Imperial College London

Imperial College London

Department of Life Sciences

Microbes: Is everything everywhere? Quantifying the contribution of dispersal and niches to diversity maintenance in a microbial system.

Author:
Amy Solman

Supervisor:
Prof. Ryan Chisholm University of Singapore

A thesis submitted for the degree of

MRes Computational Methods in Ecology and Evolution

July 9, 2020

Abstract

Microbial communities play an essential role in terrestrial and aquatic ecosystem functioning. Despite their importance, relatively little is known about the impacts of habitat size, niche availability and immigration rate on microbial species richness and diversity. The theory of island biogeography predicts that species richness increases with island area, also known as the species-area relationship (SAR). Not all empirical evidence supports the SAR theory. Small islands often show no clear relationship between species richness and area, dubbed the small-island effect (SIE). Chisholm *et al* [1] developed a unified theory of a biphasic SAR, where species richness transitions from a niche-structured regime to a colonisation-extinction regime as island area increases. This study attempts to show the presence of a biphasic island SAR in microbial communities by quantify the contributions of immigration and niches to diversity maintenance under experimental conditions.

Declaration

I decalre that this thesis has been written by myself. The data sources used in this thesis are listed in the Supplementary Materials section.

Acknowledgements

Thank you to my supervisors, Ryan, James and Tom. Thank you to all of the scientists that took the time to share their data with me. Thank you to my mother, my partner and my friends for giving me unending support outside of this project. I couldn't have done it without you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

MacArthur and Wilson's theory of island biogeography [2] is widely accepted as a fundamental ecological law. It posits that island communities are maintained by a combination of immigration and niche availability. Niches provide reduced inter- and intraspecific competition, slowing competitive exclusion. Immigration provides new species and new individuals, offsetting loses to competitive exclusion and genetic drift. Islands with a larger area and reduced distance from mainland communities or other islands, will exhibit higher species diversity at colonization-extinction dynamic equilibrium than those smaller and further away. A great deal of empirical evidence has been amassed to support the theory of island biogeography and it has been widely applied in understanding the effects of habitat fragmentation [3].

Despite the popularity of this theory, there is empirical evidence to suggest it cannot always be applied to smaller islands [4][5]. MacArthur and Wilson noted that archipelagos showed unusual SARs, with smaller island species-richness varying independent of size [2]. This exception to MacArthur and Wilson's putative ecological law has been dubbed the small island effect (SIE). Several hypotheses have been offered to explain the SIE. The 'subsidized island biogeography' (SIB) hypothesis suggests that smaller islands have a greater edge:interior ratio, thus receive a greater amount of nutrients per unit area in contrast to larger islands [6][7]. Moreover, extinction rates on islands may operate independently of area due to their environmental instability and high temporal turnover, where major episodic disturbances periodically wipe out colonizing species [1]. Thirdly, the 'habitat hypothesis' suggests a limited suite of habitats on small islands, in contrast to larger islands, limits species diversity [8].

Chisholm *et al* [1] have developed a unified theory to explain this biphasic island SAR. They posit that this pattern of species-richness is due to a transition from a niche-structured regime on smaller islands, to a colonization-extinction regime on larger islands. The niche-structured regime is characteristic of deterministic niche theories where environmental filtering, biotic interactions and interspecific trade-offs determine species richness [9]. The colonization-extinction regime is characteristic of stochastic theories such as the theory of island biogeography and neutral theory, where richness is dictated by colonization and extinction rate as well as ecological drift [10].

1.1 Microbial Communities

Whilst numerous studies have looked at macro-organism SARs, relatively little is known about how microbial species are affected by habitat size, isolation and niche availability. Understanding the factors that regulate microbial species richness and community structure is important as they play a significant role in ecosystem functioning [11]. Greater understanding of the patterns underlying microbial biogeography is also important in predicting the responses of these organisms to a changing environment [12].

Debate around the applicability of SARs to microbial systems stems from the idea that they are ubiquitous in the environment and functionally redundant. This was famously articulated by Bass-Becking [13] who wrote, 'everything is everywhere, but, the environment selects'. Some studies have shown that SAR theories are not useful in describing microbial richness patterns [14] [15].

Investigating microbial biogeography also has its practical challenges. Problems often arise in experimentally manipulating habitats to a degree useful for such a study and it is only recently that

advances in molecular tools has lead to a resurgence in the field. One study found bacterial SARs in aquatic treehole habitats comparable with that of larger organisms [16]. Organic aggregates have also been used as 'islands' to explore the biogeography of aquatic pathogens [17], showing a weak positive correlation between 'island' size, community metabolic response and functional diversity. Investigation of phytoplankton SAR in water bodies found evidence for the SIE, however they also detected a large lake effect where species richness declined beyond a threshold area as wind-induced mixing increased habitat homogeneity [18]. This indicated that small-scale niche relations were the most important determinants of species richness at the smallest spatial scales. The SIE has also been seen in benthic diatoms sampled from ponds on a wide spatial scale [19]. Increased niche dimensionality has been shown to increased functional diversity, with strong evidence for niche filtering of microbial taxa communities [20].

The majority of studies looking at microscopic island biogeography have focused on aquatic communities. This is likely due to the availability of isolated water bodies and their range of spatial scales. Bacteria also represent a major contributor to soil biodiversity and processes [11]. Despite this, little is known about belowground regulators of biodiversity. An in-situ study of ectomycorrhizal fungi communities within 'tree island' root systems showed that total species richness increased significantly with island size, but distance had little effect [21]. Despite the strong SAR found in this study, there was no good evidence linking the SAR to increases in niche variety with habitat size, a result consistent with the theory of island biogeography and neutral community models. Most investigations have manipulated habitat area only, inferring that niche availability will naturally vary with habitat size. One study directly manipulated niche dimensionality by varying resource richness and found this resulted in increased functional dissimilarity, community productivity and reduced invasion [22].

To tackle the limitations of previous in-situ research where environmental variables were difficult to control [23], this study aims to quantify the contributions of immigration and niches to diversity in a soil microbial community, by manipulating mesocosm 'island' habitats. We aim to assess whether Chisholm's unified theory of biphasic island SARs is supported by our results. The parsimonious machanistic model used by Chisholm *et al* [1] to approximate the processes of a biphasic SAR is fit to the data, along with other mathematical models. The number of microbial species maintained by niches is inferred from the asymptotic richness as immigration rates are reduced. The contribution of immigration is inferred from the rate of richness increase across treatments. This investigation seeks to identify nonlinearities in the richness versus immigration curve.

1.2 Hypotheses

It is hypothesise that:

1) Soil microbial communities will be dictated by a niche-structure regime at the smallest spatial scales, before transitioning to a colonization-extinction regime at larger spatial scales.

2) 'Islands' with lower rates of immigration will have less species at the colonization-extinction dynamic equilibrium.

Chapter 2

Methods

The model fitting process was validated by developing a simulation from which parameters of theta, migration rate and niches could be retrieved. A specifically designed laboratory experiment was commenced in January 2020 to test the model's theories in relation to microbial community dynamics. Due to the COV-19 pandemic the laboratory experiment was ended in March 2020. An alternative approach was devised, in which the model was fit to microbial species-area datasets compiled from the literature.

2.1 Laboratory Experiment

2.1.1 Study area and sample collection

Soil was collected on site at Silwood Park, Berkshire, UK. Silwood Park comprises a variety of habitats including woodland, wetlands, heathland and formerly arable land [24]. Soil types across the site consist of sandy or silty loam, with pH ranging from 4 to 6 [25]. For this experiment soil was collected from a fallow field of acidic, sandy soil [24]. Samples were acquired and sterilised during the first week of February. A total of 100 litres of soil was collected.

2.1.2 Preparation of soil and mesocosms

The soil was homogenised by sieving and sterilising three times using an autoclave to destroy any present bacteria or fungi. We used five sizes of container: 0.2ml (PCR tube), 1.5ml (Eppendorf), 50ml (Fallon), 500ml and 5000ml. For each size of container 1, 2, 3, 4, 5 and maximum number of holes were made using a heated needle. Three replicates were made for each size of container and number of holes. The containers were then sterilised using the autoclave. The containers were filled with sterilised soil and sealed with an opaque lid. The containers were then buried on site, with sealed lids exposed and left to incubate for one month. 1. If we use transparent covers for the tops of the tubes/containers this may create a gradient of light energy and moisture? Provide niches for heterotrophic and phototrophic microbes? Niche-based processes have been shown to be predominant in structuring observed aridity-related community patterns [26]. Abundance of predatory myxobacterial communities has also been correlated with temperature [**?**].

2.1.3 Soil properties measurements and climate data collection

What is the pH of the soil where the samples were collected? pH of the soil where the samples were buried? pH has been show to affect microbial diversity [27].

2.1.4 DNA extraction and sequence analysis

After one month the containers were retrieved and DNA sampling was carried out the ZR Fungal/Bacterial DNA MiniPrep^{TM} Isolation kit. Estimation of relative abundances of bacterial taxonomic groups was carried out using a previously defined PCR-based method [28]. Real-time PCR, or qPCR, allows for rapid quantitative assessment of soil microbial communities. I used 388 forward primer and 518 reverse primer as suggested for targeting all bacterial groups [28]. DNA samples from all 90 mesocosms were prepared for Illumina MiSeq 16S sequencing.

2.2 Simulation

Overview

This simulation is designed to mimic the process of island colonisation from a metacommunity. The colonisation process is constrained by migration rate and each island is characterised by number of niches and the size of each niche. The output of the simulation is the final community in each niche of each island, as well as a timeseries of species richness.

Metacommunity

A metacommunity is generated at the beginning of the simulation using *coalescence_ test* function, partially modified from a script provided by Dr James Rosindell. The function takes the input parameters: metacommunity size($J\_meta = 50\ 000$) and speciation rate ($nu = 0.001$). Each run of the function produces 20 niche communities, each of size $J\_meta/20$.

The function initialises a vector (*lineages*) of length = $J\_meta/20$ = *niche_ size* with 1 as every value. An empty vector (*abundances*) is initialised. The value of *niche_ size* is given to *N*. *Theta* is calculated as *nu*(niche_ size-1)/(1-nu)*. Then, while $N > 1$, a vector (*linvect*) is created with values 1:length(*niche_ size*). A random sample of *linvect* is made (j). A random decimal number is selected between 0 and 1 (*randnum*). If *rannum* is less than *theta/(theta+N-1)*, then the value at *lineages[j]* is appended to *abundances*. Else, another random number (i) is sampled from *linvect*, excluding the last number selected. The values at *lineages[i]* and *lineages[j]* are summed and take the position of *lineages[i]*. *lineages[j]* is then removed from *lineages*, so the vector is one value shorter. The value of N is also decreased by 1. This repeats until $N = 1$. The remaining value in the *lineages* vector is added to *abundances* and the function outputs a vector of simulated species abundances.

At the end of the *coalescence_ test* function, each of the 20 niches is assigned a letter type from A to T. A list of niche communities in the metacommunity is returned. It was chosen to generate no more than 20 niches within the metacommunity, because the simulation would not go beyond modelling 20 niches on an island. Any additional niches generated for the metacommunity would go unused.

The *coalescence_ test* function is incorporated into a second function (*metacommunity*), that generates a vector of individuals from each niche abundance vector. For example, Niche A *abundances*(5,4,2,2,1,1) would generate a community *meta$A*(1,1,1,1,1,2,2,2,2,3,3,4,4,5,6) where each unique number value represents a unique species.

Parameters

The variable parameters of the simulation are: migration rate (range: 0.003-0.06), number of niches (range: 1-20), size of niches (range:1-20). Each unit of space was assumed to host one individual, therefore, number of niches x size of niches = area = size of island population. There are a total of 8000 condition combinations applied during the simulation (20 migration rates, 20 number of niches, 20 size of niches).

Simulation Logic

At timestep i an island is selected. The first niche on that island becomes the focal niche. An individual within that niche is chosen to die. With probability m (migration rate), the dead individual is replaced with a randomly chosen propagule from the same niche type in the metacommunity (e.g. from metacommunity niche A to island niche A). With probability 1 - m, the dead individual is replaced with a local propagule from the same niche. The simulation then moves to the next niche on that island. When all niches have been simulated for timestep i, the simulation moves to the next island. When all islands have been simulated for timestep i, the simulation moves to the next timestep $i + 1$, and returns to the first island (Figure 1). The species richness for each niche is calculated, totaled across all niches for each island and stored at every 5000 timesteps.

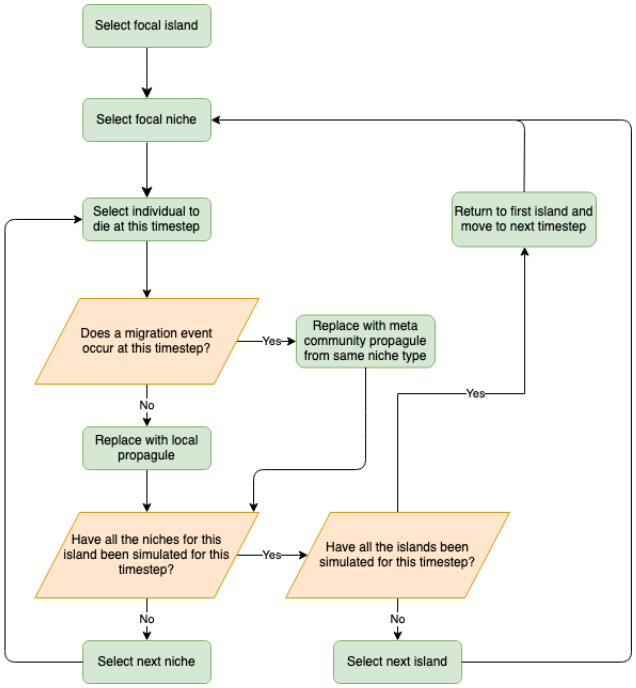

Figure 2.1: Flowchart of simulation design

High Performance Computing

The metacommunity and simulation code are contained in *ClusterSim.R*. *ClusterSim.R* functions are sourced by *ClusterCode.R* and given the input parameters: $J\_meta = 50000$, $nu = 0.001$, $num\_m\_rates = 20$, $max\_k\_num = 20$, $max\_k\_size = 20$, $wall\_time = 1380$, $output\_file\_name$ = output_file_name (where each simulation is given a unique file name "simulation_timeseries_i". *ClusterRun.sh* is used to run on the cluster, with a time limit of 24:00:00. $wall\_time$ is given as 1380 minutes (23 hours) within the function, to ensure all simulations are completed before the cluster run ends. 100 parallel simulation were run on the Imperial College London High Performance Computing service. This generated a total of 800000 islands, simulated for $>$ 30000 timesteps.

2.2.1 Data Preparation and Timeseries Plots

The 100 simulation results were imported into *DataPrep.R*. The data from each island was isolated and configured into a data frame with simulation number, migration rate, area, number of niches and number of species (*SimModelFitData.csv*). A second data frame was generated for timeseries plotting, with simulation number, island number, migration rate, timestep and species richness timeseries for each island (*SimTimeseriesPlotData.csv*).

To ensure the simulation had run long enough for each island to reach dynamic equilibrium the species richness timeseries of simulations 25, 50 and 75 were plotted(*TimeseriesPlot.R*)(Figure 2).

2.2.2 Analysis

```
chisholm_model <- function(area, theta, m0, K) {
  rho = 1
  K = K
  Js = area*rho
  J_stars = Js/K
  ms = m0/sqrt(area)
  gamma_stars = J_stars*ms/(1-ms)
  return(theta*(digamma(theta/K+gamma_stars*
```

```
  (digamma(gamma_stars+J_stars)-digamma(gamma_stars)))-digamma(theta/K)))
}
```

An analysis script (*Analysis.R*) imported the prepared data (*SimModelFitData.csv*) for each island across all simulations (800000 islands). Model estimated species richnesses were generated by giving the island parameters (*m0 = m*sqrt(area)*) and an estimated *theta* (*theta = 2*(niche_size*K)*nu*, where *niche_size * K* = size of each niche in the metacommunity from which immigration events can occur times by the number of niches contributing to the island community i.e. number of niches on the island). The results of the simulation and those estimated by the *chisholm_function* (above) were bound together in a data frame. Mean species richness results for each combination of island area, migration rate and number of niches across all 100 simulations was calculated and stored.

2.3 Data Collection and Analysis

2.3.1 Datasets

Table 2.1: Summary of datasets collected from the literature

Attribute	*Aqua*	*Terra*	*Total*
fresh	20	0	20
saline	5	0	5
in situ	24	10	34
lab	1	1	2
modified	10	3	13
natural	15	8	23
continuous	2	1	3
isolated	23	10	33
archaea	0	1	1
bacteria	12	2	14
fungi	1	6	7
algae	7	0	7
protozoa	7	0	7
pathogens	0	3	3
total	25	11	36

In-lieu of being able to generate my own dataset, I compiled 36 datasets of microbial species/taxa-area relationships. A full description of each dataset and their authors can be found in the Supplementary Material section. These datasets included a range of taxonomic groups, including archaea, bacteria, fungi, algae, protazoa and pathogens. The datasets also included aquatic and terrestrial habitats, as well as in-situ and lab based investigations. For the majority of datasets, the number of observed species, strains, operational taxonomic units or phylogroups per "island" habitat was taken. Some studies only supplied diversity indexes (gamma diversity, phylogenetic diversity, faiths dp, simpsons index, chao 1, T-RFLP, viral prevalence). The estimated number of cells/individuals per unit area was also taken from each study, or estimated from the relevant literature if unavailable.

2.3.2 Model Fitting

$$S = \theta\{\psi(\frac{\theta}{K} + \gamma(\psi(\gamma + J) - \psi(\gamma))) - \psi(\frac{\theta}{K})\} \tag{2.1}$$

For each dataset I found initial parameter estimates for m and θ. I then looped through all values of K from 1 to the maximum number of taxa recorded on one "island" and fitted the Chisholm model used NLLS fitting. From the fitting, the best-fit parameters for m and θ were selected, along with their best fit K. On "islands" where the estimated cell count per unit area was less than the value of K, estimated species richness was constrained to equal number of individuals so as not to more species than individual microorganisms.

2.3.3 Critical Area

$$ACrit = \frac{\theta(1-m)(exp(K/\theta)-1}{m\rho * log(1/m)} \tag{2.2}$$

For each successful model fitting, the critical area of transition from a niche-structured regime to an extinction-colonisation equilibrium regime was estimated. The best-fit values of K, m and θ from the model fitting were given to the critical area equation (2.1).

Chapter 3

Results

This is one of the most important components of the dissertation. It should begin with a clear statement of what the project is about so that the nature and scope of the project can be understood by a lay reader. It should summarise everything you set out to achieve, provide a clear summary of the project's background and relevance to other work and give pointers to the remaining sections of the dissertation which contain the bulk of the technical material.

3.1 Objectives

3.2 Challenges

3.3 Contributions

Chapter 4

Discussion

This is one of the most important components of the dissertation. It should begin with a clear statement of what the project is about so that the nature and scope of the project can be understood by a lay reader. It should summarise everything you set out to achieve, provide a clear summary of the project's background and relevance to other work and give pointers to the remaining sections of the dissertation which contain the bulk of the technical material.

4.1 Objectives

4.2 Challenges

4.3 Contributions

Chapter 5

Conclusion

This is one of the most important components of the dissertation. It should begin with a clear statement of what the project is about so that the nature and scope of the project can be understood by a lay reader. It should summarise everything you set out to achieve, provide a clear summary of the project's background and relevance to other work and give pointers to the remaining sections of the dissertation which contain the bulk of the technical material.

5.1 Objectives

5.2 Challenges

5.3 Contributions

Chapter 6

Supplementary Material

6.1 Datasets

Bibliography

[1] Ryan A Chisholm, Tak Fung, Deepthi Chimalakonda, and James P O'Dwyer. Maintenance of biodiversity on islands. *Proceedings of the Royal Society B: Biological Sciences*, 283(1829):20160102, 2016.

[2] Robert H MacArthur. *The theory of island biogeography*. Monographs in population biology ; 1. Princetown University Press, Princetown, 1967.

[3] Yrjö Haila. A conceptual genealogy of fragmentation research: from island biogeography to landscape ecology. *Ecological applications*, 12(2):321–334, 2002.

[4] KA Triantis, K Vardinoyannis, EP Tsolaki, I Botsaris, K Lika, and M Mylonas. Reapproaching the small island effect. *Journal of Biogeography*, 33(5):914–923, 2006.

[5] Spyros Sfenthourakis and Kostas A Triantis. Habitat diversity, ecological requirements of species and the small island effect. *Diversity and Distributions*, 15(1):131–140, 2009.

[6] Kyle Barrett, DA Wait, and WB Anderson. Small island biogeography in the gulf of california: lizards, the subsidized island biogeography hypothesis, and the small island effect. *Journal of Biogeography*, 30(10):1575–1581, 2003.

[7] WB Anderson and DA Wait. Subsidized island biogeography hypothesis: another new twist on an old theory. *Ecology Letters*, 4(4):289–291, 2001.

[8] Kostas A Triantis, Moisis Mylonas, and Robert J Whittaker. Evolutionary species–area curves as revealed by single-island endemics: insights for the inter-provincial species–area relationship. *Ecography*, 31(3):401–407, 2008.

[9] Jonathan M Chase and Jonathan A Myers. Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical transactions of the Royal Society B: Biological sciences*, 366(1576):2351–2363, 2011.

[10] Stephen P Hubbell. *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Princeton University Press, 2001.

[11] Robert I Griffiths, Bruce C Thomson, Phillip James, Thomas Bell, Mark Bailey, and Andrew S Whiteley. The bacterial biogeography of british soils. *Environmental microbiology*, 13(6):1642–1654, 2011.

[12] James A Bradley, Alexandre M Anesio, and Sandra Arndt. Microbial and biogeochemical dynamics in glacier forefields are sensitive to century-scale climate and anthropogenic change. *Frontiers in Earth Science*, 5:26, 2017.

[13] Lourens Gerhard Marinus Baas-Becking. *Geobiologie; of inleiding tot de milieukunde*. WP Van Stockum & Zoon NV, 1934.

[14] Noah Fierer and Robert B Jackson. The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*, 103(3):626–631, 2006.

[15] Michael S Henebry and John Cairns Jr. The effect of island size, distance and epicenter maturity on colonization in freshwater protozoan communities. *American Midland Naturalist*, pages 80–92, 1980.

[16] Thomas Bell, Duane Ager, Ji-Inn Song, Jonathan A Newman, Ian P Thompson, Andrew K Lilley, and Christopher J Van der Gast. Larger islands house more bacterial taxa. *Science*, 308(5730):1884–1884, 2005.

[17] MM Lyons, JE Ward, Holly Gaff, Randall E Hicks, JM Drake, and Fred C Dobbs. Theory of island biogeography on a microscopic scale: organic aggregates as islands for aquatic pathogens. *Aquatic Microbial Ecology*, 60(1):1–13, 2010.

[18] Gábor Várbíró, Judit Görgényi, Béla Tóthmérész, Judit Padisák, Éva Hajnal, and Gábor Borics. Functional redundancy modifies species–area relationship for freshwater phytoplankton. *Ecology and evolution*, 7(23):9905–9913, 2017.

[19] Ágnes Bolgovics, Éva Ács, Gábor Várbíró, Judit Görgényi, and Gábor Borics. Species area relationship (sar) for benthic diatoms: a study on aquatic islands. *Hydrobiologia*, 764(1):91–102, 2016.

[20] Kevin Lee, Stephen David James Archer, Rachel Boyle, Donnabella Castillo Lacap-Bugler, Jayne Belnap, and Steve Brian Pointing. Niche filtering of bacteria in soil and rock habitats of the colorado plateau desert, utah, usa. *Frontiers in Microbiology*, 7, 2016.

[21] Kabir G Peay, Thomas D Bruns, Peter G Kennedy, Sarah E Bergemann, and Matteo Garbelotto. A strong species–area relationship for eukaryotic soil microbes: island size matters for ectomycorrhizal fungi. *Ecology letters*, 10(6):470–480, 2007.

[22] Nico Eisenhauer, Wiebke Schulz, Stefan Scheu, and Alexandre Jousset. Niche dimensionality links biodiversity and invasibility of microbial communities. *Functional Ecology*, 27(1):282–288, 2013.

[23] Tom Fenchel and Bland J Finlay. Bacteria and island biogeography. *Science*, 309(5743):1997–1999, 2005.

[24] Michael J Crawley. *The flora of Berkshire : including those parts of modern Oxfordshire that lie to the south of the river Thames : with accounts of charophytes, ferns, flowering plants, bryophytes, lichens and non-lichenized fungi*. Brambleby Books, Harpenden, 2005.

[25] Kathryn Luckett. The biodiversity-ecosystem function relationship in natural grassland communities at silwood park, 2015.

[26] Muke Huang, Liwei Chai, Dalin Jiang, Mengjun Zhang, Yanran Zhao, and Yi Huang. Increasing aridity affects soil archaeal communities by mediating soil niches in semi-arid regions. *Science of the Total Environment*, 647:699–707, 2019.

[27] Robert I. Griffiths, Bruce C. Thomson, Phillip James, Thomas Bell, Mark Bailey, and Andrew S. Whiteley. The bacterial biogeography of british soils. *Environmental Microbiology*, 13(6):1642–1654, 2011.

[28] Noah Fierer, Jason A Jackson, Rytas Vilgalys, and Robert B Jackson. Assessment of soil microbial community structure by use of taxon-specific quantitative pcr assays. *Applied and environmental microbiology*, 71(7):4117, 2005.