, J., Fierer, N., 2013. Temporal variability in soil microbial communities across land-use types. The ISME Journal 7, 1641–1650. https://doi.org/10.1038/ismej.2013.50

Laureto, L.M.O., Cianciaruso, M.V., Samia, D.S.M., 2015. Functional diversity: An overview of its history and applicability. Natureza & Conservação 13, 112–116. https://doi.org/10.1016/j.ncon.2015.11.001

Lavorel, S., Garnier, E., 2002. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. Functional Ecology 16, 545–556. https://doi.org/10.1046/j.1365-2435.2002.00664.x

Lee, C.K., Laughlin, D.C., Bottos, E.M., Caruso, T., Joy, K., Barrett, J.E., Brabyn, L., Nielsen, U.N., Adams, B.J., Wall, D.H., Hopkins, D.W., Pointing, S.B., McDonald, I.R., Cowan, D.A., Banks, J.C., Stichbury, G.A., Jones, I., Zawar-Reza, P., Katurji, M., Hogg, I.D., Sparrow, A.D., Storey, B.C., Allan Green, T.G., Cary, S.C., 2019. Biotic interactions are an unexpected yet critical control on the complexity of an abiotically driven polar ecosystem. Commun Biol 2, 62. https://doi.org/10.1038/s42003-018-0274-5

Leibold, M.A., McPeek, M.A., 2006. Coexistence of the niche and neutral perspectives in community ecology. Ecology 87, 1399–1410.

Levin, L.A., Boesch, D.F., Covich, A., Dahm, C., Erséus, C., Ewel, K.C., Kneib, R.T., Moldenke, A., Palmer, M.A., Snelgrove, P., Strayer, D., Weslawski, J.M., 2001. The Function of Marine Critical Transition Zones and the Importance of Sediment Biodiversity. Ecosystems 4, 430–451. https://doi.org/10.1007/s10021-001-0021-4

Li, J., Hu, R., Guo, Y., Chen, S., Xie, X., Qin, J.G., Ma, Z., Zhu, C., Pei, S., 2019. Bioturbation of peanut worms Sipunculus nudus on the composition of prokaryotic communities in a tidal flat as revealed by 16S rRNA gene sequences. Microbiologyopen e802. https://doi.org/10.1002/mbo3.802

Li, J., Wang, G., Mayes, M.A., Allison, S.D., Frey, S.D., Shi, Z., Hu, X.-M., Luo, Y., Melillo, J.M., 2018. Reduced carbon use efficiency and increased microbial turnover with soil warming. Glob Chang Biol. https://doi.org/10.1111/gcb.14517

Li, Y., Adams, J., Shi, Y., Wang, H., He, J.-S., Chu, H., 2017. Distinct Soil Microbial Communities in habitats of differing soil water balance on the Tibetan Plateau. Sci Rep 7, 46407. https://doi.org/10.1038/srep46407

Liaw A. and Wiener M., 2002. Classification and Regression by randomForest. R News 2(3), 18--22.

Lindström, E.S., Langenheder, S., 2012. Local and regional factors influencing bacterial community assembly: Bacterial community assembly. Environmental Microbiology Reports 4, 1–9. https://doi.org/10.1111/j.1758-2229.2011.00257.x

Lipson, D.A., Xu, X., 2019. Integrating Soil Microbiology into Ecosystem Science, in: Hurst, C.J. (Ed.), Understanding Terrestrial Microbial Communities, Advances in Environmental Microbiology. Springer International Publishing, Cham, pp. 65–102. https://doi.org/10.1007/978-3-030-10777-2_3

Liu, K., Liu, Y., Han, B.-P., Xu, B., Zhu, L., Ju, J., Jiao, N., Xiong, J., 2019. Bacterial community changes in a glacial-fed Tibetan lake are correlated with glacial melting. Sci. Total Environ. 651, 2059–2067. https://doi.org/10.1016/j.scitotenv.2018.10.104

Lladó, S., López-Mondéjar, R., Baldrian, P., 2018. Drivers of microbial community structure in forest soils. Appl Microbiol Biotechnol 102, 4331–4338. https://doi.org/10.1007/s00253-018-8950-4

Lladó, S., López-Mondéjar, R., Baldrian, P., 2017. Forest Soil Bacteria: Diversity, Involvement in Ecosystem Processes, and Response to Global Change. Microbiol Mol Biol Rev 81. https://doi.org/10.1128/MMBR.00063-16

Louca, S., Mazel, F., Doebeli, M., Parfrey, L.W., 2019. A census-based estimate of Earth's bacterial and archaeal diversity. PLOS Biology 17, e3000106. https://doi.org/10.1371/journal.pbio.3000106

Mamet, S.D., Redlick, E., Brabant, M., Lamb, E.G., Helgason, B.L., Stanley, K., Siciliano, S.D., 2019. Structural equation modeling of a winnowed soil microbiome identifies how invasive plants re-structure microbial networks. The ISME Journal 1. https://doi.org/10.1038/s41396-019-0407-y

Martínez-García, P.M., López-Solanilla, E., Ramos, C., Rodríguez-Palenzuela, P., 2016. Prediction of bacterial associations with plants using a supervised machine-learning approach. Environ. Microbiol. 18, 4847–4861. https://doi.org/10.1111/1462-2920.13389

Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G., Neufeld, J.D., 2012. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics 13, 31. https://doi.org/10.1186/1471-2105-13-31

Menegotto, A., Dambros, C.S., Netto, S.A., 2019. The scale-dependent effect of environmental filters on species turnover and nestedness in an estuarine benthic community. Ecology e02721. https://doi.org/10.1002/ecy.2721

Momeni, B., Brileya, K.A., Fields, M.W., Shou, W., 2013. Strong inter-population cooperation leads to partner intermixing in microbial communities. eLife 2, e00230. https://doi.org/10.7554/eLife.00230

Moroenyane, I., Chimphango, S.B.M., Dong, K., Tripathi, B., Singh, D., Adams, J.M., 2019. Neutral models predict biogeographical patterns of soil microbes at a local scale in Mediterranean heathlands, South Africa. Transactions of the Royal Society of South Africa 1–12. https://doi.org/10.1080/0035919X.2019.1603126

Moulton, O.M., Altabet, M.A., Beman, J.M., Deegan, L.A., Lloret, J., Lyons, M.K., Nelson, J.A., Pfister, C.A., 2016. Microbial associations with macrobiota in coastal ecosystems: patterns and implications for nitrogen cycling. Frontiers in Ecology and the Environment 14, 200–208. https://doi.org/10.1002/fee.1262

Nielsen, U.N., Ayres, E., Wall, D.H., Bardgett, R.D., 2011. Soil biodiversity and carbon cycling: a review and synthesis of studies examining diversity–function relationships. European Journal of Soil Science 62, 105–116. https://doi.org/10.1111/j.1365-2389.2010.01314.x

Nikolenko, S.I., Korobeynikov, A.I., Alekseyev, M.A., 2013. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. BMC Genomics 14, S7. https://doi.org/10.1186/1471-2164-14-S1-S7

Ning, D., Deng, Y., Tiedje, J.M., Zhou, J., 2019. A general framework for quantitatively assessing ecological stochasticity. Proc. Natl. Acad. Sci. U.S.A. https://doi.org/10.1073/pnas.1904623116

Noda-García, L., Davidi, D., Korenblum, E., Elazar, A., Putintseva, E., Aharoni, A., Tawfik, D.S., 2019. Chance and pleiotropy dominate genetic diversity in complex bacterial environments. Nature Microbiology 1. https://doi.org/10.1038/s41564-019-0412-y

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A., Prjibelski, A.D., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., Clingenpeel, S.R., Woyke, T., McLean, J.S., Lasken, R., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J. Comput. Biol. 20, 714–737. https://doi.org/10.1089/cmb.2013.0084

Oksanen, J., Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P., O'Hara, R. B., Simpson, G., Solymos, P., Henry, M., Stevens, H., Szoecs, E., and Wagner, H. (2018). vegan: Community Ecology Package. R package version 2.5-3. https://CRAN.R-project.org/package=vegan

Pedersen TL (2019). tidygraph: A Tidy API for Graph Manipulation. R package version 1.1.2. https://CRAN.R-project.org/package=tidygraph

Peng, J., Lu, Z., Rui, J., Lu, Y., 2008. Dynamics of the Methanogenic Archaeal Community during Plant Residue Decomposition in an Anoxic Rice Field Soil. Applied and Environmental Microbiology 74, 2894–2901. https://doi.org/10.1128/AEM.00070-08

Pennings, S.C., Grant, M.-B., Bertness, M.D., 2005. Plant Zonation in Low-Latitude Salt Marshes: Disentangling the Roles of Flooding, Salinity and Competition. Journal of Ecology 93, 159–167.

Peter, H., Ylla, I., Gudasz, C., Romaní, A.M., Sabater, S., Tranvik, L.J., 2011. Multifunctionality and Diversity in Bacterial Biofilms. PLOS ONE 6, e23225. https://doi.org/10.1371/journal.pone.0023225

Qu, K., Guo, F., Liu, X., Lin, Y., Zou, Q., 2019. Application of Machine Learning in Microbiology. Front Microbiol 10. https://doi.org/10.3389/fmicb.2019.00827

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raffaelli, D., Bullock, J.M., Cinderby, S., Durance, I., Emmett, B., Harris, J., Hicks, K., Oliver, T.H., Paterson, D., White, P.C.L., 2014. Big Data and Ecosystem Research Programmes, in: Advances in Ecological Research. Elsevier, pp. 41–77. https://doi.org/10.1016/B978-0-08-099970-8.00004-X

Ramirez, K.S., Geisen, S., Morriën, E., Snoek, B.L., van der Putten, W.H., 2018a. Network Analyses Can Advance Above-Belowground Ecology. Trends in Plant Science 23, 759–768. https://doi.org/10.1016/j.tplants.2018.06.009

Ramirez, K.S., Knight, C.G., Hollander, M. de, Brearley, F.Q., Constantinides, B., Cotton, A., Creer, S., Crowther, T.W., Davison, J., Delgado-Baquerizo, M., Dorrepaal, E., Elliott, D.R., Fox, G., Griffiths, R.I., Hale, C., Hartman, K., Houlden, A., Jones, D.L., Krab, E.J., Maestre, F.T., McGuire, K.L., Monteux, S., Orr, C.H., Putten, W.H. van der, Roberts, I.S., Robinson, D.A., Rocca, J.D., Rowntree, J., Schlaeppi, K., Shepherd, M., Singh, B.K., Straathof, A.L., Bhatnagar, J.M., Thion, C., Heijden, M.G.A. van der, Vries, F.T. de, 2018b. Detecting macroecological patterns in bacterial communities across independent studies of global soils. Nat Microbiol 3, 189–196. https://doi.org/10.1038/s41564-017-0062-x

Ramirez, K.S., Leff, J.W., Barberán, A., Bates, S.T., Betley, J., Crowther, T.W., Kelly, E.F., Oldfield, E.E., Shaw, E.A., Steenbock, C., Bradford, M.A., Wall, D.H., Fierer, N., 2014. Biogeographic patterns in below-ground diversity in New York City’s Central Park are similar to those observed globally. Proc. Biol. Sci. 281. https://doi.org/10.1098/rspb.2014.1988

Raskin, L., Stromley, J.M., Rittmann, B.E., Stahl, D.A., 1994. Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. Appl Environ Microbiol 60, 1232–1240.

Rath, K.M., Fierer, N., Murphy, D.V., Rousk, J., 2019. Linking bacterial community composition to soil salinity along environmental gradients. ISME J 13, 836–846. https://doi.org/10.1038/s41396-018-0313-8

Reiss, J., Bridle, J.R., Montoya, J.M., Woodward, G., 2009. Emerging horizons in biodiversity and ecosystem functioning research. Trends in Ecology & Evolution 24, 505–514. https://doi.org/10.1016/j.tree.2009.03.018

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584. https://doi.org/10.7717/peerj.2584

Rosindell, J., Hubbell, S.P., Etienne, R.S., 2011. The Unified Neutral Theory of Biodiversity and Biogeography at Age Ten. Trends in Ecology & Evolution 26, 340–348. https://doi.org/10.1016/j.tree.2011.03.024

Schirmer, M., Ijaz, U.Z., D’Amore, R., Hall, N., Sloan, W.T., Quince, C., 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res. 43, e37. https://doi.org/10.1093/nar/gku1341

Schmid, M.W., Hahl, T., van Moorsel, S.J., Wagg, C., De Deyn, G.B., Schmid, B., 2018. Feedbacks of plant identity and diversity on the diversity and community composition of rhizosphere microbiomes from a long-term biodiversity experiment. Mol. Ecol. https://doi.org/10.1111/mec.14987

Sengupta, A., Stegen, J.C., Neto, A.A.M., Wang, Y., Neilson, J.W., Chorover, J., Troch, P.A., Maier, R.M., 2019. Assessing microbial community patterns during incipient soil formation from basalt. Journal of Geophysical Research: Biogeosciences 0. https://doi.org/10.1029/2017JG004315

Sinclair, L., Osman, O.A., Bertilsson, S., Eiler, A., 2015. Microbial Community Composition and Diversity via 16S rRNA Gene Amplicons: Evaluating the Illumina Platform. PLOS ONE 10, e0116955. https://doi.org/10.1371/journal.pone.0116955

Stahl, D.A. and Amann, R. (1991) Development and Application of Nucleic Acid Probes in Bacterial Systematics. In: Stackebrandt, E. and Goodfellow, M., Eds., Nucleic Acid Techniques in Bacterial Systematics, John Wiley & Sons Ltd., Chichester, 205-248.

Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W., Murray, C.J., Rockhold, M.L., Konopka, A., 2013. Quantifying community assembly processes and identifying features that impose them. ISME J 7, 2069–2079. https://doi.org/10.1038/ismej.2013.93

Stegen, J.C., Lin, X., Konopka, A.E., Fredrickson, J.K., 2012. Stochastic and deterministic assembly processes in subsurface microbial communities. ISME J 6, 1653–1664. https://doi.org/10.1038/ismej.2012.22

Strickland, M.S., Lauber, C., Fierer, N., Bradford, M.A., 2009. Testing the functional significance of microbial community composition. Ecology 90, 441–451. https://doi.org/10.1890/08-0296.1

Suarez, C., Piculell, M., Modin, O., Langenheder, S., Persson, F., Hermansson, M., 2019. Thickness determines microbial community structure and function in nitrifying biofilms via deterministic assembly. Sci Rep 9, 5110. https://doi.org/10.1038/s41598-019-41542-1

Sun, Y.-Q., Wang, J., Shen, C., He, J.-Z., Ge, Y., 2019. Plant evenness modulates the effect of plant richness on soil bacterial diversity. Science of The Total Environment 662, 8–14. https://doi.org/10.1016/j.scitotenv.2019.01.211

Tang, Y., Dai, T., Su, Z., Hasegawa, K., Tian, J., Chen, L., Wen, D., 2019. A Tripartite Microbial-Environment Network Indicates How Crucial Microbes Influence the Microbial Community Ecology. Microb Ecol. https://doi.org/10.1007/s00248-019-01421-8

Tarca, A.L., Carey, V.J., Chen, X.W., Romero, R. and Drăghici, S., 2007. Machine learning and its applications to biology. PLoS computational biology, 3.

Tokeshi, M., 1990. Niche Apportionment or Random Assortment: Species Abundance Patterns Revisited. Journal of Animal Ecology 59, 1129–1146. https://doi.org/10.2307/5036

Tourna, M., Freitag, T.E., Nicol, G.W., Prosser, J.I., 2008. Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. Environmental Microbiology 10, 1357–1364. https://doi.org/10.1111/j.1462-2920.2007.01563.x

van der Plas, F., 2019. Biodiversity and ecosystem functioning in naturally assembled communities. Biological Reviews undefined-undefined. https://doi.org/10.1111/brv.12499

van der Putten, W.H., Bradford, M.A., Brinkman, E.P., Voorde, T.F.J. van de, Veen, G.F., 2016. Where, when and how plant–soil feedback matters in a changing world. Functional Ecology 30, 1109–1121. https://doi.org/10.1111/1365-2435.12657

Vera-Gargallo, B., Chowdhury, T.R., Brown, J., Fansler, S.J., Durán-Viseras, A., Sánchez-Porro, C., Bailey, V.L., Jansson, J.K., Ventosa, A., 2019. Spatial distribution of prokaryotic communities in hypersaline soils. Scientific Reports 9, 1769. https://doi.org/10.1038/s41598-018-38339-z

Veresoglou, S.D., Halley, J.M., Rillig, M.C., 2015. Extinction risk of soil biota. Nature Communications 6, 8862. https://doi.org/10.1038/ncomms9862

Veresoglou, S.D., Verbruggen, E., Makarova, O., Mansour, I., Sen, R., Rillig, M.C., 2018. Arbuscular Mycorrhizal Fungi Alter the Community Structure of Ammonia Oxidizers at High Fertility via Competition for Soil NH4+. Microb Ecol. https://doi.org/10.1007/s00248-018-1281-2

Vivanco, L., Irvine, I.C., Martiny, J.B.H., 2015. Nonlinear responses in salt marsh functioning to increased nitrogen addition. Ecology 96, 936–947. https://doi.org/10.1890/13-1983.1

Wagg, C., Bender, S.F., Widmer, F., van der Heijden, M.G.A., 2014. Soil biodiversity and soil community composition determine ecosystem multifunctionality. Proc Natl Acad Sci U S A 111, 5266–5270. https://doi.org/10.1073/pnas.1320054111

Waldrop, M.P., Holloway, J.M., Smith, D.B., Goldhaber, M.B., Drenovsky, R.E., Scow, K.M., Dick, R., Howard, D., Wylie, B., Grace, J.B., 2017. The interacting roles of climate, soils, and plant production on soil microbial communities at a continental scale. Ecology 98, 1957–1967. https://doi.org/10.1002/ecy.1883

Wall, D.H., Bardgett, R.D., Behan-Pelletier, V., Herrick, J.E., Jones, T.H., Ritz, K., Six, J., Strong, D.R., Putten, W.H. van der (Eds.), 2012. Soil Ecology and Ecosystem Services. Oxford University Press, Oxford, New York.

Wall, D.H., Nielsen, U.N., Six, J., 2015. Soil biodiversity and human health. Nature 528, 69–76. https://doi.org/10.1038/nature15744

Wang, M., Veldsink, J.H., Dini-Andreote, F., Salles, J.F., 2018. Compositional and abundance changes of nitrogen-cycling genes in plant-root microbiomes along a salt marsh chronosequence. Antonie Van Leeuwenhoek 111, 2061–2078. https://doi.org/10.1007/s10482-018-1098-5

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl. Environ. Microbiol. 73, 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wang, S., Brose, U., 2018. Biodiversity and ecosystem functioning in food webs: the vertical diversity hypothesis. Ecology Letters 21, 9–20. https://doi.org/10.1111/ele.12865

Wang, X.-B., Yao, J., Zhang, H.-Y., Wang, X.-G., Li, K.-H., Lü, X.-T., Wang, Z.-W., Zhou, J.-Z., Han, X.-G., 2019. Environmental and spatial variables determine the taxonomic but not functional structure patterns of microbial communities in alpine grasslands. Sci. Total Environ. 654, 960–968. https://doi.org/10.1016/j.scitotenv.2018.11.138

Ward, D.M., Weller, R., Bateson, M.M., 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature 345, 63–65. https://doi.org/10.1038/345063a0

Whipps, J.M., 2001. Microbial interactions and biocontrol in the rhizosphere. J Exp Bot 52, 487–511. https://doi.org/10.1093/jexbot/52.suppl_1.487

Yan, G., Xing, Y., Lv, X., Xu, L., Zhang, J., Dai, G., Luo, W., Liu, G., Dong, X., Wang, Q., 2019. Effects of artificial nitrogen addition and reduction in precipitation on soil CO2 and CH4 effluxes and composition of the microbial biomass in a temperate forest. European Journal of Soil Science 0. https://doi.org/10.1111/ejss.12812

Yao, Z., Du, S., Liang, C., Zhao, Y., Dini-Andreote, F., Wang, K., Zhang, D., 2019. Bacterial community assembly in a typical estuarine marsh with multiple environmental gradients. Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.02602-18

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., Whitman, W.B., Euzéby, J., Amann, R., Rosselló-Móra, R., 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat. Rev. Microbiol. 12, 635–645. https://doi.org/10.1038/nrmicro3330

Yin, Y.L., Wang, Y.Q., Li, S.X., Liu, Y., Zhao, W., Ma, Y.S., Bao, G.S., 2019. [Effects of enclosing on soil microbial community diversity and soil stoichiometric characteristics in a degraded alpine meadow]. Ying Yong Sheng Tai Xue Bao 30, 127–136. https://doi.org/10.13287/j.1001-9332.201901.009

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

Zhang, Q., Zhou, W., Liang, G., Sun, J., Wang, X., He, P., 2015. Distribution of soil nutrients, extracellular enzyme activities and microbial communities across particle-size fractions in a long-term fertilizer experiment. Applied Soil Ecology 94, 59–71. https://doi.org/10.1016/j.apsoil.2015.05.005

Zhao, X., Hao, Y.-Q., Zhang, D.-Y., Zhang, Q.-G., 2019. Local biotic interactions drive species-specific divergence in soil bacterial communities. The ISME Journal 1–10. https://doi.org/10.1038/s41396-019-0477-x

Zhao, P., Bao, J., Wang, X., Liu, Y., Li, C., Chai, B., 2019. Deterministic processes dominate soil microbial community assembly in subalpine coniferous forests on the Loess Plateau. PeerJ 7, e6746. https://doi.org/10.7717/peerj.6746

Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., Barba, M.D., Gaucher, P., Gielly, L., Giguet-Covex, C., Iribar, A., Réjou-Méchain, M., Rayé, G., Rioux, D., Schilling, V., Tymen, B., Viers, J., Zouiten, C., Thuiller, W., Coissac, E., Chave, J., 2019. Body size determines soil community assembly in a tropical forest. Molecular Ecology 28, 528–543. https://doi.org/10.1111/mec.14919