The structure and dynamics of complex microbe-host interaction networks

Johannes Björk

Tesis presentada para la obtenció del títol de Doctor

per la Universitat Politècnica de Catalunya

Programa de Doctorat de Ciències del Mar

Biologia marina i oceanografia

Institut de Ciències del Mar

Barcelona, Maig 2016

Till mor och far

"All living beings, all animals from Amoeba to Man,
all plants from Cryptogams to Dicotyledons are constituted by an association,
the *'emboitement'* of two different beings."

P. Portier, *Les Symbiotes* (Paris: Masson 1918), vii.

Abstract

Microbes form intricate and intimate relationships with most animals and plants, many of which are crucial for host development, health and functioning. Microbe–host symbiotic associations are poorly explored in comparison with other species interaction networks. The current paradigm on symbiosis research stems from species-poor systems where pairwise and reciprocally specialised interactions between a single microbe and host that coevolve are the norm. These symbioses involving just a few species are fascinating in their own right, but more diverse and complex host-associated microbial communities, so-called microbiomes, are increasingly found, with new emerging questions that require new paradigms and approaches. In this thesis, I investigate the structure and dynamics of complex microbe-host systems, focusing on the specialisation of their ecological interactions and on identifying fingerprints of coevolution in multispecies communities. I take a holistic approach to study interactions between complex assemblages of microbes associated to multiple host species through time and space. My overreaching objective is twofold. First, to determine the structure and dynamics of host-associated cores–a set of highly specialised and complementary microbial symbionts with a disproportionately large influence on the dynamics and stability of the assemblage as a whole. Second, to increase our understanding of the underlying ecological and evolutionary processes that determine the structure and dynamics of core microbiomes, which ultimately influence the functional relationship between symbionts and hosts. In my thesis, I use marine sponges and their associated bacteria as a study system. Marine sponges are one of the invertebrate phyla harbouring the largest abundance and diversity of microbes, and it is the phylogenetically oldest, still extant metazoan phyla with the hypothesized oldest microbe-host symbiotic interactions. I show that sponge-associated microbial communities emerge as one of the most specialised, yet highly diverse ecological networks ever analysed. I show that hosts harbour a core microbiome not only common to most individuals of the same species, but that cores often consist of abundant, temporally stable microbes that persistently associate to hosts over periods of years and probably even longer, and that some hosts preserve subsets of their specialised interactions by vertically transmitting microbes to the next generation. I find that the presence of high-density core microbiomes

confers hosts a resistance against the increase in abundance of the many occasional microbes that pass through the sponge due to its filter-feeding activities. I show that intraspecific interactions through density-dependent dynamics together with weak interspecific interactions are likely key determinants of microbiome stability and fingerprints of coevolved interactions.

Resumen

Los microorganismos establecen relaciones complejas e íntimas con múltiples especies animales y vegetales, muchas de las cuales son necesarias para el desarrollo, salud y funcionamiento de sus hospedadores. Los estudios acerca de estas asociaciones simbióticas entre microbios y sus hospedadores son escasos en comparación con las investigaciones realizadas en otro tipo de redes de interacción entre especies. El paradigma actual dentro de la investigación en simbiosis proviene de sistemas con escasa diversidad biológica, donde las interacciones de especialización recíproca entre una especie de microbio y un hospedador que coevolucionan son la norma. Estas simbiosis que comprenden un escaso número de especies interaccionando entre si son fascinantes por si solas, pero se están descubriendo cada vez más comunidades microbianas mucho más diversas y complejas asociadas a otros hospedadores, conocidas como microbiomas, lo que está haciendo aparecer nuevas preguntas que requieren nuevos paradigmas y aproximaciones. En esta tesis, investigo la estructura y la dinámica de sistemas complejos de interacción entre hospedadores y microbios, centrándome en la especialización de estas interacciones ecológicas, así como en la identificación de huellas de la coevolución en comunidades con gran diversidad biológica. Adopto una perspectiva holística sobre el estudio de las interacciones entre ensamblajes complejos de microorganismos asociados a varias especies hospedadoras a lo largo del tiempo y del espacio. Tengo un doble objetivo global. En primer lugar, determinar la estructura y dinámica de los núcleos de microbios asociados a los hospedadores- definidos como aquel grupo de microbios simbiontes altamente especializados y complementarios entre si que tienen una influencia desproporcionada sobre la dinámica y estabilidad de todo el ensamblaje microbiano. En segundo lugar, mejorar nuestro conocimiento acerca de los procesos ecológicos y evolutivos responsables de la estructura y dinámica de estos núcleos del microbioma, lo que en última instancia determina las relaciones funcionales entre los simbiontes y sus hospedadores. En mi tesis utilizo las esponjas marinas y sus bacterias asociadas como sistema de estudio. Las esponjas marinas constituyen el phylum de invertebrados que albergan la mayor diversidad y abundancia de microorganismos, y es el más antiguo pero aun existente phylum, al que pertenecen las más antiguas relaciones simbiontes entre microbios y hospedadores del árbol de la vida. En

mi tesis muestro que las comunidades microbianas asociadas a esponjas son una de las más especializadas, y sin embargo diversas redes ecológicas jamás mostradas. Muestro que las esponjas contienen un núcleo de microbioma que no es sólo compartido entre individuos de la misma especie, sino que además estos núcleos están formados por microbios muy abundantes que son estables a lo largo del tiempo, de modo que estos microbios permanecen asociados a sus hospedadores durante años y seguramente periodos de tiempo mucho más extensos, y distintos hospedadores conservan un grupo de estos microbios altamente especializados gracias a la transmisión vertical de los mismos a sus descendientes. Igualmente, demuestro que la presencia de núcleos del microbioma que tienen una gran densidad de simbiontes confieren una gran resistencia a sus hospedadores frente al aumento en abundancia de los muchos microbios ocasionales presentes en las esponjas debido a sus actividades de filtrado de agua de mar. Muestro también que las interacciones entre individuos de una misma especies de microorganismo, a través de la regulación denso-dependiente, junto a las interacciones de naturaleza débil entre distintas especies de microbios, son clave para entender la estabilidad y persistencia de los microbiomas y huellas inequívocas de interacciones que han coevolucionado.

Acknowledgements

It has been a long journey. I decided to become a biologist when I was fifteen, and in June this year, I turn thirty. As of most of life's adventures, this journey has been a long winding road with tons of ups and downs, moments on the highest mountaintops and moments in deep valleys of despair.

Though only my name appears on the cover of this dissertation, my thesis represents not only work done by me, but is a result of fruitful collaborations with several great scientists. I owe my gratitude to all the people who have made my thesis possible and because of whom my PhD experience has been one that I will long remember.

First and foremost, I would like to thank my supervisor Dr. Jose Montoya for providing direction through the maze that is to undertake a PhD, but at the same time giving me the opportunity to grow and develop as a scientist. For being a source of inspiration and motivation, but above all, for his exceptionally broad view of the biotic world and of the interactions that takes place within it. It has been an almost five year journey that started in Barcelona, and that came to an end in a place that takes your imagination to J.R.R. Tolkien's Middle Earth (Moulis, France)–I hope our interaction is not lost to the periphery but persist, part of the core.

My sincere gratitude and appreciation also goes to the Bayesian wizards Dr. Robert (Bob) O'Hara and Dr. Francis Hui, for introducing me 'the other' side of statistics. Without their wizardry, much of the statistical modelling in this thesis would not have been possible.

I would also like to thank Dr. Marta Ribes and Dr. Rafel Coma, for fruitful collaborations, and without whose invaluable time-series data, my thesis would be a static representation of yet another microbiome. A special thanks also go to Dr. Rafel Coma and Eduard Serrano for helping me during my fieldwork in L'Estartit. I would also like to thank the large number of people behind the Earth Microbiome Project, and specifically all the contributing groups working on the sponge microbiome. In particular, I want to thank Dr. Torsten Thomas for brining many of us

together and coordinating the sponge division of this massive undertaking.

I also want to show my gratitude to admirable scientists at the ICM, specifically to Dr. Pep Gasol, Dr. Ramon Massana, and Dr. Ramiro Logares for productive discussions and help during various stages of my thesis. A special thanks also go to Dr. Pablo Sánchez for managing the 'BioCluster' on which most of my work have been dependent, and for always providing fast feedback and help.

I have to say, it is not easy learning a new language while taking on a PhD. One thing for certain, I utterly failed. I am therefore grateful for the friends and colleagues at the ICM who made an effort conversing in English. I am particularly thankful to Valerio, Teresa and Giulia, whose company I have enjoyed during countless of lunches, beers and other fun gatherings. On a random tangent: Valerio and I met at a conference in Cambridge, UK, some six months before starting my PhD in Barcelona, and as it turned out, he became my first roommate in 'Casa Llull', together with Cydney and Sandra–a time I will always remember.

I have also had the pleasure of working alongside my friends and fellow 'EcoNetGC' group members, Núria and Miguel, whose company and discussions I always have enjoyed.

My experience in Moulis the past year has been nothing else than great, and for that I have to thank *la famille Luzenac*: Aisha, Jarad, Jeff, Keoni, Mat, Núria and Theresa. I will especially miss our dinners and get-togethers, where science never was too far away. Thanks for feeding me (either directly, or indirectly- yes, I was stealing your food!) during my insanely intense period of thesis writing. A special thanks also go to Núria for always offering me a car seat between Moulis and Barcelona. I will almost miss the long winding journeys across the Pyrenees!

I have really grown fond of Cataluña, its people and their culture, and even if I do not speak their language, I feel at home here. And for that, I have to thank *meu amor* Tania. I owe you enormously for your tremendous patience. I love you dearly, *y em moro de ganes de compartir la meva vida amb tu*!

Last but not the least, I would like to thank my family, who always support me no matter what. My thesis is for you.

Summary

In this thesis, I investigate the structure and dynamics of complex microbe-host systems, focusing on the specialization of their ecological interactions and on identifying fingerprints of coevolution in multispecies communities. I take a holistic approach to study interactions between complex assemblages of bacteria associated to multiple host species through time and space. This approach is highly integrative, combining disciplines such as community ecology and network theory, microbiology and evolutionary biology. My overarching objective is twofold. First, to determine the structure and dynamics of host-associated cores–a set of highly specialised and complementary microbial symbionts with a disproportionately large influence on the dynamics and stability of the assemblage as a whole. Second, to increase our understanding of the underlying ecological and evolutionary processes that determine the structure and dynamics of core microbiomes, which ultimately influence the functional relationship between symbionts and hosts.

In my thesis, I use marine sponges (*Porifera*) and their associated bacteria as a study system. *Porifera* is one of the invertebrate phyla harbouring the largest abundance and diversity of microbes (Webster, Luter, *et al.* 2012; Hentschel, Piel, *et al.* 2012), and is the phylogenetically oldest, still extant metazoan phyla with the hypothesised oldest bacteria-host symbiotic interactions (Taylor, Radax, *et al.* 2007). Sponges are commonly divided into two groups depending on the microbial density found within their bodies–High Microbial Abundance (HMA) and Low Microbial Abundance (LMA) sponges. This classification in turn affects numerous host morphological and physiological traits. With their dense interior, narrower aquiferous canals and small choanoyctes, HMA sponges are well suited for harbouring dense and diverse consortia of microbes (Vacelet & Donadey 1977). In contrast, LMA sponges have an architecture more fitted for filter-feeding (Schläppy *et al.* 2010; Poppell *et al.* 2014). Correspondingly, LMA sponges have higher pumping rates and higher intake of particulate organic matter (Jiménez *et al.* 2007; Weisz *et al.* 2008; Schläppy *et al.* 2010; Freeman & Thacker 2011) compared to HMA sponges, which rely more on nutrients produced by their microbial symbionts (Freeman & Thacker 2011; Freeman, Thacker, *et al.* 2013; Freeman & Class Freeman

2014; Poppell *et al.* 2014).

The interior of larvae from many HMA sponges contain dense microbial assemblages, while larvae from LMA sponges are often found to be essentially sterile (Ereskovsky & Tokina 2004; Maldonado 2007; Schmitt, Weisz, *et al.* 2007; Gloeckner, Lindquist, *et al.* 2013). Vertical transmission is a key eco-evolutionary process for microbe-host interactions as it provides an essential mechanism by which increased levels of specialisation can evolve between microbes and hosts by preserving particular combinations of symbionts through host generations.

In Chapter I, I explore the specificity and temporal dynamics of microbe-sponge symbiosis. In microbe-host systems in particular, but also generally in ecological networks, interactions are often considered as static. This is an over-simplification that can lead to misleading results. I analyse a 12 month time-series consisting of the most abundant symbionts harboured by three sponge hosts from a local benthic community. I specifically ask, (i) How specific are the associations between the most abundant bacterial species and their host? (ii) How persistent are these associations over time? and (iii) Do host traits (HMA/LMA) determine the specificity and persistence of the associations? I find evidence for extreme microbe-host specialisation, with no symbionts shared between hosts. Each host assemblage contains unique genetic variation and phylogenetic structure. Interestingly, HMA hosts harbour related symbionts, while LMA hosts harbour taxonomically similar symbionts to those found in the surrounding seawater. Compared to LMA hosts and the surrounding seawater, HMA hosts harbour a larger fraction of persistent and abundant core symbionts whose relative abundance is more stable over time.

In Chapter II, I define and characterise the structure, dynamics and stability of the core microbiomes from several sponge host species using highly-resolved temporal series. I show that HMA hosts harbour species-richer, more abundant and temporally stable cores that likely ensure host functionality compared to core microbiomes associated to LMA hosts. Then, I ask what are the possible mechanisms for the stability of these cores. In doing so, I develop a theoretical model using time-series of microbial abundances to decompose the main sources of temporal variability. The model also allows for inferring and characterising the possible interaction network, and consequently quantifying the nature and strength of microbe-microbe interactions within each core microbiome. I hypothesise that the higher diversity of cores microbiomes harboured by HMA hosts confers a larger temporal stability compared to lower core diversity of LMA hosts. I further hypothesise that selection acts to minimise competitive interactions between symbionts, thereby simultaneously reinforcing selection for complementarity. I therefore do not expect to find strong competitive interactions within the core microbiomes, nor do I expect to find strong cooperative interactions, as

species coupling and positive feedback are likely destabilising. I analyse a 36 month time-series from highly resolved microbiomes from six sponge hosts from the same local benthic community as in Chapter I. I find that the majority of microbes in each microbiome are rare and opportunistic, but in line with the results from Chapter I, I find evidence for highly host-specific and temporally persistent core microbiomes with either a low or high density of symbionts. I find that high-density cores promote stability as hosts are less prone to invasion by occasional, likely antagonistic, microbes that can increase in abundance. I hypothesise that vertical transmission underpins the observed temporal stability of high-density cores, as the arrival order of species is known to heavily affect community assembly and stability (Chase 2010; Fukami 2015), and the process of vertical inheritance of symbionts will likely have similar outcomes as priority effects. The complementary group of symbionts that are transmitted from adult to offspring preempt the host niche by fast reaching carrying capacity. I found evidence suggesting that density-dependence stabilise high-density cores, while maintaining a higher level of temporal variability in the low-density cores.

Contrary to recent predictions, I find that, while interspecific interactions are almost negligible in LMA cores, HMA cores are characterised by weak, asymmetric interactions with similar degrees of commensalism and amensalism. Coyte and colleagues (2015) showed analytically that competition among microbes is predicted to maintain stability by counteracting the destabilising effect of high species diversity. While this might be the case when considering all the species within the microbiome, I provide evidence that this is not the case when only considering core microbiomes.

In Chapter III, I discern among key ecological processes driving variation in microbial cooccurrence across space as a function of important hosts features. In order understand the sources of variability, it is necessary to understand how a suite of processes related to host-specific features structure these communities. At the same time, to analyse such data requires a unifying, model-based framework capable of discerning among the various processes operating. Motivated by these challenges, I developed a Bayesian hierarchical joint species distribution model that analyses microbial cooccurrences across space as a function of several important host features. I specifically focus on discerning among the effects of host traits (HMA and LMA), host identity and host phylogenetic relatedness. I model data from microbial communities associated to 19 sponge species from 48 different locations worldwide. The models I build aim to simultaneously investigate a set of fundamental, but non-mutually exclusive questions of interest. I first ask three questions focusing on important host features: (i) Are microbial communities associated with hosts with similar traits more similar irrespective of their of geographic location? (ii) Are microbial communities associated with the same host species more similar regardless of geographic location? (iii) Are microbial communities associated with

phylogenetically closely related hosts more similar irrespective of geographic location? Following these three primary questions, I then asked two additional questions concerning the geographical variability present in the data, but which are not explicitly modelled: (iv) Do closely located host species harbour more similar microbial communities than hosts located farther apart? (iv) Is similarity among host-associated microbial communities distributed along a geographical and/or environmental gradients? I find that sponge traits (HMA and LMA) is the foremost important diver of microbial cooccurrences across hosts. I also find that many host species, irrespective of location and geographical distance form intraspecific clusters, suggesting vertical inheritance of symbionts. Perhaps surprisingly, I do not find any effect of host phylogenetic relatedness. In addition, I find that host-associated microbial communities are distributed along a gradient of increasing in dissimilarity. I hypothesise that this gradient reflects latitude and longitude, thus it importantly encapsulates different environmental gradients such as temperature and productivity.

Chapter IV represents a large collaborative effort part of the Earth Microbiome Project. The chapter explores specialisation in microbe-sponge interactions at an unprecedented resolution: a bipartite network consisting of 81 sponge species (from 804 samples) and close to 40.000 bacterial species from the waters of 20 countries bordering the Atlantic, Pacific and Indian Oceans as well as the Mediterranean and Red Seas. Variability of symbiont assemblages between individuals of the same host species is indicative of the nature and strength of host symbiont interactions. Low variability would indicate that only specific symbionts could interact with the host, while a relaxed pressure on the interaction would result in higher variability of symbionts among specimens of the same host species. The study finds that compared to planktonic assemblages, most hosts maintain a low variability within symbiont assemblages, indicative of selective interactions at the host species level. Moreover, the majority of symbionts are found to be specialists (i.e. only found in one or a few sponge species), while only a few are truly cosmopolitan (i.e. found across many sponge species). Interestingly, the study finds that symbiont assemblages are characterised by a combination of highly generalist and truly specialist symbionts. Generalists are cosmopolitan not only qualitatively (i.e. present in a large number of species), but also quantitatively (i.e. consistently present in a large fraction of individuals of those host species). Similarly to Chapter II, this chapter examines core structure by modeling population dynamics, but using a space-for-time substitution instead. Comparably to the results of Chapter II, the chapter finds that density-dependence is the main determinant promoting the observed high stability within hosts. Interestingly, interspecific interactions are almost negligible across modelled hosts, indicating that symbionts coexisting within a host do not influence each other.

Finally, in Chapter V, I focus on different facets of microbe-host specialisation by developing a new

sampling protocol that, for example, allows for quantifying vertical transmission from adults to offspring. The new sampling protocol naturally allows for specialisation to change across a series of networks with varying levels of microbe specificity: from networks considering all observed microbes and host, to microbes only shared between larvae and adult. In ecological networks, modules may represent coevolutionary units and in other large mutualistic networks, modules have been found to contain species with convergent morphological traits, suggesting that modularity is a consequence of coevolutionary relationships (Olesen, Bascompte, Dupont, *et al.* 2007; Jordano, García, *et al.* 2007). However, the large complexity of microbiomes requires additional approaches in order to reveal truly coevolutionary links. I hypothesise that the frequency of such links increases as the criteria of what to consider a microbe-host interaction, moving from less to more specialised networks. I show that the architecture of the analysed microbe-sponge interaction networks are characterised by extreme levels of specialisation, which likely represent the most specialised yet diverse ecological networks ever analysed.

Contents

Introduction

1.1 Interactions at the base of coevolution

Species are inextricably bound to each other, elaborately woven into what Darwin called the 'entangled bank' (Darwin, 1859). The vast majority of species have evolved in ways that require them to use not only their own genetic machinery, but also that of others. As such, most of evolution is coevolution–the process of reciprocal evolutionary change between two or more interacting species driven by natural selection–therefore interactions are not secondary attributes added onto species, but at the core of their evolution (Thompson 1994; Thompson 2005; Thompson 2013). Amongst the tropical plants producing flowers and fruits, for instance, the majority depend on animals for pollination and seed dispersal (Howe & Smallwood 1982; Willson *et al.* 1989; Bawa 1990).

The degree of coevolution between two species reflects the degree of specialisation. An extreme specialist, for example, relies on only a few related species for its survival and reproduction. For instance, the composite species lichens represent an intimate association between fungi, algae and/or cyanobacteria where the fungi depend on its photosynthetic partners for its survival (Honegger 1996). Such cases often represent old associations that are fine-tuned over long periods of evolutionary time. The oldest fossil of a lichen where both partners were recovered, dates back around 400 million years (Taylor, Remy, *et al.* 1995). Another example is the obligate association between aphids and their intracellular symbionts Buchnera. This association began about 200 million years ago, with host and symbiont coevolving in parallel ever since (Gil *et al.* 2002). Buchnera is vertically transmitted from parent to offspring (Douglas 1998), and as a result of the coevolutionary process, Buchnera has dramatically reduced the size of its genome only retaining a minimal set of genes necessary for its specialised lifestyle (Gil *et al.* 2002). As such, Buchneras' phylogenetic tree perfectly mirrors that of its aphid hosts.

However, the relationship between coevolution and specialisation is not necessarily symmetric. An apparent

specialist, for instance, may have coevolved with closely related and/or ecologically similar species. Interactions between species are constrained by the "phylogenetic baggage of structure, physiology, and behaviour that organisms inherit from their ancestors" (Thompson 1994), and species will therefore fit in where they can (Janzen 1980).The simian immunodeficiency virus (SIV), for example, coevoloved with monkeys long before it crossed to apes, and from apes to humans (Sharp & Hahn 2011).

As almost all species consist of genetically distinct populations, and most interacting species are patchily distributed across heterogeneous landscapes, species interactions depend on the spatiotemporal environment in which they occur. Pea aphids, for example, can, in addition to their obligate Buchnera symbiont, harbour up to five vertically transmitted facultative symbionts. The proportion of these symbionts greatly differs between populations across the aphid's geographical range (Tsuchida *et al.* 2002). The combination of symbionts depends on the plant species each aphid population uses as host together with the distribution of the insect parasitoids that infect them (Oliver *et al.* 2003).

1.2 Multispecies microbial symbioses

These symbioses involving just a few species are fascinating in their own right, but more diverse and complex host-associated microbial assemblages are increasingly found (Robinson *et al.* 2010; McFall-Ngai *et al.* 2013). These, so-called microbiomes often represent assemblages of hundreds to thousands of different microbes associated to a host species and/or distinct sub-habitats on a given host. As free-living microbial assemblages, microbiomes are shaped by a combination of different processes, such as selection, dispersal, diversification and drift (Nemergut *et al.* 2013). However, due to the intimacy of the relationship, the effect of coevolution and the "phylogenetic baggage" of each interacting partner is likely to play a significant role in shaping these complex interactions (Thompson 2005).

The human body, for example, are characterised by highly site-specific assemblages (Costello, Lauber, *et al.* 2009; Zhou *et al.* 2013; Oh *et al.* 2014). Interestingly, the gut microbiome can be predicted by factors such as age, culture and geography (Muegge *et al.* 2011; Yatsunenko *et al.* 2012), and across 59 mammalian lineages, humans grouped with other omnivores (Ley *et al.* 2008a). However, each mammalian host, including humans, harboured a proportion of individually unique microbes, suggesting that some members of the mammalian gut have codiversified with their respective host lineage (Ley *et al.* 2008a; Muegge *et al.* 2011; Delsuc *et al.* 2014). These microbes likely constitute a core that is crucial for the development, health and functioning of their hosts. Germ-free mice, and animals with experimentally induced dysbiosis, for instance, exhibit smaller body sizes and abnormal organ development,

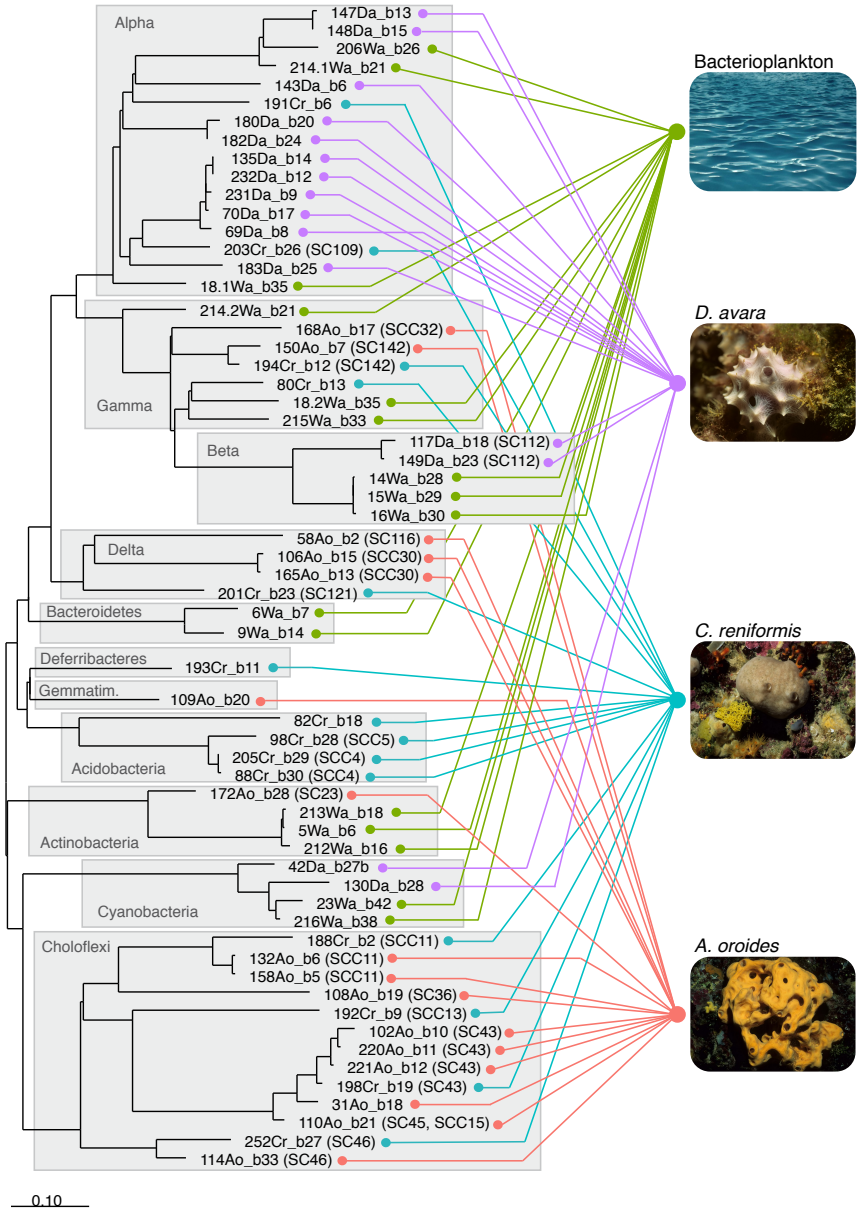


Figure 1.1: Microbe–host interaction network of three sponge hosts and their most abundant microbes (explored in detail in Chapter 2). Each sponge host is represented as a node to the right with connecting edges to nodes to the left, corresponding to tips in a phylogeny. Gray boxes correspond to different bacterial phyla. The abbreviation SC/SCC indicates if the microbe assign to sponge-/sponge-coral-specific clusters. Members of these clusters are closer related to each other than to microbes found elsewhere (Hentschel, Hopke, *et al.* 2002,Simister *et al.* 2012). Green nodes/edges correspond to microbes found in the plankton, purple to host *D. avara*, teal to host *C. reniformis*, and red to host *A. oroides*. There is an extreme specificity with no shared microbes across hosts and the bacterioplankton.

lower fecundity and longevity (see review Lee & Hase 2014). In addition, studies on humans have revealed a multitude of correlations between dysbiotic patterns and several aspects of human health, including autoimmune disorders (Round & Mazmanian 2009; Scher & Abramson 2011), diabetes (Qin *et al.* 2012), obesity (Ley, Turnbaugh, *et al.* 2006; Turnbaugh, Hamady, *et al.* 2009) and even psychiatric conditions (Foster & McVey Neufeld 2013).

Subsequent studies have, however, found little evidence for a common microbial core amongst healthy humans. The human microbiome is instead characterised by large variation between individuals and body habitats (Ley *et al.* 2008a; Dethlefsen *et al.* 2007; Arumugam *et al.* 2011; Bogaert *et al.* 2011; Huse *et al.* 2012). In fact, each individuals' microbiome is so individually unique that it can be used to identify one person among hundreds (Franzosa *et al.* 2015). Turnbaugh and Gordon (2009), nevertheless, found a common core in the human gut consisting of genes providing similar metabolic capacities.

1.3 Specialisation: Insights from complex networks

Much of the research on complex ecological networks focuses on patterns that emerge from interspecific interactions. Research on networks has a long tradition in ecology, with the first representation of a food web comprising who eats whom dating back to the end of 19th century (Camerano 1880). One particular aspect that has pervaded the literature since then is ecological specialisation: the type that occurs when a consumer specialises on a resource, and specifically reciprocal one-to-one specialisation, most famously illustrated by the Malagasy orchid and the hawk moth Darwin predicted would pollinate it (Darwin 1862). It was early noted that different consumers specialise on different resources, for instance, moths exploit different species of flowers than do bees and hummingbirds (Proctor *et al.* 2012). Conventional wisdom based on pairwise coevolution suggested that within these taxonomically related groups, reciprocal specialisation would be common.

Over the last decade, a major revolution has occurred in network ecology, in parallel to research on other complex biological and technological networks. Some universal patterns on the organisation of specialisation have been characterised. Among them, the degree distribution, i.e. the distribution of the number of links per species, is widely used to summarise the topology of ecological networks (Newman 2010). Early studies showed that the degree distribution of many ecological networks follow a power-law or a truncated power-law (Solé & Montoya 2001; Jordano, Bascompte, *et al.* 2003; Vazquez & Aizen 2003; Montoya *et al.* 2006), which are distributions heavily skewed towards many specialist and a few very generalist species. In other words, there are many species with few links and few species with many links. The skewed degree distribution of ecological networks compared to the poisson

distribution of random networks is generally assumed to be a result of ecological and/or evolutionary processes shaping species interactions (Jordano, Bascompte, *et al.* 2003; Bascompte & Jordano 2007). The degree distribution is tightly linked to the size and connectance of the network, that is, the number of realised interactions among all the possible ones.

These patterns of specialisation have seldom been explored in microbial systems. Exceptions include the marine environment (Steele *et al.* 2011), and a global microbial 16S rRNA gene database (Chaffron *et al.* 2010). While being clearly different from what would be expected by chance, interestingly, these networks are best described by the same skewed degree distributions found for other ecological networks (Solé & Montoya 2001; Jordano, Bascompte, *et al.* 2003; Vazquez & Aizen 2003; Montoya *et al.* 2006), suggesting that similar eco-evolutionary processes shape ecological specialisation in networks either comprising macro or microorganisms.

These findings of skewed degree distribution in ecological networks begged the question: how common is then reciprocal specialisation when we consider a large number of interacting species in a given ecosystem? A recent study found that reciprocal specialisation, such as the famous Darwin's orchid-moth case, is extremely rare when the overall network of interactions is examined (Joppa *et al.* 2009). The authors analysed reciprocal specialisation in a large set of mutualistic (insect pollinators and the plants they pollinate) and antagonistic (insect parasitoids and the insect hosts they attack) networks, and found no evidence for a preponderance of reciprocal specialisation in structuring these networks (Joppa *et al.* 2009). Instead, specialists tended to interact with generalists and vice versa, regardless if they looked at the consumer or resource level.

Another well-studied pattern in ecological networks is nestedness, that is, specialists tend to interact with one or a fewer of the same species generalists interact with (Bascompte, Jordano, *et al.* 2003). Taking the example of the Darwin's orchid-moth case: The Malagasy orchid with its deep corolla tube is dependent on the hawk moth's long proboscis for pollination, while on the other hand, although being specialised on the Malagasy orchid, the hawk moth can obtain nectar from all flowers with shallower tubes. In other words, the specialised interaction between the Malagasy orchid and the hawk moth is nested within a network of more promiscuous interactions. Opposite to nestedness lies modularity. Modularity is the pattern of non-overlapping subsets of species–modules–that strongly interact among themselves but only very weakly with species in other modules (Jordano 1987; Olesen, Bascompte, Dupont, *et al.* 2007; Dupont & Olesen 2009; Jordano 2010).

Both nestedness and modularity are signatures of coevolutionary mechanisms. Nestedness is commonly

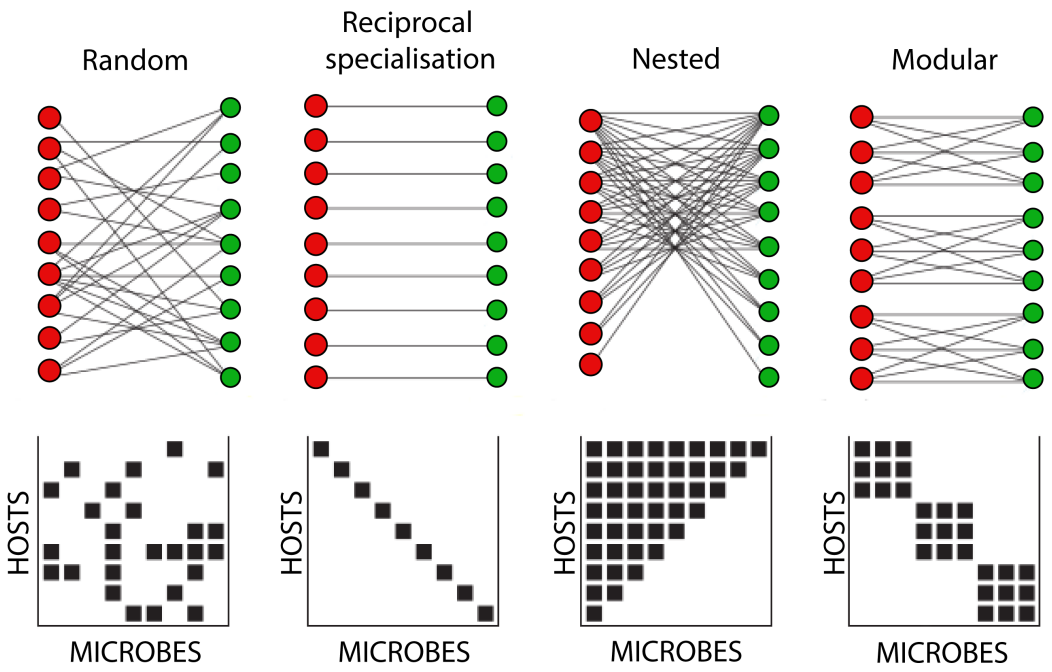


Figure 1.2: Four key patterns commonly explored in ecological networks. Red nodes correspond to hosts (i.e. resource level) whereas green nodes represent microbes (the consumer level). **Random networks:** the pattern of microbe-host interactions are not statistically different than what would be expected by chance. **Reciprocal specialisation:** Every microbe only interacts with one host and vice versa. **Nested networks:** interactions form a hierarchy where specialists interact with nodes generalists also interact with. A perfectly nested network is one in which both hosts and microbes can be ordered from 1 to S: in this ordering, host 1 interacts with the least number of microbes, e.g. due to phylogenetic constraints (i.e., only one of the S microbes interact with it), whereas host S is the most promiscuous host (i.e., all of the S microbes interact with it). The next most permissive host interact with all but one microbe, and so on. Similarly, in this ordering, microbe S is a generalist and can infect all S hosts whereas microbe 1 is a specialist and only interacts with one host. The next most specialised microbe interacts with the two most permissive hosts, and so on. **Modular networks:** contain interactions that occur among distinct groups of microbes and hosts. In the case of a maximally modular network with S hosts and S microbes, all interactions occur between microbes and hosts in the same group (i.e., 'module') rather than between groups. Figure adapted from Weitz *et al.* 2013.

observed in mutualistic networks among free-living organisms, but has also been reported in non-mutualistic systems. In phage-bacteria interaction networks, for example, nestedness emerges as a result of the hierarchical structure of infection ability among phages and resistance among hosts (Flores, Meyer, *et al.* 2011; Weitz *et al.* 2013). This structure itself is hypothesised to be a consequence of gene-for-gene coevolution (Lenski & Levin 1985; Agrawal & Lively 2003). Similarly, if microbial composition within hosts is solely a function of phylogenetic relatedness, nestedness would emerge. On the other hand, if certain microbes only associated with a unique host or a limited number of closely related hosts, a modular structure would be expected. As such, modules may represent coevolutionary units (Bascompte & Jordano 2014). Although a strictly modular network cannot be nested, and vice versa, modularity and nestedness can co-occur in some instances. For example, in phage-bacteria networks, nestedness can emerge within modules if the network consists of a large number of phages and hosts, each covering wide phylogenetic distances (Flores, Meyer, *et al.* 2011; Weitz *et al.* 2013).

1.4 Inferring species interactions

The study of ecological networks has provided insights into what determines ecosystem stability, and how stable systems can be disrupted by disturbances. The stability of ecological communities depends on their diversity, composition, and the patterning of interspecific interactions (Pimm 1991; McCann 2000). Of the many facets of stability (Pimm 1984; Donohue *et al.* 2013), in this thesis I am particularly interested in temporal variability, which largely depends on diversity and on the nature and strength of species interactions (Cottingham *et al.* 2001; Cardinale *et al.* 2002). While diversity and composition are commonly described for complex microbial communities, the pattern of interspecific interactions often remains elusive. Compared to interactions among macroscopic species, microbial interactions are for obvious reasons essentially impossible to observe outside a petri dish. Therefore, in order to decipher the nature and strength of microbial interactions, we need computational and/or statistical approaches.

Community dynamics, in particular how species abundances fluctuate over time, is dependent on a number of factors. These factors can be grouped into four major categories: responses to environmental fluctuations, demographic stochasticity, density-dependence, and interspecific interactions. Each of these components can be inferred by fitting a suitable population dynamic model to time-series of species' population abundances (Ives, B, *et al.* 2003; Ives & Carpenter 2007; Mutshinda *et al.* 2009; Mutshinda *et al.* 2011; Thibaut 2012). The most widely used approach to simulate population dynamics resulting from interspecific interactions is Lotka-Volterra type models, that directly considers the per capita effect of other species on the growth rate of a given species (Ives, B, *et al.* 2003; Ives &

Carpenter 2007; Mutshinda *et al.* 2009; Mutshinda *et al.* 2011; Thibaut 2012). While the classical Lotka-Volterra model assumes population logistic growth (Gotelli 1998), it is possible to incorporate other growth equations, such as the Gompertz function (Gompertz, Benjamin 1825). While the growth curve of the logistic equation is symmetric, Gomperz growth reaches the upper asymptote much more gradually than the lower asymptote. Interestingly, the exponential growth of microbes is often well-described by this particular function (McKellar & Lu 2003 and references within).

Mutshinda and colleagues developed a Bayesian framework that models population dynamics with a discrete-time stochastic Gompertz model including interspecific interactions of the Lotka-Volterra type, while simultaneously decomposing temporal fluctuations of species abundances into contributions from interspecific interactions, density-dependence and environmental stochasticity (Mutshinda *et al.* 2009). Modeling time-series from insects, fishes, macrocrustaceans, birds and rodents, they found that environmental stochasticity together with density-dependence are the major factors driving fluctuations in species abundances within the modelled communities. Using the same framework, Almaraz & Oro 2011 found, similarly, that environmental stochasticity accounted for most of the variability in a seabird community.

It is a common approach to infer microbial interactions from pairwise species abundance correlations (Arumugam *et al.* 2011; Barberan *et al.* 2012; Gilbert, Steele, *et al.* 2012). However, correlation does not always imply interaction (Fisher & Mehta 2014; Harris 2015). For example, if microbe i interacts with microbe j who in turn interacts with microbe k, the abundance of microbe i and k are likely to be correlated even if they do not directly interact. Another caveat is the fact that a correlation matrix between relative abundances is symmetric, hence the inferred interaction between microbe i and j will necessarily always be the same as between microbe j and i, if no time lag is introduced. One of the strengths of using a discrete-time Lotka-Volterra model is that it relates the abundance of species i at time $t-1$ to the abundances of all the species in the community at time t. Each interaction coefficient describing the per capita influence of species j on the growth rate of species i is encoded in the interaction matrix. Importantly, there is no easy linear relationship between interaction coefficients and species abundance correlations (Fisher & Mehta 2014).

Going a step further than pairwise abundance correlations, we may ask what type of interspecific interactions can we expect to take place within complex and diverse microbiomes. It is often assumed that the human microbiome, for instance, is characterised by mutualistic and cooperative interactions, both between host and microbes and among

microbes themselves (Backhed, Ley, *et al.* 2005; Dethlefsen *et al.* 2007; Schluter & Foster 2012). However, a recent mathematical analysis by Coyte *et al.* 2015 using classical ecological stability theory (May 1973; Allesina & Tang 2012) showed that cooperation between microbes tend to destabilise the microbiome as cooperation introduces coupling between species and positive feedback loops. Therefore, if one microbe decreases in abundance, it will tend to pull other species down with it and cause a community collapse. Applying their framework to the mouse gut microbiome, the authors found that a large proportion of interactions are non-cooperative and weak. In agreement, a study inferring interactions from time-series of the mouse gut microbiota using a Lotka-Volterra framework, found that the majority of the possible interactions between 17 microbes were competitive (Marino *et al.* 2014). However, as this study focused on the early colonisation phase of the gut microbiome, competitive interactions are expected (Parrish & Bazzaz 1982; Martorell & Freckleton 2014). Another study using the classical Lotka-Volterra model together with regression techniques revealed that commensal interactions within the mouse gut microbiome may likely protect against infections by pathogens (Stein *et al.* 2013). Tsai *et al.* 2015 used a rule-based network algorithm and found mostly cooperative interactions between six microbes present in the human infant gut microbiome. Trosvik & de Muinck 2015 modelled time-series using a dynamic regression model to infer interactions within the human gastrointestinal tract microbiome, and found that while competitive interactions were common among members of the same phylum, strong amensal and commensal interactions emerged between phyla. These studies illustrates the dynamic nature of microbial interactions. The sign and strength one microbe exerts on the other species in the assemblage are highly dynamic and contingent on the composition, densities and the environment (Ramsey *et al.* 2011). In order to understand how microbial interactions influences microbiome stability and functionality, it is crucial to model likely important species.

Recent developments in ecological network theory have shown that it is the coexistence of different interaction types (e.g. competition, mutualism, trophic interactions) and their relative strengths that determine the dynamics and stability of species interaction networks (Mougi & Kondoh 2012; Sauve *et al.* 2014; Lurgi, Montoya, *et al.* 2015). Therefore, to understand the unprecedented complexity of microbiomes, we need to first elucidate their underlying dynamics and resulting interactions. This will be a crucial step in order to bring microbiome research beyond describing patterns to determining processes that shape these complex systems.

1.5 Core species in a dynamical world

Microbial communities are often characterised by many orders of magnitude more species than those found in most macroscopic communities. It is common to reduce their complexity to focus on the 'important', key microbial groups, either by removing microbes below a certain abundance threshold and/or by grouping species into higher taxonomic ranks and/or functional groups. Lately, the quest for core microbiomes has received considerable attention (Turnbaugh, Ley, Hamady, *et al.* 2007; Tschöp *et al.* 2009; Turnbaugh, Ridaura, *et al.* 2009; Elli *et al.* 2010). Identifying a microbial core common to most healthy host individuals has important implications for predicting the overall microbiome response to perturbations (Shade & Handelsman 2012). However, while cores are commonly defined across individuals, few studies have explored temporally persistent cores, that is, microbes that continuously associate with their host over time (Faith *et al.* 2013). This may explain the apparent lack of core microbiomes among humans for example. In addition, adding a temporal dimension to the core concept offers insights into the dynamics, stability and associated functionality of microbiomes.

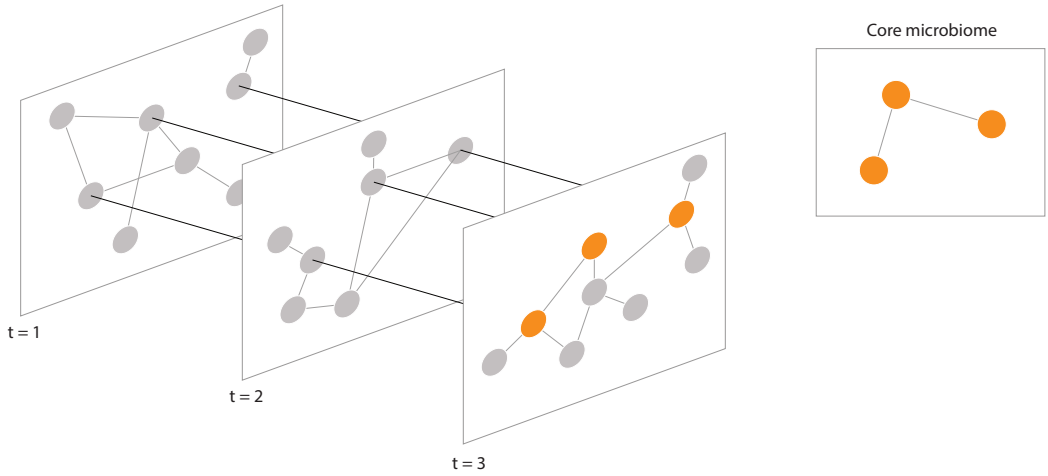


Figure 1.3: A conceptual representation of a temporal core microbiome. Each time slice is characterised by its own (static) interaction network, and it is only by incorporating a temporal dimension, we can discover species with a disproportionately large influence on the dynamics, stability and functionality of the microbiome as a whole. Similarly to modules identified in static networks, the core microbiome represents a group of microbes that interact more strongly (in terms of relative abundance) with the focal host than to other potential host species. When the core microbiome has been identified, their interspecific interactions can be inferred (see *Inferring species interactions* for more information).

The keystone species concept was introduced by ecologist Robert Paine (Paine 1969) to define those species at high trophic levels that exert a disproportionately large influence on species diversity and community dynamics. The

definition was however deemed very narrow as it only considered trophic interactions, while in fact species at lower trophic levels or other types of non-trophic relationships can be afforded keystone status (Lamont 1992). After heated debates, Mills and colleagues suggested ecologists to abandon the concept completely Mills *et al.* 1993. However, through new developments in network ecology, the concept has been revitalised. In network ecology, keystone species are those that serve as key connectors between the remaining species in a community, and whose removal can trigger secondary extinctions and loss of functionality (Montoya *et al.* 2006; Thompson *et al.* 2013).

Hanski proposed the core-satellite species hypothesis as a null model for metacommunity models Hanski 1982; Hanski 1991. The model explains the relationship between local species abundance and regional distribution. The distribution of species in a region is bimodal and falls into two groups: core and transient species. Core species reflect locally abundant and regionally common species, while transient species are rare and only occur at a few sites. Grime (1998) later made an attempt to link plant diversity and ecosystem functioning, by classifying species as dominants, subordinates and transients. In this classification, dominant species have a disproportionately large influence on ecosystem functioning. Similarly to the core-satellite hypothesis, the empirical species abundance distribution of a fish estuarine community sampled over a period of 21-years could be decomposed into two separate distributions (Magurran & Henderson 2003). Importantly, core species were persistent and abundant whereas occasional species were infrequent and rare, and while core species contributed little to species richness, their overall abundance accounted for more than 98 per cent in any given year (Magurran & Henderson 2003). Common species are essential for community structure and functioning (Gaston & Fuller 2008), as well as for dynamics and stability (Henderson & Magurran 2014). The loss of only a small proportion of these species can cause a significant reduction in the provision of several ecosystem services (Gaston & Fuller 2008).

1.6 Marine sponges: a new study system for microbiome and network ecology

Porifera is the phylogenetically oldest, still extant metazoan phyla with the hypothesised oldest bacteria-host symbiotic interactions (Taylor, Radax, *et al.* 2007). It is one of the taxa amongst invertebrates that harbours the largest diversity and abundance of microbes (Hentschel, Piel, *et al.* 2012; Webster, Luter, *et al.* 2012). Sponges are commonly classified into two groups depending on their microbial abundance–High Microbial Abundance (HMA) and Low Microbial Abundance (LMA) sponges. HMA sponges are morphologically adapted to harbour abundant microbial communities with their dense interior, narrower aquiferous canals and small choanoyctes (Vacelet & Donadey 1977). LMA sponges on the other hand, have an architecture more fitted for filter-feeding (Schläppy *et al.* 2010; Poppell

et al. 2014), and therefore have higher pumping rates and intake of particulate organic matter (Jiménez *et al.* 2007; Weisz *et al.* 2008; Schläppy *et al.* 2010; Freeman & Thacker 2011) compared to HMA sponges, which rely on nutrients produced by their symbionts (Freeman & Thacker 2011; Freeman, Thacker, *et al.* 2013; Freeman & Class Freeman 2014; Poppell *et al.* 2014). Although the HMA-LMA dichotomy likely represents a continuum, the vast majority of studied sponge species fit either one of the two classifications, suggesting two evolutionary stable strategies (Gloeckner, Wehrl, *et al.* 2014).

While the human microbiota is dominated by only a few phyla (Dethlefsen *et al.* 2007), sponges consistently harbour over 32 phyla and candidate phyla (Schmitt, Hentschel, *et al.* 2012; Reveillaud *et al.* 2014). Some of these microbes are vertically transmitted from parent to offspring, whereas others are selected from the environment (Sipkema *et al.* 2015). Interestingly, larvae from some LMA sponges are essentially sterile (Ereskovsky & Tokina 2004; Maldonado 2007; Schmitt, Weisz, *et al.* 2007; Gloeckner, Lindquist, *et al.* 2013), while the interior of some HMA sponge larvae contains dense microbial assemblages from up to 10 phyla (Schmitt, Angermeier, *et al.* 2008). Many of the microbes inhabiting sponges are more closely related each other than to microbes found elsewhere (Hentschel, Hopke, *et al.* 2002; Simister *et al.* 2012). These monophyletic clusters, so-called sponge-specific clusters span 14 bacterial and archaeal phyla, and even if present in the water column and sediment, they are only found at low abundances outside sponge hosts, as part of the rare biosphere (Taylor, Radax, *et al.* 2007; Taylor, Tsai, *et al.* 2013).

In vertebrates, the adaptive immune system efficiently detects and eliminates pathogens, but also controls which microbes to allow for symbiosis in order to attain, for example, suitable metabolic functions. The innate immune defence system of some sponges can differentiate between pathogens, food bacteria and symbionts in a manner similar to the adaptive immune system of vertebrates (Wilkinson *et al.* 1984; Wehrl *et al.* 2007; Wiens *et al.* 2007; Thomas *et al.* 2010; Yuen *et al.* 2014; Degnan 2015). In invertebrates, one of the best characterised symbiont recognition mechanism is found in the bobtail squid: it selects symbionts for the colonisation of its light organs directly from the surrounding seawater (Nyholm & McFall-Ngai 2004). Similarly, in hydrothermal deep-sea vents, invertebrates select symbionts present in the environment in order to utilise the high concentrations of sulphur, hydrogen and/or carbon dioxide (Dubilier *et al.* 2008).

Similarly, microbe-sponges symbioses are known to be at least partially underpinned by metabolic exchange between symbiont and host, including nitrogen cycling, fixation of carbon dioxide, secondary metabolite production, and uptake and conversion of dissolved organic matter (Taylor, Radax, *et al.* 2007; Fan *et al.* 2012; Goeij *et al.*

Figure 1.4: The mediterranean coralligenous, a low-turnover hard bottom community of biogenic origin produced by the accumulation of calcareous encrusting algae in dim-light conditions. The ecological relevance of this emblematic community is due to its high diversity (up to 1666 species, Ballesteros 2006)), which is a consequence of the structural complexity provided by engineering species such as sponges. Within the 20×20 cm transect shown in the photo, six of the sponge species studied in this thesis coexist. These sponges essentially "breath" the same water- why then is their microbial community so different? The photo is taken at the dive site close to Islas Medas marine reserve in the NW Mediterranean Sea. Photo taken by Dr. Rafel Coma.

2013). These metabolic processes cannot be performed by the sponge host alone, and are therefore likely crucial for host fitness. In this respect, sponge symbionts perform analogous functions to the symbionts found in mammalian guts and plants (Hacquard *et al.* 2015). Importantly, sponges can maintain highly diverse, yet specific symbiont communities, despite the constant influx of microbes from the surrounding seawater resulting from their filter-feeding activities (Taylor, Tsai, *et al.* 2013). Therefore microbe-sponge symbioses represent an ecologically relevant example of microbe-host interactions in an early-diverging metazoan clade.

Previous studies have indicated low intra-individual variation (Lee, Wong, *et al.* 2009; Webster, Luter, *et al.* 2012) and spatial stability (Hentschel, Hopke, *et al.* 2002; Taylor, Schupp, Dahllof, *et al.* 2004; Taylor, Schupp, De Nys, *et al.* 2005; Pita *et al.* 2013) of sponge-associated microbes. Variability of microbial communities between individuals of the same host species is indicative of the nature and strength of host-microbe interactions (Thacker & Freeman 2012). Low variability would indicate that only specific microbes could interact with the host, while a relaxed pressure on the interaction would result in higher variability of microbes among individuals of the same host species.

1.7 Main findings in a general context

In Chapters I and II, I corroborate earlier findings by showing that sponge-associated microbial communities are characterised by high levels of host-species specificity (Taylor, Schupp, Dahllof, *et al.* 2004; Erwin, L??pez-Legentil, *et al.* 2012; Pita *et al.* 2013). As most previous work only considers a single time-point or aggregated communities, by analysing a 12 month time-series of the most abundant microbes associated to three host species and the seawater, I extend earlier work by showing that some of these highly host-specific microbes are also persistently occurring over time. I show that HMA hosts harbour more temporally stable communities compared to those found in LMA hosts and the water column. In Chapter II, I further show that microbiome temporal stability is underpinned by the presence of high-density cores, that is, temporally persistent microbes accounting for the majority of microbiome relative abundance. As sponges are constantly invaded by microbes from the water column, high-density cores confer host a resistance against occasional microbes to increase in abundance. I suggest that the main mechanism underlying the observed temporal stability of high-density cores is vertical transmission through priority effects, further maintained by both density-dependence and weak asymmetric interactions among microbes in the form of commensalism and amensalism. The arrival order of species affects community assembly and stability (Chase 2010; Fukami 2015), and the process of vertical inheritance of symbionts will likely have similar outcomes as priority effects. In addition, strong density-dependence likely determines the dynamics of abundant and temporally stable assemblages (Henderson & Magurran 2014). However, while density-dependence can further increase stability in communities close to carrying capacity by dampening fluctuations caused by environmental stochasticity, it can also destabilise communities by facilitating self-sustained population cycles. As I observe high levels of density-dependence in both high- and low-density cores, I suggest that density-dependence may cause increased stability in high-density cores, while maintaining higher levels of temporal variation in low-density cores.

In Chapter II, I model microbe-microbe interactions within each core using a discrete-time Lotka-Volterra model. A recent mathematical analysis showed that competition among gut microbes is predicted to maintain stability by counteracting the destabilising effect of high species diversity (Coyte *et al.* 2015). While this might be the case for the whole microbiome, I show that this is not the case for the core microbiome. Instead, I show that microbial interactions within each core are infrequent, weak and asymmetric in the form of commensalism and amensalism, likely representing a blueprint of stability and a fingerprint of coevolved interactions.

In Chapter III, I develop a novel Bayesian hierarchical joint species distribution modelling framework that

analyses microbial cooccurrences across sites as a function of several important host features. For the first time in species distribution modelling, I decompose variation in microbial cooccurrences into host species trait and phylogenetic information. Host traits and phylogenetic relatedness are key factors particularly important in order to disentangle the effects of environmental filtering and biotic interactions, including those between microbes and hosts. While classical species distribution models correlate species occurrences with environmental variables, such as climate and topography, by fitting each species independently and predicting its distribution (Austin 2002), joint species distribution modelling fits multiple species simultaneously in order to assess the shared environmental response separately from other abiotic and/or biotic processes that give rise to the observed cooccurrence patterns (Clark *et al.* 2014; Pollock *et al.* 2014).

A key aspect of the models I propose in Chapter III is the inclusion of latent variables, serving a number of important purposes, including as means of performing model-based unconstrained ordination, and accounting for 'hidden' residual covariation in microbial cooccurrences not explained by the included host features (Warton, Blanchet, *et al.* 2015; Hui, Warton, *et al.* 2016).

I find that sponge classification (HMA and LMA) is the foremost important driver of microbial cooccurrences across hosts. I also find that many host species, irrespective of location and geographical distance form intraspecific clusters, suggesting vertical inheritance of symbionts. Perhaps surprisingly, I do not find an effect of host phylogenetic relatedness. In addition, I find that host-associated microbial communities are distributed along a gradient of increasing dissimilarity. I hypothesise that this gradient reflects latitude and longitude, thus importantly encapsulates different environmental gradients such as temperature and productivity.

As part of the Earth Microbiome Project, Chapter IV explores specialisation in microbe-sponge interactions at an unprecedented resolution: a bipartite network consisting of 81 host species (804 samples) and close to 40.000 bacterial species from the waters of 20 countries bordering the Atlantic, Pacific and Indian Oceans as well as the Mediterranean and Red Seas. Chapter IV corroborates earlier work by showing that compared to the microbial community found in the plankton, most hosts maintain a very low intraspecific variability, indicative of selective interactions at the host species level (Lee, Wong, *et al.* 2009; Webster, Luter, *et al.* 2012). Chapter IV further shows that similarly to studies on mutualistic networks, the structure of the analysed network differs greatly from what would be expected if connections between microbes and hosts were randomly assigned (Solé & Montoya 2001; Jordano, Bascompte, *et al.* 2003; Vazquez & Aizen 2003; Montoya *et al.* 2006), indicating that non-random eco-evolutionary

processes are shaping these interactions. This further shows that the majority of microbes are specialists (i.e., only found in one or a few sponge species), while only a few are truly cosmopolitan (i.e., found across many sponge species). Generalists are cosmopolitan not only qualitatively (i.e., present in a large number of species), but also quantitatively (i.e., consistently present in a large fraction of individuals of those host species). These patterns have not previously been observed for ecological networks, as it has traditionally been difficult to undertake repeated measures of many individuals across multiple host species. This chapter extends the model I developed in Chapter II, also modeling population dynamics, but with a space-for-time substitution. In agreement with the findings of Chapter II, this chapter shows that density-dependence is the main determinant promoting the low intraspecific variability across hosts, while interspecific interactions were found to have an almost negligible effect across the modelled hosts.

In Chapter V, I develop a novel sampling protocol that allows for the construction of a series of bipartite microbe-host networks that naturally increase in ecological specialisation. I show that these networks are among the most specialised, yet diverse mutualistic networks ever analysed. This extreme specialisation results from high values of modularity, with each module corresponding to a host species in the network. In ecological networks, modules may represent coevolutionary units and in other large mutualistic networks, modules have been found to contain species with convergent morphological traits, suggesting that modularity is a consequence of coevolutionary relationships (Olesen, Bascompte, Dupont, *et al.* 2007; Jordano 2010; Bascompte & Jordano 2014). However, the large complexity of microbiomes requires additional approaches in order to reveal such interactions. I show that the developed methodology successfully isolate links likely representing truly coevolutionary relationships. In addition, I find patterns coherent with a process of 'selective enrichment', where the host actively differentiate between symbionts and invasive microbes, allowing some symbionts to increase in density.

Specificity and temporal dynamics of complex microbe-sponge interactions

2.1 Introduction

Studies on multispecies interaction networks, in particular food webs and mutualistic networks of free-living species (e.g. plants and the insects that pollinate them), have revealed ubiquitous structural patterns of species interactions (Montoya *et al.* 2006; Ings *et al.* 2009; Bascompte 2009). Some of these patterns have challenged prevailing wisdom based on the study of isolated pairwise interactions. Reciprocal one-to-one specialisation, famously illustrated by the Madagascar star orchid and the hawk moth that pollinates it, for example, is extremely rare when the overall plant-insect pollination network is examined (Joppa *et al.* 2009). These interaction patterns help to understand the ecological and evolutionary processes that shape multispecies communities, and have strong implications for ecosystem stability, resilience and functioning (Montoya *et al.* 2006; Dunne 2006; Bascompte & Jordano 2007).

What is commonly missed in the study of species interaction networks is the link between eukaryotes and prokaryotes. Although generally agreed that prokaryotes are extremely abundant, diverse and present in all environments, that they constitute the majority of the branches in the tree of life, and that they regulate ecosystem functioning (Pedros-Alio 2006; Fuhrman *et al.* 2008), they are commonly neglected in species interaction network studies. Microbes, in general, and prokaryotes, in particular, are usually considered as a black box that interacts with almost any other species within a given ecosystem (Ings *et al.* 2009).

In parallel, and disconnected from species interaction network studies, microbe-host symbiotic associations have received considerable attention. There exists an enormous diversity of microbe associations with larger organisms (Moya *et al.* 2008), demonstrating a continuum of association strengths and types, which are often of great importance for the development, health, and functioning of the host (Robinson *et al.* 2010). Host-microbe studies are usually performed in relatively species-poor microbial systems usually characterised by extreme one-to-one or two-to-one

reciprocal specialisations, illustrated by the intimate associations between, e.g. Buchnera-aphids, Rhizobia-legumes, and Wolbachia-arthropods (see Ruby 2008 for a review).

One major challenge is to study multiple microbes associated with multiple hosts in more diverse systems. The associated microbiota to different parts of the human body is an example. It has shown evidence for an enormous complexity, host specificity, and temporal stability of the many dense and diverse microbial communities inhabiting the human biome (Parfrey & Knight 2012).

Here we investigate the specificity and temporal dynamics of multiple bacteria associated with different host species within a particular habitat. Our host species are marine sponges, *Porifera*, which represent the phylogenetically oldest, still extant metazoan phyla, with the hypothesised oldest bacteria-host symbiotic interactions (Taylor, Radax, *et al.* 2007). Moreover, it is the group amongst invertebrates having the most abundant and diverse microbial community (Webster & Taylor 2012; Hentschel, Piel, *et al.* 2012).

Sponges are commonly divided into two groups that result from the combination of a number of different ecological and evolutionary characteristics. Some [High Microbial Abundance] sponges harbour extremely abundant and diverse microbial communities, including species from nearly all major prokaryotic linages, while other [Low Microbial Abundance] sponges only host a few (Hentschel, Hopke, *et al.* 2002). The observed difference in microbial community structure and complexity is hypothesised to reflect the host-environment [morphology and physiology of the host], the former having a denser and more intricate structure than the latter. Although both types can form symbiotic relationships with microbes (Thacker & Starnes 2003; Schläppy *et al.* 2010), communities associated with low abundant species are usually very similar to those found in the ambient bacterioplankton, indicating horizontally acquisition of microbes, while communities associated with high abundant species are very different (Webster & Taylor 2012; Hentschel, Piel, *et al.* 2012). Moreover, the larvae from high abundant species have been shown to harbour a subset of the microbial community of the adult with both vertical and horizontal transmission shown to be important in the maintenance of symbiotic associations (Schmitt, Angermeier, *et al.* 2008; Lee, Chiu, *et al.* 2009; Webster, Taylor, *et al.* 2010). These two host strategies can be viewed as two different eco-evolutionary responses to their association with different symbiotic microbes. High Microbial Abundance sponges tend to depend more on nutrients produced by their associated microbes, while Low Microbial Abundance sponges tend to be highly dependent on nutrients uptaken from the water column. This is reflected by differences in water pumping rates and sponge tissue density: High Microbial Abundance sponges have lower pumping rates and a substantially denser tissue than Low

Microbial Abundance hosts (Weisz *et al.* 2008; Ribes *et al.* 2012). These differences, together with other mechanisms e.g. host defence and symbiont recognition, may suggest of two different evolutionary trajectories that have resulted in profound morphological and physiological differences.

Here we adopt an intermediate complexity approach to study microbe-host interaction networks. This means we do neither focus on a single microbe-host associations nor consider the entire 'rare biosphere' (i.e. bacteria at very low abundances) that have been described in these systems (Hentschel, Piel, *et al.* 2012). We study the most abundant members of the bacterial community associated to three Mediterranean sponges, and the free-living bacterioplankton (i.e. bacteria living in surrounding seawater not associated to any host) they filter to feed. In particular, we ask three questions: (1) How specific are the most abundant bacterial species to the host they are associated within a given habitat? (2) How persistent are host-bacteria associations over time? and (3) Does host eco-evolutionary characteristics, i.e. High- and Low Microbial Abundance sponges determine bacterial specificity and persistence of the associations?

2.2 Methods

Bacteria-sponge interactions

Our study system is the Mediterranean coralligenous, a low-turnover hard bottom community of biogenic origin produced by the accumulation of calcareous encrusting algae in dim-light conditions. The ecological relevance of this emblematic community is due to its high diversity (up to 1666 species, Ballesteros 2006)), which is a consequence of the structural complexity provided by engineering species such as calcareous algae and sponges. Sponges are ancient metazoans abundant in most hard bottom substrata around the world. They are potential key species in benthic-pelagic coupling due to their ability to continuously clear large water volumes, particularly from picoplankton (Gili & Coma 1998).

We collected our samples close to Islas Medas, in the north-western Mediterranean Sea $42°3'0''N, 3°13'0''E$. Specimens of sponges Agelas oroides, Chondrosia reniformis and Dysidea avara were collected monthly for a year (March 2009 – February 2012) by scuba-diving (depths 5-10 metres). These sponges represent a subset of the most abundant species in this sub-littoral ecosystem. The two former are categorised as High Microbial Abundance (HMA) species, while the latter is categorised as a Low Microbial Abundance (LMA) species (Vacelet & Donadey 1977; Schläppy *et al.* 2010).

During dives, specimens were placed in separate plastic bottles and brought to the surface where we care-

fully transferred them to jars containing 2 L of filtered seawater (0.22 μm). In parallel, but independent from specimen collection, samples of the ambient seawater at 5 metre depth were taken to analyse free-living bacterioplankton. Samples were transported in an insulated cooler to the laboratory (¡ 2h) where specimens were cut into small pieces and frozen at -80°C until DNA extraction. Triplicate aliquots of seawater (300–500 ml each) were filtered through 0.2 μm using polycarbonate filters, which subsequently were submerged in lysis buffer and frozen at -80°C.

The same individual was not sampled twice. This was done to avoid observing an effect in bacterial community composition caused by a decline in host condition. Stressed induced individuals have been shown to change their community composition (Mohamed *et al.* 2008; Webster, Cobb, *et al.* 2011). In addition intra-specific variation in host bacterial communities associated to sponges is generally small (Taylor, Radax, *et al.* 2007; Schmitt, Hentschel, *et al.* 2012).

Sequence and data analysis

Tissue samples from each sponge specimen were dissected into small sections using a sterile scalpel. Total genomic DNA was extracted from sponges and from filters using a commercial kit (DNeasy Blood and Tissue Kit, Qiagen). Purified DNA was quantified using spectrophotometry (Nanodrop, Thermo Scientific) and subsequently used as a template in the PCR. General 16S rDNA bacterial primers, GC-358F and 907RM, were used (Muyzer & Smalla 1998. Denaturing Gradient Gel Electrophoresis (DGGE) separate the PCR amplicons ($\approx$ 586 base pairs), generating band fingerprints mirroring the most abundant members of the bacterial community (Muyzer & Smalla 1998; Fromin *et al.* 2002). An individual discrete band corresponds to a unique sequence type or phylotype (hereafter species), which are treated as a discrete bacterial population. The abundance of the most abundant subset of each bacterial population was estimated using band image analysis. Band intensity has been shown to be correlated to the relative abundance of the corresponding bacterial species (Fromin *et al.* 2002). Although, used before (Iwamoto *et al.* 2000; Schauer *et al.* 2003; Pascoal *et al.* 2009), this approach of infering relative abundance is debated since it has been shown to produce biased estimates (W. *et al.* 2013). This is why we use it here as a mere proxy for relative abundance. Moreover, we use it as a proxy for comparisons among samples that might be affected by the same bias. Bands within the same position in the gel in different months correspond to identical species. This allows for assessing variation on their relative abundances over time. A subset from the total number of bands, the most representative (i.e. the most intense) were excised and re-amplified with the original primer set (without the GC-clamp). The products were submitted for automated DNA sequencing (Macrogen Inc., Amsterdam, Netherlands) for direct sequencing (Figure A.1, Table A.2).

DGGE images were analysed using the Quantity-One software package (Bio-Rad). The contribution of individual bands was calculated relative to that of the total absorbance of each lane. Sequences were aligned against the SILVA reference database (Pruesse et al. 2007) using the kmer search option and the Needleman-Wunch algorithm. The alignment was filtered, using the default settings, to remove non-informative gaps introduced in the previous step. This was performed using the Mothur package (Schloss *et al.* 2009). Gene sequences were deposited in Genbank under accession numbers KC200485-KC200545.

Specificity and temporal dynamics of bacteria-host interactions

To assess different dimensions of specificity and temporal dynamics of bacteria-host associations in different hosts, we used two different temporal scales. First, we used monthly-resolved host-bacteria associations. This resulted in 12 different bacterial community structures for each host and for the bacterioplankton. We called these monthly-resolved communities. Second, we pooled all the bacterial species present over the year in each host and in the bacterioplankton. This resulted in one community structure for each host, hereafter called the aggregated community.

At both temporal scales, we performed a number of analyses to determine differences in the level of host-bacteria specificity between different hosts and the bacterioplankton. To determine temporal dynamics we used monthly-resolved communities.

Community similarity

First, we aimed at comparing how similar bacterial communities are among them in terms of the presence/absence of bacterial species. The similarity of community membership between hosts i and j is defined by the Jaccard similarity index J_{ij}^{α}

$$J_{ij}^{\alpha} = \frac{A_{ij}^{\alpha}}{A_{ij}^{\alpha} + B_{ij}^{\alpha} + C_{ij}^{\alpha}} \tag{2.1}$$

where A is the total number of sequences present in both hosts, B is the number of sequences present in i but not in j, and C is the number of sequences present in j but not in i, and α is the distance cut-off to consider two sequences as the same. By changing α it is possible to compare community similarity at different levels of taxonomic resolution. This is important if we aim at determining whether specificity occurs at fine or broad bacterial taxonomic resolutions. Here we use individual non-overlapping DGGE bands as a proxy for species level, and an α of 5% and 20% to indicate genus and phylum level, respectively.

Sponge-specific bacterial clusters

Second, we aimed at determining whether our bacterial species were specific of sponge hosts, and not widespread distributed among other hosts or free-living environments. To do so, we compared our sequenced bacterial species with previously sponge-derived bacterial species to determine whether these were specific to sponge hosts. This was done by assigning our sequences to already defined sponge-specific monophyletic clusters by incorporating sequences into existing phylogeny using the quick-add parsimony tool as implemented in the ARB software package (Ludwig 2004). A sponge-specific cluster is a sponge-derived group of at least three 16S rRNA gene sequences, which *(i)* are more similar to each other than to sequences from other, non-sponge sources; *(ii)* are found in at least two host sponge species and/or in the same host species but from different geographic locations; and *(iii)* cluster together irrespective of the phylogeny inference method used (Hentschel, Hopke, *et al.* 2002; Simister *et al.* 2012).

Bacterial genetic diversity

Third, we tested whether different hosts have different bacterial genetic diversity associated to them. This provides a finer resolution by introducing genetic information into how similar/different host-specific bacteria are. We did so for both aggregated and monthly-resolved communities, testing whether bacterial genetic diversity is different from that which would result from pooling the diversity from any 2 host species. We used Analysis of MOlecular VAriance (AMOVA) (Excoffier et al. 1992). This method is widely used in population genetics, but can be extended to community ecology (Schloss 2008). It is a nonparametric analogue of ANOVA, based on a distance matrix of sequences retrieved from different communities. We used the uncorrected p-distance, which calculates the distance between pairs of sequences as the proportion of unique nucleotide positions. This distance does not correct for multiple nucleotide substitutions and should therefore be regarded as a raw distance of sequence divergence. We performed AMOVA with 100.000 randomisations using Mothur (Schloss *et al.* 2009).

Bacterial phylogenetic diversity and structure in each host

Fourth, we included phylogenetic information about bacteria to calculate bacterial phylogenetic diversity and structure to get an even finer resolution into how unique and phylogenetically diverse bacterial communities are among hosts and over time. Two phylogenies (aggregated and monthly-resolved) were estimated, using RAxML (v.7.3.5), by maximum-likelihood interference and GTR+CAT approximation with 1.000 bootstraps (Stamatakis 2006). The two trees were rooted by including a sponge associated Archaeal 16S rRNA gene sequence (accession number: EF529650)

as an outgroup. We calculated phylogenetic structure using the unweighted UniFrac distance U (Lozupone & Knight 2005). The UniFrac distance U is a phylogenetic extension of the Jaccard index, attempting to capture the amount of evolution unique to one community compared to another. It does so by calculating the fraction of branch length between two communities in the phylogeny, leading exclusively to one or the other, but not to both, as

$$U = \frac{\sum_{i=1}^{N} l_i |A_i - B_i|}{\sum_{i=1}^{N} max(A_i, B_i)} \tag{2.2}$$

where N is the number of nodes in the phylogenetic tree consisting of all bacterial sequences across hosts and months, l_i is the branch length between node i and its parent, and A_i and B_i are indicators equal to 0 or 1 as descendants of node i are absent or present in communities A and B, respectively. This method tests the hypothesis that lineages from two or more communities are undergoing equal rates of evolution. We performed the unweighted UniFrac with 10.000 randomisations using Mothur (Schloss et al. 2009).

To calculate phylogenetic diversity across hosts and months we used Faith's phylogenetic diversity index (Faith 1992). Faith's PD calculates the diversity in a community by summing the lengths of all those branches that are members of the corresponding minimum spanning path connecting species of the same community. It was calculated with 10.000 randomisations using Mothur (Schloss *et al.* 2009).

Persistence of bacteria-host associations

Finally, we used 3 measures of temporal persistence for the bacterial communities associated with each host and the bacterioplankton. The first was calculated by counting the number of months each bacterial species was associated to a particular host. Second, we measured the coefficient of variation (CV) of each DGGE band intensity over time associated with each host. Band intensity is correlated to the relative abundance of the corresponding bacterial species (Fromin *et al.* 2002). We also calculated the CV of the phylogenetic diversity (PD) of the bacterial community over time for each host and the bacterioplankton.

Bacteria-bacteria association networks

We applied Local Similarity Analysis (LSA) (Ruan *et al.* 2006) on the monthly resolved data set to explore contemporaneous and time-lagged correlations between bacterial relative abundances within and between hosts and the bacterioplankton. LSA is a complement to ordinary correlation analysis since it has the ability of finding potentially time-delayed correlations between two species varying over a time-series, which cannot otherwise be identified. To

account for any seasonal effects on our dataset, we chose a time-lag limit for detecting correlations of three months. This technique provides a score for each pairwise correlation, with an associated p-value (see (Ruan *et al.* 2006) for full details). To avoid the risk of false discovery rate (FDR) the software implements q-values to adjust for multiple testing. We chose a stringency level of 0.05 for both the P- and the q-values. The analysis was permuted 1.000 times.

2.3 Results

Table 2.1: Bacterial genetic differentiation in terms of raw 16S rRNA sequence dissimilarity analysed using AMOVA and bacterial phylogenetic composition analysed using UniFrac. For each analysis, pairwise comparisons are made between hosts and the bacterioplankton for the monthly resolved and aggregated communities.

	Monthly resolved communities					Aggregated communities				
	AMOVA			UniFrac distance		AMOVA			UniFrac distance	
Comparison	Fs	df	P	U	P	Fs	df	P	U	P
Ao-Cr	18.94	1,320	<0.0001†	0.723	<0.0001†	1.38	1,28	0.14743	0.745	0.3374
Ao-Da	44.47	1,298	<0.0001†	0.916	<0.0001†	5.28	1,29	0.0001	0.905	0.0012
Ao-Bp	36.54	1,323	<0.0001†	0.879	<0.0001†	3.76	1,30	0.0001	0.878	0.0016
Cr-Da	19.95	1,250	<0.0001†	0.902	<0.0001†	2.57	1,27	0.00762	0.894	0.0013
Cr-Bp	25.25	1,275	<0.0001†	0.906	<0.0001†	2.93	1,28	0.00039	0.918	0.0004
Da-Bp	14.98	1,253	<0.0001†	0.678	<0.0001†	2.39	1,29	0.02854	0.697	0.0126

Notes: Ao corresponds to *A. oroides*, Cr to *C. reniformis*, Da to *D. avara*, and Bp to bacterioplankton. Each P value is presented with its corresponding statistic (i.e., Fs and U). † Significant difference between hosts after multiple testing using Bonferoni corrections (P <0.0083)

DGGE analysis yielded a total of 149 bands (each representing a bacterial species) throughout the year, whereof 35, 30, 33 and 51 were associated with host *A. oroides*, *C. reniformis*, *D. avara* and the bacterioplankton, respectively. Of these bands, 61 were of good enough quality to sequence and of those, 16, 14, 15 and 16 were associated with host *A. oroides*, *C. reniformis*, *D. avara* and the bacterioplankton, respectively. Each sequenced band (i.e. species) correspond to the most abundant subset of the bacterial community within each host.

Intraspecific variability within a month was very low for host-associated bacterial communities and the bacterioplankton, as reported earlier for different sponge species (Mohamed et al. 2008, Webster et al. 2011). For the bacterioplankton, replicates of the same month clustered together in a dendrogram using Jaccard similarity index (A.2). For host-associated communities we could not perform cluster analysis using dendrograms because monthly replicates were done in a different gel than temporal dynamics. However, for example, for *A. oroides*, the mean CV of band intensities (i.e. relative abundance) within monthly replicates was significantly lower (0.36) than the mean CV of band intensities across the time series (0.68) (ANOVA: $F_{1,66}$=4.051, P=0.048). Importantly, although some small

qualitative (presence/absence) differences among individuals within a month were observed, the relative abundance of the species responsible for these differences was very small. In all cases, these species lay within the 33% of the least abundant species, suggesting low intraspecific variability across hosts and the bacterioplankton.

Specificity and persistence of bacteria-host interactions

We did not find any overlap on bacterial species composition between hosts and the bacterioplankton, indicating an extreme reciprocal specialisation at the bacterial species level (individual DGGE bands). This was consistent at both temporal scales considered, i.e., aggregated (Table 2.1, Figure 2.1) and monthly resolved (Table 2.1). When assigning sequences to genus level (i.e. by grouping sequences at a 5% distance) in the aggregated community, hosts *A. oroides* and *C. reniformis* shared 4% of sequences and *D. avara* shared 13% with the bacterioplankton. Increasing the level of taxonomic aggregation (e.g. at the phylum level) did not change the results qualitatively (A.1). However, monthly-resolved communities only showed overlap for distance cut-offs above phylum level, with bacterial communities across months showing a very high similarity within each host (Table 2.1, Figure 2.2).

In addition, we found that the aggregated community associated with the two HMA hosts (i.e. *A. oroides* and *C. reniformis*) could more frequently be assigned to pre-defined sponge-specific monophyletic clusters (Simister et al. 2012), than could the aggregated community associated with the LMA host and the bacterioplankton (Pearson chi-square test: χ^2_3 =38.17, N_{Ao}=14, N_{Cr}=11, N_{Da}=2, N_{Bp}=0, P<0.001) (A.2). By definition, bacterial species assigned to these clusters are more closely related to each other than to bacterial species from non-sponge sources (Hentschel et al. 2002).

Similarly, bacterial genetic differentiation was very different between hosts and between hosts and bacterioplankton, when monthly-resolved communities were analysed (Table 2.1). Only the aggregated community of both HMA hosts (AMOVA F_{pseudo} =1,38, P=0,147) and the LMA host and the bacterioplankton (AMOVA F_{pseudo} =2.387, P=0.028) showed similar levels of bacterial genetic differentiation.

Results from the phylogenetic community structure point in the same direction. The UniFrac analysis of the aggregated community showed a strong similarity between bacterial communities associated with the HMA hosts and between bacterial communities associated with the LMA host and the bacterioplankton (Table 2.1). The monthly-resolved communities showed higher similarities within than between communities associated with hosts and the bacterioplankton. Figure 2.2 illustrates this, with bacterial communities clustering by host identity, with no overlap

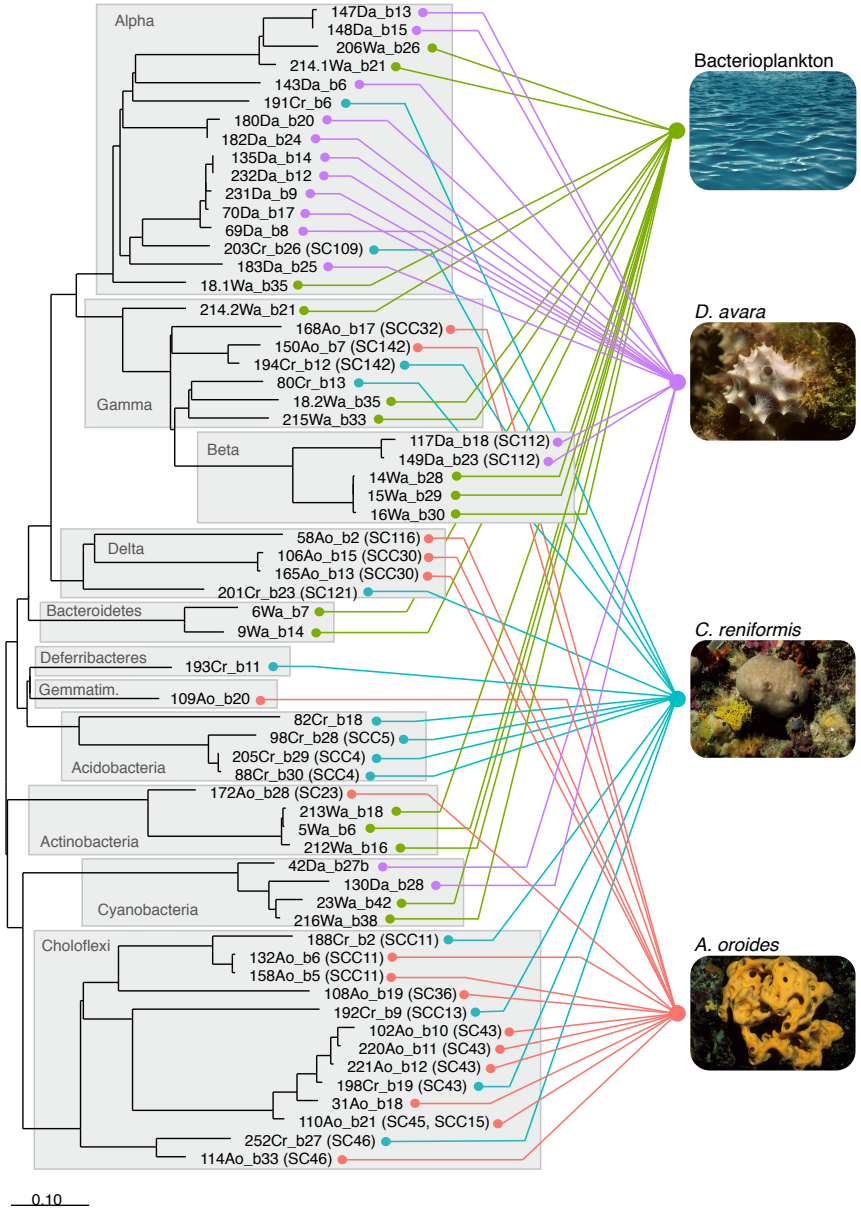


Figure 2.1: Bacteria–host interaction network for the aggregated community. Each sponge host is represented as a node (to the right) with connecting edges to nodes (to the left) corresponding to bacterial tips in a phylogeny. Each gray box corresponds to a different bacterial phylum. The abbreviation SC/SCC indicates sequences assigning to sponge-/sponge-coral-specific clusters, respectively (see *Sponge-specific bacterial clusters* for details; Table 2.2, A.2). Green nodes/edges correspond to sequences found in the bacterioplankton, purple to *D. avara*, teal to *C. reniformis*, and red to *A. oroides*. Each bacterial species corresponds to an individual excised and sequenced DGGE band.

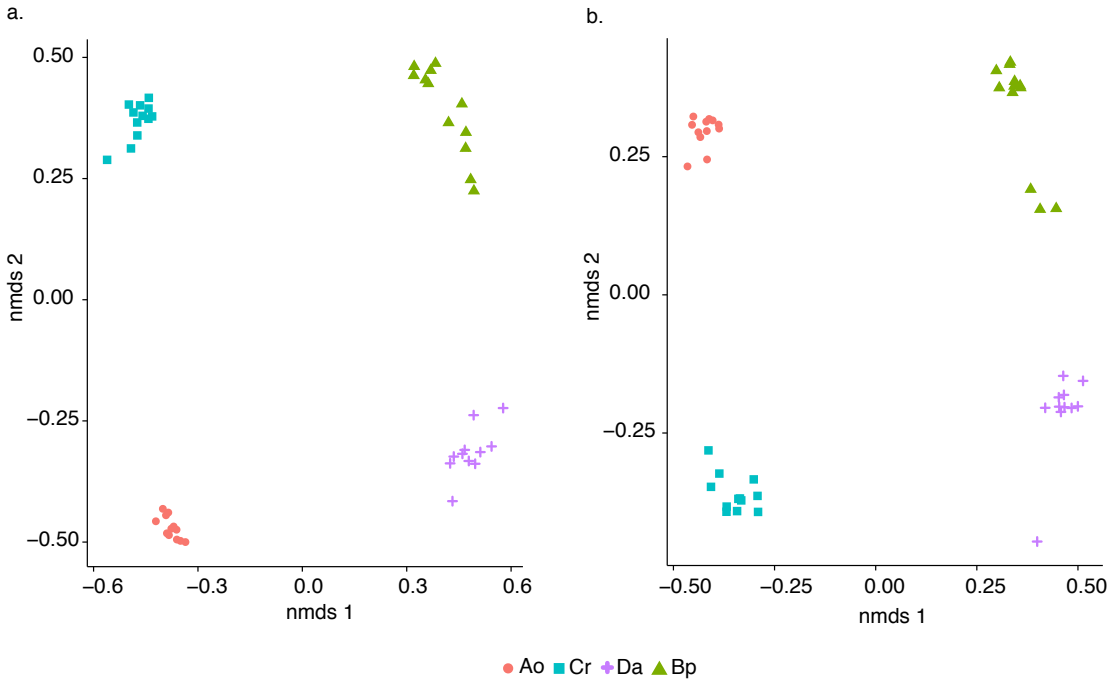


Figure 2.2: a. Nonmetric multidimensional scaling (NMDS) plot for the Jaccard index clustering species at a distance α =5% corresponding to genus level for monthly resolved bacterial communities; lowest stress, 0.163; r^2 =0.894. b. NMDS plot for the UniFrac distance for monthly resolved bacterial communities; lowest stress, 0.144; r^2 =0.919. Red circles correspond to *A. oroides* (Ao), teal squares to *C. reniformis* (Cr), purple crosses to *D. avara* (Da), and *green triangles* to the bacterioplankton (Bp).

Table 2.2: Permanent, i.e. only absent three months or less, and temporary, i.e. absent more than three months, bacterial species and their relative contribution to total bacterial abundance (mean ± SD).

Host	Number of species	Permanent species (%)	Contribution of permanent species to abundance (%)	Temporary species (%)	Contribution of temporary species to abundance (%)
A. oroides	35	74	93 ± 3	26	7 ± 3
C. reniformis	30	50	83 ± 5	50	17 ± 5
D. avara	33	30	71 ± 12	70	29 ± 12
Bacterioplankton	51	20	48 ± 6	80	52 ± 6

Note: Number of species corresponds to the total number of DGGE band for each sponge host and the bacterioplankton.

between either HMA hosts or LMA host and bacterioplankton.

However, we found support for the HMA-LMA classification of sponges when we analysed the phylogenetic diversity of bacterial communities at both temporal scales. We found the highest diversity in the HMA hosts *A. oroides* ($PD_{aggregated}$ =3.67, $PD_{monthly-resolved}$ =3.45) and *C. reniformis* ($PD_{aggregated}$ =4.04, $PD_{monthly-resolved}$ =3.14)

and the lowest diversity in LMA host *D. avara* ($PD_{aggregated}$ =2.29, $PD_{monthly-resolved}$ =1.98). Interestingly, but not surprisingly, the bacterioplankton showed a similar diversity as to the HMA hosts ($PD_{aggregated}$ =3.68, $PD_{monthly-resolved}$ =3.14).

Temporal persistence of bacteria-host interactions, shown as monthly re-occurrences of bacterial species within each host, was significantly higher for *A. oroides* (9.86 $\pm$ 0.56) and *C. reniformis* (8.27 $\pm$ 0.78) compared to *D. avara* (5.76 $\pm$ 0.62) and the bacterioplankton (5.71 $\pm$ 0.52) (Table 2.3, Figure 2.3a). We further divided bacterial species into two categories according to the number of months they appeared. Permanent species were defined as those always present or only absent 3 months or less, and temporary species, defined as those absent at least more than 3 months. Percentages of permanent species varied across hosts, from 74% for *A. oroides* to 20% for bacterioplankton. In all hosts and bacterioplankton, permanent species contributed disproportionally more to their total bacterial abundance, while temporary species contributed disproportionally less to bacterial abundance (Table 2.2). For example, temporary species in *C. reniformis* (50% of observed species) only contributed to 17 $\pm$ 5% of total abundance. Temporary species then tend to be rare among the most abundant members of the microbial community.

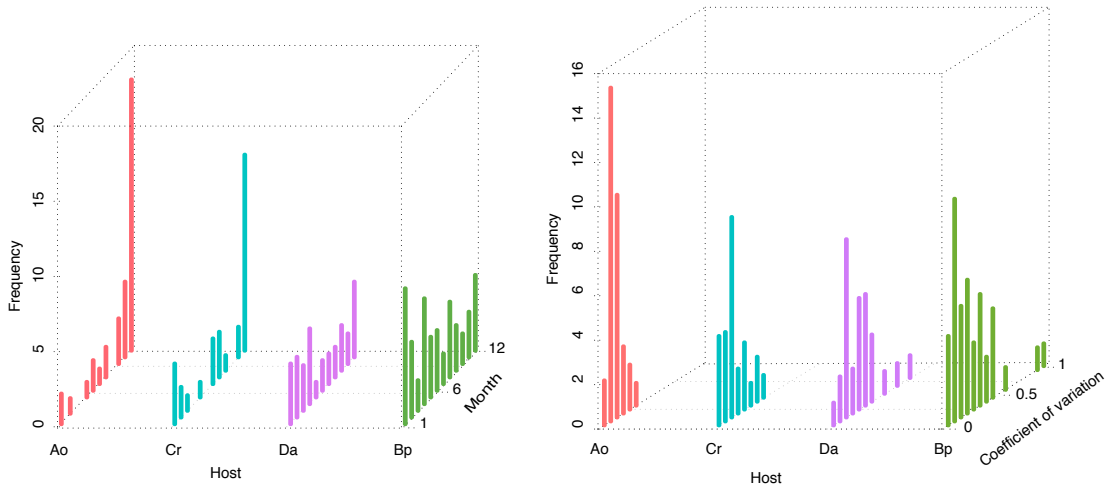


Figure 2.3: Temporal persistence of bacteria–host interactions, shown as a. Monthly reoccurrences of bacterial species within each host (i.e., the number of months a bacterial species is present), and b. Coefficient of variation (CV) of DGGE band intensities as a proxy for species relative abundances. Red corresponds to *A. oroides* (Ao), teal to *C. reniformis* (Cr), purple to *D. avara* (Da), and green to the bacterioplankton (Bp).

Similarly, the average coefficient of variation for species relative abundances was lower for *A. oroides* ($\hat{CV}$ =0.27) and *C. reniformis* ($\hat{CV}$ =0.31) compared to *D. avara* ($\hat{CV}$ =0.41) and the bacterioplankton ($\hat{CV}$ =0.36) (Table 2.2, Figure 2.3 b). However, the observed difference between *C. reniformis* and the bacterioplankton was not

significant. Also, the coefficient of variation for the phylogenetic diversity across months was lower for bacterial communities associated with *A. oroides* (CV=0.04) and *C. reniformis* (CV=0.06) than for *D. avara* (CV=0.12) and the bacterioplankton (CV=0.20). Overall, we found that communities associated with the HMA hosts had a higher temporal persistence than did communities associated with the LMA host and the bacterioplankton.

Bacteria-bacteria interaction networks

Local Similarity Analysis (LSA) for the 12 months only revealed 12 significant correlations (P and q-value ≤ 0.05) out of the 1830 pairwise possible correlations for the sequenced species across all hosts and bacterioplankton. This represents 0.67% when all interactions are considered, which is a much lower percentage than previously reported percentages from LSA studies from bacterioplankton communities, typically showing around 15% of correlated bacterial abundances (Ruan *et al.* 2006). This indicates bacterial relative abundances within hosts change more independently from each other than what is observed in free-living bacterioplankton.

Table 2.3: Temporal persistence of bacteria-host interactions

	Monthly reoccurrences		CV	
Comparison	D	P	D	P
Ao-Cr	0.243	0.296	0.268	0.200
Ao-Da	0.531	0.001∗	0.533	0.001∗
Ao-Bp	0.550	0.001∗	0.372	0.008∗
Cr-Da	0.433	0.006∗	0.362	0.040∗
Cr-Bp	0.340	0.023∗	0.225	0.328
Da-Bp	0.098	0.991	0.301	0.068

Note: A Kolmogorov-Smirnov test was performed for pairwise comparisons of the different distributions shown in Figure 2.3. Each P value is presented with the Kolmogorov-Smirnov statistic D. $*$ 0.05.

2.4 Discussion

Multispecies bacteria-host symbiotic associations are poorly explored in comparison with other species interaction networks. The current paradigm on symbiosis research stems from species-poor system studies where pairwise and reciprocally specialised interactions between a single microbe and a single host that coevolve are the norm (Ruby 2008). These symbioses involving just a few species are fascinating in their own right, but more diverse and complex host-associated microbial communities are increasingly found (Robinson *et al.* 2010), with new emerging questions that require new paradigms and approaches. Here we have tried to answer some of these questions related to the

specificity and temporal variability of the most abundant bacteria associated to different coexisting host species with diverse eco-evolutionary characteristics.

We showed that host-microbe specificity depends on both the level of microbial taxonomic aggregation and the time-scale considered. Without any taxonomic aggregation, specificity is extreme among the most abundant host-associated bacteria. Reciprocal one-to-one specialisation is the norm, in marked contrast to what is found in host-parasitoid systems and plant-pollination networks (Joppa *et al.* 2009), although in both network types rare species are also considered. This host idiosyncratic effect dissipates when microbes are clustered at different levels of taxonomic aggregation. This is because when relaxing taxonomic aggregation and going up in taxonomic rank, one is broadening the symbionts' host-range. We found that hosts with similar eco-evolutionary characteristics share more microbes among them than with hosts with different characteristics. In particular, our two High Microbial Abundance sponge species were more similar in phylogenetic community structure, while the Low Microbial Abundance sponge and the free-living bacterioplankton shared more phylogenetically similar microbes. This suggestive pattern needs to be confirmed in future studies using different host sponges.

Although our study focused on sponges as hosts, their eco-evolutionary characteristics allow for extrapolations to other complex microbe-host interactions. High- and Low Microbial abundance sponges generally reflect differences in a number of morphological and physiological characteristics (e.g. water pumping rates, sponge tissue density) and in a number of factors that affect host-microbe interactions (e.g. host defence-, symbiont recognition and transmission- mechanisms). This perspective of different eco-evolutionary characteristics affecting host-microbe interactions could be extended to other host multi-species systems where related hosts differ in relation to microbe acquisition/transmission and/or feeding ecology. These include, for example, complex gut microbial communities of mammals, termites, beetles, Lepidoptera, and coral-associated microbes (see Robinson *et al.* 2010 for a review).

The different temporal dynamics exhibited by different host species and the free-living bacterioplankton highlights the importance of considering temporally-resolved interaction networks (Olesen, Bascompte, Elberling, *et al.* 2008; Ings *et al.* 2009; Lurgi, López, *et al.* 2012). Otherwise, we might not capture the true differences in phylogenetic community structure. We only observed similarities in phylogenetic community structure across hosts when we aggregated bacterial communities over time. However, when time-resolved (monthly, in our case) communities were analysed independently, no differences were observed, and bacterial communities clustered by host species. This is due to the stability of bacterial communities associated with HMA sponges: in some cases communities belonging

to different months have the same phylogenetic structure, acting almost as replicates, and thus clustering together. This stability, combined with the higher temporal variability of bacterioplankton and LMA sponges, makes difficult interpreting patterns using single-time snapshot interaction networks, because this snapshot would be a good reflection of stable microbial communities but a poor reflection of continuously changing host-associated microbiota. Caution must be exercised when studying any interaction network involving species with high turnover rates. In particular, future network studies incorporating microbes would greatly benefit from adopting a temporal perspective.

This temporal perspective would allow inferring possible bacteria-bacteria associations. In contrast to hosts containing a single symbiont, in hosts containing a complex microbiota, microbe-microbe interactions can have a strong influence on the resulting microbial community (McFall-Ngai 2008; Robinson *et al.* 2010). However, microbe-microbe interactions are very difficult to observe directly in comparison to e.g. predator-prey interactions, and indirect methods need to be used. Temporal series analyses of relative abundances are useful here. We used local similarity analyses, finding that bacterial relative abundances within our hosts tend to be uncorrelated among them, in contrast to previous findings on free-living microbes (Ruan *et al.* 2006; Fuhrman 2009). Other methods can be used in combination to local similarity analyses to determine the importance of microbial interspecific interactions in shaping community structure, like the one developed in (Mutshinda *et al.* 2009) to decompose temporal fluctuations in species relative abundances into contributions from environmental stochasticity and inter-/intraspecific interactions. Future studies might greatly benefit from including environmental variables to disentangle its relative role in regulating bacterial dynamics.

We used DGGE band intensities as a proxy for species relative abundance, in agreement with other studies (Iwamoto *et al.* 2000; Schauer *et al.* 2003; Pascoal *et al.* 2009). However, caution must be exercised when interpreting relative abundances from any PCR-based method, because it can produce biased estimates of quantitative diversity measures (W. *et al.* 2013). For comparative (i.e. temporal, across host species) studies as the one presented here, these biases might affect all samples in a similar way, so changes on band intensities are likely to reflect real changes. Moreover, it is important that qualitative (based on presence/absence) and quantitative (using relative abundance proxies) results point in the same direction.

The inclusion of phylogenetic information in network studies is increasingly recognised as a necessary step towards understanding interaction patterns and the eco-evolutionary mechanisms shaping them (Bascompte & Jordano 2007). In our case the use of molecular phylogenies based on 16S rRNA bacterial sequences allowed us

for a fine-grained characterisation of phylogenetic community structure. We chose an intermediate level of network complexity, focusing on the most abundant bacterial species. As massive sequencing technologies are increasingly available to characterise both the abundant and rare biosphere, we ask whether our results would hold when less abundant microbes are considered. As in other network studies, food webs in particular, focusing on intermediate complexity levels- food web modules- allows for understanding and predicting system dynamics before addressing the structure and dynamics of the broader food web (McCann 2012). The ultimate goal is to introduce microbes into ecological network studies and a network perspective into host-microbe interactions, and here we suggest as a starting point restricting the analyses to the most abundant and likely important microbes for host functioning.

[1]

[1]This chapter represents a collaboration with C. Diez, M. Ribes, R. Coma and JM. Montoya. It is published in Ecology. **J R. Björk**, C Díez-Vives, R Coma, M Ribes, and J M. Montoya. 2013. Specificity and temporal dynamics of complex bacteria-sponge symbiotic interactions. Ecology, 94 (12), 2781-2791. DOI: 10.1890/13-0557.1

The dynamic core microbiome

3.1 Introduction

Microbes form intricate relationships with most animals and plants, with symbiosis postulated as one of the driving forces behind diversifications across the tree of life (McFall-Ngai *et al.* 2013). Research on host-microbe symbiosis is typically restricted to highly specialised reciprocal interactions, with one or a few microbes interacting with a single host, leading to mutual benefits for the microbe and the host (Gil *et al.* 2002; Nyholm & McFall-Ngai 2004). However, more diverse and complex host-associated microbial communities (*hereafter microbiomes*) are increasingly found in different plant and animal taxa (McFall-Ngai *et al.* 2013). This poses a challenge, because the pairwise specificity, co-evolution and reciprocity of host-microbe interactions might not explain the structure, dynamics and functioning of microbiomes. The mere existence of multiple microbes interacting with a host means that microbe-microbe interactions might also be an important driver regulating the overall composition and abundance of microbiomes and their associated functionality.

The diversity, complexity and highly dynamic nature of microbiomes makes them difficult to understand. We thus require approaches that embrace microbiome complexity and dynamics but still let us find meaningful ecological patterns. The quest for core microbiomes is a promising avenue that lately has received considerable attention (Turnbaugh, Ley, Mahowald, *et al.* 2006; Tschöp *et al.* 2009; Turnbaugh, Hamady, *et al.* 2009; Elli *et al.* 2010). However, a temporal dimension has rarely been incorporated into the core concept (Faith *et al.* 2013; Flores, Caporaso, *et al.* 2014). A core, broadly defined as the set of microbes consistently present over long periods of time, is likely to have a large effect on both the presence and abundance of occasional visitors as well as on host performance. Studies on human cohorts have revealed a collective effect of the microbiome on health and disease, showing a multitude of correlations between dysbiotic patterns and several aspects of human health, including autoimmune disorders (Round & Mazmanian 2009; Scher & Abramson 2011), diabetes (Qin *et al.* 2012), obesity (Ley, Turnbaugh, *et al.*

2006; Turnbaugh, Hamady, *et al.* 2009; Turnbaugh & Gordon 2009 and even psychiatric conditions (Foster & McVey Neufeld 2013). Therefore, the long-term stability of the core microbiome is likely critical, as the persistent occurrence of beneficial symbionts and their associated functions ensure host health and well-being, whereas major shifts in the microbiome is hypothesised to lead to ill-health (Hartman *et al.* 2009; Lozupone, Stomabaugh, *et al.* 2012; Relman 2013; Cho & Blaser 2012; Missaghi *et al.* 2014).

Despite recent realisations that complex microbiomes pervade the tree of life, little is known about microbiome dynamics beyond humans. We studied the microbiomes of six coexisting marine sponges (*Porifera*) belonging to different orders that were sampled over 36 consecutive months. Unlike humans, sponges are amongst the most ancient extant metazoans with the hypothesised oldest bacterial symbiotic interactions (Webster 2007; Taylor, Radax, *et al.* 2007; Freeman, Thacker, *et al.* 2013). As sponges filter large volumes of water, they are key-species in regulating primary and secondary production and transferring energy between the pelagic and benthic zones (Goeij *et al.* 2013; Coppari *et al.* 2016). We sampled sponges belonging to two different groups that are hypothesised to differ markedly in numerous traits that illustrate their dependence upon their associated microbes. The classification is based on the abundance and diversity of microbes they harbour – High and Low Microbial Abundance hosts (HMA and LMA), respectively. This classification pervades host morphology and physiology: LMA hosts have an interior architecture fitted for high pumping rates, whereas HMA hosts are morphologically adapted to harbour dense microbial assemblages within their tissue (Weisz *et al.* 2008; Schläppy *et al.* 2010). Therefore, LMA hosts are more dependent on nutrient uptake from the water column (Jiménez *et al.* 2007; Weisz *et al.* 2008; Schläppy *et al.* 2010; Freeman & Thacker 2011; Maldonado, Ribes, *et al.* 2012), whereas HMA hosts rely more heavily on nutrients produced by their microbial symbionts (Freeman & Thacker 2011; Maldonado, Ribes, *et al.* 2012; Ribes *et al.* 2012; Freeman, Thacker, *et al.* 2013; Freeman, Easson, *et al.* 2014; Poppell *et al.* 2014). These two sets of hosts provide an ideal system to explore whether the structure and temporal dynamics of complex microbiomes differs across hosts with different eco-evolutionary characteristics and life styles.

In this chapter we describe the structure, dynamics and stability of complex microbiomes across hosts with different eco-evolutionary characteristics and life styles. We specifically focus on core microbiomes, although, in order to fully understand each microbiome, we also characterise the large number of temporally transient microbes that occur within each host. The aggregated abundance of these transient microbes is likely to affect core dynamics and stability. In order to elucidate microbial dynamics and interactions between constituent core members, we fitted a population dynamic model to the time-series of each core microbiome.

In particular, we hypothesise that HMA hosts have larger, more abundant and stable core microbiomes that ensures host functionality compared to the core microbiomes of LMA hosts. Additionally, we expect that HMA cores consists of a network of complementary symbionts with selection acting to minimise competition, thereby simultaneously reinforcing selection for complementarity. Therefore, we do not expect to find strong competitive interactions, nor do we expect to find strong cooperative interactions, as species coupling and positive feedback are likely destabilising.

3.2 Methods

Sponge collection

Sponge specimens from species *Agelas oroides*, *Chondrosia reniformis*, *Petrosia ficiformis*, *Axinella damicornis*, *Dysidea avara* and *Crambe crambe* were collected monthly from March 2009 until February 2012 close to the Islas Medas marine reserve in the NW Mediterranean Sea $42°3'0''N$, $3°13'0''E$ by SCUBA at depths between 5-10 m. The collected sponge species belong to six different orders, and represent common members of the Mediterranean benthic community. Each species were identified based on distinct morphological features. Replicates were carefully placed in separate plastic bottles and brought to the surface. Three replicates per sponge species were sampled and frozen in liquid nitrogen until DNA extractions.

DNA extraction and sequencing

DNA was extracted from ≈25 mg of sponge tissue per sample using the DNeasy tissue kit (Qiagen, Valencia, CA) and prepared for 16S rRNA gene amplicon (454) sequencing following standard protocols.

Analysis of sequencing data

454 reads were processed in mothur v.1.29.2 (Schloss *et al.* 2009). Raw reads were pooled from replicates belonging to the same sponge species. Fasta, qual and flow files were extracted from binary sff files; sffinfo(. . .,flow=T). Flow files were then filtered based on barcodes to speed-up the proceeding de-noising process; trim.flow. Sequences were de-noised; shhh.flows(. . ., lookup= LookUp_ Titanium.pat). The LookUp-file is necessary and specific to the 454 technology used. Next the barcode and primer sequences were removed together with sequences shorter than 200bp and/or contained homopolymers longer than 8bp; trim.seqs(. . ., pdiffs =2, bdiffs =1, maxhomop =8, minlength =200). In order to minimise computational effort, files were reduced to non identical sequences; unique.seqs.

Non redundant sequences were aligned to SILVA 102 reference alignment with default kmer search and Needleman-Wunsch algorithm; align.seqs(. . ., flip =F). Non overlapping sequences were removed; screen.seqs(. . ., optimize= end, start= 1044, criteria = 95), in addition to empty columns that were introduced from the alignment process; filter.seqs(. . .,vertical =T, trump =.). Aligned sequences were reduced to non redundant sequences; unique.seqs. To further reduce amplification errors, less abundant sequences were binned to more abundant sequences if they were within 2bp of a difference; pre.cluster(. . ., diffs =2). Chimeric sequences were identified; chimera.uchime(. . ., dereplicate =T) and removed; remove.seqs. Sequences were classified using the RDP reference taxonomy; classify.seqs(. . ., template =trainset9_ 032012.pds.fasta, taxonomy =trainset9_ 032012.pds.tax, cutoff =80), and non bacterial lineages were removed; remove.lineage(. . ., taxon= Mitochondria-Chloroplast-Archaea-Eukaryota-unknown). We calculated pairwise distances between aligned sequences; dist.seqs(. . ., cutoff =0.050).

We analysed sequence distribution across the eight 96-well plates in order to detect any potential plate effect. As samples belonging to the same sponge species were distributed across two plates, and were found to differed in their sequence count depending on plate, we subsampled 1500 sequences from each monthly host sample. This number corresponded to the average of the three lowest host-plate averages. This was followed by clustering sequences into OTUs; classify.otu(. . ., label=0.030) and creating an OTU-table (.shared-file); make.shared(. . ., label=0.030). All post analyses were conducted in R v.3.2.1 (R Development Core Team, 2008).

Identification of sponge-specific and sponge-specific clusters

A representative sequence from each OTU was taxonomically assigned using a BLAST 62 search against a curated ARB-SILVA database containing 178 previously identified sponge-specific clusters (SC) (Simister *et al.* 2012). For each BLAST search, the 10 best hits were aligned in order to determine sequence similarities. The most similar OTU sequence to the respective reference sequence within the database was then assigned to an SC based on application of a 75% similarity threshold (i.e. a sequence read was only assigned to a cluster if it was more similar to the members of that cluster than to sequences outside the cluster and its similarity to the most similar sequence within that cluster was above 75%). In cases where the assignment of the most similar sequences was inconsistent, a majority rule was applied, and the OTU sequence was only assigned to an SC if at least 60% of the reference sequences were affiliated with this cluster.

Null model

A randomised realisation was created for each host in order to assess if observed diversity patterns could have been generated by random assembly processes from a species regional pool. In our null model, a regional pool was created for each 36 month by pooling all species and their occurrences from the corresponding month across hosts, including the water column. A randomised realisation was created by randomly sample the same number of individuals (sequences) present in each month (max 1500) from the corresponding monthly regional pool. This meant that species were randomly sampled from each monthly regional pool in proportion to their abundance. We ran the null model 999 times for each host.

Temporal variability

Temporal turnover

We applied a newly developed measure of temporal turnover that describes the extent to which individual species and consequently the community changes over time (Shimadzu *et al.* 2015. Importantly, this measure of temporal turnover decomposes abundance fluctuations into two additive contributions of changes due to: species composition and total abundance.

Total turnover, D, between times t and u, $(u > t)$ as

$$D(t:u) = \sum_{i=1}^{S} d_i(t:u) = \sum_{i=1}^{S} \log\left(\frac{\lambda_{i,u}}{\lambda_{i,t}}\right) p_{i,t} \tag{3.1}$$

$$= -\sum_{i=1}^{S} \log\left(\frac{p_{i,t}}{p_{i,u}}\right) p_{i,t} + \left(\frac{\lambda_u}{\lambda_t}\right) \tag{3.2}$$

$$= D_1(p_t : p_u) + D_2(\lambda_t : \lambda_u) \tag{3.3}$$

where λ_t represent the expected total abundances of the species in the community $\lambda_t = \sum_{i=1}^{N} \lambda_{i,t}$. The expected total abundance $\lambda_t, i = 1, 2, \ldots, s$, is unknown and therefore needs to be estimated from observed time-series. $p_{i,t}$ represent the relative abundance of the $i - th$ species at time t and is calculated as $p_{i,t} = \left(\frac{\lambda_{i,t}}{\lambda_t}\right)$. As such, total turnover, D, can be decomposed into two additive components, D_1 which is related to the amount of change in community composition, and D_2 reflecting the amount of change in total abundance.

As noted above, the expected abundance needs to be estimated from observed time-series data. We modeled

each observed time-series, $N_{i,t}$, as a Poisson process with a time varying mean parameter equal to $\lambda_{i,t}$

$$N_{i,t} \sim \text{Pois}(\lambda_{i,t}) \tag{3.4}$$

$$\log(\lambda_i(t)) = x_j' \beta_j \tag{3.5}$$

where X_j' is a matrix of explanatory variables (e.g. temperature and salinity). The regression coefficient β_j represents the expected change in the log of the mean per unit change in the predictor x_j.

Temporal invariability

We applied two newly developed measures of invariability in order to assess the aggregated stability (Haegeman et al. in review) at both population and community (ecosystem) level.

Invariability at the community (ecosystem) and population level is defined as follows

$$I_{eco} = \frac{1}{CV(X_{tot})^2} = \frac{(\bar{X}_{tot})^2}{Var(X_{tot})} \tag{3.6}$$

$$I_{pop} = \frac{1}{\left(\sum_i \frac{\bar{X}_i}{\bar{X}_{tot}} CV(X_i)\right)^2} = \frac{(\bar{X}_{tot})^2}{\sum_i \sqrt{Var(X_i)^2}} \tag{3.7}$$

where X_{tot} and X_i denotes total and species-specific abundance at time t, respectively, and $\bar{X}_{tot}$ and $\bar{X}_i$ their respective averages.

Interestingly, these two measures are connected. The ratio of I_{pop} and I_{eco} equals community (ecosystem)-wide synchrony

$$\frac{I_{pop}}{I_{eco}} = \frac{Var(X_{tot})}{\sum_i \sqrt{Var(X_i)^2}} \tag{3.8}$$

with 0 and 1 for a perfectly asynchronous and synchronous ecosystem, respectively. This implies that $I_{pop} \leq I_{eco}$, and that the equality $I_{pop} = I_{eco}$ represents a perfectly synchronous ecosystem.

Core dynamics and ecological interactions

We developed a statistical model based on Mutshinda *et al.* (2009) for modeling core dynamics and in order to infer interactions between microbes from temporal series data.

Process model If we denote $n_{i,t}*$ as the expectation of $n_{i,t}$ which is the natural logarithm of the observed time-series $N_{i,t}$, then on the natural logarithmic scale we have the expected number of individuals of core species i in time t within any given host species described by

$$n_{i,t} * |n_{i,t-1} = n_{i,t-1} + r_i \left[1 - \sum_{j=1}^{S} \frac{\alpha_{i,j} n_{j,t-1}}{k_i} \right] + \epsilon_{i,t} \tag{3.9}$$

where r_i and k_i represent the intrinsic growth rate and the carrying capacity (on natural logarithmic scale) of core species i, respectively. r_i is assumed $\sim \mathcal{N}(0, 10)$ while k_i is assumed $\sim Exp(1)$ in order to limit k_i to positive values. $\alpha_{i,j}$ represents the interaction coefficient between core species i and j and expresses the per capita effect of species j on the growth rate of core species i from $t-1$ to time t. Finally, $\epsilon_{i,t}$ represents the effect of unexplained (latent) environmental stochasticity on the population dynamics of species i and enters the model through the term $\eta_{t,h}\theta_{i,h}$, $h = 1, \ldots, 2$. $\eta_t = \eta_{t,1}, \eta_{t,2}$ represent the two latent variables, and $\theta_i = \theta_{i,1}, \theta_{i,2}$ represent the two corresponding loadings quantifying each species response to each latent variable. In fact, if the latent variables were known covariates, their loadings would simply correspond to standard regression coefficients. All the species-specific loadings related to the latent variables, $\theta_{i,h}; h = 1, \ldots, 2$, were assigned normal priors $\mathcal{N}(0, 1)$ while taking to account the appropriate constraints for parameter identifiability (see Hui, Taskinen, *et al.* 2015, for details).

Observation model Each time-series of core species is modeled as a Poisson process, where $y_{i,t}$ denotes the number of observed individuals of core species i in time t

$$n_{i,t} = \mathcal{MVN}(n_{i,t}*, \sum) \tag{3.10}$$

$$y_{i,t} = \text{Pois}(\lambda_{i,t}) \tag{3.11}$$

$$\log \lambda_{t,i} = n_{i,t} + \log N_t + \Lambda_t \tag{3.12}$$

where N_t and Λ_t are offsets representing the total abundance at time t. The latter is treated as a random variable assumed $\sim \mathcal{N}(0, 100)$.

The total variance V_i of individual species abundances can be decomposed into additive contributions from interspecific interactions, intraspecific interactions (density-dependence) and environmental stochasticity, respectively.

$$V_i = \left(\left(\frac{r_i}{k_i} \right)^2 \sum_{j \neq i} \alpha_{i,j}^2 v_{j,j} \right) + \left(\left(\frac{r_i}{k_i} \right)^2 v_{i,i} \right) + \epsilon_{i,i}^2 \tag{3.13}$$

where $v_{i,i}$ is the stationary variance for n_i, so that the proportion of variation attributed to e.g. intraspecific interactions (density-dependence) can be calculated as follows

$$var(intra)_i = \left(\frac{r_i}{k_i}\right)^2 \alpha_{i,i}^2 v_{j,j}/V_i \tag{3.14}$$

We used Gibbs Variable Selection (GVS) (O'Hara & Sillanpaa 2009) method to constrain the model to only use interspecific interaction coefficients, $\alpha_{i,j}$, for which there were strong support in the data. This was achieved by introducing a binary indicator variable $\gamma_{i,j}$ for $i \neq j$, and assuming $\gamma_{i,j} \sim Bernoulli(p)$, such that $\gamma_{i,j} = 1$ when species j is included in the dynamics of species i, and $\gamma_{i,j} = 0$, otherwise. Where there was low support for $\alpha_{i,j}$ in the time-series data, $\gamma_{i,j} = 0$, and the interaction was excluded from the model. When $\gamma_{i,j} = 1$, $\alpha_{i,j}$ is freely estimated from the time-series data. We set p to 0.1 as we did not expect more than 10 per cent of all possible interspecific interactions to be realised.

Finally, to build core networks, we analysed the interaction and sign structure of the posterior distribution for the interaction coefficient α_{ij}. α_{ij} is a full probability distribution, hence it contains the probability of OTU j having a per capita effect on the growth of OTU i (interaction strength), and vice versa. Using all information in α_{ij}, we constructed core networks for each HMA host as a means of visualising the most 'credible' network structure. This was done by mapping the posterior average number of links onto α_{ij}, and in doing so, extracting the links with the highest probability of non-zero interactions. This was done by custom-written R scripts. As a way of validating the structure of each core network, we compared the connectance of each network to the posterior average connectance for α_{ij} for each host. The networks were plotted using the *igraph* package in R v.3.2.1.

3.3 Results

Core microbiomes, transient and opportunistic assemblages

We divided each microbiome into three different temporal assemblages based on the persistence of individual microbes over the 36 consecutive months. Core microbiomes consisted of microbes present in more than (or equal to) 70% of each total time-series (i.e., persisting $\geq$ 26 months), whereas opportunistic assemblages included microbes present in less than (or equal to) 30% of each the total time-series (i.e., persisting $\leq$ 11 months). The transient assemblages consisted of intermediately persistent microbes (i.e., those persisting between 12 and 25 months). We analysed six microbiomes: three belonging to hosts classified as HMA and three from hosts classified as LMA (HMA

hosts: *Agelas oroides*, *Chondrosia reniformis*, *Petrosia ficiformis*; LMA hosts: *Axinella damicornis*, *Dysidea avara*, *Crambe crambe*) (Gloeckner, Wehrl, *et al.* 2014; Erwin, Coma, *et al.* 2015).

Table 3.1: Microbiome richness by temporal assemblage

	HMA			LMA		
	A. oroides	*C. reniformis*	*P. ficiformis*	*A. damicornis*	*D. avara*	*C. crambe*
Core	45 38.4 ± 11.7	33 27.4 ± 11.1	40 32.1 ± 12.3	8 6.7 ± 2.1	6 6.4 ± 2	8 6.4 ± 2.3
Transient	90 41.3 ± 14.4	54 26.4 ± 12.5	140 67.9 ± 30.8	31 12.7 ± 5.9	44 21.4 ± 7.4	41 17.7 ± 6.2
Opportunistic	2658 140 ± 77.7	2436 129.9 ± 66.7	2580 141.4 ± 30.2	2443 126.8 ± 78.1	2763 126.3 ± 42.8	3465 160.1 ± 60.4
Total biome	2793	2523	2760	2482	2813	3514

Notes: For each assemblage, the first row shows the total number of unique species, while second row shows the monthly average (mean±sd) number of coexisting species. The last row, Total biome shows the total number of species present in each microbiome.

Species richness and compositional overlap

The cores of HMA hosts had 4 times more species (defined as 16S rRNA sequences 97% similarity) than the cores of LMA hosts (Table 3.1). Core species richness only represented a small fraction (1.24%) of the overall richness that resulted from the aggregation of all species found throughout the time-series, with a clear dominance of opportunistic species. Species richness saturated fast for both core and transient assemblages, while the rarefaction curve for opportunistic assemblages did not reach an asymptote, indicating that further sampling will reveal additional opportunistic species (Figure B.3). Additionally, the HMA hosts had a higher core richness than the LMA hosts (table 3.1).

A detailed analysis of monthly coexisting species showed that cores actually represented an important fraction of monthly diversity, especially in HMA hosts (Table 3.1). Here, opportunistic assemblages only had between 3.7 to 4.8 times more species than the core microbiomes, whereas LMA hosts harbored between 18 to 27 times more opportunistic than core species (Table 3.1).

Microbiomes differed markedly across hosts, with very little overlap in species composition (ANOSIM: R =0.767, P<0.001, Figure 3.1, Table B.1, Table B.2) and phylogenetic community structure (ANOSIM: R =0.675,

P<0.001, Figure B.4, Table B.1). Core microbiomes had no species in common, with the exception of a single species shared by two LMA hosts (Table B.6). Amongst the transient assemblages, the maximum overlap was found between hosts *A. damicornis* and *P. ficiformis* (6.5%), as well as between *A. damicornis* and *C. crambe* (6.5%) (Table B.5). The same pairs shared 31 and 30%, respectively, of their opportunistic microbes (Table B.4). Comparing overlap across assemblages, we observed a general tendency for hosts to become increasingly similar in both species composition (Figure 3.2) and phylogenetic community structure (Figure B.5) when moving from core microbiomes and transient assemblages to the opportunistic assemblages .

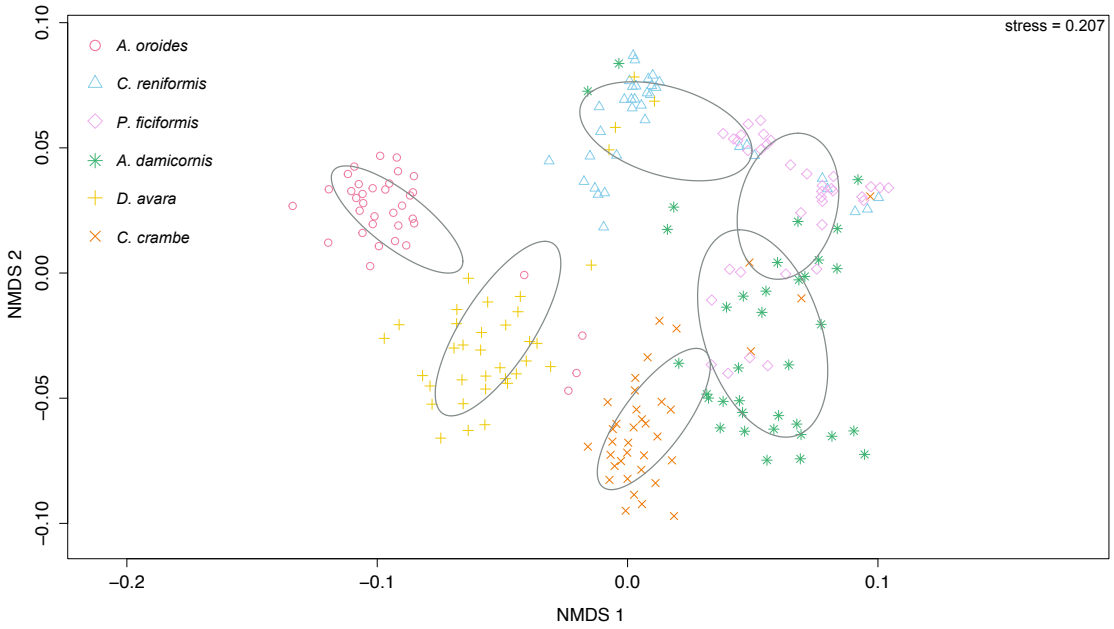


Figure 3.1: Non-metric multidimensional scaling (NMDS) calculated from Jaccard distances between monthly samples from each host in terms of species present. Colours and shapes denote all (36) monthly samples from a given host species. Host samples are surrounded by an ellipse showing the intraspecific variability across time. *A. oroides*: Red circle, *C. reniformis*: Blue triangle, *P. ficiformis*: Purple diamond, *A. damicornis*: Green star, *D. avara*: Yellow cross (+), *C. crambe*: Orange cross (×).

Analysing overlap by grouping host species based on host classification showed that HMA and LMA hosts overlapped in species composition (ANOSIM: R =0.268, P<0.001, Figure B.6) but less in phylogenetic community structure (ANOSIM: R =0.463, P<0.001, Figure B.7). Analysing overlap at higher taxonomic ranks showed that HMA cores harbored three dominant phyla (*Chloroflexi*, *Acidiobacteria* and *Gammaproteobacteria*) that accounted for roughly half of their diversity. These hosts still kept a unique taxonomic profile by harboring other known symbiotic

Figure 3.2: Non-metric multidimensional scaling (NMDS) calculated from Jaccard distances between monthly samples from each host and assemblage in terms of species present. Colours denote all monthly samples from a given host and shapes denote all monthly samples for a given assemblage. Circle, *C. reniformis*: triangle, *P. ficiformis*: diamond, *A. damicornis*: star, *D. avara*: cross (+), *C. crambe*: cross (×) and colours denote assemblages; purple: core microbiomes, green: transient assemblages, pink: opportunistic assemblages.

phyla such as *Actinobacteria*, *Nitrospira*, *Spirochaetes* and *SAUL* (Figure 3.3), whereas LMA cores were largely dominated by members of *Proteobacteria* (Figure 3.3).

In order to assess whether random assembly processes from a species-rich microbial regional pool generated empirical patterns in species richness and overlap across hosts, we created a simple stochastic null model that replicated each microbiome over the temporal series (i.e., on a monthly basis). A stochastic realisation was created for each microbiome by randomly sampling the same number of individuals present in any given month from a regional pool consisting of all individuals present in the corresponding month across all microbiomes taken together. In this simple model, species were randomly sampled in proportion to their abundances. We found that each stochastic realisation had a much higher richness, but far fewer unique species and hence a higher compositional overlap than their empirical counterparts. In fact, the random realisations almost overlapped 100% in species composition (ANOSIM, R=-0.017, P=1, Figure B.8). This suggests that a suite of non-random assembly processes are responsible for the observed patterns, for instance, host selection of specific microbes and microbe-microbe interaction dynamics.

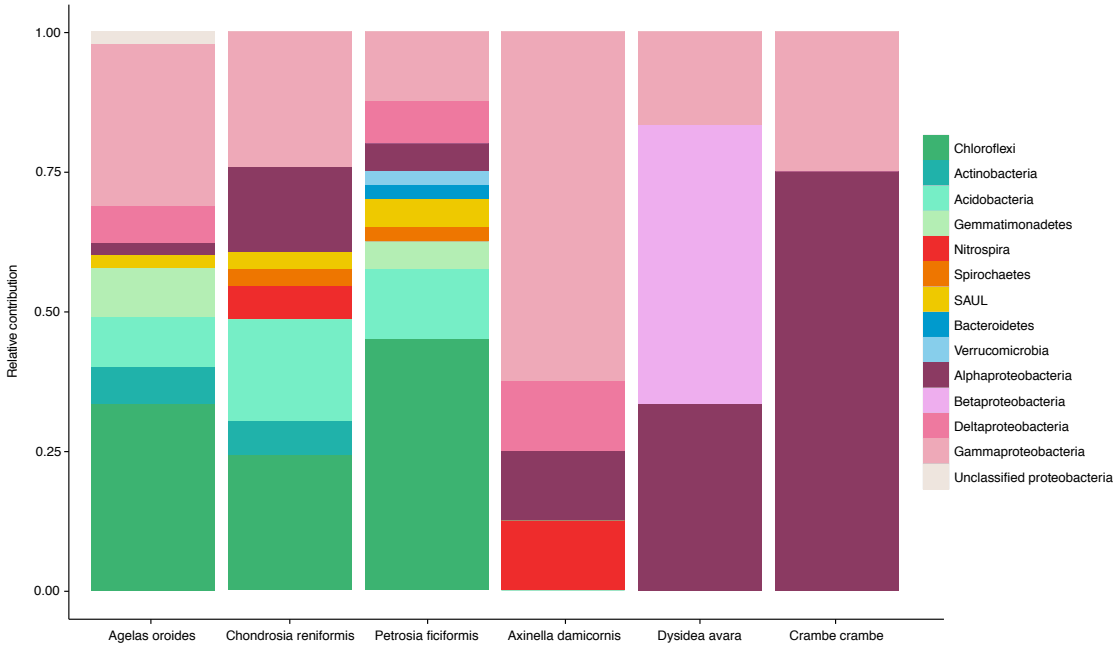


Figure 3.3: Shows the taxonomic classification (at the phylum level) of core species and their relative contribution to the total species richness of each core across hosts. It is clear that HMA cores contain a higher phylogenetic diversity than LMA cores.

Sponge-specific microbes

Sponges are known to harbour certain microbes that show high specificity to the phylum Porifera. These microbes form monophyletic 'sponge-specific' clusters that range over 14 bacterial and archaeal phyla (Hentschel, Hopke, *et al.* 2002; Simister *et al.* 2012). Even if they are commonly discovered in the sediment and seawater, they are only detected at very low abundances outside sponge hosts (Taylor, Radax, *et al.* 2007; Taylor, Tsai, *et al.* 2013). Sponge-specific lineages, together with the direct observation of vertical transmission of some microbes, provide strong support for microbe-sponge coevolution and cospeciation (Thacker & Freeman 2012). These microbes are hypothesised to be crucial for host health and functionality (Webster, Cobb, *et al.* 2011). We found that HMA hosts, and in particular their cores, harboured a larger proportion of microbes that assigned to 'sponge-specific' clusters than LMA hosts (Figure 3.4, Table B.7). These microbes were also, on average, much more abundant in the cores compared to transient and opportunistic assemblages (Table B.8).

Next, we used our null model to investigate the proportion of randomly sampled microbes mapping to any of the 'sponge-specific' clusters revealed in the observed data. Interestingly, we found no apparent difference between

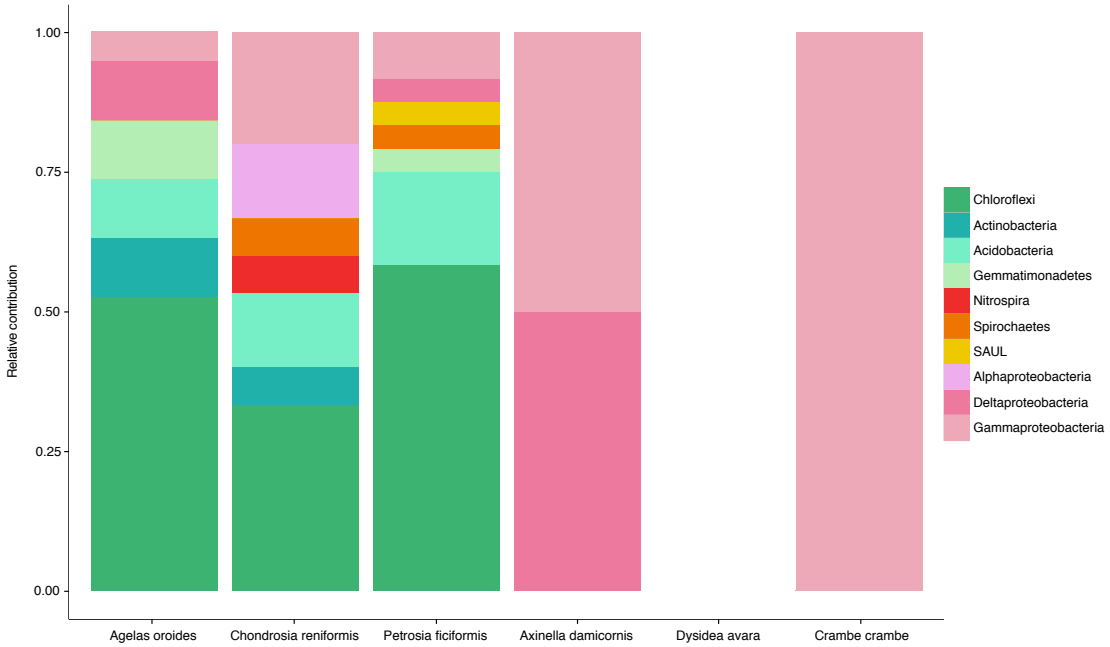


Figure 3.4: Shows the taxonomic classification (at the phylum level) of core species mapping to sponge-specific clusters and their relative contribution to the total species richness of each core across hosts. It is clear that HMA cores contain a lager phylogenetic diversity of species belonging to sponge-specific clusters than LMA cores.

HMA and LMA hosts (Table B.9). While, the proportion of microbes that assigned to 'sponge-specific' clusters in the random realisations declined faster and more consistently from the core microbiomes to the opportunistic assemblages, we observed a rapid increase in their average abundances (Table B.10). The latter contrasts with the observed data, where 'sponge-specific' microbes were most abundant in the core microbiomes and the least abundant in the opportunistic assemblages. These results indicate that the observed prevalence of 'sponge-specific' clusters in sponge hosts is not a result of random assembly processes, but rather implies that some selection mechanisms are favouring these microbes, thus permitting them to accumulate in high richness and densities within the core microbiomes.

Temporal variability

Across hosts, we observed that individual core and transient species were more stable (measured as co-efficient of variation, CV) (Kruskal-Wallis test: H =2198, df=2, P<0.001 two-tailed; Dunn's post-hoc test with bonferroni correction; P <0.001) and abundant (Kruskal-Wallis test: H =1694, df= 2, P<0.001; Dunn's post-hoc test with bonferroni correction; P< 0.001 two-tailed) than opportunistic microbes (Figure 3.5).

Temporal turnover is an inherent property of our definition of core, transient and opportunistic assemblages. For example, as opportunistic microbes persist less than 30% of the total time-series, these assemblages are bound to heavily fluctuate in microbial composition. However, in order to quantify the temporal turnover of each assemblage, we applied a newly developed measure that disentangles the two additive determinants of temporal turnover: total abundance and species composition (Shimadzu *et al.* 2015). As expected, we found that cores were mainly driven by changes in abundance (Mann-Whitney U-Test: U =16676, P<0.001 two-tailed). All cores had a similar degree of turnover, except LMA host *A.damicornis* whose core had a significantly larger turnover than all the other cores (Kruskal-Wallis test: H=84.965, df=5, P<0.001 two-tailed; Dunn's post-hoc test with bonferroni correction; P<0.001 two-tailed). Both transient and opportunistic assemblages were as predicted, govern by changes in microbial composition (Mann-Whitney U-Test: U=10223, P<0.001 two-tailed; Mann-Whitney U-Test: U=696, P<0.001 two-tailed, respectively. (Figure B.12).

When aggregating species abundance for each assembly, we observed two somewhat different temporal dynamics across hosts (Figure 3.6). In the microbiomes of *A. oroides*, *C. reniformis* and *C. crambe*, cores were very dense, i.e., they accounted for the majority of microbiome relative abundance. In contrast, the cores of *D. avara*, *A. damicornis* and *P. ficiformis* were sparser and instead transient and/or opportunistic assemblages dominated microbiome relative abundance (Figure 3.6, Table B.11). Interestingly, we observed that dense cores tended to be more stable over time than sparse cores, both at the assembly and population level, respectively (*A. oroides*: 7.9|1.8, *C. reniformis*: 4.0|2.1, *C. crambe*: 5.0|4.0, and *D. avara*: 2.9|1.0, *A. damicornis*: 1.5|1.1 and *P. ficiformis*: 2.1|1.2) (Figure 3.6).

Core dynamics and ecological interactions

Next we asked what mechanisms, other than density could be responsible for the observed patterns of core temporal variability. The individually unique, but temporal stable microbiome of the human gut, for example, is hypothesised to result from, among other factors, interactions among its constituent members (Faith *et al.* 2013).

The observed dynamics may have resulted from inter- and/or intraspecific (i.e., density dependence) interactions among core members. In order to disentangle the complexity of microbe-microbe interactions within each core, we applied a statistical model that assumes Lotka-Volterra population dynamics. The model allows for decomposing temporal variation in microbial population abundances into contributions of interspecific interactions, intraspecific interactions (i.e. density dependence), and environmental stochasticity. Importantly, the model estimates the interaction

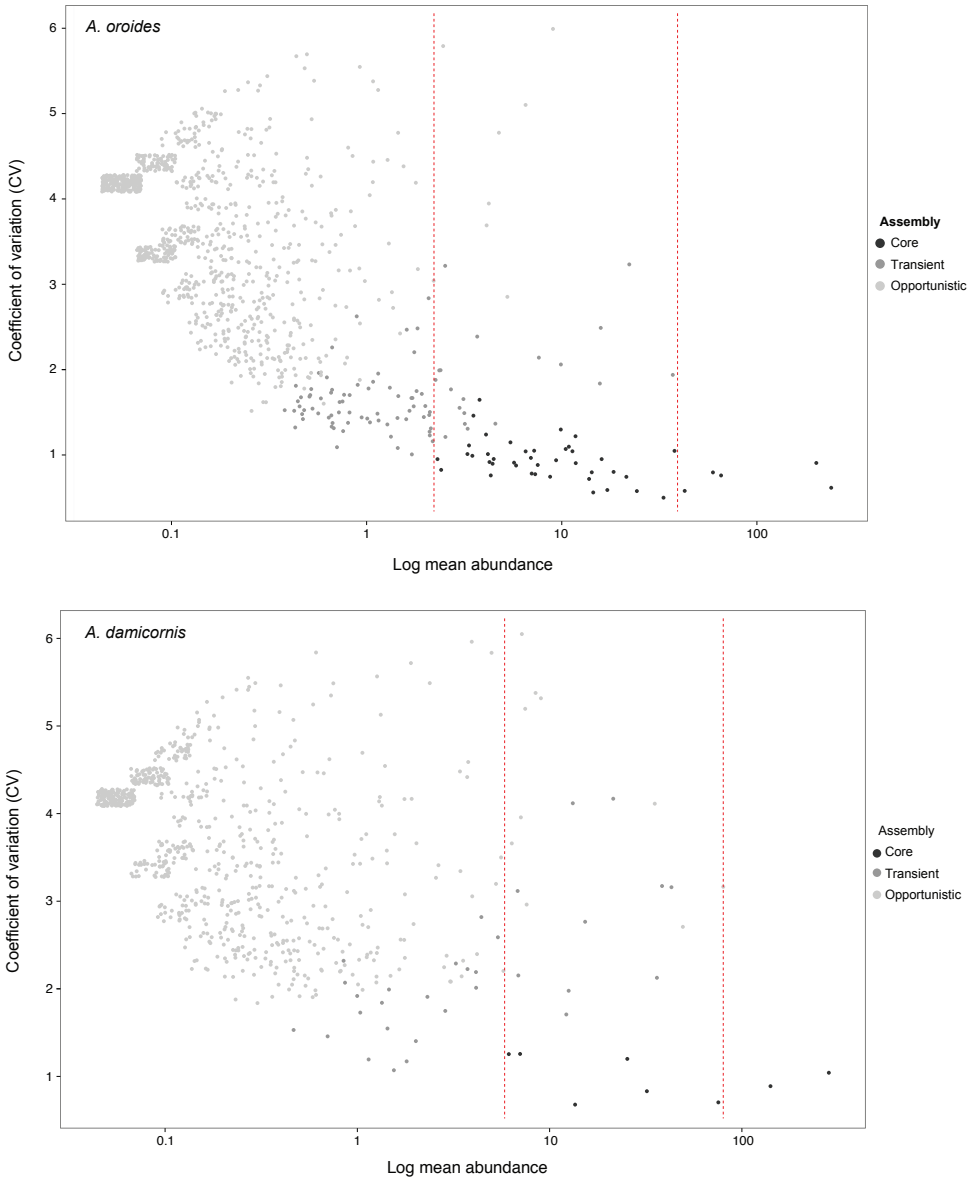


Figure 3.5: The relationship between coefficient of variation (CV) and (log) mean abundance of individual species. Overlaying points have been separated by adding jitter (random noise) of 0.1 in both y and x direction. Opportunistic, transient and core species are each shown by an increasing grey scale. The top panel shows the relationship for host *A. oroides* and the bottom panel for *A. damicornis*. It is interesting to note that *A. oroides* had the densest and most stable core microbiome across hosts, while the core microbiome of *A. damicornis* was among the sparsest and most unstable. The red dashed lines mark a potentially critical area by which you can predict microbiome temporal stability. If there are a few abundant and occasional species relative to the number of core species, stability is predicted to be high (as in the case of *A. oroides*, whereas, if there are many abundant and occasional species compared to core species, stability is predicted to be low. See Figure B.1 and B.2 for the other host microbiomes

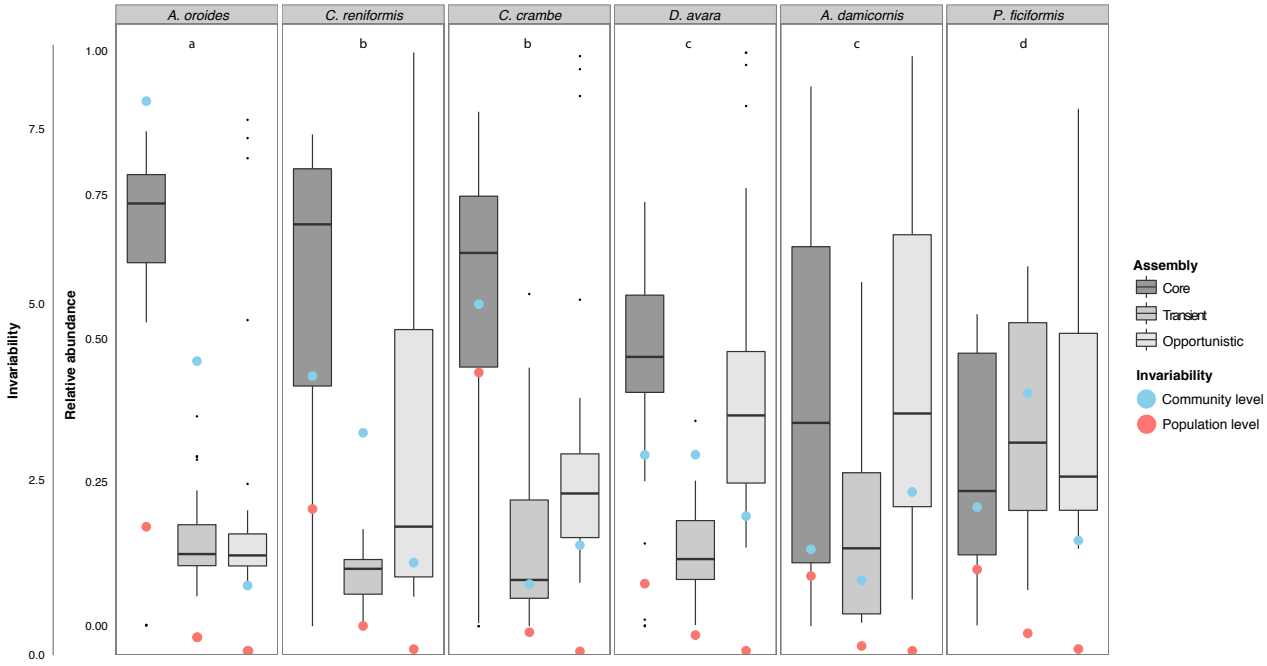


Figure 3.6: The inner y-axis shows the contribution of each assembly and host to the monthly aggregated relative species abundance. Each box shows the median including the first and third quartiles (the 25th and 75th percentiles), representing monthly variation. The outer y-axis shows the invariability (temporal stability) at the level of population (red dots) and community (blue dots) for each assembly and host. The figure is ordered from the highest to the lowest in terms of core density. Lowercase letters denote different significant scenarios (Dunn's post-hoc test for Kruskal-Wallis rank sum test (see Table B.11 for more details). a. The core microbiome was significantly different from the transient and opportunistic assemblages, but transient and opportunistic assemblages were not significantly different from each other. b. All assemblages were significantly different. c. Core microbiomes and opportunistic assemblages were not significantly different, but core microbiomes and transient assemblages and transient and opportunistic assemblages were significantly different. d. No significant differences between any assemblages. What emerged was three high-density (a, b) and three low-density (c, d) cores, respectively.

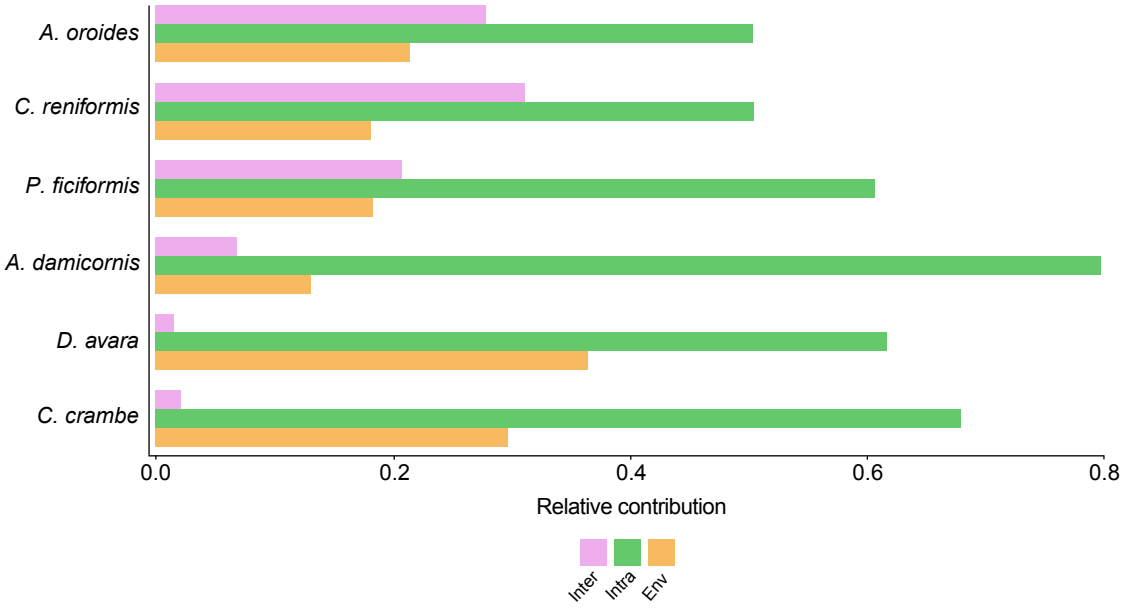


Figure 3.7: Relative contribution of inter- and intraspecific (i.e. density-dependence) interactions and environmental stochasticity to the total temporal variation in microbial population abundances across hosts. Across hosts, dynamics were mainly driven by intraspecific interactions (i.e. density-dependence). While hosts were relatively equally affected by environmental stochasticity, an important driver of the dynamics in HMA cores were interspecific interactions which was almost negligible in LMA cores.

matrix and its encoded interaction coefficients describing the per capita influence of microbe j on the growth rate of microbe i. We used the interaction coefficients in order to construct and characterise core interaction networks.

Somewhat surprisingly, model results showed similar dynamics not based on high and low density cores, but based on host eco-evolutionary characteristics (HMA and LMA). These two classifications were largely congruent, except in the case of LMA host *C.crambe* and HMA host *P.ficiformis*, which harboured high and low density cores, respectively.

Density-dependent processes were the main factor driving the dynamics across all cores (Figure 3.7). Much of the remaining variation in HMA cores was explained by interspecific interactions, followed by environmental stochasticity. In LMA cores, environmental stochasticity was the second largest determinant, followed by interspecific interactions that had an almost negligible contribution to total variance (Figure 3.7). Although, environmental stochasticity differed in its relative importance between HMA and LMA cores, it still explained similar proportions of variation in both host types (Figure 3.7).

As interspecific interactions had an almost negligible effect on the core dynamics within LMA hosts (Figure 3.7), we now focus only in describing the network structure for the HMA cores. We found that only a small fraction of the interspecific interactions among the possible ones were likely to occur (i.e. low core connectance), ranging from 5 to 7%. These interactions were mostly very weak, and for each core we observed a skewed distribution of interactions strengths towards many weak and a few strong interactions (Figure B.12). Amongst the most probably interactions, we found largely asymmetric interactions in the form of commensalism $\{+/0\}$ and amensalism $\{-/0\}$. Cooperative $\{+,+\}$, competitive $\{-,-\}$, and exploitative $\{+/-\}$ interactions were exceptionally rare (Figure 3.8).

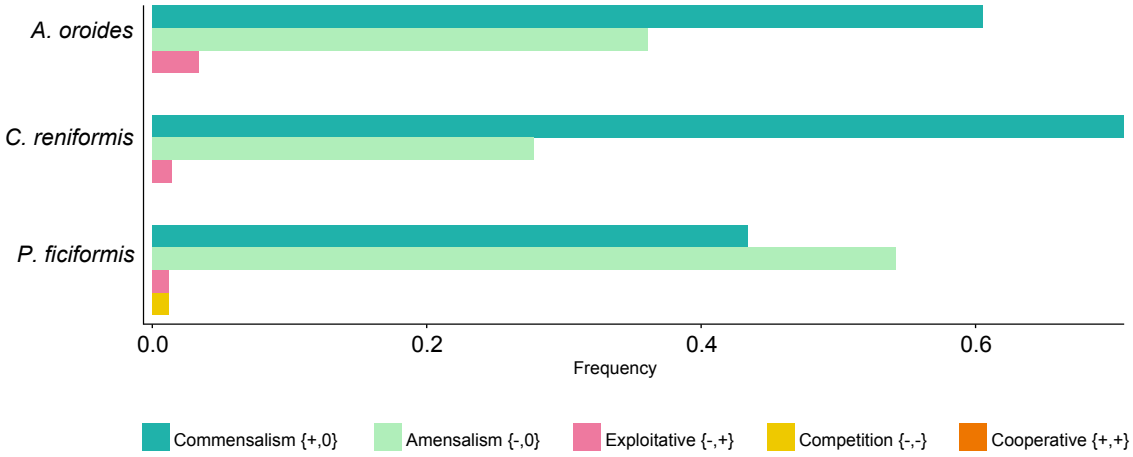


Figure 3.8: Relative frequency of all possible interaction types within HMA cores. Commensalism $\{+,0\}$ and amensalism $\{-,0\}$ were the most frequent interaction types acorss HMA cores. Competitive $\{-,-\}$ and exploitative $\{+/-\}$ interactions were exceptionally rare. Noteworthy, cooperative interactions $\{+,+\}$ were never observed.

Each core network had a mixture positive and negative interactions, and nodes experiencing higher and lower degrees, i.e. the number of in- and out-going links (Figure 3.9, 3.10, 3.11). The core network of, for example, *A. oroides* was dominated by nodes mapping to the phyla *Chloroflexi* and *Gammaproteobacteria*, and some of these nodes also had the highest degree (Figure 3.9). Similarly, *C. reniformis* also had a high proportion of species from the same phyla, although, nodes mapping to the phylum *Chloroflexi* had an overall low degree while nodes mapping to the phylum *Gammaproteobacteria* generally were more connected (Figure 3.9). The core network of *P. ficiformis* was also dominated by *Chloroflexi*, some highly connected. Other well connected nodes in this network mapped to the phyla *SAUL* and *Verrucomicrobia* (Figure 3.10). As noted above, all HMA cores had a high proportion of microbes assigning to 'sponge-specific' clusters. However, we did not observe that nodes mapping to 'sponge-specific' clusters were more connected than other nodes in the networks (Figure 3.9, 3.10, 3.11).

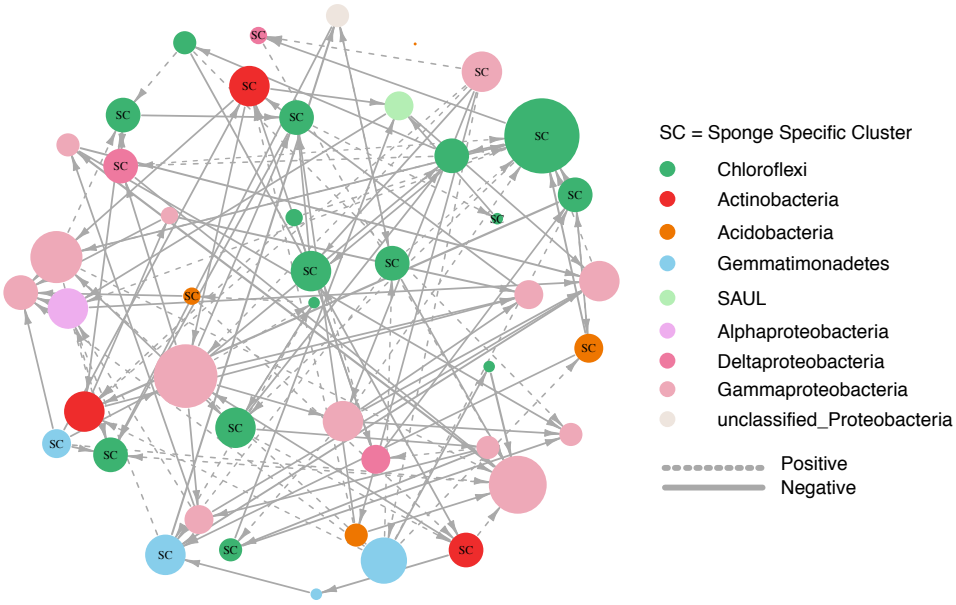


Figure 3.9: Core network for HMA host *A. oroides*. Nodes represent core symbionts and links their interactions. Colours correspond to different bacterial phyla, and dash and solid lines represent positive and negative interactions, respectively

3.4 Discussion

In order to increase our understanding of processes that govern microbiome assembly, stability and functionality, it is critical to incorporate temporal dynamics. Recent studies have shown that microbiomes are examples of highly diverse dynamical systems. Most of these studies have focused on the human microbiome (Caporaso, Lauber, Costello, *et al.* 2011; Flores, Caporaso, *et al.* 2014; Backhed, Roswall, *et al.* 2015), but a few studies have explored dynamics in other microbe-host systems (Björk *et al.* 2013; Pitta *et al.* 2014).

Microbes not continuously present in our microbiomes usually occurred at smaller densities than persistent core microbes. We found a high proportion of occasionally rare microbes within the sponge microbiomes, in agreement with that observed in non-microbial assemblages (Magurran & Henderson 2003). However, we observed many orders of magnitude more occasional than persistent core microbes, and sometimes these occasional microbes were present at large abundances. This suggests that occasional microbes can become very abundant at times, potentially affecting microbiome dynamics and stability.

We applied two related measures of invariability to describe temporal stability at the population- and community-

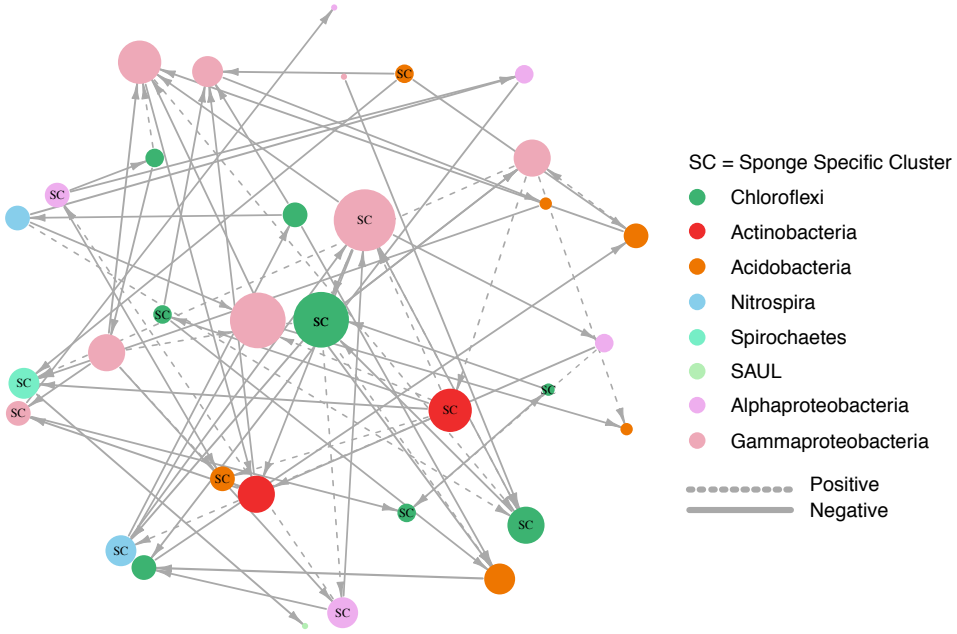


Figure 3.10: Core network for HMA host *C. reniformis*. Nodes represent core symbionts and links their interactions. Colours correspond to different bacterial phyla, and dash and solid lines represent positive and negative interactions, respectively

level. Although we did not focus on measuring synchrony per se, it is interesting to note that the ratio of the two invariability measures reflect community-wide synchrony (Haegeman et al. in review). Synchrony is linked to community stability, where a larger diversity tends to increase the potential for species asynchrony, thus stabilising community-level properties, while simultaneously decreasing population-level stability (Loreau 2010). We hypothesised that the higher diversity found in the cores of High-Microbial Abundance (HMA) hosts when compared to lower diversity observed in the cores of Low-Microbial Abundance (LMA) hosts would confer larger temporal stability to HMA hosts.

However, we observed two different temporal dynamics across sponge microbiomes that did not perfectly match the HMA-LMA dichotomy. In three hosts (two HMA and one LMA), cores accounted for the majority of microbiome relative abundance, resulting in high-density cores. In the remaining three hosts (two LMA and one HMA), the relative abundances of the cores were similar to that of the transient and/or opportunistic assemblages, resulting in low-density cores. Low-density cores tended to vary more over time than high-density cores, suggesting that core stability was negatively affected by the presence of occasionally abundant microbes. In other words, it is likely that high-density cores conferred hosts a resistance against occasional microbes to increase in abundance.

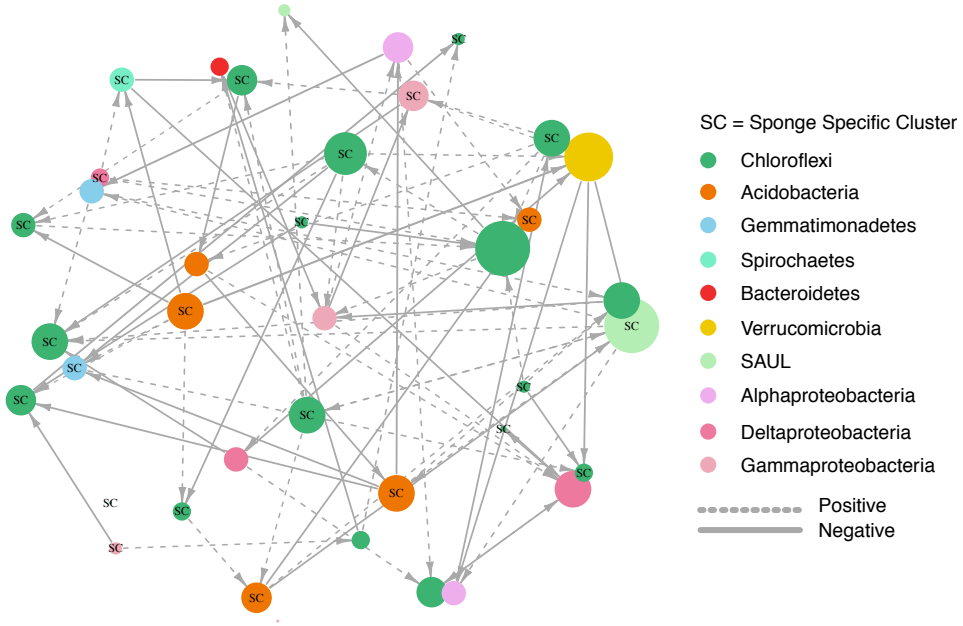


Figure 3.11: Core network for HMA host *P. ficiformis*. Nodes represent core symbionts and links their interactions. Colours correspond to different bacterial phyla, and dash and solid lines represent positive and negative interactions, respectively

Interestingly, high-density cores belong to sponge species known to transmit symbionts vertically from adult to offspring (Levi and Levi 1976; Schmitt, Angermeier, *et al.* 2008; Uriz *et al.* 2001), while low-density cores correspond to host with larvae deprived of microbes (Lepore et al. 1995; Riesgo & Maldonado 2008; Maldonado & Riesgo 2009). Vertical transmission provides an evolutionary mechanism for preserving particular combinations of symbionts, including their associated functioning (Thompson 2005). For example, the high-density core of LMA host *C. crambe*, although temporally stable, was dominated by a single *Alphaproteobacteria*. In agreement with the HMA-LMA dichotomy, the metabolic profiles of *C. crambe* and *P. ficiformis* match those of other LMA and HMA hosts, respectively (Morganiti et al. in review). The low-density core of HMA host *P. ficiformis* was temporally variable, but harboured a high diversity of microbes, whereof the majority assigned to sponge-specific clusters. In fact, *P. ficiformis* had the highest overall proportion of sponge-specific clusters, indicating that *P. ficiformis* displays HMA characteristics by horizontally selecting symbionts from the water column. It further suggests that sponge-specific clusters, even though present in the water column and sediment as part of the rare biosphere (Webster, Taylor, *et al.* 2010; Taylor, Tsai, *et al.* 2013), can be crucial for host functionality. As the planktonic microbial community is typically characterised by high variability and seasonality (Gilbert, Steele, *et al.* 2012; Cram *et al.* 2015), therefore by selecting symbionts from the water column, it is inevitable that the core of *P. ficiformis* partially mimics the variability

present in the plankton.

Our results suggest that high-density cores conferred hosts a resistance against temporally occasional microbes to increase in abundance. We hypothesise that vertical transmission underpinned the observed temporal stability of high-density cores. The arrival order of species is known to heavily affect community assembly and stability (Chase 2010, Fukami 2015), and the process of vertical inheritance of symbionts will likely have similar outcomes as classical priority effects. The complementary group of symbionts that are transmitted from adult to offspring preempt the host niche by fast reaching carrying capacity. Microbes are also known to readily modify their host niche (Nogueira *et al.* 2009; McNally *et al.* 2014), thereby further inhibiting the colonisation of some microbes, while facilitating the establishment of others (e.g. horisontally selected microbes) (Fukami 2015). Although, the high-density core of *C. crambe* was dominated by one microbe, its total microbiome richness was the highest across all studied hosts. This indicates that priority effects, through vertical transmission, may strongly affect the temporal stability of microbe-host systems.

We found that individually persistent microbes generally occurred at larger densities and displayed higher temporal stability than occasional microbes. Henderson & Magurran 2014 observed similar patterns in a non-microbial community. The authors showed that density-dependence underpinned the temporal stability of core species, while suggesting that the dynamics of occasional species were governed by environmental stochasticity. In other non-microbial communities, environmental stochasticity has been shown to be the single most important determinant of community dynamics (Mutshinda *et al.* 2009; Mutshinda *et al.* 2011; Almaraz & Oro 2011; Almaraz, Green, *et al.* 2012; Martorell & Freckleton 2014; Crone 2016). While environmental stochasticity explained a relatively low proportion (between 11 and 25%) of temporal variation in microbial abundances across all cores, we observed that density-dependence explained the largest proportion (between 51 and 80%). While density-dependence can increase the stability of communities close to carrying capacity by dampen fluctuations caused by environmental stochasticity, it can also destabilise communities by facilitating self-sustained population cycles (Stenseth Nils Chr *et al.* 2003; Ims *et al.* 2008). All together, it is possible that density-dependence helped to further stabilise high-density cores, whereas, in the low-density cores, density-dependence may have increased variability together with fluctuations caused by non-modelled stochasticity (e.g. occasional species attaining large abundances). Core dynamics of LMA host *A. damicornis*, for example, was largely driven by density-dependence, and simultaneously displayed one of the most variable microbiomes.

While interspecific interactions had an almost negligible contribution to the dynamics of LMA cores, it explained a comparable and relatively large proportion (between 21 and 31%) of the dynamics in HMA cores. Interestingly, these cores were characterised by similar interactions, both in terms of interaction strength and nature. Theoretical results have shown that that exploitation $\{+/-\}$, cooperative $\{+/+\}$ and competitive $\{-/-\}$ interactions differ in their effects on population and community stability (Mougi & Kondoh 2012; Sauve *et al.* 2014; Lurgi, Montoya, *et al.* 2015). A recent mathematical analysis showed that cooperative interactions are predicted to decrease microbiome stability by introducing strong species coupling and positive feedbacks. On the other hand, competition among microbes should help to maintain stability by counteracting the destabilising effect of high species diversity (Coyte *et al.* 2015). While this might be the case for the whole microbiome, we hypothesised that selection acts to minimise competitive interactions between core symbionts, thereby simultaneously reinforcing selection for complementarity.

In all HMA hosts, we observed a skewed distributions of interaction strengths towards a few strong and many weak interactions. This pattern has been found in many empirical networks (Paine 1992; Fagan William F. & Hurd L E. 1994; Wootton 1997; Emmerson & Raffaelli 2004; Raffaelli & Hall 1996; Rooney & McCann 2012; Vazquez, Lomascolo, *et al.* 2012) and has been suggested to promote community and ecosystem stability (Emmerson & Yearsley 2004; Wootton & Emmerson 2005).

We found that HMA cores were characterised by asymmetric interactions with similar degrees of commensalism and amensalism. While the cores of hosts *A. oroides* and *C. reniformis* were largely dominated by commensal $(+, 0)$ interactions, the core of host *P. ficiformis* had a higher frequency of amensalism $(-, 0)$, and it was the only core that displayed competitive interactions among its members, although at very low frequencies. It is likely that the observed difference is related to the mode of symbiont acquisition. In the case of hosts *A. oroides* and *C. reniformis*, cores most likely correspond to vertically inherited symbionts, while the core of host *P. ficiformis* likely represent environmentally acquired symbionts. Microbial interspecific interactions are highly dynamic, with sign and strength largely contingent on the exchange of metabolic by- and exo products (Estrela & Brown 2013; Estrela, Whiteley, *et al.* 2014). While metabolic by-products are used as nutrients by other microbes, exo-products are often harmful toxins excreted in chemical warfare between competing microbes and/or molecules secreted in interspecific communication stimulating, for example, biofilm formation (Bassler & Losick 2006). During biofilm formation, microbes undergo a drastic physiological change from a free-living (planktonic) to a sessile state, also radically changing their gene and protein expressions (Peters *et al.* 2012). The formation of complex multispecies biofilms facilitate cross-talk between

microbes, enhancing both functionality and protection (James *et al.* 1995; Burmolle *et al.* 2014). Interestingly, Zan *et al.* 2012 found that symbionts in the sponge *Mycale laxissima* used cross-talk to activate their locomotion when population densities became too high, limiting their own biofilm formation. Furthermore, we observed a skewed distribution of interactions strengths towards many weak and a few strong interactions. Microbes are known to modulate their interaction strength by increasing or decreasing their spatial distance to neighbouring species and the level of interspecific cell-mixing (Stacy *et al.* 2014).

It would be interesting to experimentally validate the type of interactions our model infers and deems as the most credible. However, to our knowledge, such experiments have not been performed. This might be due to the difficulty of cultivating microbes that are abundant in sponges (i.e. likely core microbes) (T. Thomas and M Ribes, personal communication). Assuming it would be possible growing core microbes in the lab, the highly dynamic nature of microbial interactions make it essentially impossible to ”recreate” their interactions. In addition, without the coevolved environment of the sponge, the interactions between core microbes may likely look very different (S. Brown, personal communication).

Core symbionts of host *P. ficiformis* likely experience higher levels of competition with other invasive, non-symbiotic microbes, suggesting that the higher observed frequency of amensal interactions among these symbionts may simply reflect collateral damage.

The innate immune defense of some sponges can differentiate between pathogens, food bacteria and symbionts in a manner similar to the adaptive immune system of vertebrates (Wilkinson 1978; Wehrl *et al.* 2007; Wiens *et al.* 2007; Thomas *et al.* 2010; Yuen *et al.* 2014; Degnan 2015). Interestingly, some sponges have been found to compartmentalised symbionts in specialised cells called bacteriocytes (Boury-Esnault & Rützler 1997; Maldonado 2007). As symbionts metabolise various inorganic compounds, including waste products produced by the host, compartmentalising symbionts within bacteriocytes allow the host to easy transport the desired combination of symbionts to the required location where they can be released (Maldonado 2007). This further illustrate the selective capacity of some sponge hosts.

Our study highlights the importance of defining core microbiomes through time, rather than only across individuals or species. Our results show that the microbe-sponge system consist of highly dynamic interactions, with core microbiomes belonging to HMA hosts consisting of complementary symbionts only weakly and indirectly interacting with on another. Our results further suggest that the observed microbe-microbe interactions are a result of

a mutual dependency between microbes and host. Both parties have the capacity to actively modify their interactions. The sponge host and core symbionts have likely coevolved in ways which allows for maintaining functionality and fitness over ecological and, probably, evolutionary, time scales.

[1]

[1]This chapter represents a collaboration with R. O'Hara, M. Ribes, R. Coma and J M. Montoya

Uncovering the drivers of microbe-host interactions with joint species distribution models

4.1 Introduction

What drives observed variability in host-associated microbial communities?

Baas Becking early put forward the hypothesis that "everything is everywhere, but, the environment selects" (Baas Becking 1934, see also De Wit & Bouvier 2006). This statement reflects a longstanding contradiction – the idea that microbes due to their small size, large population abundances, and high dispersion rates, are cosmopolitan, whereas culture dependent observations early on showed that some microbes are specific to certain environments. The idea that microbes are uniformly distributed persisted into the 21th century (Finlay 2002) and was only recently challenged (Hillebrand *et al.* 2001; Horner-Devine *et al.* 2004; Bell *et al.* 2005; Reche *et al.* 2005; Green & Bohannan 2006; Bell 2010; Martiny *et al.* 2011; Astorga *et al.* 2012; Ranjard *et al.* 2013; Wang, Shen, *et al.* 2013; Zinger, Boetius, *et al.* 2014). These studies demonstrated that, for example, the distance-decay and taxa-area relationship that are often observed for animals and plants also exist for free-living microbial assemblages in both terrestrial and aquatic environments.

Microbes form intricate relationships with most animal and plants, many of which show high degrees of mutual dependency and high specificity. Such intimate associations include, but are not limited to, Buchnera-aphids, Rhizobia-legumes, and Wolbachia-arthropods (see Ruby 2008 for a review). In these cases of extreme pairwise reciprocity, microbe and host spatial distributions are tightly coupled due to phylogenetic constraints that reduce the potential for any biogeography. However, more diverse and complex host-associated microbial communities are increasingly found (Robinson *et al.* 2010), perhaps best illustrated by the Human Microbiome Project (The Human Microbiome Project Consortium 2012), studying the diverse consortia of microbes associated to the human body.

Studies have found, for instance, that microbial communities on humans cohabiting are more similar compared to humans not living together (Song *et al.* 2013), and although humans are born with a sterile gastrointestinal tract (Mackie *et al.* 1999), relatives harbour more similar gut microbiomes compared to unrelated individuals (Turnbaugh, Hamady, *et al.* 2009; Yatsunenko *et al.* 2012; Faith *et al.* 2013), irrespective of cohabiting or not (Faith *et al.* 2013).These examples illustrate well the potentially confounding effects of shared environment, lifestyle and host genetics on structuring host-associated microbial communities compared to free-living species assemblages. However, little is known about the processes driving variability in host-associated communities beyond humans. In parasite communities, such as helminths and fleas parasitising mammals, community similarity decays with geographical distance as host vagility and environmental continuity strongly affect parasite dispersal (Poulin 2003; Krasnov *et al.* 2005; Brouat & Duplantier 2007).

Discerning amongst processes through joint species distribution models

In order understand the drivers of variability in host-associated microbial communities, we need to understand how a suite of processes related to host–specific features structure these communities. At the same time, to analyse such data we require a unifying, model-based framework capable of discerning among the various processes acting on the species community. Motivated by these challenges, we developed a Bayesian hierarchical joint species distribution modelling framework that analyses microbial cooccurrences across sites as a function of several important host features.

Classical species distribution models correlate species occurrences with environmental variables, such as climate and topography, by fitting each species independently and predicting its distribution (Austin 2002). On the other hand, in joint species distribution modelling, multiple species are fitted simultaneously in order to assess the shared environmental response separately from other abiotic and/or biotic processes that give rise to the observed cooccurrence patterns (Clark *et al.* 2014; Pollock *et al.* 2014).

We focus on discerning among the effects of host traits, host identity and host phylogenetic relatedness. To our knowledge, host traits and host phylogenetic relatedness have not explicitly been considered in joint species distribution models before, but are particularly important in order to disentangle the effects of environmental filtering and biotic interactions, including those between microbes and hosts. A key aspect of the models we proposed is the inclusion of latent variables, serving a number of important purposes. Firstly, they allow for performing model-based unconstrained ordination (Hui, Taskinen, *et al.* 2015; Hui, Warton, *et al.* 2016, and secondly, they properly account

for 'hidden' residual covariation in microbial cooccurrences not explained by the included host features (Warton, Blanchet, *et al.* 2015). Latent variables model residual covariation as a linear combination of underlying, missing predictor variables with associated coefficients (loadings) quantifying individual species' responses to the latent predictors. If the latent variables were known predictors, their loadings would simply correspond to standard regression coefficients (Warton, Blanchet, *et al.* 2015). Latent variables capture the effect of biotic interactions, missing environmental parameters and/or underlying gradients (Hui, Taskinen, *et al.* 2015). Compared to traditional distanced-based methods, employing a model-based framework for studying host-associated microbial community data offers a number of advantages, with the central one being a straightforward interpretation of the various parameters in the model reflecting different biological processes in play.

The models we build aim to simultaneously investigate a set of fundamental, but non-mutually exclusive questions of interest. First, we ask three questions focusing on important host features: (i) Are microbial communities associated with hosts with similar traits more similar irrespective of their of geographic location? (ii) Are microbial communities associated with the same host species more similar regardless of location? (iii) Are microbial communities associated with phylogenetically closely related hosts more similar irrespective of geographic location? Following these three primary questions, we then asked two additional questions concerning the geographical variation present in our data, but which we do not explicitly model: (iv) Do closely located host species harbour more similar microbial communities than hosts located farther apart? (v) Is similarity among host-associated microbial communities distributed along a geographical and/or environmental gradients?

Sponge-associated microbial communities as a study system

We applied our models to sponge host-associated microbial communities. We conducted a meta-analysis of 16S rRNA sponge-derived sequences submitted to GenBank. Our modelled data represent the most abundant microbes from 19 sponge hosts sampled across 48 different locations worldwide.

Marine sponges (*Porifera*) are sessile filter-feeding metazoans common to all benthic habitats across the world. They represent one of the invertebrate phyla harboring the largest microbial diversity (Webster & Taylor 2012; Hentschel, Piel, *et al.* 2012), and compared to the human microbiota which is dominated by only a few phyla (Dethlefsen *et al.* 2007), sponges consistently harbour over 32 phyla and candidate phyla (Schmitt, Hentschel, *et al.* 2012; Reveillaud *et al.* 2014).

Sponges are commonly divided into two groups reflecting various morphological and physiological traits– High Microbial Abundance (HMA) and Low Microbial Abundance (LMA) sponges, respectively. While this division, in a strict sense, refer to the abundance of microbes harboured by the sponge, HMA sponges have a denser interior, including narrower aquiferous canals and smaller choanoyctes compared to LMA sponges (Vacelet & Donadey 1977). LMA sponges, on the other hand, have an architecture fitted for high pumping rates. As a result, HMA sponges harbour denser microbial communities and rely on nutrients produced by their symbionts, whereas LMA sponges have a higher intake of particulate organic matter (Jiménez *et al.* 2007; Weisz *et al.* 2008; Schläppy *et al.* 2010; Freeman & Thacker 2011; Freeman, Thacker, *et al.* 2013; Freeman & Class Freeman 2014; Poppell *et al.* 2014).

Many HMA sponges vertically transmit microbes from parent to offspring, whereas larvae from most LMA sponges are essentially sterile (Ereskovsky & Tokina 2004; Maldonado 2007; Schmitt, Weisz, *et al.* 2007; Gloeckner, Lindquist, *et al.* 2013). In addition, microbes associated to LMA sponges also tend to be present in the plankton (Erwin, Olson, *et al.* 2011; Giles *et al.* 2013; Moitinho-Silva *et al.* 2014). However, as both modes of microbial transmission co-occur in species from either classification, these sponge host-associated microbial communities are likely a mixture of vertically inherited and horizontally selected microbes. Although, some sponge-associated microbial communities may better reflect past evolutionary history, others may mirror current ecological processes (Taylor, Radax, *et al.* 2007). As a consequence, we expect LMA hosts to display a higher similarity among themselves within local communities when compared to HMA hosts.

Most previous studies on geographical variation in sponge host-associated microbial communities have focused on one or a few sponge species across narrow geographical ranges (Hentschel, Hopke, *et al.* 2002; Taylor, Schupp, Dahllof, *et al.* 2004; Webster, Negri, *et al.* 2004; Taylor, Schupp, De Nys, *et al.* 2005; Lee, Wong, *et al.* 2009; Anderson *et al.* 2010; Yang *et al.* 2011; Pita *et al.* 2013; Burgsdorf *et al.* 2014). These studies have either reported high spatial stability (Hentschel, Hopke, *et al.* 2002; Taylor, Schupp, Dahllof, *et al.* 2004; Webster, Negri, *et al.* 2004; Taylor, Schupp, De Nys, *et al.* 2005; Pita *et al.* 2013) or large levels of differentiation among microbial communities depending on location (Lee, Wong, *et al.* 2009; Anderson *et al.* 2010; Yang *et al.* 2011; Burgsdorf *et al.* 2014). In order to understand which factors drive such variation, we need to go beyond describing patterns and develop quantitative approaches that allow for a mechanistic insight into the key ecological processes driving this type of communities.

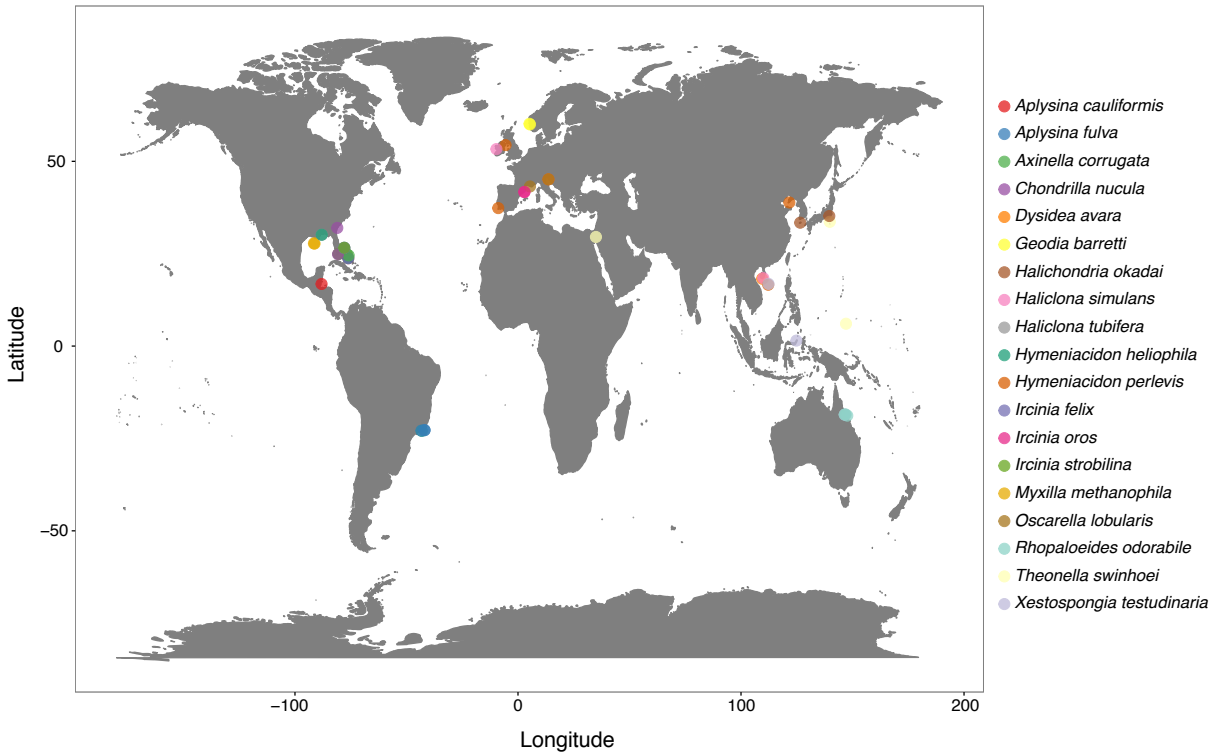


Figure 4.1: Geographic location of all sponge derived 16S rRNA gene sequences coloured by host species. Note that some sites contain multiple host species hence their points are overlaid. See Table 4.1 for more detailed information.

4.2 Methods

Sequence meta-analysis

To assess variation in microbial cooccurrences across different host species sampled in different locations, we conducted a meta-analysis of sponge-associated 16S rRNA gene sequences available in GenBank between September 2007 and August 2014. Only sequences which were above 350 base pairs in length and had sufficient meta data were extracted. The meta data that we deemed necessary were scientific host species identity, sample location down to coordinates, and that samples had been sequenced using either culture, DGGE and/or clone-libraries, as these methods only consider the most abundant microbes. These criteria yielded more than 17.000 sequences from almost 250 host species. For the ease of data management, sequences and corresponding meta data were stored and filtered in a PostgreSQL ($www.postgresql.org$) database. For each sequencing method, sequences were filtered by host species and sample location, creating three primary data sets. To be included in any of the data sets, each host species had to

be represented by at least two different sample locations in which the host species was required to be associated with at least 10 different sequences. Of these three data sets, we only analysed the one containing clone-library sequences as it was the largest in terms of the number of host species meeting the above criteria. The analysed data set contained a total of 3874 16S rRNA sequences with an average length of 996 base pairs. Sequences were derived from 9 High and 10 Low Microbial Abundance sponges, sampled from a total of 48 different locations (n_{HMA} =28, n_{LMA} =20) across the Atlantic, Pacific Ocean, Mediterranean and Red Seas (Figure 4.1, Table 4.1). Sequences were aligned and clustered into Operational Taxonomic Units (OTUs) using mothur v.1.32.1 (Schloss *et al.* 2009). We clustered sequences at three thresholds of nucleotide similarity representing order (85%), family (90%), and genus (95%) level, respectively (Webster, Taylor, *et al.* 2010; Schmitt, Tsai, *et al.* 2012). At higher and lower sequence similarity, OTU clusters tended to become either too specific or too broad, thus generating too sparse data.

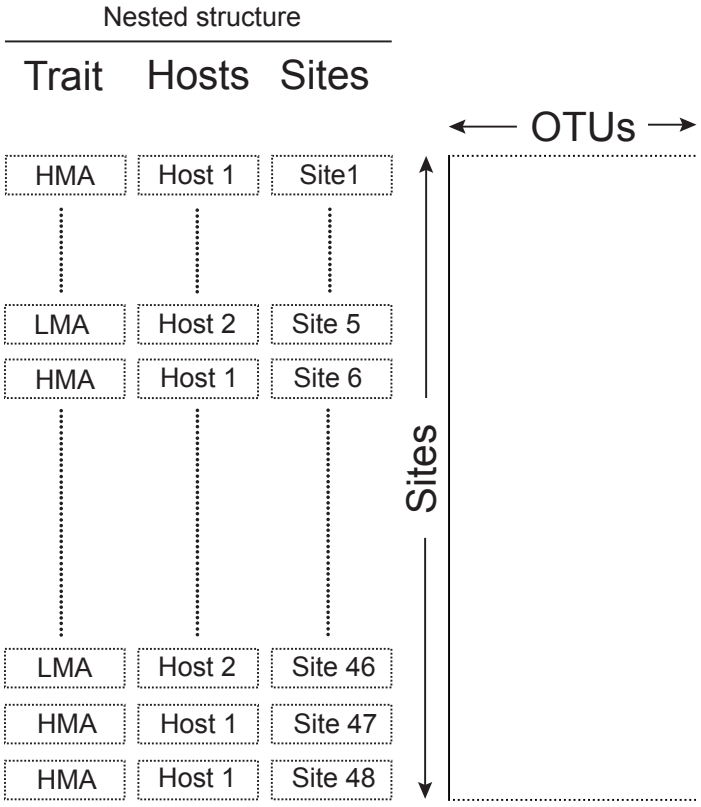


Figure 4.2: Schematic figure of the modelled OTU-site matrix. While columns correspond to OTUs, rows have a nested structure; sites are nested under host species which are nested under host trait. In this example, Site 1, 6, 47 and 48 are nested under Host 1 which is nested under host trait HMA. Site 4 and 46 are nested under Host 2 which is nested under host trait LMA. This is an inherent structure of most host-associated microbial community and microbiome data that can be accounted for.

Joint Species Distribution Models

We developed a Bayesian joint species distribution modelling framework (Clark *et al.* 2014; Pollock *et al.* 2014; Warton, Blanchet, *et al.* 2015) that jointly models the cooccurrence of microbes across sites while accounting for host species identity, host traits (HMA and LMA, hereafter termed *ecotype*) and host phylogenetic relatedness. The model incorporates latent variables as a means of performing model-based ordination (Hui, Taskinen, *et al.* 2015; Hui, Warton, *et al.* 2016) and accounting for residual correlation between OTUs not explained by the processes included in the model, for example, biotic interactions, geographical distance, and/or missing environmental parameters (Morales-Castilla *et al.* 2015).

The response matrix being modeled consisted of counts of n OTUs observed at m sites. Due to their lack of information, OTUs with a total abundance of less than 3, and/or OTUs with less than 3 presences across sites were removed. This resulted in 83, 120, and 117 OTUs at order (85%), family (90%), and genus (95%) level, respectively, across 48 sites and 19 host species. It is important to highlight that the data features a nested structure which needs to be taken in account in the model (Figure 4.2). Specifically, the $m = 48$ sites are nested within the 19 host species, with the 19 host species nested within one of two ecotypes (Figure 4.2).

Due to the presence of overdispersion in the counts, a negative binomial distribution with quadratic mean-variance relationship was assumed for the response y_{ij}, such that $\text{Var}(y_{ij}) = \mu_{ij} + \phi_j \mu_{ij}^2$ where ϕ_j is the OTU-specific overdispersion parameter. Mean abundance was related to the covariates using a log link function. We denote the response and mean abundance of OTU j at site i by y_{ij} and μ_{ij} respectively. Let $\mathcal{N}(\mu, \sigma)$ denote a normal distribution with mean μ and standard deviation σ, and analogously, let $\mathcal{MVN}(\mu, \Sigma)$ denote a multivariate normal distribution with mean vector and covariance matrix Σ. Then, we have the model formulation (e.g., for the family-level data set) as follows

$$y_{ij} \sim \text{Negative-Binomial}(\mu_{ij}, \phi_j); i = 1, \ldots, 48; \quad j = 1, \ldots, 120 \tag{4.1}$$

$$\log(\mu_{ij}|\boldsymbol{z}_i) = \alpha_i + \beta_j + \boldsymbol{z}_i^T \boldsymbol{\theta}_j \tag{4.2}$$

$$\beta_j \sim \mathcal{N}(0, 10)$$

$$\alpha_i \sim \mathcal{N}(\mu_i, \sigma(\text{host}))$$

$$\mu_i = \mu(\text{host})_{s[r]} + \lambda * \mu(\text{phylo})_s; \quad r = 1, 2; \quad s = 1, \ldots, 19 \tag{4.3}$$

$$\mu(\text{host})_{s[r]} \sim \mathcal{N}(\mu(\text{ecotype})_r, \sigma(\text{ecotype}))$$

$$\mu(\text{ecotype})_r \sim \mathcal{N}(0, 10)$$
$$\boldsymbol{\mu}(\text{phylo})_s \sim \mathcal{MVN}(\mathbf{0}, \boldsymbol{\Sigma}(\text{phylo})).$$

In Equation *4.1*, the quantity β_j represent the OTU effects, which account for differences in species richness (columns totals in the response matrix). The quantities α_i represent optional site effects, whose inclusion serves two main purposes. First, it means that the resulting ordination constructed by the latent variables, $\boldsymbol{z_i}$, are in terms of species composition only, as opposed to site relative abundance and species composition when site effects are not included (Hui et al. 2014). Second, including α_i allows us to account for the nested structure of the data and its potential effect on site total abundance. In particular, to account for site i being nested within host species s, which in turn is nested within ecotype r, the row effect α_i is drawn from a normal distribution with a mean that is a linear function of both a host-specific mean $\mu(\text{host})_{s[r]}$ and a host-specific phylogenetic effect $\boldsymbol{\mu}(\text{phylo})_s$ (Equation *4.3*) (phylo effect $\mu(phylo)_s$; see Section *Phylogenetic reconstructions* for more details on the phylogenetic effect.). Furthermore, the host effects themselves are drawn from a normal distribution with a ecotype-specific mean $\mu(\text{ecotype})_r$. To clarify the above formulation, the subscript r indexes the ecotype, while s indexes the host species, so that "$s[r]$" means "host species s nested within ecotype r". Note that, in contrast to the means (μ), the variance parameters $\sigma(\text{host})$ and $\sigma(\text{ecotype})$ are common across hosts and ecotypes. This implies that, *a-priori*, the host and ecotypes can differentiate in location but not in dispersion.

We identified the model with and without site effects included, so that two types of ordinations were constructed. In the former, the ordination is only in terms of OTU *composition*, whereas in the latter, it encompasses OTU *composition* and *abundance*. For the model without site effects, α_i and its associated nested structure were removed from Equation *4.2*, such that $\log(\mu_{ij}|\boldsymbol{z}_i) = \beta_j + \boldsymbol{z}_i^T\boldsymbol{\theta}_j$. As discussed above, we include a vector of q latent variables $\boldsymbol{z}_i$ as means of performing model-based ordination. As is conventional with ordination, we set $q = 2$ so that, once fitted, the latent variables $\boldsymbol{z}_i = (z_{i1}, z_{i2})$ can be plotted on a scatterplot to visualize the main patterns in the data (Hui, Taskinen, *et al.* 2015; Warton, Blanchet, *et al.* 2015). The corresponding vector of loadings $\boldsymbol{\theta}_j = (\theta_{j1}, \theta_{j2})$ quantifies each individual OTUs' response to the latent variables e.g., increasing z_{i1} by one unit changes the log of the mean response by a value θ_{j1}.

To complete the above formulation, we assigned priors to the appropriate parameters. For the OTU-specific overdispersion parameters, ϕ_j (Equation *4.1*), we chose to assign a weakly-informative Gamma prior, Gamma$(0.1, 0.1)$. Both the standard deviations for host, $\sigma(\text{host})$, and ecotype, $\sigma(\text{ecotype})$ (Equation *4.2* and *4.3*), were assigned uni-

form priors Unif$(0, 100)$. All the OTU-specfic coefficients related to the latent variables, $\theta_{jk}; k = 1, \ldots, 2$, were assigned normal priors $\mathcal{N}(0, 100)$ while taking to account the appropriate constraints for parameter identifiability (see Hui et al. 2014, for details). Finally, for the scale parameter related to the phylogenetic effect, λ, we assigned a weakly-informative exponential prior with a rate parameter of 0.1.

One advantage of the different effects in the hierarchy in Equation *4.3* is that we can calculate the total variance of the μ_i's, and partition this variance into components reflecting variation in site total abundance attributable to differences in host ecotype $\mu(\text{ecotype})_r$, differences in host species identity $\mu(\text{host})_s$, and differences in host phylogeny $\mu(\text{phylo})_s$, respectively. Such a variance decomposition is analogous to sum-of-squares and variance decompositions seen in Analysis of Variance (ANOVA) and linear mixed models (Nakagawa & Schielzeth 2013; Faraway 2014), although to our knowledge this is the first time such a decomposition has been done with joint species distribution models. Importantly, such a partitioning offers a model-based approach for directly answering how certain certain host features drive the cooccurrence of microbes across hosts (questions i, ii and iii outlined in the *Introduction*).

Let V_{total} denote the total variance of the μ_i's, while $V_{\text{ecotype}}, V_{\text{host}}$, and V_{phylo} denote the variances due to ecotype, host identity and host phylogeny respectively. Then we have,

$$V_{\text{total}} = (\mu(\text{ecotype})_{LMA} - \mu(\text{ecotype})_{HMA})^2 + V_{\text{ecotype}} + V_{\text{host}} + V_{\text{phylo}}, \quad \text{where} \tag{4.4}$$

$$V_{\text{ecotype}} = \sigma^2(\text{ecotype}) \tag{4.5}$$

$$V_{\text{host}} = \sigma^2(\text{host}) \tag{4.6}$$

$$V_{\text{phylo}} = \lambda^2 \tag{4.7}$$

In Equation *4.4*, $(\mu(\text{ecotype})_{LMA} - \mu(\text{ecotype})_{HMA})^2$ is the difference between LMA and HMA sponges in terms of intraspecific variation among host species within each ecotype. Therefore, $(\mu(\text{ecotype})_{LMA} - \mu(\text{ecotype})_{HMA})^2 / V_{\text{total}}$ represents the proportion of total variation in site total abundance driven by host ecotype. On the other hand, $\sigma^2(\text{ecotype})$ accounts for intraspecific variation among host species *within* each ecotype. That is, if the proportion $V_{\text{ecotype}}/V_{\text{total}}$ is small compared to $(\mu(\text{ecotype})_{LMA} - \mu(\text{ecotype})_{HMA})^2/V_{\text{total}}$, then host species are more similar within compared to between ecotype. Next, $\sigma(\text{host})$ reflects the intraspecific variation among locations within host species, and hence small values of $V_{\text{host}}/V_{\text{total}}$ imply that sites within the same host species are more similar within than between host species. Finally, λ^2 corresponds the intraspecific variation among sites within host species that can be attributed to host phylogenetic relatedness, meaning small values of $V_{\text{phylo}}/V_{\text{total}}$ provide evidence that host phylogeny

has little influence on variation in site total abundance (see the Section *Phylogenetic reconstructions* on how $\Sigma(\text{phylo})$ was constructed).

We used Markov Chain Monte Carlo (MCMC) simulation method by running JAGS (Plummer 2003) in *R* through the *rjags* (Plummer *et al*., 2016) and *runjags* (Denwood, *2015*) packages to sample from the joint posterior distribution of the model parameters. We ran 5 independent chains with dispersed initial values for 100.000 iterations saving every 50th sample and discarding the 100.000 samples of each chain as burn-in. This resulted in 1 million samples. We evaluated convergence of model parameters by visually inspecting trace and density plots using the R packages coda (Plummer *et al.*, 2015) and mcmcplots (McKay Curtis *et al.*, 2015).

Phylogenetic reconstructions

For each host species, 18S rRNA gene sequences were downloaded from Genbank and aligned using the default options in ClustalW (1.83). The phylogenetic relationship between sponges were reconstructed by implementing a HKY + Γ_4 substitution model using BEAST (1.7.4) (Drummond and Rambaut, 2007). For a few host species (*I. oros, H. simulans, M. methanophila and X. testudinaria*), we did not retrieve any 18S rRNA sequences, and in these cases, we constrained the host species to the clade containing its genera.

A posterior distribution of trees were sampled using Markov Chain Monte Carlo (MCMC) simulations as implemented in BEAST. We ran four chains of 20 million generations sampled every 100.000 steps with a burn-in of 25 per cent. We evaluated convergence using Tracer (v1.5). We transformed a set of 2500 trees from the posterior distribution of the sponge phylogeny to variance-covariance matrices. We assumed Brownian motion so that each covariance between host i and host j were proportional to their shared branch length from the most recent common ancestor. We used the distribution of phylogenetic variance-covariance matrices, $\Sigma(\text{phylo})$, as a prior in Equation *4.3*, where $\boldsymbol{\mu}(\text{phylo})_s \sim \mathcal{MVN}(\mathbf{0}, \Sigma(\text{phylo}))$. Note that as the variance i.e., the diagonal elements in each variance-covariance matrix is scaled to one by the construction of $\Sigma(\text{phylo})$, we multiplied it with a scaling factor λ as seen in the formulation in Section *Phylogenetic reconstructions*.

4.3 Results

In this section, we present results for each of the five questions we posed in the *Introduction*. Have in mind that the LVM-based ordinations properly displays the effect of location and dispersion, while we only test for location in the variance decomposition, as we assigned a common variance across hosts and ecotypes (Equation *4.3*).

Table 4.1: Detailed information about geographical location for each host. Each host species is represented by an host ID, its ecotype, sample site, ocean, and latitude and longitude.

Host ID	Host species	Ecotype	Site	Ocean	Lat	Lon
1	Aplysina cauliformis	HMA	Carrie Bow Cay, Belize	Caribbean Sea	16.803	-88.082
1	Aplysina cauliformis	HMA	Lee Stocking Island, Bahamas	North Atlantic Ocean	23.769	-76.099
2	Aplysina fulva	HMA	Caboclo Island, Rio de Janeiro, Brazil	South Atlantic Ocean	-22.755	-41.890
2	Aplysina fulva	HMA	Lee Stocking Island, Bahamas	North Atlantic Ocean	23.769	-76.099
2	Aplysina fulva	HMA	Rio de Janeiro, Brazil	South Atlantic Ocean	-22.875	-43.278
2	Aplysina fulva	HMA	Sweetings Cay, Bahamas	North Atlantic Ocean	26.600	-77.900
2	Aplysina fulva	HMA	Tartaruga beach, Rio de Janeiro, Brazil	South Atlantic Ocean	-22.756	-41.904
3	Axinella corrugata	LMA	Conch Reef, Key Largo, Florida USA	Caribbean Sea	24.950	-80.454
3	Axinella corrugata	LMA	Little San Salvador Island, Bahamas	North Atlantic Ocean	24.548	-75.934
4	Chondrilla nucula	HMA	Grays Reef, USA	North Atlantic Ocean	31.984	-81.019
4	Chondrilla nucula	HMA	Limski Canal, Croatia	Adriatic Sea	45.131	13.663
4	Chondrilla nucula	HMA	Mangrove channel, Florida Keys, USA	North Atlantic Ocean	24.863	-80.717
5	Dysidea avara	LMA	Limski Canal, Croatia	Adriatic Sea	45.131	13.663
5	Dysidea avara	LMA	Sanya Island, China	South China Sea	18.233	109.489
6	Geodia barretti	HMA	Korsfjord, Norway	North Atlantic Ocean	60.153	5.148
6	Geodia barretti	HMA	Langenuen, Norway	North Atlantic Ocean	59.978	5.382
7	Halichondria okadai	LMA	Jeju Island South Korea	East China Sea	33.390	126.540
7	Halichondria okadai	LMA	Miura peninsula, Japan	Pacific Ocean	35.199	139.586
8	Haliclona simulans	LMA	Galway, Ireland	North Atlantic Ocean	53.316	-9.669
8	Haliclona simulans	LMA	Sanya Island, China	South China Sea	18.402	109.994
9	Haliclona tubifera	LMA	Gulf of Mexico, USA	Gulf of Mexico	30.138	-88.002
9	Haliclona tubifera	LMA	Sweetings Cay, Bahamas	North Atlantic Ocean	26.600	-77.900
10	Hymeniacidon heliophila	LMA	Gulf of Mexico, USA	Gulf of Mexico	30.138	-88.002
10	Hymeniacidon heliophila	LMA	Praia Vermelha, Brazil	South Atlantic Ocean	-22.955	-43.163
11	Hymeniacidon perlevis	LMA	Ballyhenry Island, Ireland	North Atlantic Ocean	54.393	-5.575
11	Hymeniacidon perlevis	LMA	Dalian City, China	Yellow Sea	38.867	121.683
11	Hymeniacidon perlevis	LMA	Portugal	North Atlantic Ocean	37.342	-8.852
11	Hymeniacidon perlevis	LMA	Yongxing Island, China	South China Sea	16.600	112.200
12	Ircinia felix	HMA	Exumas, Bahamas	North Atlantic Ocean	24.881	-76.792
12	Ircinia felix	HMA	Sweetings Cay, Bahamas	North Atlantic Ocean	26.560	-77.884
13	Ircinia oros	HMA	Blanes, Spain	Mediterranean	41.673	2.804
13	Ircinia oros	HMA	Tossa de Mar, Spain	Mediterranean	41.720	2.941
14	Ircinia strobilina	HMA	Conch Reef, Key Largo, Florida USA	Caribbean Sea	24.950	-80.454
14	Ircinia strobilina	HMA	Exumas, Bahamas	North Atlantic Ocean	24.881	-76.792
14	Ircinia strobilina	HMA	Sweetings Cay, Bahamas	North Atlantic Ocean	26.600	-77.900
15	Myxilla methanophila	LMA	Bush Hill, USA	Gulf of Mexico	27.783	-91.507
15	Myxilla methanophila	LMA	Green Canyon, USA	Gulf of Mexico	27.740	-91.222
16	Oscarella lobularis	LMA	Limski Canal, Croatia	Adriatic Sea	45.131	13.663
16	Oscarella lobularis	LMA	Marseille, France	Mediterranean	43.197	5.364
17	Rhopaloeides odorabile	HMA	Davies Reef, Australia	Great Barrier Reef, Coral Sea	-18.826	147.641
17	Rhopaloeides odorabile	HMA	Pelorus island, Australia	Great Barrier Reef, Coral Sea	-18.545	146.488
17	Rhopaloeides odorabile	HMA	Rib Reef, Australia	Great Barrier Reef, Coral Sea	-18.492	146.878
18	Theonella swinhoei	HMA	Eilat, Israel	Red Sea	29.531	34.957
18	Theonella swinhoei	HMA	Hachijo-jima Island, Japan	Pacific Ocean	33.633	139.800
18	Theonella swinhoei	HMA	Western Caroline Islands, Palau	Pacific Ocean	6.050	147.083
19	Xestospongia testudinaria	HMA	Manado Bay, Indonesia	Celebes Sea	1.486	124.835
19	Xestospongia testudinaria	HMA	Orpheus Island, Australia	Great Barrier Reef, Coral Sea	-18.560	146.485
19	Xestospongia testudinaria	HMA	Yongxing Island, China	South China Sea	16.833	112.333

(i) Are microbial communities associated with hosts with similar traits more similar irrespective of their of geographic location?

The latent variables were able to capture differences in microbial community structure, i.e. without controlling for site total abundance, present in the *location* between HMA and LMA sponges, with the clearest separation at the family-level (90% similarity) (Figure 4.3 *right*). This is also visible from the variance decomposition, where the proportion of variation attributed to $V_{\text{ecotype}}/V_{\text{total}}$ was smaller than the proportion explained by $(\mu(\text{ecotype})_{LMA} - \mu(\text{ecotype})_{HMA})^2/V_{\text{total}}$ (Table 4.2), showing that sites were consistently more similar in terms of abundance within than between ecotypes.

When controlling for site total abundance, HMA and LMA hosts generally clustered much closer, although, still preserving some degree of separation. The ordinations show a small effect of dispersion but not of location (Figure 4.3 *left*).

Table 4.2: Variance decomposition into components reflecting variation in site total abundance attributable to differences in host ecotype $\mu(\text{ecotype})_r$, differences between ecotype $(\mu(\text{ecotype})_{LMA} - \mu(\text{ecotype})_{HMA})^2$, differences in host species identity $\mu(\text{host})_s$, and differences in host phylogeny $\mu(\text{phylo})_s$, respectively.

	Order	Family	Genus
$V_{ecotype}$	32.3%	33.8%	32.4%
$V_{LMA-HMA}$	38.7%	37.4%	38.9%
V_{host}	27.2%	27.3%	27.1%
V_{phylo}	1.8%	1.5%	1.5%

Note: $V_{LMA-HMA}$ represents the proportion of total variation in site total abundance driven by host ecotype. On the other hand, $V_{ecotype}$ represents the proportion of intraspecific variation among host species *within* each ecotype. That is, if the proportion $V_{ecotype}$ is smaller than $V_{LMA-HMA}$, then host species are more similar within compared to between ecotype. Next, V_{host} reflects the proportion of intraspecific variation among locations within host species, and if small this implies that sites within the same host species are more similar within than between host species. Finally, V_{phylo} represents the proportion of variation attributed to host phylognenetic relatedness. If small, then there is little evidence that host phylogeny influence variation in site total abundance).

(ii) Are microbial communities associated with the same host species more similar regardless of location?

In the model without site effects (i.e. without controlling for site total abundance), sites did not generally cluster based on host species, indicating that sites contained different abundances irrespective of host identity (Figure 4.3 *right*). This was also further supported by a relatively large proportion of variation explained by $V_{\text{host}}/V_{\text{total}}$ (Table 4.2). Interestingly, however, controlling for site total abundance, the latent variables revealed that sites were more similar in terms of microbial composition within than between host species (Figure 4.3 *left*).

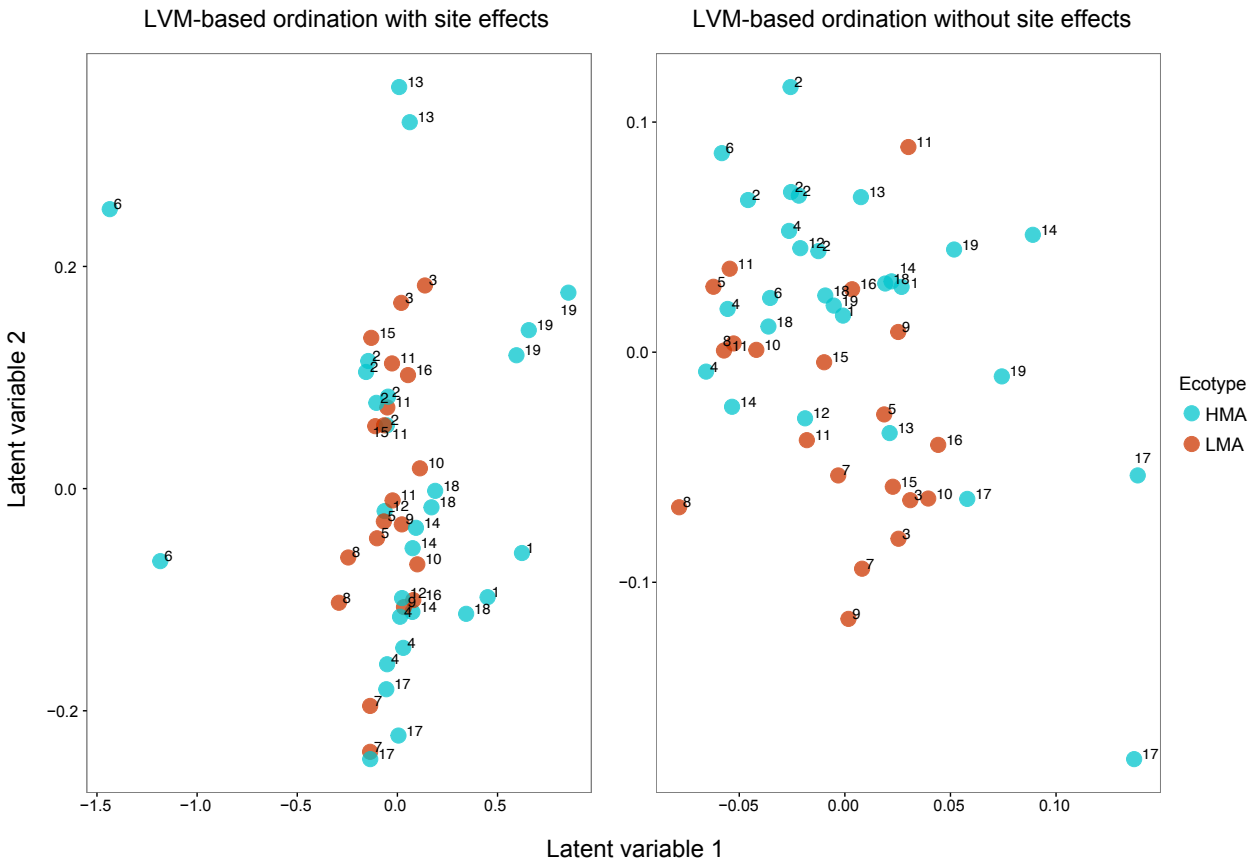


Figure 4.3: Model-based unconstrained ordination with (left) and without (right) site effects included. Teal and red correspond to HMA and LMA hosts, respectively. Host-associated microbial communities are modelled at order-level (85% sequence similarity cutoff), see *Appendix 3* for family- and genus-level plots. Numbers (host IDs) correspond to host species in Table 4.1

(iii) Are microbial communities associated with phylogenetically closely related hosts more similar irrespective of geographic location?

Differences seen in the ordinations could be a result of phylogenetic constraints. However, we found that host phylogenetic relatedness explained an almost negligible proportion of variation in site total abundance across taxonomic levels (Table 4.2).

(iv) Do closely located host species harbour more similar microbial communities than hosts located farther apart?

We found that some of the hosts that formed intraspecific sub-clusters in Figure 4.3 *left* were in fact located in geographically distant sites, in some cases even from different ocean basins (Table 4.1).This shows that some host species harboured similar microbial composition regardless of the geographical distance between their sites. Contrary to our expectations, this was observed for both HMA and LMA hosts. It is specially noteworthy that host species with sites separated by more than 5.000 km (Figure 4.5, Table 4.1), such as HMA host *Chondrilla nucula* (host ID 4) and LMA host *Dysidea avara*, quite consistently formed intraspecific sub-clusters across taxonomic levels (Figure 4.3 *left*, C.1 *left*).

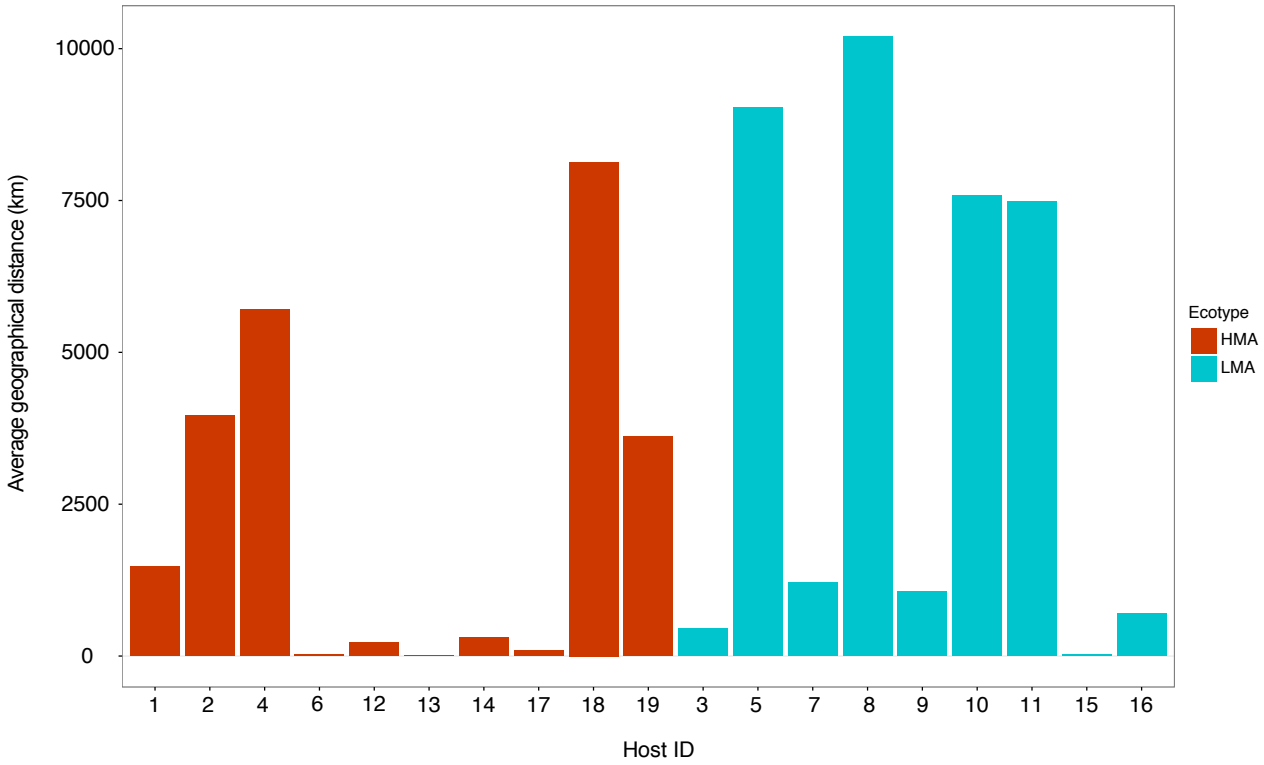


Figure 4.4: Average geographical distance (km) among sites from the same host species. Teal and red correspond to HMA and LMA hosts, respectively. Numbers (host IDs) correspond to host species in Table 4.1

(v) Is similarity among host-associated microbial communities distributed along a geographical and/or environmental gradients?

Controlling for site total abundance, the latent variables collapsed sites onto a clear gradient, where many host species formed intraspecific sub-clusters (Figure 4.3 *left*, C.1 *left*). While this suggests a low intraspecific variability in microbial composition, it also shows that similarity in microbial composition across hosts decreased along a

gradient. Although, intraspecific sub-clusters remained fairly persistent across taxonomic levels, moving from order to genus level, increased the overall similarity among sites, thus reducing the effect of the detected gradient (Figure 4.3 *left*, C.1 *left*).

Moreover, the host species that clustered at each end of the gradient were consistently different across taxonomic levels (Figure 4.3 *left*, C.1 *left*). For example, at the order-level, HMA host *Ircinia oros* (host ID 13) and *Rhopaloeides odorabile* (host ID 17) formed their own intraspecific sub-clusters at either end of the gradient (Figure 4.3 *left*). While both host shared several phyla, they each harboured a unique taxonomic profile. *Ircinia oros* contained several OTUs from *Alphaproteobacteria*, *Cyanobacteria* and *Bacteroidetes* which *Rhopaloeides odorabile* did not house, whereas *Rhopaloeides odorabile* harboured multiple OTUs from *Actinobacteria*, *Chloroflexi* and *Gammaproteobacteria* which where not present in *Ircinia oros*. At the family level, LMA host *Hymeniacidon perlevis* (host ID 11) and HMA host *Ircinia felix* (host ID 12) formed the intraspecific sub-clusters with the largest distance between them (Figure C.1 *left*). While *Hymeniacidon perlevis* housed several OTUs from *Alphaproteobacteria* and *Cyanobacteria* not present in *Ircinia felix*. *Ircinia felix* on the other hand, harboured multiple OTUs from the phyla *Acidobacteria*,*Chloroflexi*, *Gammaproteobacteria* and *Gemmatimonadetes* which *Hymeniacidon perlevis* did not house. At the genus level, most host species appeared in a common cluster less affected by the underlying gradient. Although, not forming closely-knit intraspecific clusters, *Rhopaloeides odorabile* and *Aplysina cauliformis* represented the host with the largest distance between them (Figure C.1 *left*). Interestingly, *Geodia barretti* was the only host that across taxonomic levels never clustered within the detected gradient (Figure 4.3 *left*, C.1 *left*).

It is possible that the detected gradient represents latitude and/or longitude and their underlying environmental gradients. Analysing the latent variables in relation to latitude and longitude revealed that sites collapsed onto a similar gradient as in Figure 4.3 (Figure C.1) when plotting *latent variable 1* against either latitude and longitude. This suggests that the detected gradient corresponds to latitude and longitude, thus importantly encapsulates different environmental gradients, such as temperature and productivity.

4.4 Discussion

The complexity of host-associated microbial communities, particularly microbiomes often preclude their understanding, and therefore we currently lack a mechanistic view of the processes shaping these systems. Here we have presented a Bayesian hierarchical joint species distribution modelling framework that allows for mechanistic insight into some of the important processes governing the cooccurrence of microbes across different hosts. Our

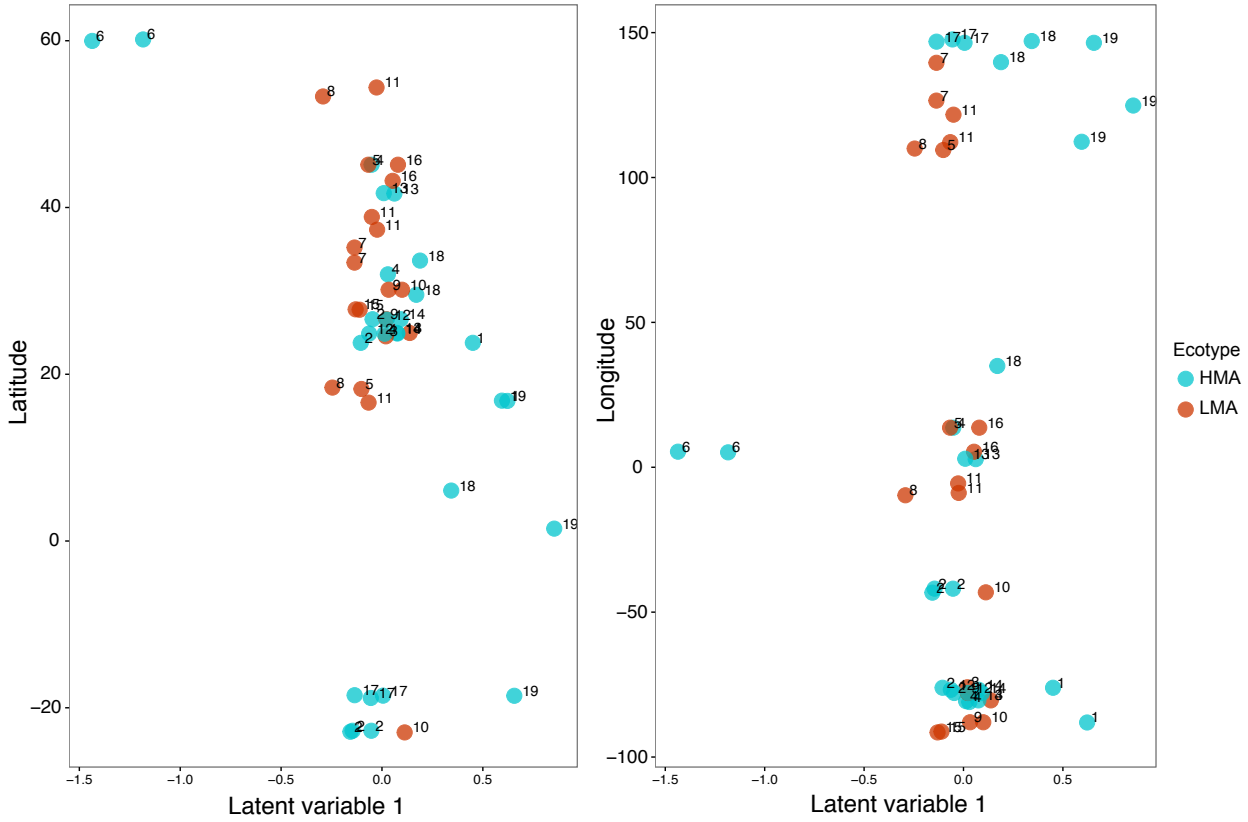


Figure 4.5: Model-based unconstrained ordination with site effects included. Latent variable 1 plotted against latitude (right) and longitude (left). Teal and red correspond to HMA and LMA hosts, respectively. Host-associated microbial communities are modelled at order-level (85% sequence similarity cutoff), see *Appendix 3* for family- and genus-level plots. Numbers (host IDs) correspond to host species in Table 4.1

framework attribute observed variation in microbial cooccurrences to three host features reflecting important processes that are likely driving variation in sponge host-associated microbial communities. The processes we have aimed to distinguish between are; host convergent evolution and/or phylogenetic relatedness, vertical inheritance of symbionts and/or environmental acquisition.

A key aspect of our modeling framework is the inclusion of latent variables as a means of performing model-based unconstrained ordination to visualise the main structure in the data (Hui, Taskinen, *et al.* 2015), and to properly account for uncertainty resulting from extraneous processes not explicitly modelled that would otherwise lead to erroneous inference (Warton, Blanchet, *et al.* 2015). Latent variables capture the effect of biotic interactions, missing environmental parameters (Morales-Castilla *et al.* 2015), and/or as in our case, a geographical gradient.

Although we focused on modelling location, i.e. mean differences (μ), our framework can easily be extended to formally detect dispersion effects by including host and ecotype specific variances. Warton, Wright, *et al.* 2012 showed that many studies that had reported a dispersion effect, for example in species abundance, in fact was due to unequal mean-variance relationships. While traditional distance-based approaches confound location and dispersion effects, our framework properly models the mean-variance relationship, and therefore can accurately detect differences between the two.

While Bayesian hierarchical modelling of species has emerged over the past several decades as a popular approach for analysing single species (Royle & Dorazio 2008; Kery & Royle 2015), particularly given that a hierarchical framework naturally allows for the incorporation of various hypothesised ecological processes and their related uncertainty (Cressie *et al.* 2009), extending them to modelling species assemblages has been largely hampered by the high-dimensional nature of multivariate abundance data. That is, given the number of taxa recorded often has the same order or exceeds the number of sites, as is the case with our data set where we had between 83 and 117 OTUs which was larger than the 48 sites, modeling the covariation between all taxa using an unstructured correlation matrix is often unreliable due to the sheer number of elements in the matrix that need to be estimated (Warton, Blanchet, *et al.* 2015. Indeed, models that do attempt to employ an unstructured residual correlation matrix either consider only a subset of abundant species or require a substantial number of sites (Clark *et al.* 2014; Pollock *et al.* 2014). As reviewed in Warton, Blanchet, *et al.* 2015, latent variables provide an alternative, parsimonious method of modelling species covariation: they assume a low rank representation of the residual correlation matrix, in a manner analogous to principal components analysis and factor analysis, so that most of the covariation is represented by a small number of latent axes. As mentioned above, these axes have a natural interpretation as ordination axes, and in turn, latent variables can be viewed as a model-based approach to ordination. Compared to the unstructured correlation matrix, substantially less parameters are required to model species covariance, with this reduction being greater the more species are modelled.

Over the past two years, various applications of latent variable models have emerged in community ecology. Some examples include spatially explicit latent variable models that jointly estimate the distributions of multiple species without the need for measured environmental parameters (Ovaskainen, Roy, *et al.* 2015; Thorson *et al.* 2015), latent variables models for inferring association networks (Ovaskainen, Abrego, *et al.* 2015), and for determining the environmental factors governing species coexistence (Letten *et al.* 2015). To our knowledge, our analysis is the first to apply latent variable models to analyse microbial data, including host-associated microbial communities. It is also, as

far as we know, the first study that explicitly incorporates host traits, host phylogenetic relatedness in a manner similar to Ives & Helmus 2010, and a variance decomposition framework similar to Mutshinda *et al.* 2009.

Our modelling framework successfully identified important factors reflecting key ecological processes shaping sponge host-associated microbial communities. We found that both the LVM-based ordination (without site effects) and the variance decomposition assigned ecotype (HMA/LMA) as an important determinant of microbial relative abundance across sites. This is consistent with the literature, commonly stating the difference in microbial abundance between HMA and LMA sponges. HMA sponges typically harbour 10^8 to 10^{10} microbes per gram tissue, and as a consequence microbes can make up for as much as 35 per cent of the sponge biomass. LMA sponges on the other hand, typically harbour several orders of magnitude less microbes per gram tissue (Reiswig 1981; Webster, Webb, *et al.* 2001; Hentschel, Piel, *et al.* 2012). While we did observe somewhat of a separation between HMA and LMA hosts in terms of composition, it is clear that their location does not differ much. However, it looks like there could be a difference in dispersal between the two ecotypes. HMA and LMA sponges are commonly found to harbour distinct microbial composition (Schmitt, Hentschel, *et al.* 2012; Moitinho-Silva *et al.* 2014). However, as we modelled the most abundant microbes (only representing a small fraction of the diversity commonly found in sponges) from a large number of sites, overlap in OTU composition among hosts is expected. Including more species, particularly more rare species would likely change these results. One of the major strengths with latent variable models is that a larger number of species can be modelled more effectively, specially rare ones (Warton, Blanchet, *et al.* 2015).

Controlling for site total abundance, thus constructing the latent variables in terms of microbial composition, revealed that many sites were more similar within than between host species. This is likely to be a consequence of either sites being connected by ocean currents or sponges species vertically transmitting microbes from adult to offspring. Interestingly, we found that sites within some of the intraspecific sub-clusters were separated by large geographical distances, even from different basins, indicating vertical transmission of microbes. Perhaps surprisingly, we observed this pattern in both HMA and LMA hosts. Vertical transmission is known to occur in several sponge species, but has so far been discovered in more HMA than LMA sponges (Ereskovsky & Tokina 2004; Maldonado 2007; Schmitt, Wehrl, *et al.* 2007; Schmitt, Angermeier, *et al.* 2008; Gloeckner, Lindquist, *et al.* 2013).

We did not find any effect of host phylogenetic relatedness on site total abundance. This is expected as we did not find an overall tendency for sites to cluster based on host species, but instead, we found a rather large proportion of variation explained by differences among host species. This result contrast that of Easson & Thacker 2014

who found that closely related host species harboured more similar abundance patterns (in terms of the inverse of the Simpson index D) than expected by chance. However, the discrepancy may be a consequence of several methodological differences. Our model framework is conceptually very different from the permutation approach used by Easson & Thacker 2014. Also, Easson & Thacker 2014 only considered OTUs with at least 500 sequences, while we used a lower limit of three.

As we did observe an effect of ecotype on site total abundance, but not of host phylogenetic relatedness, this suggests that ecotype is not a phylogentically conserved trait. This is congruent with the findings of Gloeckner, Wehrl, *et al.* 2014 who performed a larger phylogenetic analysis of sponges classified as either HMA or LMA. Therefore, as these two trait does not seem to be phylogenetically conserved, but correspond to convergent morphological and physiological host features, the observed similarity (abundance and/or composition) is likely due to convergent interactions (Bittleston *et al.* 2016). Such convergence are often observed across unrelated microbial taxa that inhabit convergent morhphological structures, such as the hindgut (Ley *et al.* 2008b; Bittleston *et al.* 2016).

The LVM-based ordination controlling for site effects revealed not only that many sites were more similar within than between host species, but that sites appeared along a clear gradient. While it is impossible to precisely know what each latent variable represent without additional information, they often resemble missing gradients (Hui, Taskinen, *et al.* 2015). In our case, as we did not explicitly model the geographical distance between sites, it is likely that the latent variables detected such a gradient. In fact, we found suggestive evidence that the detected gradient represent the different environmental gradients, such as temperature and productivity, encapsulated by latitude and longitude.

We have successfully applied our developed modelling framework to sponge-associated microbial communities from sites across the globe. However, while our joint species distribution models can be applied to any microbe-host system, they are specifically useful for complex microbiomes, as they can efficiently model a large number of species, particularly the many rare species inhabiting microbiomes, the so-called rare biosphere. In order to move research on host-associated microbial communities, including microbiome research, beyond describing patterns to discern the processes that are shaping them, we need new quantitative modelling approaches that allow for discerning among the multiple determinants shaping complex microbiomes.

[1]

[1]This chapter represents a collaboration with F KC. Hui, R. O'Hara, and J M. Montoya

Global Sponge Microbiome: Diversity, structure and convergent evolution of symbiont communities across the phylum Porifera

5.1 Introduction

Microbial symbionts are essential for the function and survival of multicellular eukaryotes, ranging from humans to invertebrates to plants (Kau *et al.* 2012; Cabreiro & Gems 2013; McFall-Ngai *et al.* 2013; Tkacz & Poole 2015). Most symbioses involve complex communities of microorganisms, often comprising a large phylogenetic breadth of microbial diversity associated with a single host organism. Many factors, including host-derived nutrients, chemico-physical characteristics (e.g. pH) and host properties (e.g. immune response), determine the composition and structure of symbiont communities over time and space. However the evolutionary and ecological drivers of symbiont composition in animals and plants remain largely unknown (Hacquard *et al.* 2015).

Sponges are among the most ancient living Metazoa and generally form symbiotic relationships with complex communities of microorganisms (Taylor, Thacker, *et al.* 2007; Hentschel, Piel, *et al.* 2012; Yin *et al.* 2015). Sponges can maintain highly diverse, yet specific symbiont communities, despite the constant influx of seawater microorganisms resulting from their filter-feeding activities (Fan *et al.* 2012). These symbioses are known to be at least partially underpinned by metabolic exchange between symbiont and host, including nitrogen cycling, CO_2 fixation, secondary metabolite production, and uptake and conversion of dissolved organic matter (Taylor, Radax, *et al.* 2007; Goeij *et al.* 2013). In this respect, sponge symbionts perform analogous functions to the symbionts found in mammalian guts and plants (Hacquard *et al.* 2015). Therefore sponge-microbe symbioses represent an ecologically relevant example of host-microbe interactions in an early-diverging metazoan clade.

While the diversity of sponge symbionts has been extensively addressed using molecular tools, comparative work has been hindered due to methodological differences in sampling, sample processing and data analyses. Large-

scale efforts, such as the Human Microbiome Project (The Human Microbiome Project Consortium 2012; Taylor, Tsai, *et al.* 2013) and the Earth Microbiome Project (Gilbert, Jansson, *et al.* 2014), have standardised these technical aspects to reliably and consistently describe patterns of microbial diversity and composition. These efforts have generated a large knowledge base for host-associated microbiomes of vertebrates, and especially humans, but equivalent datasets for invertebrates are missing. To gain critical insights into the evolution and complexity of symbiotic interactions, we require a greater understanding of the properties and origins of microbial symbioses in early-divergent Metazoa. Furthermore, microbiome research has primarily focused on within-species comparisons, in particular humans, or the comparative analysis of microbiomes of very disparate host organisms (e.g. plants versus mammalian guts) (Hacquard *et al.* 2015). However, to define important aspects for the evolution of microbial symbiosis, a deeper understanding of symbiont communities in closely related host species within defined phylogenetic clades (e.g. a single phylum) is required.

Here, we provide a comprehensive analysis of microbial symbiont communities associated with 81 species from the phylum Porifera. Through a community effort, a total of 804 sponge samples were collected from the waters of 20 countries bordering the Atlantic, Pacific and Indian Oceans as well as the Mediterranean and Red Seas, primarily from shallow water habitats. For environmental comparison, we simultaneously collected 133 seawater and 36 sediment samples as potential sources or sinks of microorganisms associated with sponges (Fan *et al.* 2012). Microbial community composition for each sample was determined using standardised DNA extraction and 16S rRNA gene sequencing protocols established by the Earth Microbiome Project (Gilbert, Jansson, *et al.* 2014). With this extensive data set we aimed to define the diversity, variability, specificity and similarity of symbiont communities across the phylum *Porifera* and determine the interaction patterns and evolutionary forces that shape their complexity and composition.

5.2 Results and Discussion

The complexity of symbiont communities varies greatly across the phylum Porifera

Richness of microbial symbiont communities varies widely across different host species within the phylum Porifera (Figure 5.1). Complexity (as assessed by number of OTUs) ranges from 50 to 3820 genetically distinct symbionts per host. Seawater OTUs were removed from sponge samples as they were considered likely to represent "environmental contaminants" obtained during filter feeding and sampling (see *Methods* for details). The large richness estimates are unlikely to be inflated by sequencing errors as approximately one third of samples reached complete

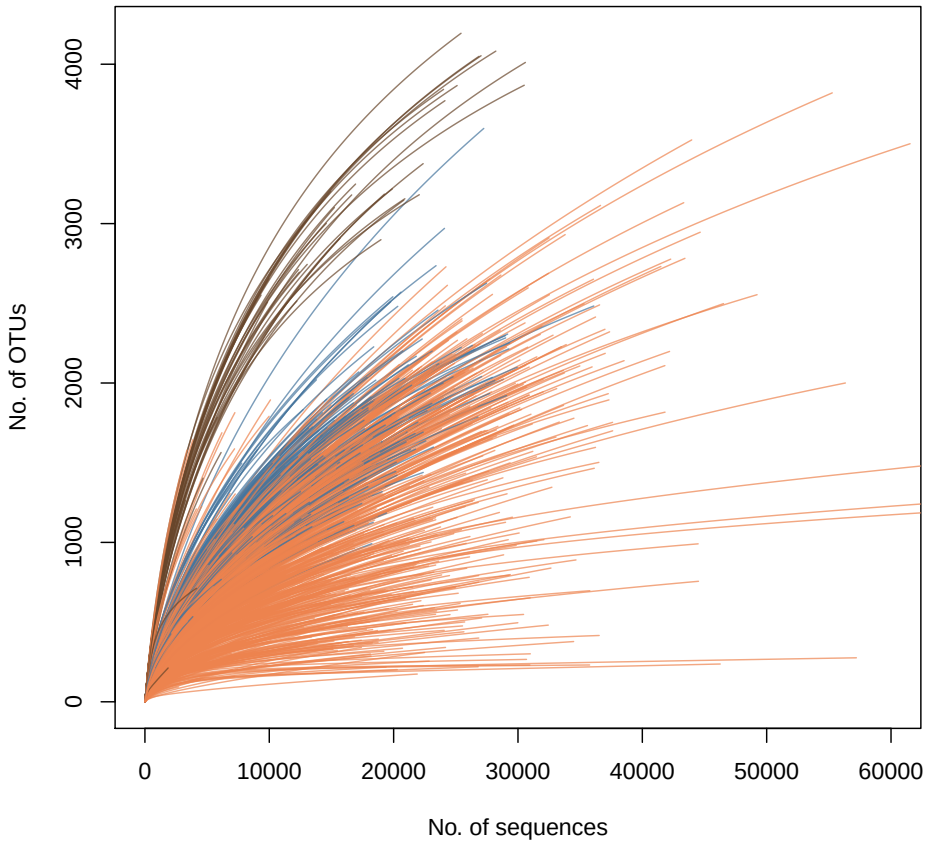


Figure 5.1: Rarefaction curves of 16S rRNA gene diversity of seawater (blue), sediment (brown) and sponge (orange) samples.

saturation (Figure 5.1). Variation of richness across the sponge samples contrasted with the more consistent richness estimates found within seawater and sediment samples (Figure 5.1). The most diverse sponge samples approach the microbial richness found in seawater or sediment, however most sponge species appear to have somewhat less complex communities than the other two habitats.

For symbiont communities of the phylum Porifera we observed a continuum of intraspecific dissimilarities across all species investigated (Figure 5.2). Variability of symbiont communities between individuals of the same host species is indicative of the nature and strength of host-symbiont interactions, ranging from obligate to facultative (Dethlefsen *et al.* 2007; Schmitt, Tsai, *et al.* 2012). Thus low variability would indicate that only specific symbionts can

interact with the host (high specificity), while a relaxed pressure on the interaction would result in higher variability of symbionts among specimens of the same sponge species. Compared to planktonic communities, most sponges maintain low variability within communities (Figure 5.2). Variability was also found to be independent of symbiont diversity or richness (Figure D.1). This indicates a generally restrictive or selective habitat or interactions at the host species level, for both diverse and more depauperate symbiont communities.

The human microbiome is dominated by four phyla (*Firmicutes*, *Bacteroidetes*, *Actinobacteria* and *Proteobacteria*) (The Human Microbiome Project Consortium 2012) and this phylum-level trend has also been observed in other mammals (Ley *et al.* 2008b). In contrast, only the phylum *Proteobacteria* (especially classes *Alpha-* and *Gammaproteobacteria*) was dominant in most sponges species analysed here, with *Chloroflexi*, *Cyanobacteria* and *Crenarchaeota* occasionally reaching high relative abundances (10%). Nevertheless, sponges host a high diversity of phyla (albeit at low relative abundances), with over 32 phyla and candidate phyla regularly reported to associate with sponges (Schmitt, Hentschel, *et al.* 2012) and a further 6 phyla and 14 candidate phyla recently reported as part of the rare community using a deep Illumina sequencing approach (Reveillaud *et al.* 2014). In the current study, we detected 41 phyla (including candidate phyla) with all sponges hosting members of at least 13 different phyla (Figure 5.3).

Sponges harbour an exceptional diversity of marine microorganisms

High sample replication (n>20) employed in this study facilitated estimation of total microbial richness for specific sponge species and seawater. Analysis of the 133 surface seawater samples (collected here from disparate geographic areas, including Spain, Florida, Puerto Rico, Sweden, Mexico, Bahamas and Australia) showed that the combined planktonic richness in these regions approaches 15,000 OTUs (at 97% sequence identity) (Figure 5.4). This estimate lies between the ≈20,000 and ≈9000 predicted OTUs (at 97% sequence identity) found in surface waters of the costal and open ocean, respectively, as part of the International Census of Marine Microbes (ICoMM) (Zinger, Amaral-Zettler, *et al.* 2011), which was based on pyrosequencing analysis of the V6 region of the 16S rRNA gene. However, the estimated planktonic richness in this and our study is lower than the 29,457 OTUs (at 97% sequence identity) recently reported using Illumina amplicon sequencing of seawater (Reveillaud *et al.* 2014) or the 37,470 OTUs estimated from metagenomic sequencing of the global Tara Oceans samples (Sunagawa *et al.* 2015), with the higher richness in the latter two studies likely explained by the inclusion of deep-water samples. Remarkably though, richness estimates show that a single sponge species can harbour as many different OTUs as might be expected from the surrounding seawater. For example, *Carteriospongia foliascens* and *Ircinia variabilis* ($n \geq 50$ individuals across their

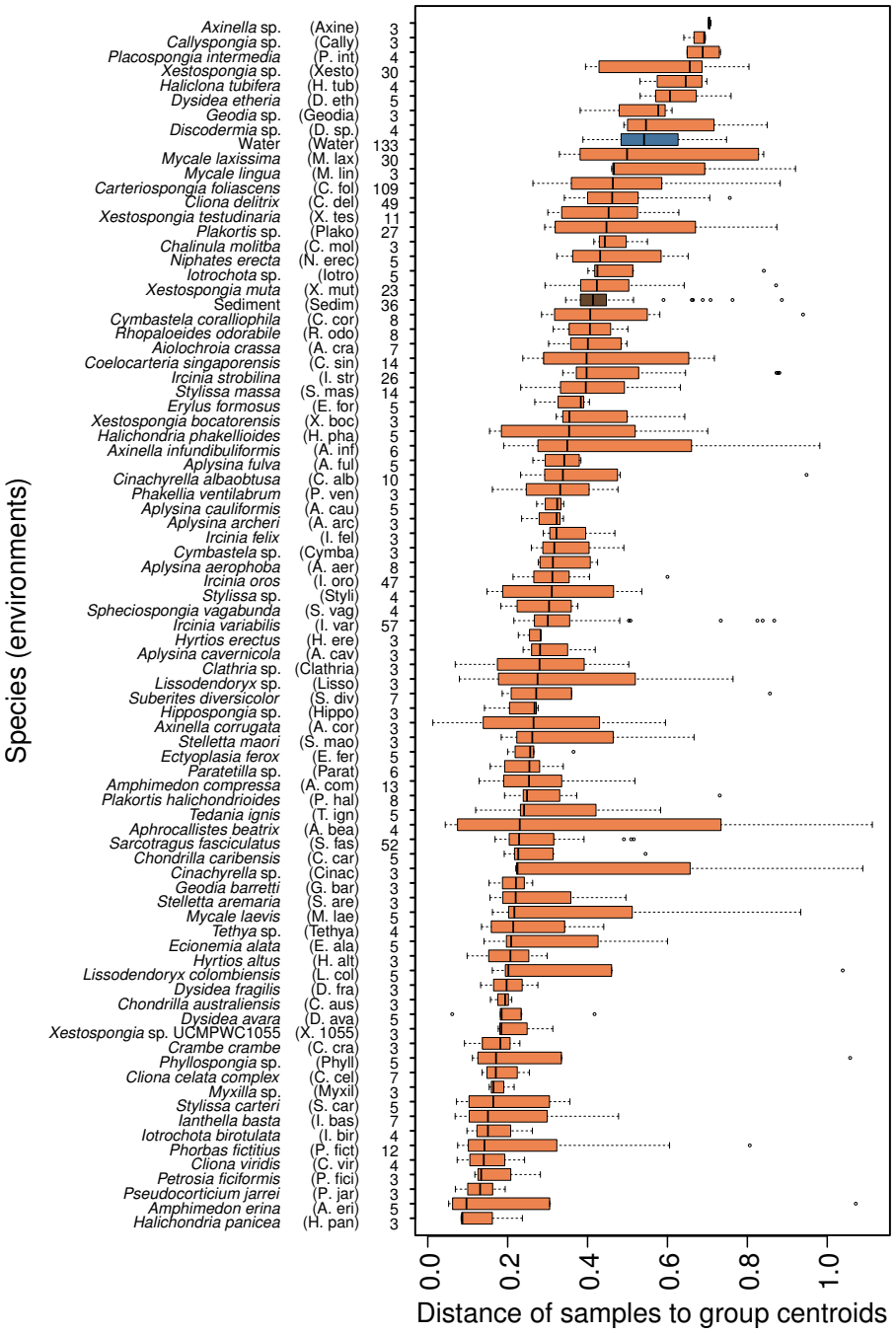


Figure 5.2: Intraspecific community dissimilarity measured as distance of samples to group centroids for 16S rRNA gene composition of different sponge species (orange) and habitats (blue: seawater; brown: sediment). Vertical bar represent the median, the box represent the first to third quartiles and whiskers show the lowest or highest datum within 1.5 times the inter-quartile range of the lowest and upper quartile, respectively. Names in brackets represent the abbreviations used in all subsequent figures. The number behind the brackets refers to the number of individual samples analysed per sponge taxon, seawater or sediment.

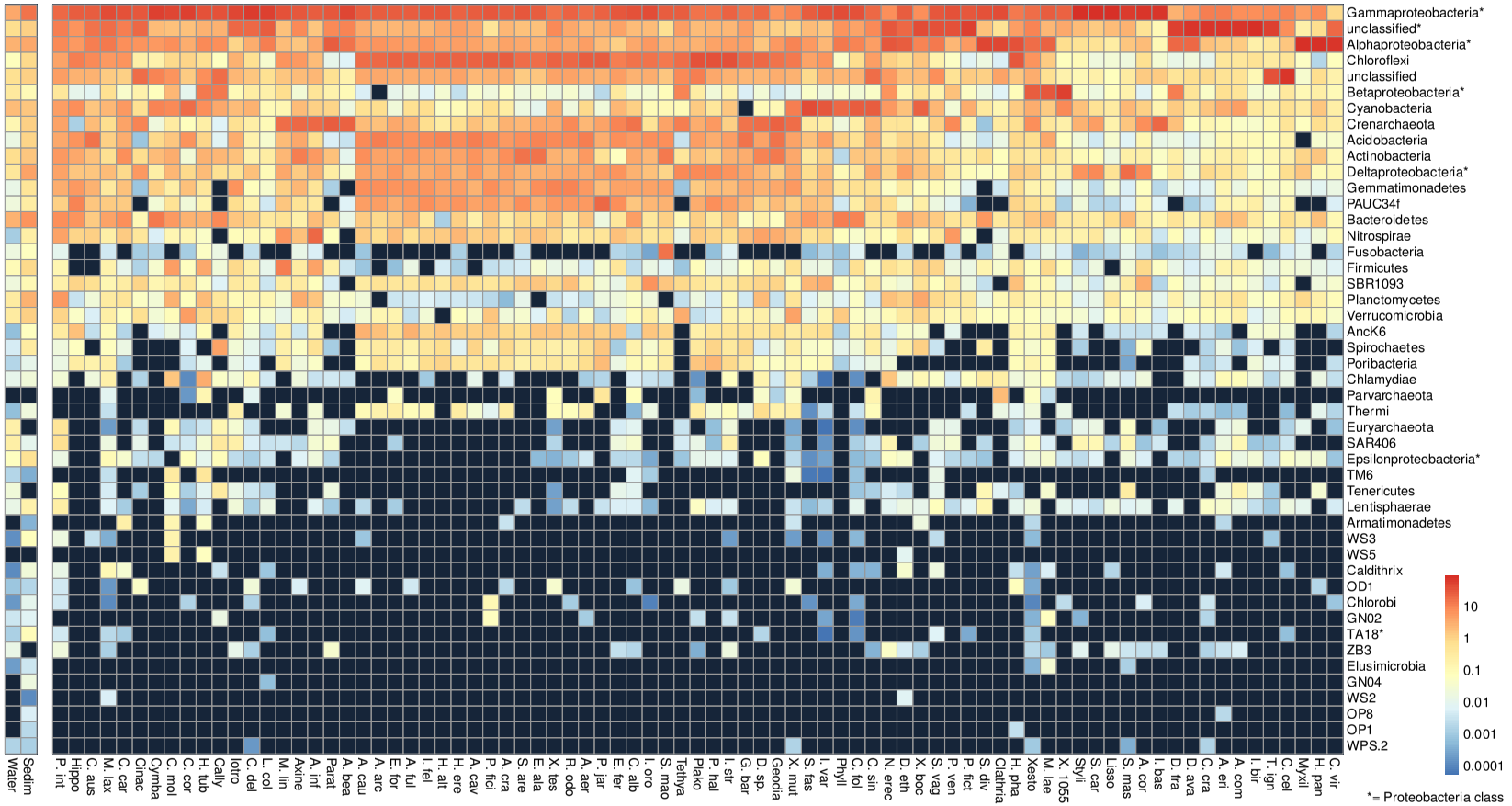


Figure 5.3: Average phylum-level taxonomic profile of microbial symbiont communities in 81 different sponge species, seawater and marine sediments. Colour scale shows relative abundance in percentage within each host species. The phylum *Proteobacteria* is shown as individual classes (including unclassified *Proteobacteria*), which are indicated by an asterisk. Black squares indicate zero counts. Columns and rows of the heatmap are ordered by Bray-Curtis dissimilarity of their taxonomic profiles (except for seawater and sediments). Sponge species abbreviations are outlined in Figure 5.2.

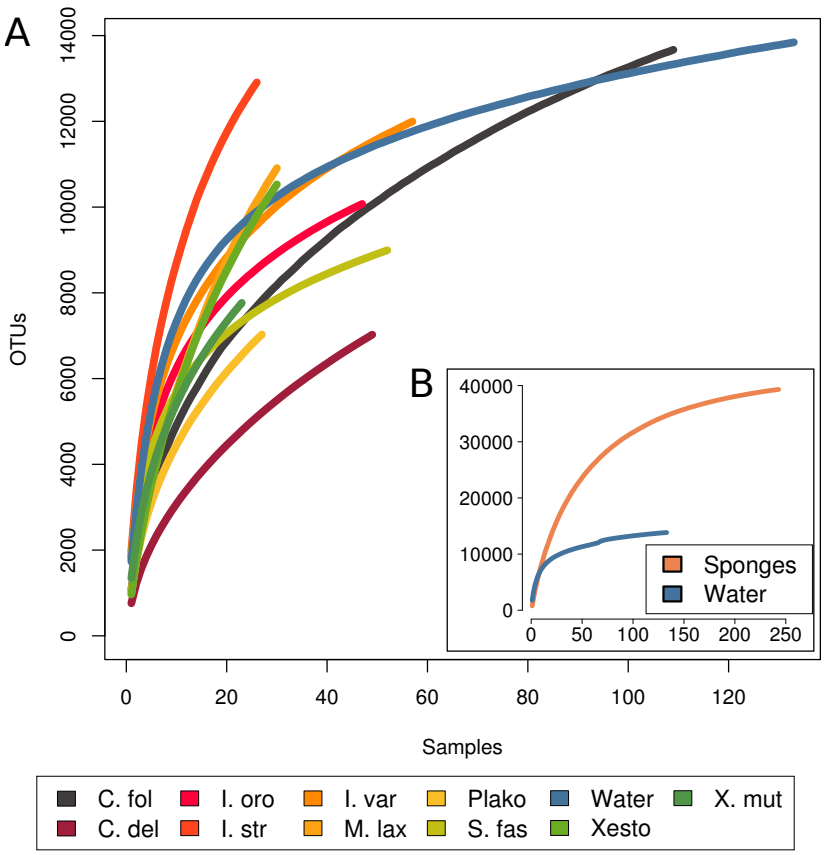


Figure 5.4: Rarefaction analysis of 16S rRNA gene diversity of microbial communities in sponges and seawater. (a) Rarefaction curves for sponge species with more than 20 replicate samples as well as seawater from all sampled geographic regions. OTU diversity is at 97% sequence identity cut-off. (b) Rarefaction analysis of all sponge species with three randomly selected samples per sponge species. Sponge species abbreviations are outlined in Figure 5.2

sampled biogeographic distribution) contain more than 12,000 OTUs (Figure 5.4). Similar richness projections were observed for the species *Cliona delitrix*, *Ircinia strobilina*, *Ircinia oros*, *Mycale laxissima*, *Plakortis halichondrioides*, *Sarcotragus fasciculatus*, *Xestospongia sp.* and *Xestospongia muta*, which were each sampled between 20 and 50 times (Figure 5.4).

Limited overlap in microbiome structure was observed between different sponge species or between sponges and the seawater and sediment samples (Figure 5.5). Thus, considering all OTUs discovered across the 804 sponge samples that included 81 different species, richness estimates approach a value of 40,000 OTUs (see inset Figure 5.4). The 81 sponge species analysed here represent only a tiny fraction of the 8,553 described sponge species (and likely a higher number when considering undescribed species) (van Soest *et al.* 2012) suggesting that sponge-associated (and likely other host-associated) communities are a significant global source of unique microbial diversity.

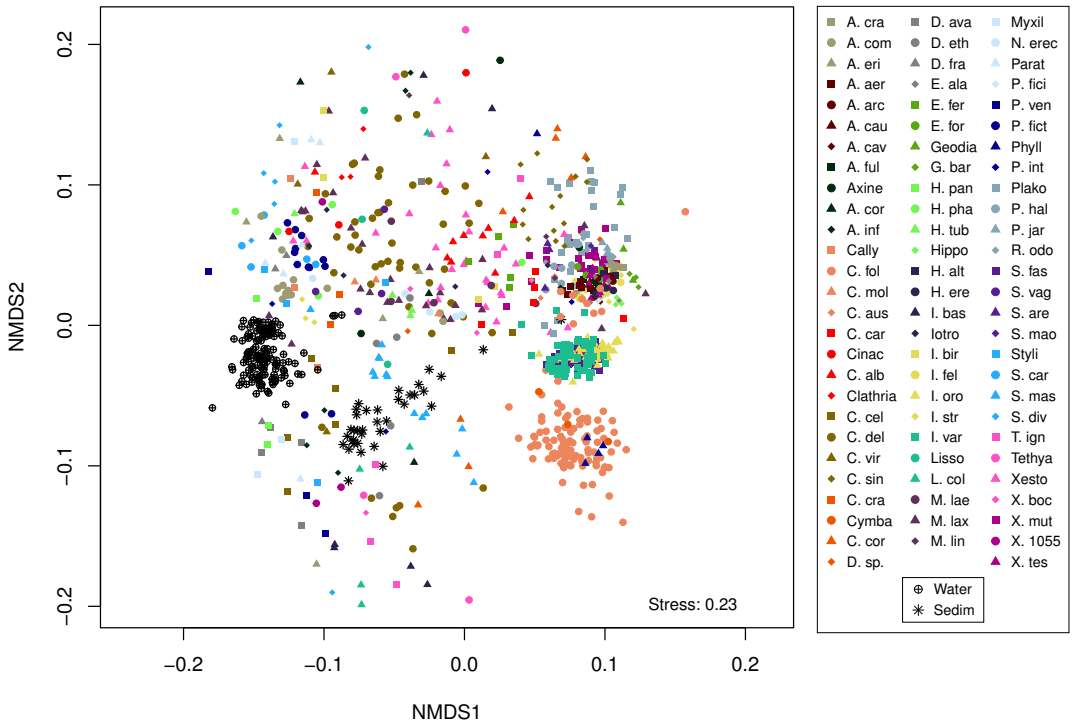


Figure 5.5: Community similarity for microbial communities in sponges, seawater and sediments. Clustering was performed using multi-dimensional scaling of Bray-Curtis distances. Sponge species abbreviations are outlined in Figure 5.2

Symbiont communities consist of generalists and specialists

To better understand the distribution of symbionts across the Porifera, we constructed a global bipartite network using the associations between OTUs and individual sponge species. The structure of this network differs greatly from what would be expected if connections between sponges and OTUs were randomly assigned (Figure 5.6). This suggests that assembly mechanisms (such as ecological and evolutionary processes) are behind the structure of this network of interactions, as has been suggested for other types of networks of ecological interactions (Montoya *et al.* 2006).

The cumulative probability of finding an OTU in the network with k or less associated hosts revealed a skewed degree distribution following a truncated power-law with an exponential cut-off at 7.44, almost half the number of host species an OTU is expected to interact with (the average number of hosts a given OTU is found in is 12.13) (Figure 5.6). This shows that the majority of symbiont OTUs have a small number of connections and only a few OTUs are very well connected. The majority of OTUs are thus specialists (i.e. found in only one or a few sponge

species), while only a few are truly cosmopolitan (i.e. found across many sponge species). Importantly, the degree distribution for the subset of OTUs belonging to previously defined sponge-specific sequence clusters follows the same distribution as the whole (see below).

The cumulative probability of finding a sponge host with k or less associated OTUs also follows a skewed degree distribution with exponential decay. A large fraction ($>$50%) of species harbour a symbiont diversity between $\approx$60 and $\approx$1800 distinct taxa, while a small fraction of sponge species can harbour up to $\approx$7000 OTUs (see also above). Skewed degree distributions have been identified in several types of ecological networks, and are linked to important properties of ecological communities, such as their robustness to species loss and their stability over time (Montoya *et al.* 2006). Our results suggest that ecological communities formed between microbial symbionts and their sponge hosts display similar patterns, which may be linked to their ability to maintain important functions at both the host and ecosystem levels (Taylor, Radax, *et al.* 2007).

To further investigate the specialisation of OTUs in our interaction network we analysed how consistently they are found across individual replicates of any given host species. Both highly specialised (defined here as those found in less than 5 different host species) and generalist OTUs (defined here as those found in more than 50 different host species) are present in a large fraction of the biological replicates of their respective host species (Figure 5.7 and D.2). In contrast, a large proportion of OTUs with an intermediate degree of host association (between 5 and 50 host species) can be considered as opportunistic taxa, associated with only a few biological replicates of multiple host species. Thus, symbiont communities within the phylum *Porifera* are characterised by a combination of highly generalist and truly specialist community members. Our analysis showed that generalists are cosmopolitan not only qualitatively (i.e. present in a large number of species), but also quantitatively (i.e. consistently present in a large fraction of individuals of those host species). To our knowledge, such patterns have not previously been observed for ecological networks, as it has traditionally been difficult to undertake repeated measures of many individuals across multiple host species.

Generalist symbionts comprise the core microbiome of sponge hosts

Considering the existence of generalist (i.e. cosmopolitan) sponge OTUs, we queried their relative contribution to the core microbiome of any individual species. Here, we define a core membership as any OTU that is present in at least 85% of the replicates for any single host species. In order to effectively model population dynamics of these OTUs, we identified host species with a sufficiently large number of replicates (here $\geq$47) across the entire dataset.

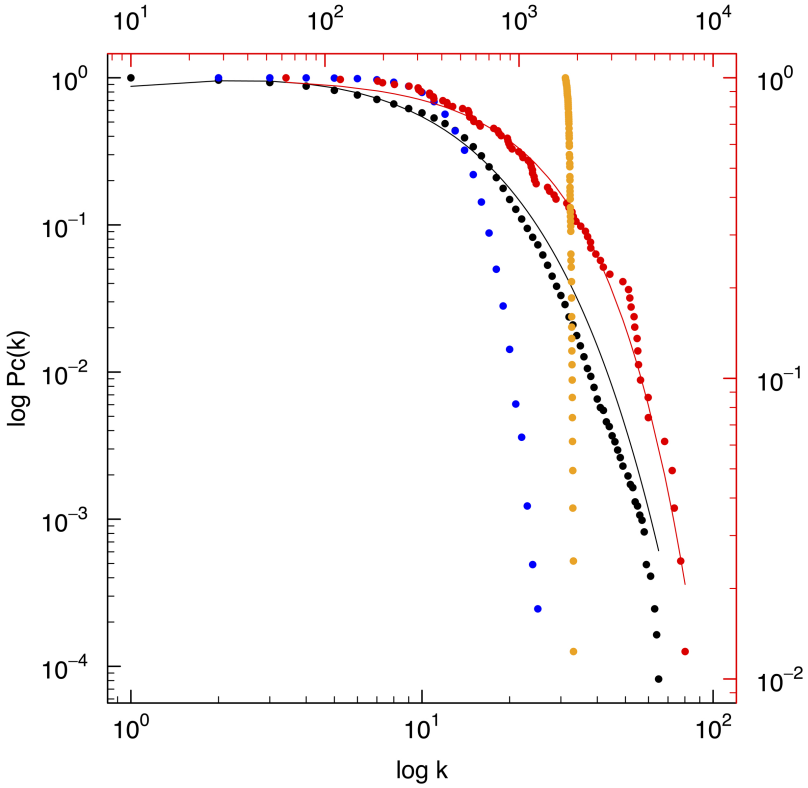


Figure 5.6: Cumulative degree distributions for OTUs (black dots, bottom and left axes) and sponges (red dots, top and right axes). Black dots correspond to the number of different host species (k) that contain a given OTU, represented as the cumulative probability of finding an OTU in the network with k or less associated hosts ($Pc(k)$). Red dots correspond to the number of different OTUs (k) found in a given host species, represented as the cumulative probability of finding a sponge host with k or less associated OTUs ($Pc(k)$). The OTU degree distribution followed a truncated power-law $Pc(k) = k - 0.32 * e - (k/7.44)$, while the sponge degree distribution followed an exponential given by $Pc(k) = e - (k/1849)$. Blue and orange dots correspond to random degree distributions for OTUs and sponges, respectively, where the number of nodes and links from the empirical distribution is kept constant.

We identified five host species (*Carteriospongia foliascens*, *Cliona delitrix*, I*rcinia oros*, *Ircinia variabilis*, and *Sarcotragus fasciculatus*) that fit this requirement and observed cores ranging in size from 7 to 20 OTUs. The proportion of OTUs with a certain degree (number of connections to different sponge species) or higher and the frequency distribution of degrees were compared for all OTUs present in the global bipartite network and the aggregated subset of OTUs present in all five core microbiomes (Figure 5.8). The core OTUs aggregated from all five sponge species showed an uneven distribution of degree frequencies. Core OTUs are primarily generalist and cosmopolitan (high degree) OTUs, while specialist (low degree) and intermediate degree OTUs are under-represented. This shows that highly connected OTUs in the global bipartite network also tend to comprise a larger fraction of the core for each of the host species

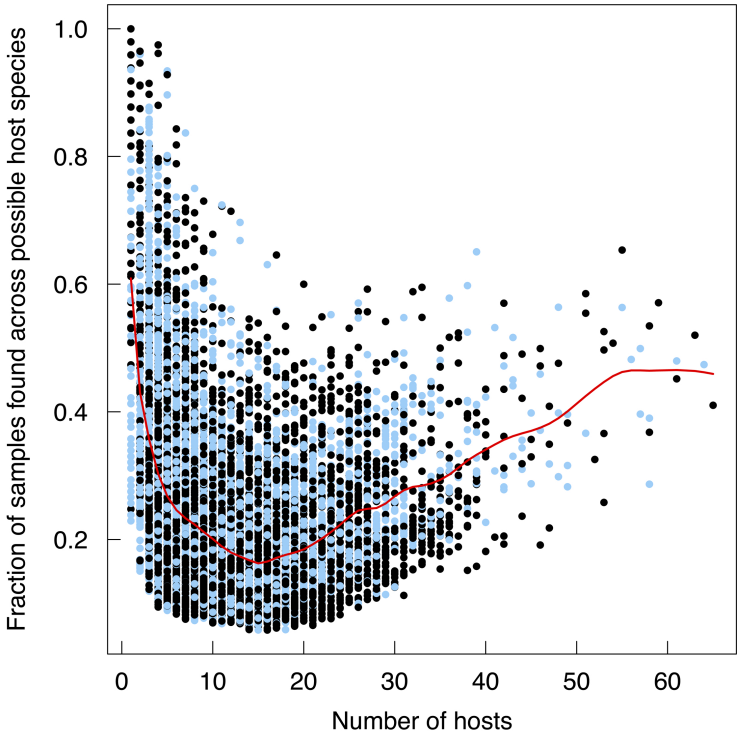


Figure 5.7: Number of host species (degree) containing a given OTU in the bipartite sponge versus symbiont OTU network plotted against the fraction of individual samples where each OTU has been found among all the samples from their known host species. Each point represents an OTU and the red line is a smoothing spline fit to the data (see *Methods*). Blue dots represent OTUs that assign to 'sponge-specific' clusters

investigated here.

Core symbionts have strong density dependence and weak, unidirectional interactions

Of particular interest is whether these core OTUs and their local interactions are important for the overall dynamics of the symbiont populations within each host species. For instance, density dependence (i.e. the growth rate of a population is controlled by its density) has a strong effect on community dynamics, with stabilising effects on population fluctuations (Henderson & Magurran 2014). In order to disentangle the complex nature of microbe-microbe interactions within our five sponge hosts described above, we applied a statistical framework (Mutshinda *et al.* 2009) which models population dynamics of the Lotka-Volterra type and allows us to decouple the variation in relative abundance of populations into contributions of i) inter-specific interactions, ii) density dependence and iii) environmental

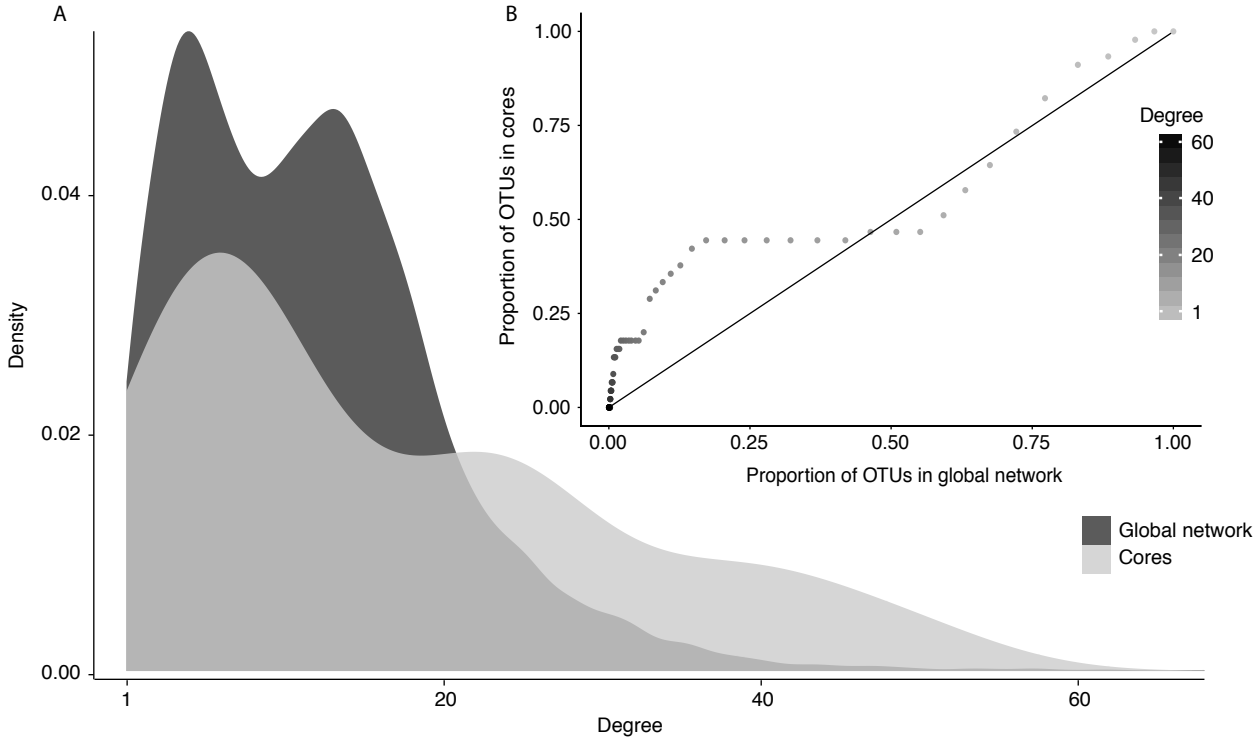


Figure 5.8: Frequency (density) distribution of degrees for the global bipartite network (dark grey) and the aggregated OTU cores (light grey) Panel A shows the proportion of OTUs with a certain degree or higher present in both sets. In panel B, the x-axis shows the proportion of OTUs in the global bipartite network, while the y-axis shows the proportion of OTUs in the aggregated cores. The 1:1 line indicates the expected distribution, if degrees were evenly distributed across the global bipartite network and the aggregated cores. This analysis reveals an over-representation of generalist and cosmopolitan OTUs within the aggregated cores, with a break occurring at $k > 12$

stochasticity. Population dynamics are sensitive to both stochasticity and fluctuating environmental conditions (Mutshinda *et al.* 2009). However, in this study, the environment is considered as fixed due to replicates being sampled from similar environments during the same time period, hence population dynamics and species interactions are considered to be influenced solely by stochastic processes.

Density-dependent processes were found to explain the majority of variation in the relative abundance of core OTUs across biological replicates, followed by stochastic mechanisms (Figure 5.9). Only a small proportion of variance (3 to 8% across hosts) is explained by inter-specific interactions (Figure 5.9). It should, however, be noted that the contribution of inter-specific interactions may be larger because we are missing those interactions excluded from the cores (i.e. interactions with more opportunistic OTUs).

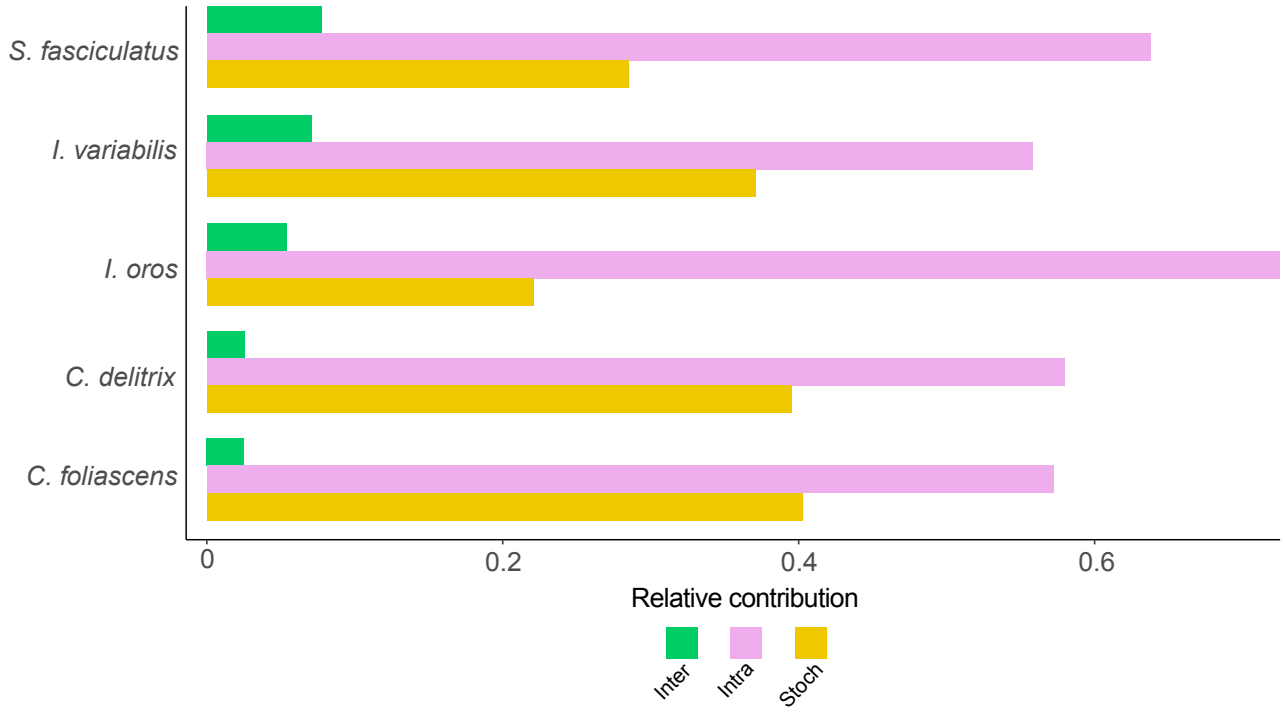


Figure 5.9: Relative contribution of inter- and intraspecific (i.e. density-dependence) interactions and stochasticity to the total temporal variation in microbial population abundances across hosts. Across hosts, dynamics were mainly driven by intraspecific interactions (i.e. density-dependence).

Although inter-specific interactions contribute little to the dynamics of the core microbiomes, it is still important to investigate the nature and strength of these interactions as, for example, antagonism (i.e. competition) and mutualism are known to differ in how they affect population and community stability (Thébault & Fontaine 2010). Both empirical and theoretical studies in community ecology demonstrate that distributions skewed towards many weak and a few strong interactions enhance both population and community stability, and may arise during the assembly of persistent communities (McCann 2000; Rooney & McCann 2012). Similarly, mutualism or skewed interactions only affecting one interacting partner (i.e. amensalism and commensalism) have been shown to promote diversity and lead to community stability ("The architecture of mutualistic networks minimizes competition and increases biodiversity." 2009; Lurgi, Montoya, *et al.* 2015).

A number of indices were calculated for each core microbiome (Table 5.1). Despite some variability in OTU number and linkages across different hosts, connectance (defined as the fraction of realised links among all possible links) was consistently low, ranging between 4.5 and 7.5%. We find that all cores are characterized by very few strong

and many weak interactions (Figure D.3). Moreover, cores are distinguished by a mixture of positive and negative interactions with amensalism and commensalism as a signature rather than competition and/or mutualism (Figure 5.10 illustrates this using the example of Ircinia oros; see Figure D.3 and D.3 for further details and other sponge species). Across hosts, we observe that the most probable links are generally negative, although as the core size increases, the fraction of positive inter-specific interactions increases. Interestingly, we find that some OTUs, which are highly connected in the global bacteria-sponge (bipartite) network, are also highly connected within the core network. This suggests that OTUs that are present in a large number of different host species tend to be important for population dynamics within each particular host.

The low connectance, weak, and amensal and/or commensal interactions, together with strong density dependence found in most sponge species, suggest that symbiont communities in the phylum *Porifera* have stable cores. However, whether these stable cores play a role in the dynamics of remaining OTUs within individual microbiomes, and more importantly, whether this stability guarantees the homeostasis of host functionality requires further investigation.

Table 5.1: Number of microbes assigning to 'sponge-specific' clusters within modules of the partial networks under instant and starved treatments.

	Number of OTUs in core	Number of links	Connectance	Interaction strength
S. fasciculatus	10	6.73 ± 2.13	$7.47\% \pm 2.4\%$	$0.03\% \pm 0.08$
I. variabilis	13	9.80 ± 2.86	$6.28\% \pm 1.85\%$	$0.02\% \pm 0.03$
I. oros	20	17.99 ± 4.31	$4.70\% \pm 1.10\%$	0.01 ± 0.04
C. delitrix	8	3.25 ± 1.60	$5.80\% \pm 2.85\%$	0.02 ± 0.05
C. foliascens	7	2.06 ± 1.28	$4.90\% \pm 3.04\%$	0.02 ± 0.06

Sponge-associated diversity is enriched in specific sequence clusters

Many of the microbes inhabiting sponges have previously been found to fall into monophyletic clusters of 'sponge-specific' or 'sponge and coral-specific' 16S rRNA gene sequences, with these clusters spanning 14 bacterial and archaeal phyla (Hentschel, Hopke, *et al.* 2002; Taylor, Radax, *et al.* 2007; Simister *et al.* 2012; Taylor, Tsai, *et al.* 2013). The ecological and evolutionary significance of these monophyletic clusters remains unclear, yet it is noteworthy that this phenomenon has not been reported outside the phylum *Porifera*. Over 43% of all sponge-derived sequences from this global sponge analysis were assigned to previously defined monophyletic sponge-specific clusters. However, using deep sequencing and our extensive sampling, 2.7% of seawater sequences and 8.7% of sediment sequences were also assigned to these clusters, demonstrating some clusters are not strictly 'sponge-specific', but

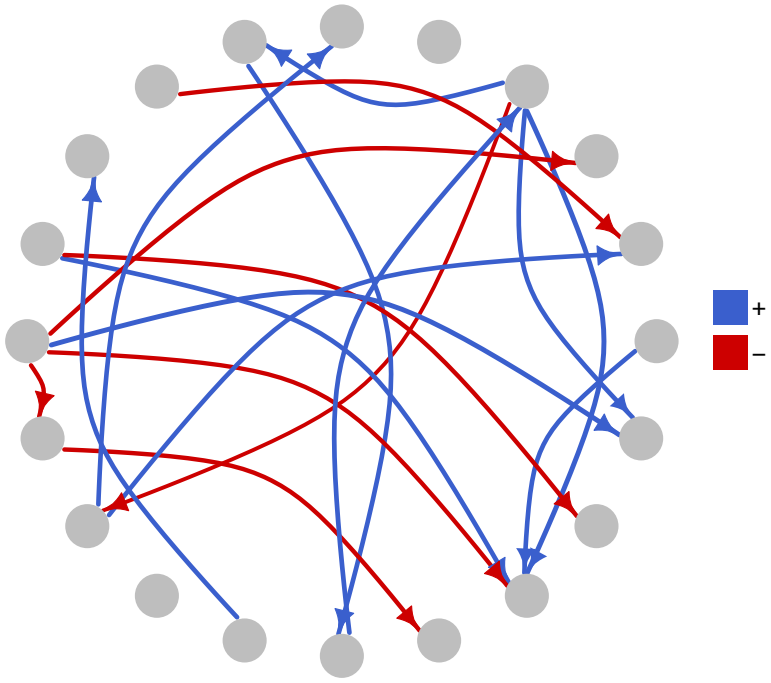


Figure 5.10: Representative network for the core microbiome of *Ircinia oros*. Each node corresponds to a single OTU, and links illustrate the most credible interspecific interactions. Positive and negative interactions are depicted in blue and red, respectively. None of the interspecific interactions are bidirectional, indicating either commensal $\{+, 0\}$ or amensal $\{-, 0\}$ interactions.

better described as 'sponge-enriched' (Figure S8) (Moitinho-Silva *et al.* 2014). Importantly, these clusters contain generalists, specialists and opportunists (Figure 5.7) indicating that the sponge-specific/enriched microbial sequence clusters have evolved multiple times, either early (i.e. core) or late (i.e. specialists and opportunists) in the assembly of symbiont communities.

Host phylogeny and identity concomitantly structure symbiont communities

Environmental and host factors are known to influence the composition of host-associated communities (Taylor, Radax, *et al.* 2007; Costello, Stagaman, *et al.* 2012; Easson & Thacker 2014); however, the impact of host evolutionary history on the structure and composition of symbiont comunities has only recently been explored (Easson & Thacker 2014). Considering the phylogenetic breadth of sponge species sampled here we were able to evaluate the relationship between host phylogeny and microbial diversity. Diversity was assessed using the inverse Simpson's index (D), while Blomberg's K was calculated using the phylosignal function in the R package *picante* (Kembel *et al.*, 2010) (Figure 5.11; see *Methods* for details). K values of 1 correspond to a random process, values closer to sero correspond

to patterns of convergent or random evolution and values greater than 1 indicate phylogenetic conservatism (Kembel *et al.*, 2010). We observed a significant value of K for the inverse Simpson's index (K =0.151, P=0.027), supporting a significant host evolutionary signal. Pagel's lambda (Harmon *et al.*, 2008) was calculated to further compare the similarity of covariances among species to the covariances expected given a random process. The lambda value of 0.216 (AICc=623.3; with λ fixed at 0, AICc=627.0) was significantly larger than what would be expected if there was no phylogenetic signal. Combined, these findings indicate a significant signal of convergent evolution in community structure, whereby sponges hosting more diverse communities are more phylogenetically related than expected by chance.

Beta-diversity analysis of symbiont communities (using Bray-Curtis distance) also indicated significant differences among species, with the factor "host species" accounting for approximately 64% of the observed variation among specimens. A partial Mantel test showed that host phylogeny was significantly correlated with Bray-Curtis distance (r =0.442, R^2 =0.195, P=0.001), as was host identity (r=0.706, R^2 =0.498, P=0.001). Testing for the effect of host phylogeny given host identity greatly reduced the explanatory power of host phylogeny (r=0.223, R^2 =0.050, P= 0.001), although host phylogeny still had a significant effect.

Overall, the evolutionary history of the host plays a significant role in structuring the diversity of symbiont communities, but only a minor role in structuring community composition (i.e., identity of microbial symbionts), where host identity (reflective of species-level forces) is the more important determinant. Thus, the evolutionary history of the host exerts a significant influence on microbial diversity despite strong selective forces for divergent microbiome composition, which might be critical for niche differentiation among closely related hosts (Freeman & Class Freeman 2014).

5.3 Conclusion

This global microbiome survey of an early-diverging metazoan phylum has revealed that sponges are a reservoir of exceptional microbial diversity and a major contributor to the total microbial diversity found in the world's oceans. Across the Porifera, symbiont communities exhibit little commonality in species composition or structure although a number of emerging properties related to community organisation are evident. For instance, sponge symbiont communities are characterised by a predominance of both specialists and generalists (as opposed to opportunists) and the core microbiomes are characterised by generalist symbionts with an underrepresentation of specialists. These communities represent dynamic systems, with the interacting members featuring all possible ecological interaction

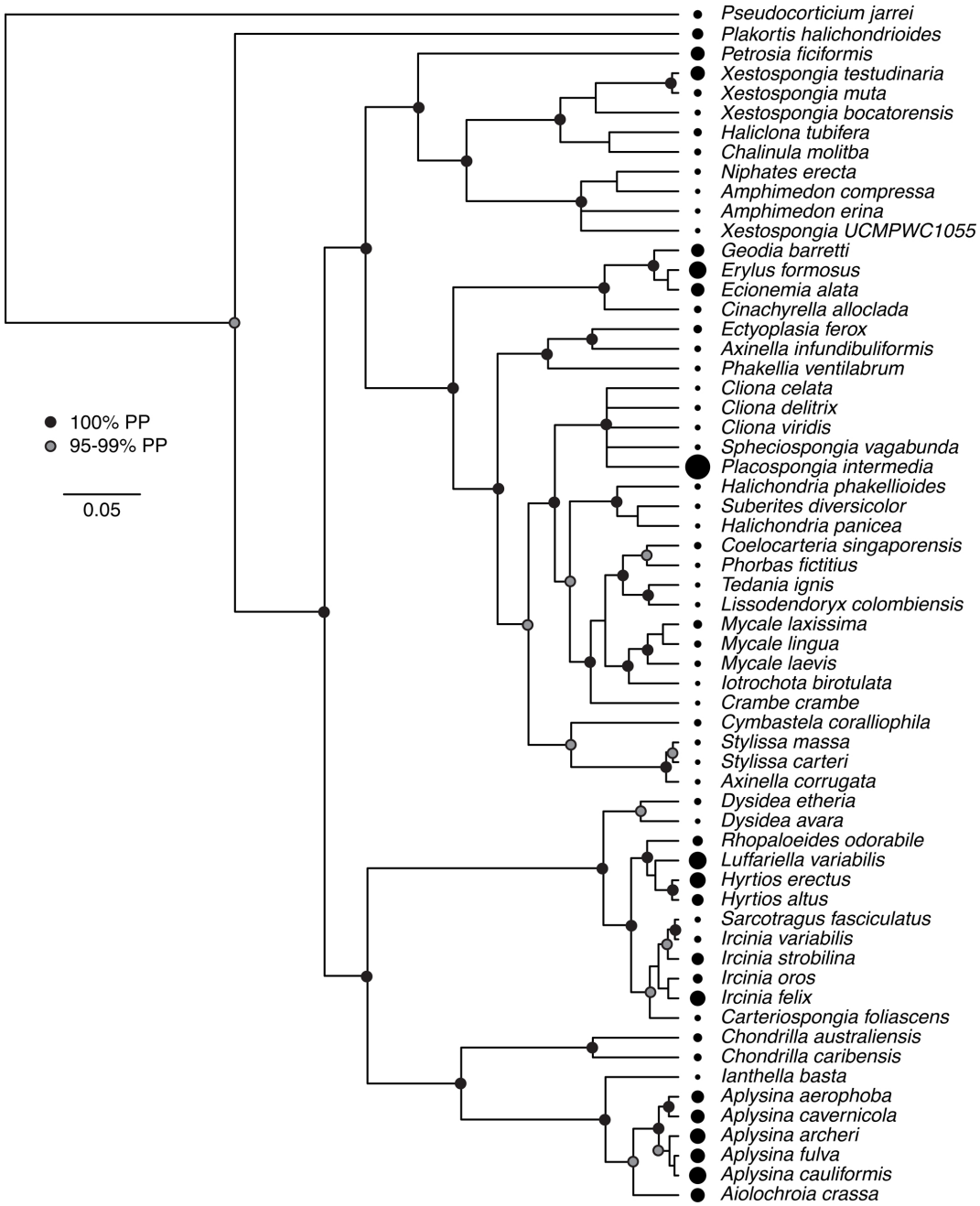


Figure 5.11: Phylogenetic signal of the inverse Simpson's index (D). In this multi-gene phylogeny of host sponge species, 100% Bayesian posterior probabilities (PP) are indicated by black circles at internal nodes, while gray circles indicate 95 to 99% PP, respectively. Nodes with less than 95% PP are not labeled. Black circles at the tips of the phylogeny are sized in proportion to the mean value of D calculated for the symbiotic microbial community associated with each host species. Multiple clades of sponges contain either high (e.g., *Aplysina*) or low (e.g., *Mycale*) values of D

types (positive, negative and neutral) (Faust & Raes 2012). The sign and strength of species interactions among community members has previously been shown to be highly dynamic and contingent on species composition, species densities and the environment (Ramsey *et al.* 2011). Here we show that the core symbiont communities in sponges are strongly density dependent, have few and weak interactions, low connectance and amensal and/or commensal interactions indicative of stable core symbionts within the Porifera (McCann 2000; "The architecture of mutualistic networks minimizes competition and increases biodiversity." 2009; Rooney & McCann 2012; Lurgi, Montoya, *et al.* 2015). Perhaps surprisingly, symbionts that appear to be phylogenetically unique to sponges (i.e. having previously been defined as 'sponge-specific') did not disproportionally contribute to the core microbiome or to any class of symbionts (i.e. specialist, generalist or opportunist), indicating that symbiont communities have independently assembled or evolved across the Porifera and that convergent forces have resulted in the analogous community organisation and interactions (Fan *et al.* 2012). Although the evolutionary history of the host is undoubtedly a driving force in this process, we show here that host phylogeny primarily impacts the complexity rather than the composition of the symbiont community. These findings further support a model of convergent evolution in symbiont communities across the entire host phylum (Fan *et al.* 2012).

5.4 Methods

Sampling and sample processing

Samples were taken and processed according to standard operating procedures to ensure maximum comparability. At least three different specimens of each sponge species were collected into sterile bags and species identities were confirmed by microscopic examination of morphological characters following established protocols (Hooper & Soest 2002). Specimens were either processed directly or after freezing, depending on logistical constraints of each sampling event. Specimens were cleaned of external growth (e.g. barnacles), washed three times with sterile seawater to remove planktonic or loosely-associated microorganisms and cut into small pieces from which a random sub-sample of pieces was used for subsequent DNA extraction. Sediment samples were collected under water in close proximity to sponges. Sediments were scooped into sterile containers using sterile spatulas to avoid laboratory contamination. Seawater was drained from the containers upon surfacing and prior to freezing. Sponges and sediment samples were immediately frozen and kept on dry ice or at -80°C until further processing. DNA was extracted from ≈25 g of sponge tissue or sediment using the PowerSoil DNA Extraction kit (MoBio) according to standard protocols ($http://press.igsb.anl.gov/earthmicrobiome/emp-standard-protocols/dna-extraction-protocol/$).

Microbial communities in seawater were collected by passing two litres of seawater through 0.2μm Sterivex filters and DNA was extracted from the filters as previously described (Webster, Taylor, *et al.* 2010). All samples were extracted in one of three laboratories (Australian Institute of Marine Sciences, Townsville, Australia; University of Wuerzburg, Germany; Nova Southeastern University, Dania Beach, Florida, USA) to minimise shipment of frozen specimens and between-laboratory variability. Aliquots of the specimens and DNA were kept at the three locations (and are available on request) and an aliquot of the extracted DNA was shipped to the University of Colorado, Bolder, Colorado, USA for sequencing of the 16S rRNA gene using standard procedures of the Earth Microbiome Project ($http://www.earthmicrobiome.org/emp-standard-protocols/16s/$). Briefly, the V4 region of the 16S rRNA gene was amplified using the primer 515f–806rB and sequenced using the HiSeq2500 platform (Illumina)(Caporaso, Lauber, Costello, *et al.* 2011). Sequencing data are publicly available through the Qiita website ($http://qiita.ucsd.edu/$) under Project ID 1740.

Analysis of sequencing data

Illumina reads were processed in mothur v.1.31.2 46. Firstly, quality-filtered, demultiplexed fastq sequences were trimmed according to quality (using the trim.seqs command: parameters qwindowaverage=30, qwindowsize=5, maxambig=0, maxhomop=8, minlength=100). To minimise computational effort, files were reduced to non-identical sequences (unique.seqs and count.seqs). Non-redundant sequences were aligned (align.seqs: flip=T) to a trimmed reference SILVA 102 (Quast *et al.* 2013) bacteria database (pcr.seqs: start=11894, end=25319, keepdots=F), which was provided by mothur (Schloss *et al.* 2009). Only sequences that aligned to the expected position were kept (screen.seqs: start=1968, end=4411; filter.seqs: vertical=T, trump=.). Aligned reads were reduced to non-redundant sequences (unique.seqs). Chimeric sequences were detected using Uchime (chimera.uchime: dereplicate=T) (Edgar *et al.* 2011), and filtered out (remove.seqs). Pairwise distances between aligned sequences were calculated (dist.seqs: cutoff=0.05) and were used for clustering. Prior to clustering, aligned sequences were phylogenetically classified based on the trimmed SILVA database (classify.seqs). Sequences were clustered (cluster.split: taxlevel=4, cutoff=0.03, hard=T, method=furthest) and converted to .shared file format (make.shared: label=0.03). Finally, OTU representative sequences were retrieved based on the distance among the cluster sequences (get.oturep: label=0.03) and were further classified based on Greengenes, SILVA and RDP taxonomies (classify.seqs: cutoff=60) ("Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB" 2006; Quast *et al.* 2013; Cole *et al.* 2014). Furthermore, Fastq sequences from additional samples (n=340) that were generated at a later time point (using the same sequencing procedure as described above) were processed with the same pipeline. These sequences were inte-

grated into the shared file using QIIME 1.8 (Caporaso, Kuczynski, *et al.* 2010), based on their similarity to the OTU representative sequences (parallel_pick_otus_uclust_ref.py: -similarity 0.985 -optimal_uclust). Sequences that were not similar to the OTU representative sequences were separately clustered with mothur and integrated into the previous files (.shared and taxonomy files). The integrated OTU table (.shared file) was filtered to remove low-abundance sequences (sequences less than 0.001% across the whole dataset) and chloroplasts (according to SILVA or Greengenes). Additionally, counts from seawater-like OTUs (>0.01% across all seawater samples) were removed from sponge samples. File manipulation and processing was carried out with python scripts ($www.python.org$).

Calculation of community metrics

Rarefaction curves were generated using the R package *vegan* 2.2-1 (Dixon *et al.*, 2003). Inter-sample rarefaction curves were generated by mothur (rarefaction.shared). Distances of the samples in a group (sponge species, seawater or sediment) and their respective group centroids were calculated based on Bray-Curtis distances by the function betadisper from the *vegan* package in R. Richness indicators (Chao, Ace, Sobs) were also calculated with *vegan*. Nonmetric multidimensional scaling (NMDS) was calculated with *vegan* package based on Hellinger transformed OTU counts (Legendre & Gallagher 2001). Taxonomic profiles were obtained based on Greengenes, which provided more phylum-level assignments than the SILVA or RDP databases. Briefly, percentage OTU counts were averaged by species/environment with the R package analogue 0.16-0 56. Phylum percentages were calculated by summing averaged OTU percentages. Bray-Curtis dissimilarities were calculated and heatmap was obtained using the package *pheatmap* 1.0.7 (Kolde, 2012).

Sponge-Bacteria bipartite network analysis

A bipartite interaction network was constructed using the presence of specific OTUs within each of the sponge species in the dataset. OTUs were considered part of the network only if they were found in at least 25 distinct samples from the whole dataset. In this bipartite host-microbe interaction network, nodes represent sponge species (on one side) and OTUs (on the other); and links among them represent the presence of an OTU in the microbial community of the sponges to which it is linked. The network was constructed using a software script developed in R using the package *igraph* 0.7.1 (Csardi and Nepusz, 2006) and interrogated using statistical tools to describe its properties.

The degree distribution of sponges and OTUs was analysed in order to assess the heterogeneity of the

network in terms of node connectivity. Degree distributions depict the statistical probability distribution of finding nodes with a certain degree (number of other nodes it is connected to). A variant of the degree distribution was employed: the cumulative degree distribution, which has the same probability distribution, but shows the probability of finding nodes with that degree or less. These probability distributions (one for the OTUs and another for the sponges) were obtained using the cumsum function in R. Additionally, we fitted truncated power law and exponential functions to the OTUs and sponges cumulative degree distributions, respectively. This was achieved using the non-linear least squares (nls) function provided by R. This analysis reveals the pattern of connectivity between sponges and OTUs and facilitates determination of the balance between generalist and specialist species. The thresholds for specialism and generalism were chosen arbitrarily, but following basic requirements for this type of network analysis. Firstly, neither of the groups contains the parameter that provides the characteristic scale at which the exponential cutoff occurs in the truncated power-law distribution. In our case this value is 7.44 ($Pc(k) = k_{-0.32} * e_{-(k/7.44)}$), so specialists need to have a number of links below that threshold, and generalist species should have a number of links that is several times this number–for this purpose we selected 7 times this number. Secondly, the average number of links in the specialist and generalist groups should be very different from the mean number of links in the network, and the difference in this ratio (mean group/mean network) should be similar for both groups. The mean number of network links is 12.13, the mean number of links for specialists is 2.5 and for generalists is 60, with specialists thus having a mean number of links approximately 4.47 times smaller than the average, and generalists 4.66 times larger than the mean.

The relationship between the fraction of samples within which a given OTU is found for a particular sponge species vs. the total number of sponges in which that OTU was found (degree of the OTU in the network) was also assessed. This was achieved by obtaining the fraction of sponge samples in which a given OTU was found out of all the samples available for the sponge hosts to which that OTU is connected in the host-microbe network. This information was plotted against the degree of the OTU. To better visualise this relationship a smoothing spline was fitted using the smooth.spline function provided by R. This relationship is used to analyse the true shape of specialisation vs. generalism in ecological networks.

Bacteria-bacteria network analysis

The bacteria-bacteria network analysis utilised host species with more than 47 individual replicates. If more than 47 replicates were available, random subsampling to 47 was performed. Cores were created for each host species by extracting OTUs occurring in at least 85% of the 47 replicates and were further filtered by removing OTUs with

a relative abundance less than 0.01. This was achieved using filter.shared(minpercentsamples=85, minpercent=1, makerare=f) in mothur v.1.31.2. The statistical model developed in Mutshinda *et al.* 2009 for inferring interactions from temporal series data was adapted to substitute time for space such that spatial replicates for each host were used rather than temporal samples. Spatial replicates here refer to the different geographic locations from which the 47 selected replicate samples for each sponge species were obtained. If we denote $n_{i,m}$ as the natural logarithm of $N_{i,m}$, then on a natural logarithmic scale we have the number of individuals of core OTU i in replicate m within any given host species described by

$$n_{i,m} * |n_{i,m-1} = n_{i,m-1} + r_i \left[1 - \sum_{j=1}^{S} \frac{\alpha_{i,j} n_{j,m-1}}{k_i} \right] + \epsilon_{i,m} \tag{5.1}$$

where r_i and k_i represent the intrinsic growth rate and the carrying capacity of OTU i, respectively, $\alpha_{i,j}$ represent the interaction coefficient between OTU i and j and expresses the per capita effect of OTU j on the growth rate of OTU j from replicate $m-1$ to replicate m. Finally,$\epsilon_{i,t}$ represents the effect of unexplained, latent, stochastic noise on the population dynamics of species i. Data are then modelled using a Poisson distribution, where $y_{i,m}$ denotes the number of individuals of OTU i in replicate m

$$n_{i,m} = \mathcal{MVN}(n_{i,m}*, \sum) \tag{5.2}$$

$$y_{i,m} = \text{Pois}(\lambda_{i,m}) \tag{5.3}$$

$$\log \lambda_{i,m} = n_{i,m} + \log N_m + \Lambda_m \tag{5.4}$$

where N_m and Λ_m are offsets representing the total abundance in replicate m. The latter is treated as a random variable assumed $\sim \mathcal{N}(0, 100)$.

The total variance V_i of individual species abundances can be decomposed into additive contributions from interspecific interactions, intraspecific interactions (density-dependence) and environmental stochasticity, respectively.

$$V_i = \left(\left(\frac{r_i}{k_i} \right)^2 \sum_{j \neq i} \alpha_{i,j}^2 v_{j,j} \right) + \left(\left(\frac{r_i}{k_i} \right)^2 v_{i,i} \right) + \epsilon_{i,i}^2 \tag{5.5}$$

where $v_{i,i}$ is the stationary variance for n_i, so that the proportion of variation attributed to e.g. intraspecific interactions

(density-dependence) can be calculated as follows

$$var(intra)_i = \left(\frac{r_i}{k_i}\right)^2 \alpha_{i,i}^2 v_{j,j}/V_i \tag{5.6}$$

Gibbs Variable Selection (GVS) method (O'Hara & Sillanpaa 2009) was used to constrain the model to only use inter-specific interaction coefficients,$\alpha_{i,j}$, for which there were strong support in the data. This is achieved by introducing a binary indicator variable $\gamma_{i,j}$ for $i \neq j$, and assuming $\gamma_{i,j} \sim Bernoulli(p)$, such that $\gamma_{i,j} = 1$ when species j is included in the dynamics of species i, and $\gamma_{i,j} = 0$, otherwise. Where there is low support for $\alpha_{i,j}$ in the data, $\gamma_{i,j} = 0$, and the interaction is excluded from the model. When $\gamma_{i,j} = 1$, $\alpha_{i,j}$ is freely estimated from the data. We set p to 0.1 as we did not expect more than 10% of all possible interspecific interactions to be realised.

Markov chain Monte Carlo (MCMC) simulation methods in R using the *runjags* package (Denwood in press) were used to sample from the joint posterior distribution of the model parameters. We ran 10 independent chains with dispersed initial values for 5 million iterations, discarding the first 2 million samples of each chain as burn-in and thinned the remainder to every 50th sample. The convergence of model parameters was assessed by visually inspecting trace and density plots. In addition, to ensure good mixing of $\alpha_{i,j}$ we calculated the number of jumps $\gamma_{i,j}$ between its two states (0 and 1).

Finally, to build the representative networks, we analysed the interaction and sign structure of the posterior distribution for the interaction coefficients $\alpha_{i,j}$. $\alpha_{i,j}$ is a full probability distribution, hence it contains the probability of OTU j having a non-zero per capita effect on the growth of OTU i (interaction strength), and vice versa. Using all information in $\alpha_{i,j}$, we constructed a representative network for each host species as a mean of visualising the most 'credible' network structure. This was done by mapping the posterior average number of links onto $\alpha_{i,j}$, and in doing so, extracting the links with the highest probability of a non-zero interaction. This was done by R scripts. As a way of validating the structure of each consensus network, we compared the connectance of each representative network to the posterior average connectance for each host species. The representative networks was plotted using the *igraph* package in R.

Identification of sponge-specific and sponge/coral-specific clusters

A representative sequence from each OTU was taxonomically assigned using a BLAST 62 search against a curated ARB-SILVA database containing 178 previously identified sponge-specific clusters (SC) and 32 sponge/coral-

specific clusters (SCC) (Simister *et al.* 2012). For each BLAST search, the 10 best hits were aligned in order to determine sequence similarities. The most similar OTU sequence to the respective reference sequence within the database was then assigned (or otherwise) to an SC or SCC based on application of a 75% similarity threshold (i.e. a sequence read was only assigned to a cluster if it was more similar to the members of that cluster than to sequences outside the cluster and its similarity to the most similar sequence within that cluster was above 75%). In cases where the assignment of the most similar sequences was inconsistent, a majority rule was applied, and the OTU sequence was only assigned to an SC or SCC if at least 60% of the reference sequences were affiliated with this cluster.

Phylogenetic analysis of host species and correlation with symbiont communities

Our phylogenetic analysis considered 61 sponge species for which at least one of three gene sequences (small subunit of nuclear ribosomal RNA [18S], large subunit of nuclear ribosomal RNA [28S], or mitochondrial cytochrome oxidase subunit 1 [cox1]) could be obtained from GenBank. For 39 of the 61 species (64%), sponge gene sequences were also obtained from at least one identical specimen collected for the current study.

For each gene, sequences were aligned using the default options of MAFFT 7.017 (Katoh *et al.*, 2002). Each alignment was analysed using the Gblocks Server (Castresana *et al.*, 2000) to eliminate non-conserved regions; the resulting three alignments were then concatenated using the Geneious software (version 6.1.8, Biomatters Limited). The phylogeny was constructed with MrBayes version 3.2.1 (Ronquist *et al.*, 2012), using the computational resources provided by CIPRES. Within MrBayes, five partitions (18S, 28S, and the three codon positions of cox1) were specified and separate general time reversible models of evolution for each partition were estimated, incorporating a gamma distribution of substitution rates among sites and a proportion of invariant sites (GTR+I+G) as suggested by (Huelsenbeck *et al.* 2004). The *Homoscleromorpha Pseudocorticium jarrei* and *Plakortis halichondrioides* (the only non-Demospongiae sponges of the taxon set) were constrained as an outgroup and the independent gamma rates relaxed-clock model with a birth-death process was implemented. The phylogenetic analysis included three parallel runs of 10 million generations, each utilising four Markov chains and sampling every 100 generations. At the end of the runs we assessed convergence by the average standard deviation of split frequencies, which was 0.03, and the potential scale reduction factors of all parameters, which ranged from 1.00 to 1.01. Following a burn-in of 25%, the trees sampled by each of the three runs were summarised into a consensus tree. The diversity of OTUs associated with each host species was evaluated by calculating OTU richness, the Shannon index of diversity, and the inverse Simpson index of diversity. All indices demonstrated clear differences among sponge species ($P<0.001$). Beta-diversity analy-

sis was conducted by calculating the Bray-Curtis distance among specimens, and testing for host species differences in this distance using the function *adonis* in the R package vegan.

[1]

[1]This chapter represents a large collaboration, part of the Earth Microbiome Project. It is accepted in Nature Communications, but not yet published. T. Thomas, L. Moitinho-Silva, M. Lurgi, **J R. Björk**, C. Easson, C. Astudillo, J B. Olson, P M. Erwin, S. López-Legentil, H. Luter, A. Chaves-Fonnegra, R. Costa, P. Schupp, L. Steindler, D. Erpenbeck, J. Gilbert, R. Knight, G. Ackerman, J V. Lopez, M W. Taylor, R W. Thacker, J M. Montoya, U Hentschel, N Webster. Global Sponge Microbiome: Diversity, structure and convergent evolution of symbiont communities across the phylum Porifera

Specialisation across networks of varying symbiotism

6.1 Introduction

The last two decades have been characterised by a revolution in the study of large species interaction networks, including food webs (i.e., who eats whom) (Solé & Montoya 2001; Dunne *et al.* 2002; Krause *et al.* 2003; Tylianakis *et al.* 2007; Ings *et al.* 2009) and mutualistic networks of free-living species, as those describing plants and their pollinators or seed dispersers (Olesen, Eskildsen, *et al.* 2002; Bascompte, Jordano, *et al.* 2003; Jordano, Bascompte, *et al.* 2003; Vazquez & Aizen 2004; Bascompte & Jordano 2007). Numerous theoretical and empirical studies have identified universal patterns and mechanisms in the way species interact across different environments and their consequences for community dynamics and stability (Montoya *et al.* 2006; Bascompte 2009). Some of these patterns have challenged prevailing wisdom about the outcome of coevolution based on pairwise interactions. For instance, reciprocal specialisations that occur when a consumer specialises on a particular resource, illustrated most famously by the Malagasy orchid and the hawk moth Darwin predicted would pollinate it (Darwin, 1862) are extremely rare when the whole interaction network is considered (Joppa *et al.* 2009). Ecological network studies, however, systematically ignore the largest component of biodiversity–microbes, in particular prokaryotes and their associations with larger organisms.

In parallel, but largely disconnected from research on ecological networks, new findings on microbe-host symbiosis have revolutionised the way we view the biotic world (McFall-Ngai 2008; Fraune & Bosch 2010; Hentschel, Piel, *et al.* 2012; McFall-Ngai *et al.* 2013). Symbiosis is postulated as one of the driving forces in the diversification of different plant and animal groups (Moya *et al.* 2008) and is crucial for the development, health and functioning of many hosts (Lee & Hase 2014).

Interactions between species are constrained by the "phylogenetic baggage of structure, physiology, and

behaviour that organisms inherit from their ancestors" (Thompson 1994). In contrast to networks of free-living species, interactions between host species and their symbionts (i.e. parasites, parasitoids, commensalists and/or mutualists) often involve long-lasting and sometimes extremely intimate relationships (see Ruby 2008 for a review). Coevolution acting on these interactions is therefore predicted to lead to greater reciprocal specialisation than coevolution among free-living species (Thompson 2005; Blüthgen, Menzel, Hovestadt, *et al.* 2007). Intimate and non-intimate ant-plant interaction networks, for instance, drastically differ in their network structure and degree of specialisation (Guimarães *et al.* 2007). In particular, intimate ant–plant networks are highly specialised and modular–groups of ants species only interact with certain groups of plant species, while non-intimate ant-plant networks are less specialised and highly nested—specialist species interact with a subset of the species generalists interact with (Guimarães *et al.* 2007). However, most studies on symbiotic networks, including microbe-host interactions, only address limited taxonomic groups and relatively species-poor networks of interacting species. However, a recent study by Toju *et al.* 2014 found that contrary to expectations, the structure of a complex plant-fungus network was characterised by relatively low levels of specialisation, reflecting the high plasticity of the relationship between plants and fungus.

Here we analyse a large next-generation sequencing data set of microbe-sponge associations from a local benthic habitat in the Northwest Mediterranean Sea. Marine sponges are fundamental to the functioning of marine ecosystems as they transfer energy between the benthic and pelagic zones (Goeij *et al.* 2013; Coppari *et al.* 2016). Marine sponges represent evolutionary ancient metazoans (Taylor, Radax, *et al.* 2007) that achieve functional complexity (Hillerislambers *et al.* 2012) and increased longevity (Taylor, Radax, *et al.* 2007; Freeman, Thacker, *et al.* 2013) by harbouring diverse and abundant assemblages of prokaryotic microbes (Simister *et al.* 2012).

We specifically focus on different facets of microbe-host specialisation by analysing the phylogenetic composition of microbes and the patterning of ecological associations with sponge hosts. In addition, by developing a new sampling protocol, we analyse how specialisation changes across a series of networks with varying levels of microbe specificity: from networks considering all observed microbes within a given host, to microbes only shared with larvae and indicative of symbiont vertical transmission. We analyse how different facets of ecological specialisation change as we modify the criteria for considering a microbe-host interaction. We also compare observed specialisation with equivalent random networks, and finally we discuss our results in the context of previously studied large ecological networks.

6.2 Methods

Sponge host collection

A total of 36 sponge species (Table E.1) were collected between July and August 2012 close to the Islas Medas marine reserve in the NW Mediterranean Sea $42°3'0''N$, $3°13'0''E$ by SCUBA at depths between 5-25 m. The collected species represent common Mediterranean sponges and were identified based on their distinct morphological features. Sponge HMA-LMA classification was taken from the literature, and if missing for any species, classifications were assigned based on the focal species genus.

We sampled three replicates per sponge species. Replicates were carefully placed in separate plastic bottles and brought to the surface. On the boat, the two first replicates were each cut into two pieces: one for the instant and one for the starved treatment. Each piece going to the starved treatment, including the third replicate was placed in separate 2 L jars containing filtered seawater (0.20 μm filter) for a period of four hours. The idea behind the starvation period is to remove microbes that serve as food for the host and/or any loosely associated microbes. Specimens for the instant treatment were cut (still on the boat) into pieces of approximately $2mm^2$ endosome tissue using sterilised scalpels and placed in tubes containing RNA later. All jars and tubes were transported in insulated coolers (containing ice) to the laboratory ($<$2 hours). After the four-hour starvation period, the same protocol was applied as to samples from the instant treatment. All samples were stored at -80°C until DNA extraction. Aliquots of seawater (300-500 mL each, 1 aliquot per sample jar) were concentrated on 0.2 μm polycarbonate filters, submerged in lysis buffer and stored at -80°C until DNA extraction.

Sponge larvae collection

We constructed larvae traps by modifying the traps used in (Lindquist 1996) (Figure 6.1, E.1). In order to capture dispersing larvae, traps were mounted over individual sponges by SCUBA. To prevent exaggerated levels of stress to each sponge individual, traps were only mounted for a maximum period of one week and then removed. During this time, sample bottles were collected and replaced each day. Collected bottles were placed in insulated coolers containing ice and transported to the laboratory ($<$2 hours). Larvae were identified using a stereolupe, and in order to remove loosely associated microbes, larvae were carefully rinsed with filtered seawater (0.20 μm filter) before conserved in RNA later. All sampled were stored at -80°C until DNA extraction. If larvae had been successfully identified from any sponge, that individual adult host was later sampled as described in *Sponge host collection*.

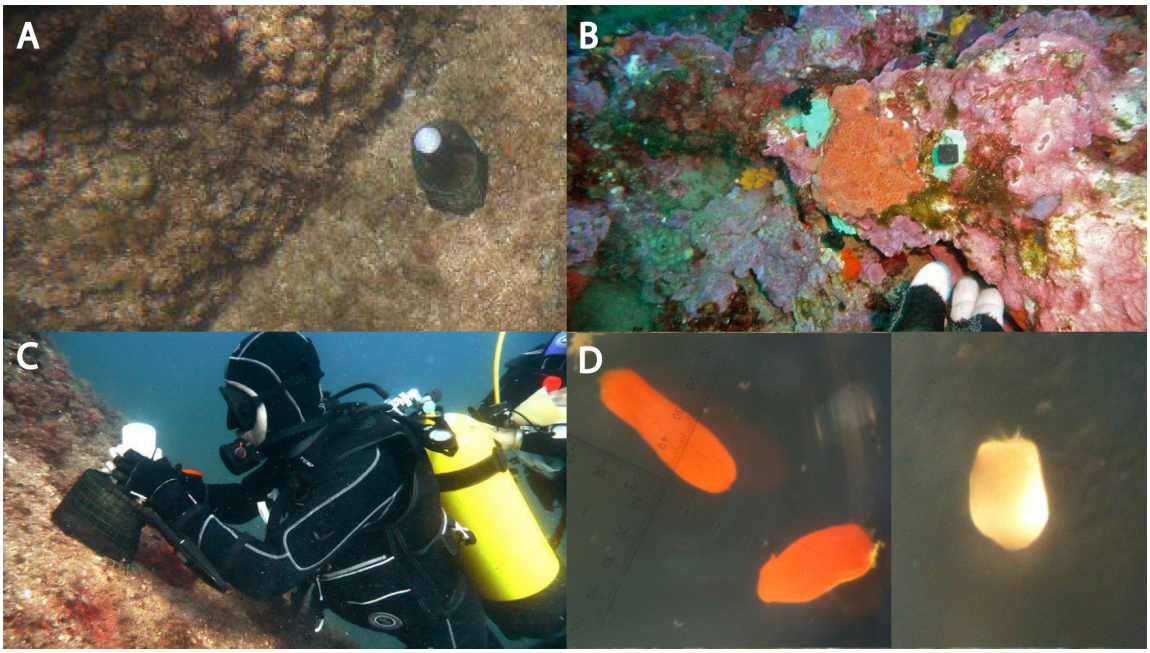


Figure 6.1: Photos displaying fieldwork from the dive site close to Islas Medas marine reserve in the NW Mediterranean Sea. Panel A. shows a larvae trap that is mounted on top of an adult hosts, B. shows screws that were attached around a host individual in order to mount the larvae trap. Panel C. shows me exchanging larvae sample bottles and panel D shows two types of parenchymella larvae.

DNA extraction and sequencing

DNA was extracted from ≈0.25 g of sponge tissue using the PowerSoil DNA Extraction kit (MoBio) according to standard protocols ($http://press.igsb.anl.gov/earthmicrobiome/emp-standard-protocols/dna-extraction-protocol/$). Microbial communities in seawater were collected by passing 2 L of seawater through 0.2μm Sterivex filters, and DNA was extracted from the filters as previously described by Webster et al. (2010). To gain insight into the prokaryotic composition, the V4 region of the 16S rRNA gene was amplified using the primer 515f-806r (Caporaso, Lauber, Walters, *et al.* 2012) and sequenced using the HiSeq2500 platform (Illumina). Sequencing and core amplicon data analysis were performed by the Earth Microbiome Project ($www.earthmicrobiome.org$) (Gilbert, Jansson, *et al.* 2014).

Analysis of sequencing data

Illumina reads were processed in mothur v.1.30.1 (Schloss *et al.* 2009). Firstly, quality-filtered, demultiplexed fastq sequences were trimmed according to quality (using the trim.seqs command: parameters keepfirst=100).

To minimise computational effort, files were reduced to non-identical sequences (unique.seqs and count.seqs). Non-redundant sequences were aligned (align.seqs: flip=T) to a trimmed reference SILVA 102 (Quast *et al.* 2013) bacteria database, which was provided by mothur. Only sequences that aligned to the expected position were kept (screen.seqs: start=13862, end=21287; filter.seqs: vertical=T, trump=.). Aligned reads were reduced to non-redundant sequences (unique.seqs). Chimeric sequences were detected using Uchime (chimera.uchime: dereplicate=T) (Edgar *et al.* 2011), and subsequently removed (remove.seqs). Pairwise distances between aligned sequences were calculated (dist.seqs: cutoff=0.05) and were used for clustering. Prior to clustering, aligned sequences were phylogenetically classified based on the trimmed SILVA database (classify.seqs) (Wang, Garrity, *et al.* 2007). Sequences were clustered (cluster.split: splitmethod=classify, taxlevel=4, cutoff=0.03, hard=T, method=furthest) and converted to .shared file format (make.shared:label=0.03). All samples were rarified to 56860 sequences. Finally, OTU representative sequences were retrieved based on the distance among the cluster sequences (get.oturep: label=0.03) and were further classified based on the SILVA reference taxonomy; classify.otus(cutoff=60).

Network construction

We created full and partial bipartite networks (Figure 1). A bipartite network describes the association between host (resource) species I and microbial (consumer) species J and is commonly displayed as $I \cdot J$ contingency table. In our case, each cell entry depicted the number of sequences belonging to a particular OTU found in a certain sponge species, thus reflecting the number of times the focal sponge-microbe association was observed. The full network represented all 36 host species, whereas the partial network reflected the subset of sponge species which we had successfully sampled larvae from. For the former, two bipartite networks were created (Figure 1): (A) the overall network containing all hosts and all their associated microbes, including those found in the water column, and (B) the host-specific network containing all hosts and their associated microbes, excluding those found in the water column. For the partial network, (A) and (B) were constructed from the subset of hosts which we had larvae from. In addition, a third network was created: (C) the symbiotic network only containing microbes shared between hosts and larvae (Figure 1). We created these networks for hosts under both instant and starved treatments.

We measured the degree of specialisation in each bipartite network by calculating (i) connectance (Pimm 1982; Dunne *et al.* 2002; Estrada 2007) and a series of quantitative indices: (ii) interaction diversity (Bersier *et al.* 2002), (iii) interaction evenness (Bersier *et al.* 2002), (iv) generality and vulnerability (Bersier *et al.* 2002), (v) H_2' index of specialisation (Blüthgen, Menzel & Blüthgen 2006) and (vi) nestedness (Almeida-Neto *et al.* 2008) using

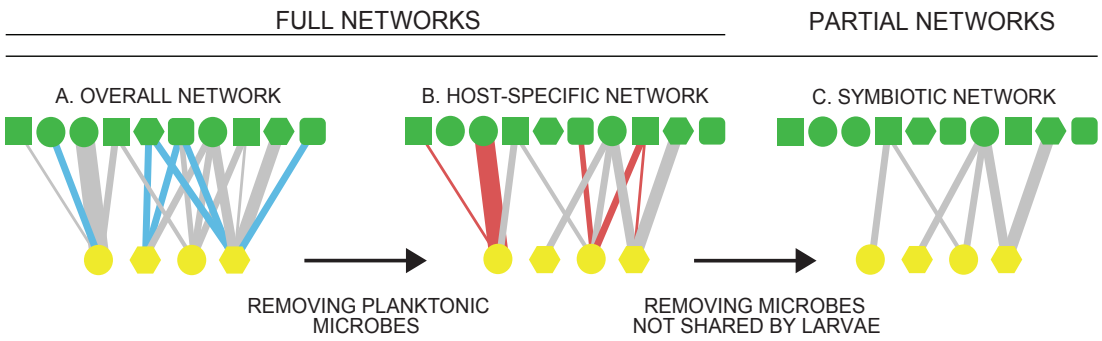


Figure 6.2: Networks after applying different criteria of what to consider a microbe-host interaction (link). Networks are increasing in specialisation from left to right. Higher levels (green shapes) correspond to microbes whereas lower levels (yellow shapes) represent sponge species. In the overall network (A) links correspond to microbes found within hosts, but can also be present in the water column (blue links). Links in the host-specific network (B) are microbes only present within hosts and not in the water column (red links). In the symbiotic network (C), links represent coevolutionary relationships and correspond to vertically inherited symbionts shared between larvae and hosts (gray links). The full network include all 36 sampled sponge species, while the partial networks only include those host species for which larvae was successfully collected from. All networks are built for both the instant and starved treatment.

the 'bipartite' v.2.05 package (Dormann & Strauss 2014) of R v.3.2.1 (R Core Team 2015). We also calculated (vii) modularity using the recently published LPAwb+ algorithm (Beckett 2016) implemented in R v.3.2.1. As network indices are sensitive to species only observed once (i.e. singletons) (Blüthgen, Fründ, *et al.* 2008; Dormann, Fründ, *et al.* 2009), those were removed from all contingency tables.

Interaction diversity is based on Shannon diversity, but applied to the number of links in a network. It is calculated as

$$H_2 = -\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij} \ln p_{ij} \tag{6.1}$$

where p_{ij} represents the proportion of links belonging to the interaction among sponge i and microbe j relative to the total number of links in the network. p_{ij} is defined as $\frac{a_{ij}}{m}$ where a_{ij} is the number of links between sponge i and microbe j, and m is the total number of interactions in the contingency table, calculated as $\sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij}$.

Interaction evenness is derived from Shannon's equitability and is calculated by dividing interaction diversity (H_2) by the natural logarithm of the total number of links L. There is a debate concerning what number of L to use: the total number of cells of the contingency table ($I \cdot J$), which is the default option in the 'biparite' package (Dormann 2015)

or the realised number of links, as suggested by Bersier *et al.* 2002 and Tylianakis *et al.* 2007. Here we defined L as the former, following

$$I_E = \frac{H_2}{\ln L} \tag{6.2}$$

Generality and vulnerability correspond to the weighted average number of sponge hosts per microbes and the weighted average number of microbes per sponge hosts, respectively.

$$\sum_{i=1}^{I} = \frac{A_i}{m} e^{H_i} \tag{6.3}$$

where $A_i = \sum_{j=1}^{J} a_{ij}$ and m is the total number of interactions in the contingency table. H_i is the Shannon diversity of interactions for each host species i, $H_i = \sum_{j=1}^{J} \left[\left(\frac{a_{ij}}{A_i} \right) \cdot \ln \left(\frac{a_{ij}}{A_i} \right) \right]$. For vulnerability, we simply replaced i with j and I with J.

The H_2' index of specialisation is a standardised version of H_2 (Equation 6.1) ranging from 0 to 1. It increases as the observed interaction frequencies depart from the expected values under a null distribution of interactions that assumes fixed marginal (interaction) totals. H_2' is calculated as

$$H_2' = \frac{H_{2max} - H_2}{H_{2max} - H_{2min}} \tag{6.4}$$

where H_{2min} and H_{2max} correspond to a perfectly quantitatively nested and modular matrix, respectively (Blüthgen, Menzel & Blüthgen 2006; Blüthgen, Fründ, *et al.* 2008).

Network null models

A null model for a given network represents a randomisation of the distribution of links across the network, following some set of rules. Null models deliberately exclude biological mechanisms in order to test whether the observed network structure can be explained by chance (Gotelli 2001; Gotelli & McCabe 2002). Even though null models never can point out an exact mechanism, they provide valuable information to what type of processes are underlying the observed network structure (Bascompte & Jordano 2014).

Null models are a debated topic in ecological network theory as it is not completely clear which randomisation algorithm is the most appropriate benchmark (Ulrich & Gotelli 2007; Blüthgen, Fründ, *et al.* 2008; Dormann,

Fründ, *et al.* 2009; Joppa *et al.* 2009; Fortuna *et al.* 2010). Without proper null models, the observed structure could solely be a consequence of the link density and/or the connectance of the network. The current consensus around which null model to use is to compare the observed network structure with a suit of different null models (Bascompte & Jordano 2014). We compared our observed networks with two null models. The first represents a scenario where each sponge host i can associate with every potential microbe j, only constrained by their marginal total observed number of interactions (r2dtable, Patefield 1981). While this algorithm represents the least conservative of commonly used null models (Blüthgen, Fründ, *et al.* 2008; Dormann, Fründ, *et al.* 2009), it has been shown to generate very similar results to the more conservative algorithms (Blüthgen, Fründ, *et al.* 2008; Toju *et al.* 2014). In this scenario, a low interaction frequency between host i and microbe j is only due to incomplete sampling (Blüthgen, Fründ, *et al.* 2008). For partial networks, we also generated randomisations using a more conservative algorithm that constrains the connectance to that of the original network (Vazquez, Poulin, *et al.* 2005; Vazquez, Melian, *et al.* 2007). In this scenario, interactions are distributed accordingly to species-specific probabilities that are (approximately) proportional to the relative abundance of microbe j in host i.

We evaluated significance by computing the probability that the observed network structure lied within the 95% credible interval of the random distribution of corresponding values (Blüthgen, Fründ, *et al.* 2008; Dormann, Fründ, *et al.* 2009). We generated equivalent random networks using the nullmodel function with 100 randomisations from the *bipartite* package (Dormann & Strauss 2014 and calculated the credible interval using the bayes.t.test function from the *BayesianFirstAid* package (Baath, 2014), in R v.3.2.1.

Phylogenetic composition

In order to further understand the observed network structure, we analysed phylogenetic beta-diversity across adult host, network types, and treatments. While traditional beta-diversity metrics do not account of the possible similarity of species not shared among communities, phylogenetic beta-diversity metrics, such as the UniFrac distance, measures species turnover among communities based on the shared branch lengths on their phylogeny (Lozupone & Knight 2005). An other advantage of phylogenetic diversity metrics is that they do not require *a priori* classification of species. The unweighted UniFrac distance attempts to capture the amount of evolution unique to one community compared to another. It does so by calculating the proportion of branch length that is unique to one community or the

other.

$$U = \frac{\sum_{i=1}^{N} l_i |A_i - B_i|}{\sum_{i=1}^{N} max(A_i, B_i)} \tag{6.5}$$

where N is the number of nodes in the phylogeny, l_i is the branch length between node i and its parent, and A_i and B_i are indicator variables equal to 0 or 1 when descendants of node i are absent or present in community A and B, respectively. The metric was calculated using the UniFrac function from the *phyloseq* package (McMurdie and Holmes, 2013) in R v.3.2.1, and location and dispersion effects were evaluated using the adonis function in the *vegan* package (Oksanen *et al*, 2016) in R v.3.2.1, along with the amova and homova function in mothur v.1.30.1 (Schloss *et al.* 2009). A distance-based bacterial phylogeny was built using the relaxed neighbor-joining algorithm (Evans *et al.*, 2006) implemented in the clearcut function in mothur v.1.30.1 (Schloss *et al.* 2009).

6.3 Results

Phylogenetic composition

Amongst host species identity, host classification (HMA/LMA) and treatment (instant/starved, including larvae), host species identity explained the largest proportion of variation driving phylogenetic composition across host-associated microbial communities (F_{pseudo} =1.288, R^2=0.461, P< 0.001 two-tailed). While host classification and treatment explained less variation, they still had a significant and visible effect on host-associated microbial phylogenetic composition (host classification: F_{pseudo} =1.375, R^2=0.015, P< 0.001 two-tailed; treatment: F_{pseudo} =1.891, R^2=0.042, P< 0.001 two-tailed) (Figure 6.3, E.2). Microbial communities harboured by adult hosts under instant and starved treatment differed in location (F_{pseudo} =1.691, P< 0.001 two-tailed) and dispersion (B=0.258, P <0.001, Figure 6.3). Although significantly different in location (F_{pseudo} =1.494, P= 0.004 two-tailed), the microbial communities associated to larvae clustered closer to their adult hosts under the starved treatment (Figure 6.3).

Network diversity and connectance

Sponge-microbe networks had a much larger microbial than sponge species richness. This resulted in a higher mean number of microbes interacting per sponge host (i.e., vulnerability), than the mean number of hosts interacting per microbe (i.e., generality) (Table 6.1,E.4). Contrary to our expectation, we observed a larger number of interacting microbes under the starved treatment in comparison with the instant networks (Table 6.1,E.4). Partial overall and host-specific networks, for example, had more than twice as many interactions per sponge host in the starved

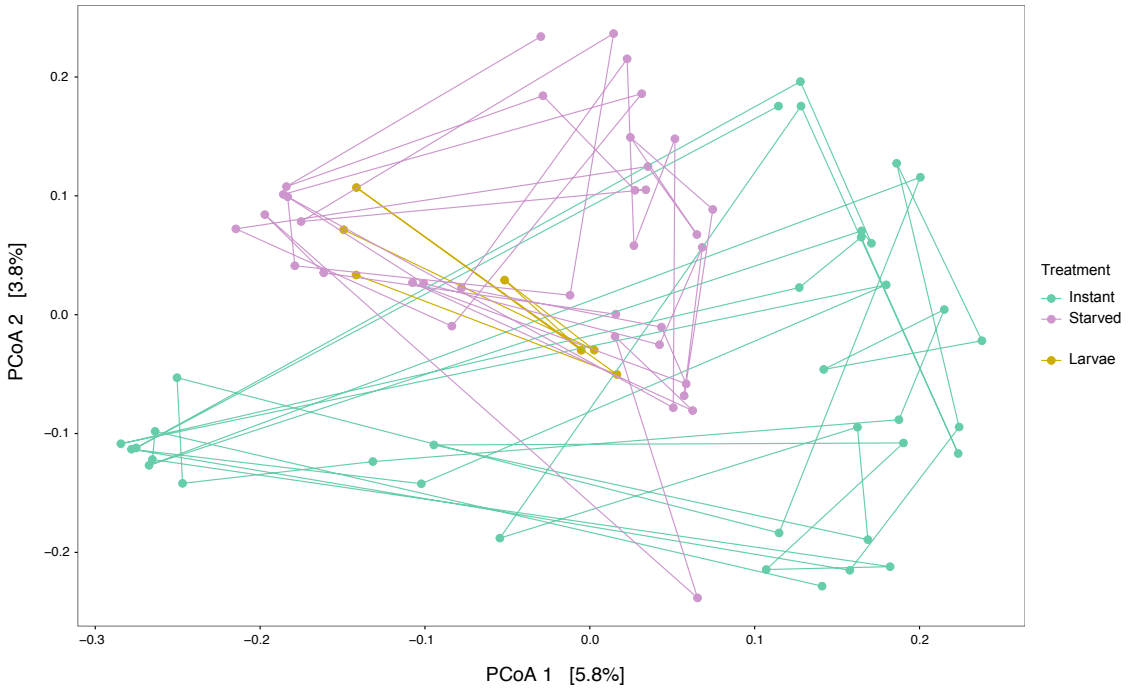


Figure 6.3: Principal coordinates analysis of unweighted UniFrac distances across all adult host-associated microbial communities, including their larvae. Dots and connecting lines coloured in green represent instant treatment, purple starved treatment and yellow larvae. Lines are included as a way of depicting each treatments dispersion.

compared to the instant treatment (Table 6.1,E.4). In comparison to the null expectation, the starvation treatment lead to an increase in specialisation of microbes on hosts (Figure E.3) that was not observed in the instant treatment (Figure E.3). Partial networks had a larger connectance than that reported for other mutualistic networks (Jordano 1987; Bascompte & Jordano 2007). Similarly, our partial networks displayed more interactions than other microbe-host networks (Toju *et al.* 2014).

Network architecture

Sponge-bacteria networks had less even distributions of link weights than expected under both null models (Table 6.1, 6.2, Figure E.3, E.4, E.5,E.6). All networks showed extreme levels of specialisation (H_2'), equal or very close to the maximum possible value (Table 6.1,6.2). This resulted from the higher observed modularity and lower nestedness than expected by chance (Table 6.1, 6.2, Figure E.3, E.4, E.5,E.6). Networks were organised in distinct clusters with interconnected modules that corresponded to the number of focal hosts in the networks (Figure 6.4, E.7, E.8, E.9, E.10), many of which showed distinctive taxonomic profiles (for symbiotic instant Figure E.11 and starved

Table 6.1: Network architecture for each partial network: overall host-specific and symbiotic networks, under instant and starved treatments, respectively. Number of hosts and OTUs in each network, vulnerability and generality, interaction diversity and evenness, and specialisation (H_2), modularity and nestedness (weighted NODF). Note that the modularity is standardised in order to make it comparable across networks.

	Instant			Starved		
	Overall	Host-specific	Symbiotic	Overall	Host-specific	Symbiotic
N_{Host}	7	7	7	7	7	7
N_{OTU}	2153	2011	210	2940	2866	276
Vulnerability	14.66	14.66	4.57	33.31	33.23	6.24
Generality	1.01	1.01	1.03	1.01	1.01	1.02
Diversity	4.47	4.28	2.80	4.94	4.90	2.80
Evenness	0.47	0.45	0.38	0.50	0.49	0.37
Connectance	0.16	0.16	0.16	0.15	0.15	0.16
Specialisation	1.00	1.00	0.98	0.99	1.00	0.99
Modularity	1.00	1.00	0.99	1.00	1.00	0.99
Nestedness	1.49	1.39	1.95	1.17	1.10	3.32

Figure 6.5,). We further explored how similar modules were within each network by calculating the phylogenetic (UniFrac) distance among them. Pairwise comparisons revealed that all modules within instant partial (overall, host-specific and symbiotic) networks contained distinctive phylogenetic profiles ($P < 0.001$) (for symbiotic: Table E.2), except for the two modules containing the HMA hosts *A. aerophoba* and *I. oros*. While similar patterns were observed for starved partial overall and host-specific networks, interestingly, modules within the starved symbiotic networks contain more similar phylogenetic profiles (for symbiotic: Table E.3).

Focusing on the partial networks, we found 'sponge-specific' clusters within all modules, although in varying numbers and densities. By simply counting the number of unique clusters within each module, it was largely possible to identify which modules interacted with an HMA host (Table 6.2). Although, these modules were characterised by a larger number of unique 'sponge-specific' clusters, analysing their relative abundance revealed a dominance of a few abundant ones (Figure 6.6). Comparing modules from networks under instant and starved treatment showed an overall increase in the number of 'sponge-specific' present in the starved treatment across networks (Table 6.2), and comparing their abundance revealed an enrichment of specific clusters in some modules (Figure 6.6).

6.4 Discussion

In this chapter we have presented the development of a sampling protocol that allowed for building a series of bipartite sponge host-microbe networks that increase in biotic specialisation. We expected the starved treatment

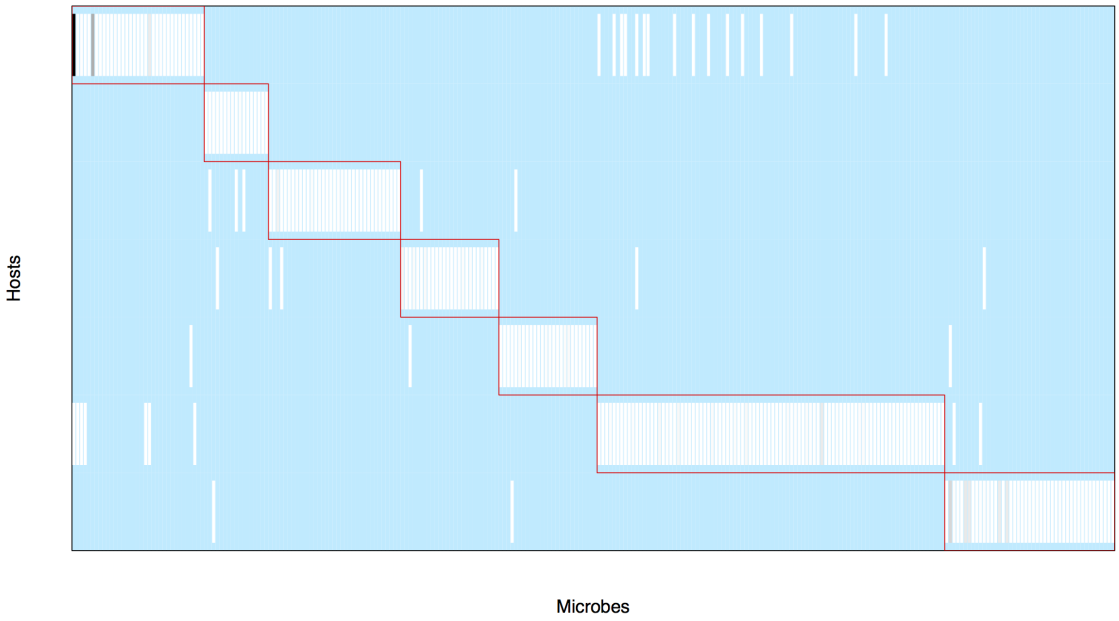


Figure 6.4: The networks are organised in modules comprising nodes interacting more among themselves than with the rest of the network. This figure corresponds to the partial symbiotic network under starved treatment. Modules were detected using the LPAwb+ algorithm for weighted networks (Beckett 2016). Rows represent hosts and columns microbes.

Table 6.2: Number of microbes assigning to 'sponge-specific' clusters within modules of the partial networks under instant and starved treatments.

	Instant			Starved		
	Overall	Host-specific	Symbiotic	Overall	Host-specific	Symbiotic
A. oroides	41	39	19	36	36	11
I. oros	38	36	15	54	54	19
C. viridis	2	2	2	30	30	6
C. crambe	11	12	2	19	18	8
D. avara	6	6	2	21	20	5
O. lobularis	3	3	2	32	32	6
H. columella	5	5	4	31	26	5

to reduce microbial richness in hosts by phagocytosis. Marine sponges display a fast clearance rate of pico and nano particles, including bacteria present in the surrounding water (Reiswig 1975; Frost 1978; Jiménez *et al.* 2007). We observed the opposite. The starved treatment had a higher microbial richness. This was likely due to a process of 'selective enrichment', where host phagocytise on some microbes while selecting others. The innate immune defence

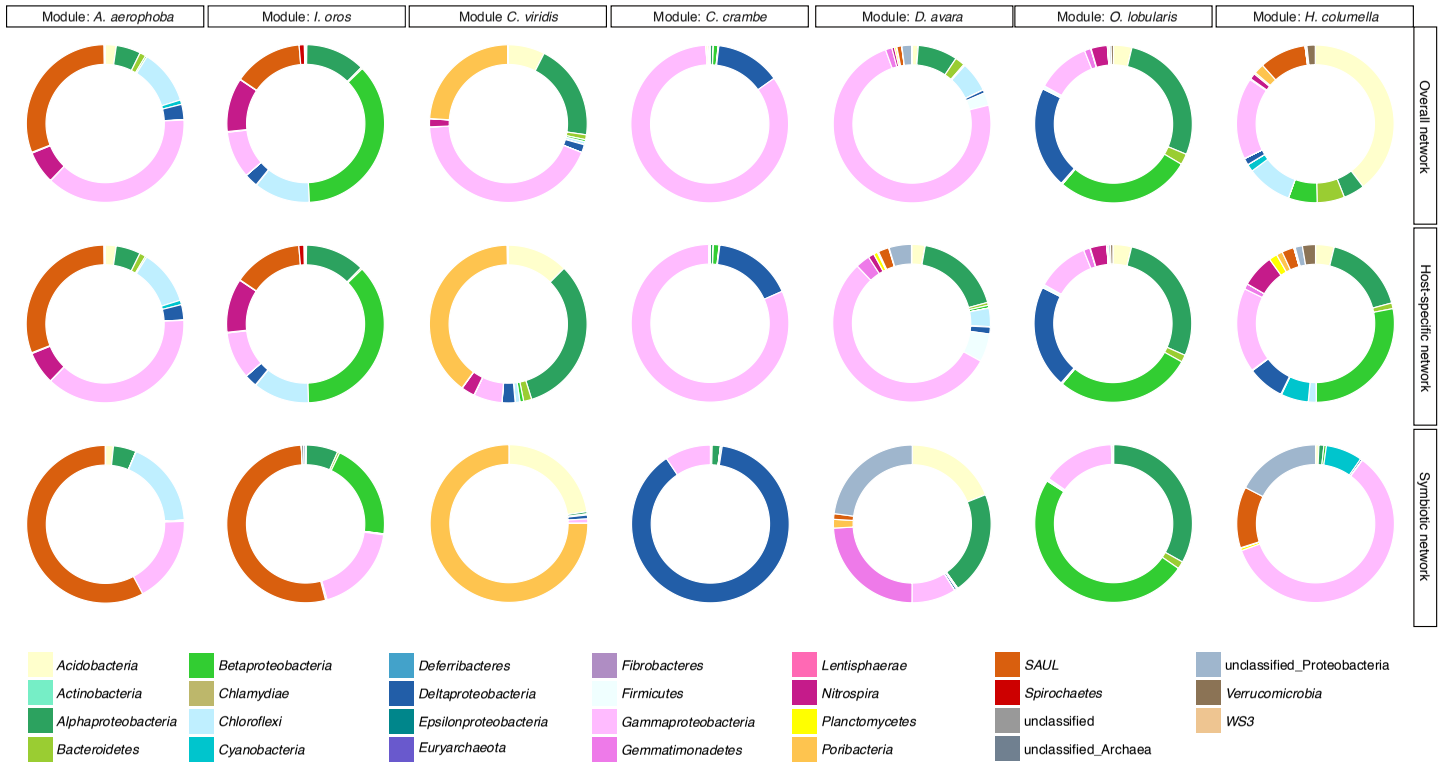


Figure 6.5: Taxonomic profile for each module within the different partial networks under starved treatment. Colours represent different phyla and their width correspond to their relative abundance within each module. Rows represent network type; overall, host-specific, and symbiotic networks respectively, and each column corresponds to a given module within these networks.

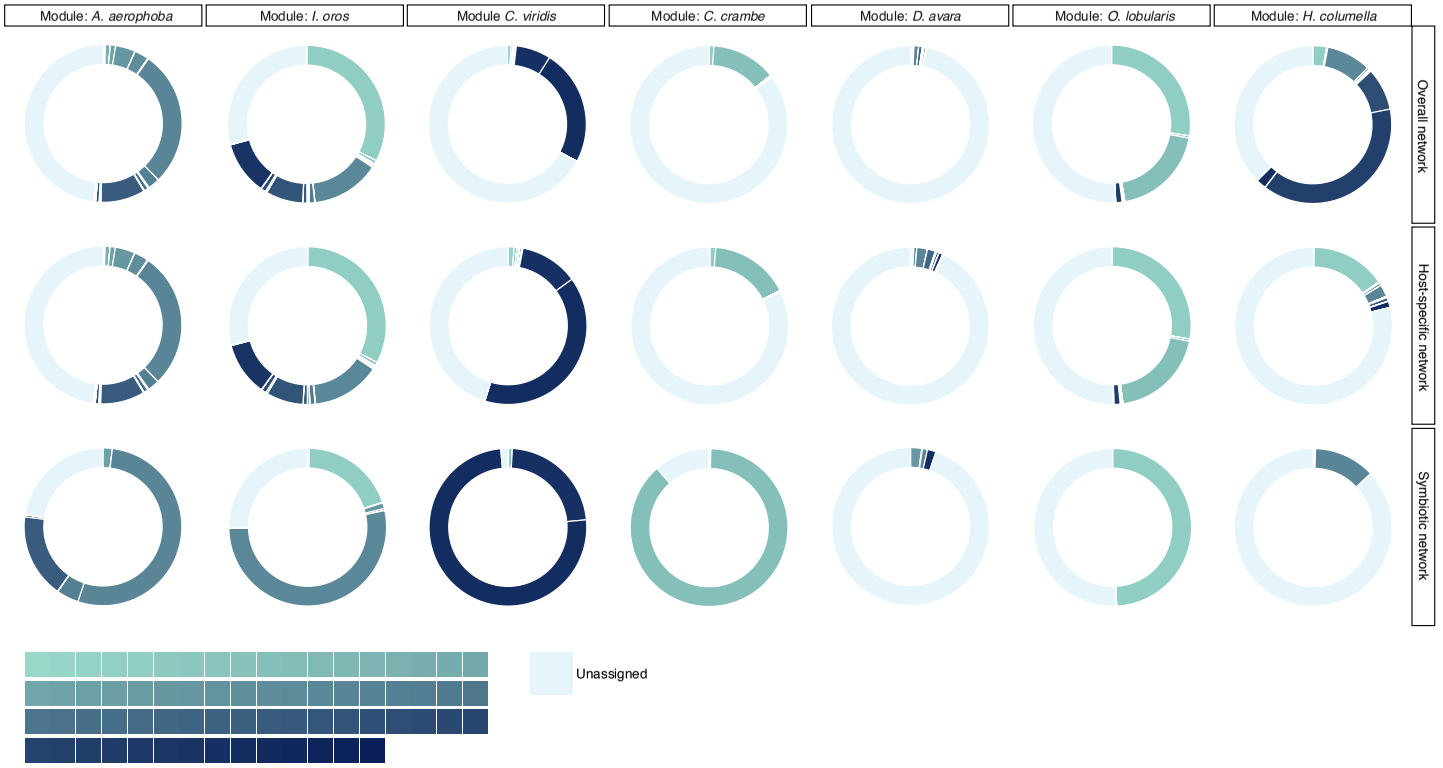


Figure 6.6: 'Sponge-specific' profile for each module within the different partial networks under starved treatment. Colour shades represent different 'sponge-specific' clusters and their width correspond to their relative abundance within each module. The lightest shade represent the relative abundance of microbes not assigning to any 'sponge-specific' cluster. Rows represent network type; overall, host-specific, and symbiotic networks respectively, and each column corresponds to a given module within these networks.

of some sponges can differentiate between pathogens, food bacteria and symbionts in a manner similar to the adaptive immune system of vertebrates (Wilkinson *et al.* 1984; Wehrl *et al.* 2007; Wiens *et al.* 2007; Thomas *et al.* 2010; Yuen *et al.* 2014; Degnan 2015). A recent study identified a large and diverse suit of pattern recognition receptor genes in the coral reef sponge Amphimedon queenslandica (Yuen *et al.* 2014), which play critical roles in detecting invasive microbes, including differentiating between friend and foe (Kaparakis *et al.* 2007; Robertson *et al.* 2012).

In addition, we hypothesise that microbes that are able to sustain their populations within hosts and do not require individuals arriving from the surrounding water are also more likely to increase in abundance than those microbes whose permanence largely depends on dispersal from the water column. We observed a strong community turnover from instant to starved hosts, with starved communities similar to those found in larvae. Our proposed mechanism of 'selective enrichment' where hosts actively differentiate between symbionts and invasive microbes is essentially the same mechanism in which hosts deposit symbionts into oocytes or developing embryos for vertical transmission (Thacker & Freeman 2012). It has been suggested that these early encounters between the embryo and vertically inherited symbionts is the onset for the sponge innate immune system to recognise and appropriately respond to these "good" microbes also in later stages of the sponge' life cycle (Degnan 2015).

Network analysis revealed that all networks were more specialised than expected by chance. In fact, to our knowledge, these are the most specialised large mutualistic networks ever analysed. We found that the broad suit of network properties analysed here did not quantitatively differ between partial and full networks. This suggests that the observed network patterns were not a consequence of network size. The high values of specialisation resulted from the high modularity observed, where the number of modules corresponded to the number of host in each network. However, as the number of interacting species within each module did not perfectly correspond to the number of microbes present in each individual host, there was some degree of overlap among hosts.

Modules may represent coevolutionary units (Bascompte & Jordano 2014). In other large mutualistic networks, modules have been found to contain species with convergent morphological traits, suggesting that modularity is a consequence of coevolutionary relationships (Olesen, Bascompte, Dupont, *et al.* 2007; Jordano 2010). However, the larger complexity of our system requires additional steps in order to reveal truly coevolutionary links. We hypothesise that the frequency of these links increases as you move from the overall network to the host-specific and finally to the symbiotic network. In support of our 'selective enrichment' hypothesis, comparisons between modules within each partial network revealed a more uniform phylogenetic composition within starved compared to instant networks.

Further analyses on the proportion of 'sponge-specific' clusters within modules between instant and starved symbiotic networks, revealed an enrichment of not all, but particular—likely functionally important 'sponge-specific' clusters. The module containing host *Cliona viridis*, for instance, did not display any detectable abundance of 'sponge-specific' clusters in the instant symbiotic network. However, in the corresponding starved network, this particular module was dominated by two abundant 'sponge-specific' clusters belonging to the phyla *Chloroflexi* and *Poribacteria*. Members of these phyla, specifically *Poribacteria*, are hypothesised to represent functionally important symbionts unique to marine sponges (Fieseler *et al.* 2004; Kamke *et al.* 2014).

Next-generation sequencing have revealed an underappreciated wealth of microbial diversity. As a consequence, the architecture of many ecological networks may have to be reconsidered in new light. However, an increased network complexity also introduces additional noise, cloaking the true coevolutionary links. Here we have presented a sampling protocol that successfully revealed such links, and their general structural patterning in comparison with other ecological networks.

[1]

[1]This chapter represents a collaboration with J M. Montoya

Conclusion

In this last section I would like to summarise the main conclusions of my thesis as a number of take-home messages that emerge from the results of the different chapters.

1.Microbe-sponge interactions are extremely diverse yet highly specialised. Sponge-associated microbial communities emerge as one of the most specialised, yet highly diverse ecological networks ever analysed. I have shown that the complex microbial communities associated to sponge hosts are characterised by high levels of host-species specificity with each host connecting to a group of microbes often comprised of related lineages with very few connections to other hosts and the water column. Each host species harbours a core microbiome common to most individuals of the same species, and that some host species preserve a subset their specialised interactions by vertically transmitting microbes to the next generation. Taken together, I have provided evidence for fingerprints of coevolution and likely cospeciation between sponge hosts and their microbial symbionts.

2. The identification of core microbiomes requires a temporal dimension that links ecological stability and host functionality, where vertical transmission and host selection of these core microbiomes are key eco-evolutionary mechanisms. I have further shown that the specialised nature of these interactions persists over long periods of time. By extending the core concept to include a temporal dimension, I have shown that hosts do not only harbour highly host-species specific core microbiomes common to most individuals of the same species, but that some of these cores consist of abundant, temporally stable microbes that persistently associate to hosts over periods of years and probably even longer. Whether hosts display stable dynamics depends on the aggregated abundance of the microbes comprising the core and its contribution to the overall microbiome abundance. In particular, the presence of high-density core microbiomes confers hosts a resistance against the increase in abundance of the many occasional microbes that pass through the sponge due to its filter-feeding activities. I have suggested that the main mechanism underlying the observed temporal stability of high-density cores is vertical transmission of microbes that

in turn triggers priority effects. However, I found suggestive evidence that host functionality may be maintained in hosts displaying unstable core microbiome dynamics through a highly selective acquisition of symbionts from the water column. Some sponges have an innate immune system similar to that of vertebrates. By starving hosts in order to remove loosely associated microbes and/or microbes used as food, I found patterns consistent with a process of 'selective enrichment' where some hosts actively differentiate between symbionts and invasive microbes, allowing some particular symbionts to increase in density.

3. Intraspecific interactions (i.e., density-dependence) together with weak interspecific interactions are key determinants of microbiome stability and fingerprints of coevolved interactions. I found strong evidence for density dependence as the main mechanism promoting stability in both spatially- and temporally defined core microbiomes. However, density-dependence can beget either stability or instability. While density-dependence likely increased stability in high-density core microbiomes close to carrying capacity by dampening fluctuations caused by environmental stochasticity, it may have maintained higher levels of variability in low-density cores by facilitating self-sustained population cycles. Compared to density-dependence, interspecific interactions were weak and infrequent, supporting the hypothesis that core microbiomes consist of highly complementary symbionts that have emerged over evolutionary time. However, I found that interspecific interactions in the form of commensalism and amensalism explained a relatively large proportion of the dynamics in some of the temporally defined core microbiomes. Importantly, these types of interactions are neglected in research on the structure, dynamics and functioning of ecological networks. My results suggests that they need to be included if we aim to gain a deeper understanding of the stability of complex microbiomes.

4. High Microbial Abundance (HMA) and Low Microbial Abundance (LMA) hosts likely constitute two opposite extremes on a continuum, where the stability of their core microbiomes adds a new fundamental dimension. Most studied sponge species to date fit the HMA-LMA dichotomy where HMA sponges harbour highly abundant, diverse and host-species specific consortia of microbes, whereas LMA sponges are commonly found to harbour less abundant microbial assemblages that resemble the free-living microbial community found in the plankton. As these classifications in turn reflect sponge morphology and physiology, they likely correspond to two evolutionary stable strategies (Gloeckner, Wehrl, *et al.* 2014). I found that the HMA-LMA dichotomy explained variation in cooccurrences of microbes in hosts across large spatial scales. However, while most studies only consider the diversity and composition of their associated microbes when comparing HMA and LMA hosts, I have shown that the distinction goes further and requires additional dimensions. HMA adult hosts and larvae harbour a much larger proportion of

related microbial lineages, i.e. 'sponge-specific' clusters, compared to LMA sponges. This also pervades their core microbiomes. While I observed that HMA hosts generally are characterised by higher temporal stability than LMA hosts, I also found notable exceptions to this rule. Importantly, however, the exceptions I found still conform to the HMA-LMA dichotomy in terms of metabolic functioning. I found that some hosts can maintain a metabolic profile common to many HMA hosts by horizontally selecting symbionts from the water column. I speculate that the same mechanism of microbial selection used by some hosts to identify and vertically transmit beneficial symbionts to their larvae is used by other hosts in order to selectively acquire symbionts from their environment.

Chapter I

Table A.1: Table showing the Jaccard index for different distance cut offs, corresponding to different taxonomic levels. The Jaccard index for different distance cut offs, α, in the aggregated bacterial community for each host. Ao corresponds to *A. oroides*, Cr to *C. reniformis*, Da to *D. avara*, and Bp to bacterioplankton. 0% represent the individual bands from the DGGE analysis.

Distance cut-off α	Represent	Ao-Cr	Ao-Da	Ao-Bp	Cr-Da	Cr-Bp	Da-Bp
0%=bands	Species	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
5%	Genus	4.17%	0.00%	0.00%	0.00%	0.00%	13.33%
20%	Phylum	31.25%	0.00%	21.43%	15.39%	20.00%	22.22%

Table A.2: Table summarising information about excised and sequenced DGGE bands, including GeneBank accesion numbers, taxonomic affiliation, and membership to sponge-specific clusters. Acc. corresponds to the Accession Number of each sequence in GeneBank. Band ID. corresponds to excised and the sequenced DGGE bands for each sponge host and the bacterioplankton (see Figure A.1). Phylogenetic affiliation corresponds to the known taxonomy of each sequenced DGGE band. SC or SCC clusters corresponds to if the sequenced band assign to any sponge-/sponge-coral-specific cluster.

Acc.	Band ID.	Phylogenetic affiliation*	SC or SCC clusters
		α-Proteobacteria	
KC200496	147Da_b13	Rhodobacterales, Ruegeria a	no
KC200497	148Da_b15	Rhodobacterales, Ruegeria a	no
KC200513	191Cr_b6	Rhizobiales, OCS116 clade	no
KC200495	143Da_b6	Sphingomonadales, Erythrobacter b	no
KC200509	180Da_b20	Rhodospirillales, Thalassospira a	no
KC200510	182Da_b24	Rhodospirillales, Thalassospira a	no
KC200519	203Cr_b26	Rickettsiales a	SC109
KC200538	69Da_b8	Rickettsiales a	no
KC200494	135Da_b14	Rickettsiales a	no
KC200531	232Da_b12	Rickettsiales a	no
KC200530	231Da_b9	Rickettsiales a	no
KC200540	70Da_b17	Rickettsiales a	no
KC200511	183Da_b25	Rickettsiales a	no
KC200521	206Bp_b26	Rhodobacterales, Roseobacter a	no
KC200524	214.1Bp_b21	Rhodobacterales a	no
KC200507	18.1Bp_b35	Rhodospirillales, Defluviicoccus a	no
		γ-Proteobacteria	
KC200516	194Cr_b12	KI89A clade b	SC142
KC200500	150Ao_b7	KI89A clade b	SC142
KC200541	80Cr_b13	Oceanospirillales, Endozoicomonas a	no
KC200504	168Ao_b17	HOC36 a	SCC32
KC200525	214.2Bp_b21	Thriotrichales, Thiothrix a	no
KC200526	215Bp_b33	Oceanospirillales, Pseudospirillum a	no
KC200508	18.2Bp_b35	Alteromonadales, OM60(NOR5) clade a	no

Table A.3: Continuation of Table A.2

Acc.	Band ID.	Phylogenetic affiliation*	SC or SCC clusters
		β-Proteobacteria	
KC200491	117Da_b18	Nitrosomonadales b	SC112
KC200498	149Da_b23	Nitrosomonadales b	SC112
KC200499	14Bp_b28	Hydrogenophilales, Thiobacillus a	no
KC200502	15Bp_b29	Hydrogenophilales, Thiobacillus a	no
KC200505	16Bp_b30	Hydrogenophilales, Thiobacillus a	no
		δ-Proteobacteria	
KC200518	201Cr_b23	Desulfurellales a	SC121
KC200536	58Ao_b2	Myxococcales a	SC116
KC200486	106Ao_b15	Sh765B-TzT-29 a	SCC30
KC200503	165Ao_b13	Sh765B-TzT-29 a	SCC30
		Bacteroidetes	
KC200539	6Bp_b7	Flavobacteriales, NS9 marine group a	no
KC200545	9Bp_b14	Flavobacteriales, NS9 marine group a	no
		Deferribacteres	
KC200515	193Cr_b11	PAUC34f marine benthic group a	no
		Gemmatimonadetes	
KC200488	109Ao_b20	PAUC43f marine benthic group a	no

Table A.4: Continuation of Table A.2, A.3

Acc.	Band ID.	Phylogenetic affiliation*	SC or SCC clusters
		Acidobacterias	
KC200544	98Cr_b28	Acidobacteriales a	SCC5
KC200543	88Cr_b30	Acidobacteriales a	SCC4
KC200520	205Cr_b29	Acidobacteriales a	SCC4
KC200542	82Cr_b18	Acidobacteriales a	no
		Actinobacteria	
KC200506	172Ao_b28	Acidimicrobiales a	SC23
KC200523	213Bp_b18	Acidimicrobiales a	no
KC200537	5Bp_b6	Acidimicrobiales a	no
KC200522	212Bp_b16	Acidimicrobiales a	no
		Cyanobacteria	
KC200535	42Da_b27	Synechococcus a	no
KC200492	130Da_b28	Prochlorococcus a	no
KC200532	23Bp_b42	Prochlorococcus a	no
KC200527	216Bp_b38	Prochlorococcus a	no
		Chloroflexi	
KC200512	188Cr_b2	Anaeorilineales a	SCC11
KC200493	132Ao_b6	Anaeorilineales a	SCC11
KC200501	158Ao_b5	Anaeorilineales a	SCC11
KC200514	192Cr_b9	SAR202 clade a	SCC13
KC200489	110Ao_b21	SAR202 clade a	SC45, SCC15
KC200517	198Cr_b19	SAR202 clade a	SC43
KC200529	221Ao_b12	SAR202 clade a	SC43
KC200485	102Ao_b10	SAR202 clade a	SC43
KC200528	220Ao_b11	SAR202 clade a	SC43
KC200534	31Ao_b18	SAR202 clade a	no
KC200533	252Cr_b27	TK10 a	SC46
KC200490	114Ao_b33	TK10 a	SC46
KC200487	108Ao_b19	Caldilineales, Caldilinea b	SC36

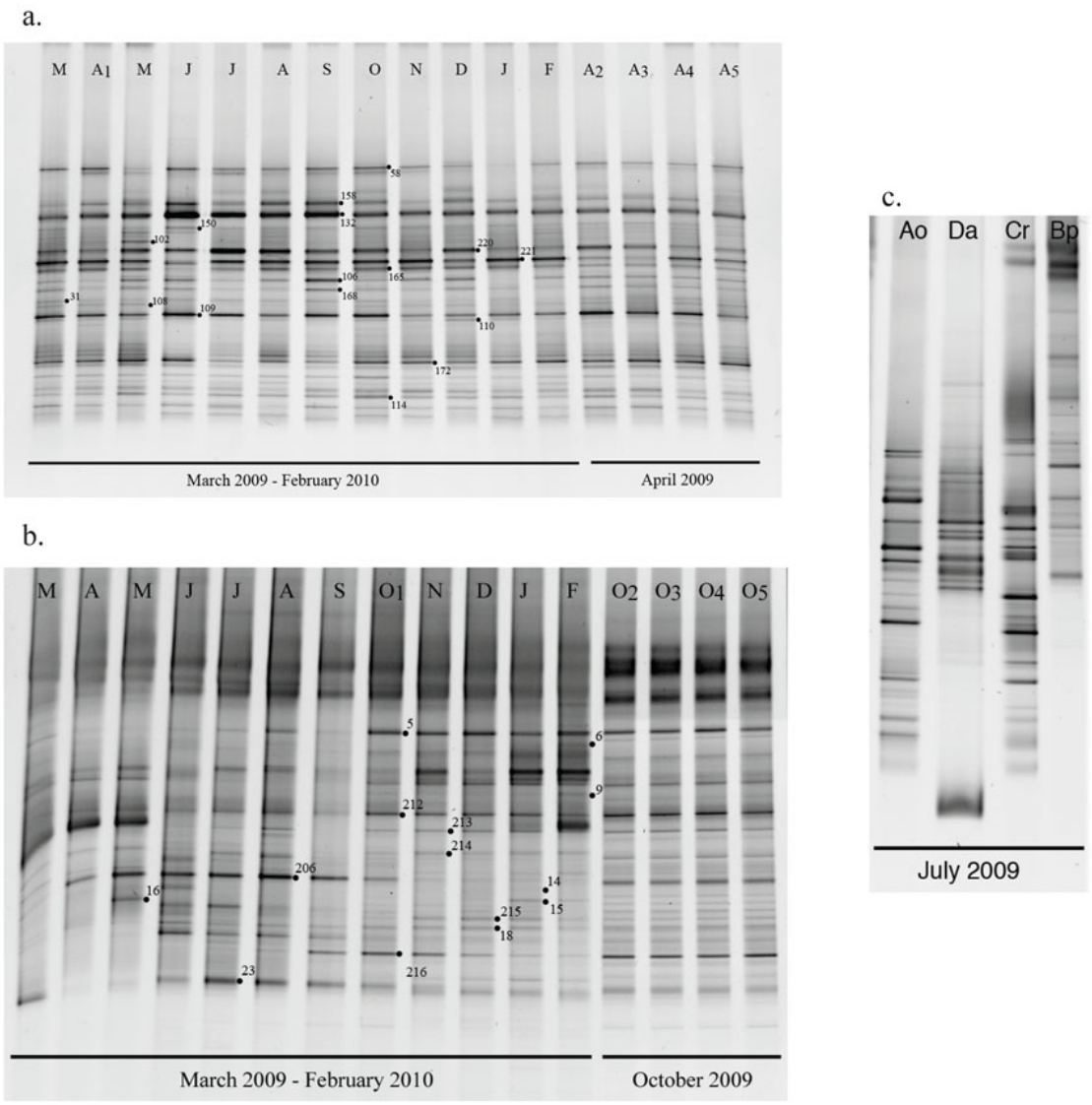


Figure A.1: DGGE analysis of 16S rRNA amplicons using 16S rDNA bacterial primers. (a) Monthly samples over one year of *Agelas oroides* with 4 replicates in March, and (b) the bacterioplankton with 4 replicates in October. Numbered dots mark the bands that were further excised and sequenced (see Table **??**). (c) DGGE image showing the relative distribution of the bands for the three sponge hosts and the bacterioplankton.

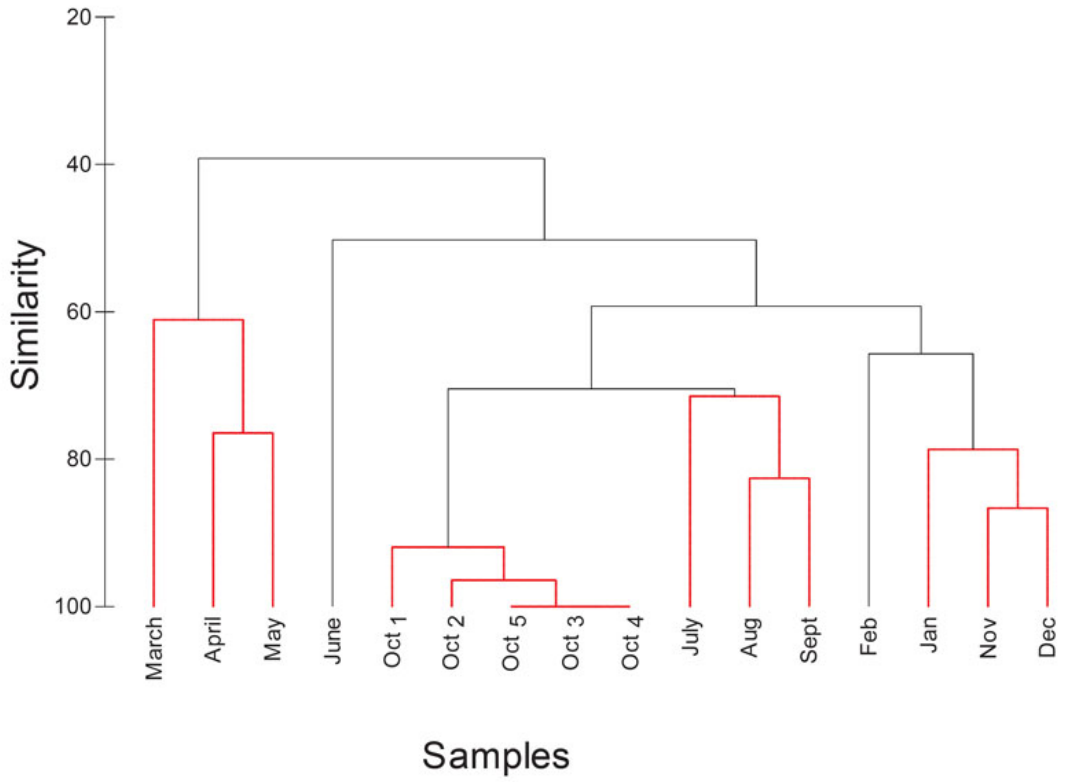


Figure A.2: Dendrogram showing the monthly Jaccard index of similarity for the bacterioplankton. October has 5 replicates, more similar between themselves than to any other month (i.e., the individual variability within months are low). Red branches indicate that the SIMPROF analysis could not find any statistical evidence for any sub-structure within these clusters.

Chapter II

Table B.1: Pairwise comparisons between hosts. The upper triangular matrix corresponds to pairwise comparisons of Jaccard distances, whereas the lower triangular matrix represents pairwise comparisons of unweighted UniFrac distances.

	Ao	Cr	Pf	Ad	Da	Cc
Ao		R=0.910 P<0.0001	R=0.845 P<0.0001	R=0.893 P<0.0001	R=0.606 P<0.0001	R=0.935 P<0.0001
Cr	R=0.795 P<0.0001		R=0.399 P<0.0001	R=0.673 P<0.0001	R=0.565 P<0.0001	R=0.904 P<0.0001
Pf	R=0.703 P<0.0001	R=0.431 P<0.0001		R=0.314 P<0.0001	R=0.411 P<0.0001	R=0.731 P<0.0001
Ad	R=0.700 P<0.0001	R=0.572 P<0.0001	R=0.376 P<0.0001		R=0.556 P<0.0001	R=0.742 P<0.0001
Da	R=0.742 P<0.0001	R=0.707 P<0.0001	R=0.832 P<0.0001	R=0.690 P<0.0001		R=0.667 P<0.0001
Cc	R=0.850 P<0.0001	R=0.889 P<0.0001	R=0.735 P<0.0001	R=0.558 P<0.0001	R=0.762 P<0.0001	

Notes: ANOSIM is a nonparametric permutation method testing for differences among groups. The method, similarly to NMDS, operates on the rank order of dissimilarity values and not directly on the dissimilarity object. Ao corresponds to *A. oroides*, Cr to *C. reniformis*, Pf to *P. ficiformis*, Ad to *A. damicornis*, Da to *D. avara*, and Cc to *C. crambe*. Each P value is presented with its corresponding R value. An R value near 1 means that there is dissimilarity between the hosts, while an R value close to 0 indicates no significant dissimilarity between the hosts. Significant difference between hosts after multiple testing using Bonferoni correction (P<0.00333)

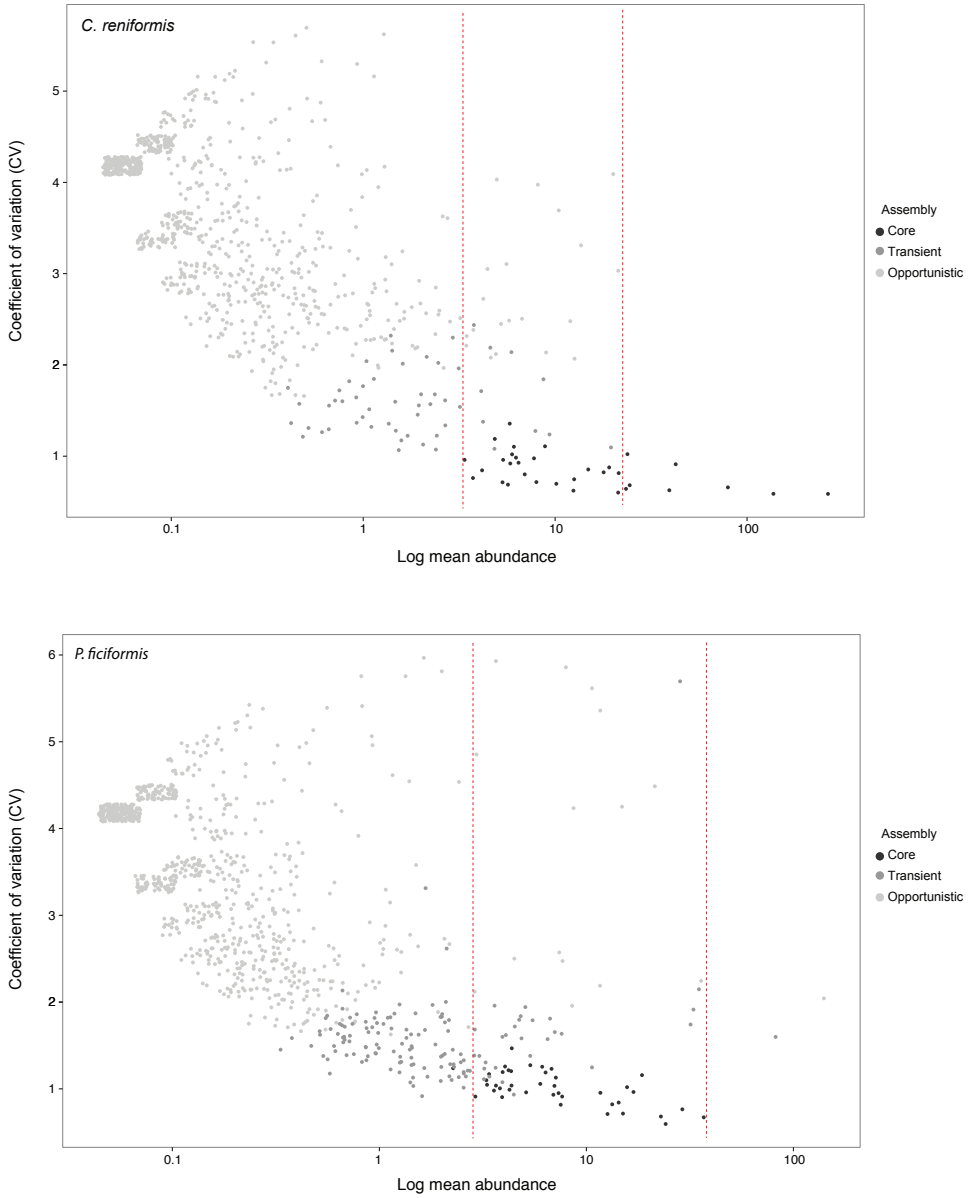


Figure B.1: The relationship between coefficient of variation (CV) and (log) mean abundance of individual species for the microbiomes of host HMA hosts *C. reniformis* and *P. ficiformis*. Overlaying points have been separated by adding jitter (random noise) of 0.1 in both y and x direction. Opportunistic, transient and core species are each shown by an increasing grey scale. The red dashed lines mark a potentially critical area by which you can predict microbiome temporal stability. If there are a few abundant and occasional species relative to the number of core species, stability is predicted to be high, whereas, if there are many abundant and occasional species compared to core species, stability is predicted to be low.

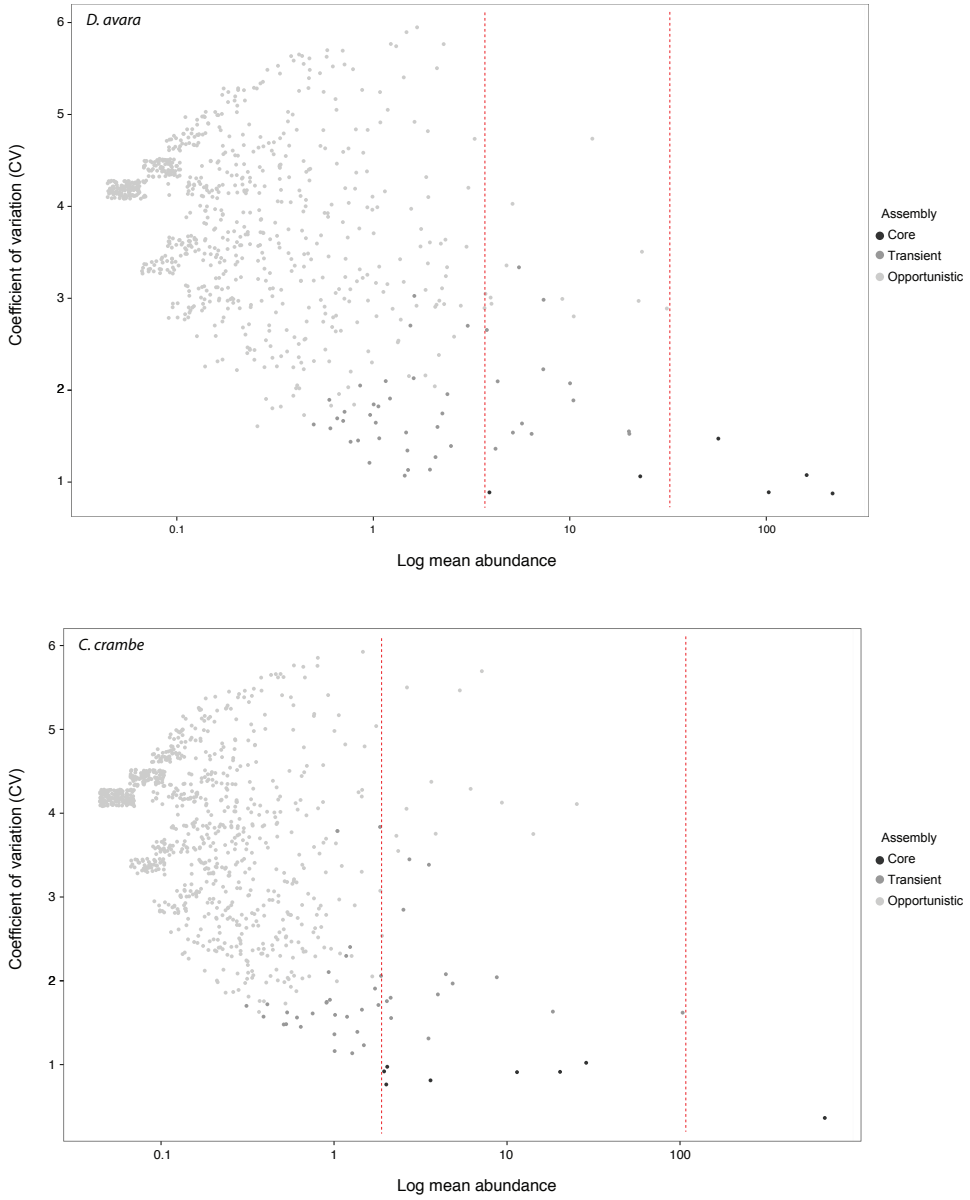


Figure B.2: The relationship between coefficient of variation (CV) and (log) mean abundance of individual species for the microbiomes of host LMA hosts *D. avara* and *C. crambe*. Overlaying points have been separated by adding jitter (random noise) of 0.1 in both y and x direction. Opportunistic, transient and core species are each shown by an increasing grey scale. The red dashed lines mark a potentially critical area by which you can predict microbiome temporal stability. If there are a few abundant and occasional species relative to the number of core species, stability is predicted to be high, whereas, if there are many abundant and occasional species compared to core species, stability is predicted to be low.

Table B.2: Detailed analysis of species overlap between hosts across assemblages. Each table shows the percentage of species overlap between a. Total microbiomes, b. Opportunistic assemblages c. Transient assemblages and d. Core microbiomes. Note that the lower and upper diagonal differ due to differences in total microbiome richness across host and assemblages (See Table *1*).

Table B.3: Total microbiomes

	Ao	Cr	Pf	Ad	Da	Cc
Ao		8.7	7.8	10.4	16.4	13.9
Cr	9.6		38.5	28.6	26.2	20.5
Pf	7.9	35.2		34.7	16.2	25.4
Ad	11.7	29.0	38.6		18.8	32.7
Da	16.3	23.5	15.9	16.6		21.8
Cc	11.0	14.7	20.0	23.1	17.4	

Table B.4: Opportunistic assemblages

	Ao	Cr	Pf	Ad	Da	Cc
Ao		8.2	8.0	10.7	14.5	14.0
Cr	8.9		38.5	28.6	26.2	20.5
Pf	8.2	28.1		29.5	15.7	20.3
Ad	11.6	26.2	31.2		18.4	30.4
Da	13.9	19.2	14.7	16.3		21.3
Cc	10.7	14.1	15.1	21.4	17.0	

Table B.5: Transient assemblages

	Ao	Cr	Pf	Ad	Da	Cc
Ao		0.0	0.0	1.1	1.1	1.1
Cr	0.0		1.9	0.0	1.9	0.0
Pf	0.0	0.7		1.4	0.0	0.0
Ad	3.2	0.0	6.5		0.0	6.5
Da	2.3	2.3	0.0	0.0		2.3
Cc	2.4	0.0	0.0	4.9	2.4	

Table B.6: Core microbiomes

	Ao	Cr	Pf	Ad	Da	Cc
Ao		0.0	0.0	0.0	0.0	0.0
Cr	0.0		0.0	0.0	0.0	0.0
Pf	0.0	0.0		0.0	0.0	0.0
Ad	0.0	0.0	0.0		0.0	12.5
Da	0.0	0.0	0.0	0.0		0.0
Cc	0.0	0.0	0.0	12.5	0.0	

Table B.7: Percentage of species within each assemblage and host assigning to sponge-specific clusters.

	HMA			LMA		
	A. oroides	*C. reniformis*	*P. ficiformis*	*A. damicornis*	*D. avara*	*C. crambe*
Core	42.2	45.6	60	25	0	12.5
Transient	43.3	40.7	47.1	25.8	9.1	9.8
Opportunistic	22.8	33.9	35.6	22.2	9.5	11.8

Table B.8: Median (including 1^{st} and 3^{rd} quintiles) of average monthly abundances of species assigning to sponge-specific clusters shown for each assemblage and host

		HMA			LMA		
		A. oroides	*C. reniformis*	*P. ficiformis*	*A. damicornis*	*D. avara*	*C. crambe*
Core	$1^{st}Qu.$	6.7	6.6	4.2	29.2	0	1.8
	Median	**11.3**	**19.8**	**6.7**	**50.3**	**0**	**16.5**
	$3^{rd}Qu.$	19.2	32.8	13.8	71.5	0	39.0
Transient	$1^{st}Qu.$	0.6	1.1	1.0	3.1	0.6	1.0
	Median	**1.0**	**1.6**	**1.7**	**3.9**	**0.8**	**2.1**
	$3^{rd}Qu.$	2.3	3.1	2.5	5.2	1.2	33.9
Opportunistic	$1^{st}Qu.$	0.0	0.0	0.0	0.0	0.0	0.0
	Median	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.1**
	$3^{rd}Qu.$	0.1	0.1	0.1	0.1	0.1	0.1

Table B.9: Percentage (mean $\pm$ SD) of species within each randomized realisation assemblage and host assigning to sponge-specific clusters

	HMA			LMA		
	A. oroides	*C. reniformis*	*P. ficiformis*	*A. damicornis*	*D. avara*	*C. crambe*
Core	48.5 ± 0.21	48.6 ± 0.20	50.9 ± 0.22	48.6 ± 0.19	50.8 ± 0.23	48.2 ± 0.20
Transient	30.1 ± 1.6	29.8 ± 1.6	30.2 ± 1.8	29.7 ± 1.6	30.2 ± 1.7	29.8 ± 1.6
Opportunistic	2.9 ± 0.1	2.8 ± 0.1	3.4 ± 0.1	2.8 ± 0.1	3.3 ± 0.1	2.8 ± 0.1

Note: Mean $\pm$ SD were calculated from 999 randomised realizations. See Methods Null model.

Table B.10: Median (including 1^{st} and 3^{rd} quintiles) (mean $\pm$ SD) of the average monthly abundance of species belonging to sponge-specific clusters across assemblages within each random host realization.

		HMA			LMA		
		A. oroides	*C. reniformis*	*P. ficiformis*	*A. damicornis*	*D. avara*	*C. crambe*
Core	$1^{st}Qu.$	1.8 ± 0.2	1.7 ± 0.2	1.4 ± 0.1	1.8 ± 0.2	1.4 ± 0.2	1.9 ± 0.2
	Median	$\mathbf{2.6 \pm 0.2}$	$\mathbf{2.5 \pm 0.2}$	$\mathbf{1.9 \pm 0.1}$	$\mathbf{2.6 \pm 0.2}$	$\mathbf{2.1 \pm 0.2}$	$\mathbf{2.6 \pm 0.2}$
	$3^{rd}Qu.$	3.4 ± 0.2	3.3 ± 0.2	2.7 ± 0.2	3.3 ± 0.2	3.0 ± 0.2	3.4 ± 0.2
Transient	$1^{st}Qu.$	3.2 ± 0.3	3.1 ± 0.2	2.4 ± 0.2	3.2 ± 0.2	2.4 ± 0.2	3.3 ± 0.2
	Median	$\mathbf{4.4 \pm 0.3}$	$\mathbf{4.3 \pm 0.3}$	$\mathbf{3.4 \pm 0.2}$	$\mathbf{4.4 \pm 0.3}$	$\mathbf{3.6 \pm 0.2}$	$\mathbf{4.5 \pm 0.3}$
	$3^{rd}Qu.$	5.9 ± 0.4	5.8 ± 0.3	4.7 ± 0.3	5.8 ± 0.4	5.1 ± 0.3	6.0 ± 0.4
Opportunistic	$1^{st}Qu.$	16.0 ± 1.0	15.4 ± 1.0	11.5 ± 0.8	15.6 ± 1.1	11.2 ± 0.8	16.3 ± 1.1
	Median	$\mathbf{20.2 \pm 1.2}$	$\mathbf{20.0 \pm 1.2}$	$\mathbf{17.3 \pm 1.1}$	$\mathbf{20.0 \pm 1.2}$	$\mathbf{17.1 \pm 1.0}$	$\mathbf{20.3 \pm 1.3}$
	$3^{rd}Qu.$	23.7 ± 1.4	23.6 ± 1.4	21.9 ± 1.2	23.8 ± 1.5	24.8 ± 1.3	24.2 ± 1.4

Note: Mean $\pm$ SD were calculated from 999 randomised realizations. See Methods Null model.

Table B.11: Mean rank sums for relative abundance across assemblages and hosts. For each host, a Kruskal-Wallis rank sum test was done to test for differences between assemblages (Ao: H=40.103, df=2, P<0.001 two-tailed; Cr: H=31.222, df=2, P<0.001 two-tailed; Cc: H=40.696, df=2, P<0.001 two-tailed; Da: H=48.474, df=2, P<0.001 two-tailed; Ad: H=17.747, df=2, P<0.001 two-tailed; Pf: H=3.567, df=2, P=0.168 two-tailed). Where there was a significant difference, Dunn's post-hoc test for pairwise comparisons with bonferroni correction was used.

	HMA			LMA		
	A. oroides	*C. reniformis*	*P. ficiformis*	*A. damicornis*	*D. avara*	*C. crambe*
Core	81.47	74.33	78.33	73.06	59.32	46.50
Transient	40.15	33.17	31.25	25.17	37.11	59.28
Opportunistic	41.88	56.00	53.92	65.28	67.07	57.72
	a	b	b	c	c	d

Note: The lowercase letters indicate different significant scenarios (see Figure *5* in Chapter *2*). a. The core assembly was significantly different from the transient and opportunistic assembly, but transient and opportunistic assemblages were not significantly different from each other. b. All assemblages were significantly different. c. Core and opportunistic assemblages were not significantly different, but core and transient assemblages and transient and opportunistic assemblages were significantly different. d. No significant differences between any assemblages.

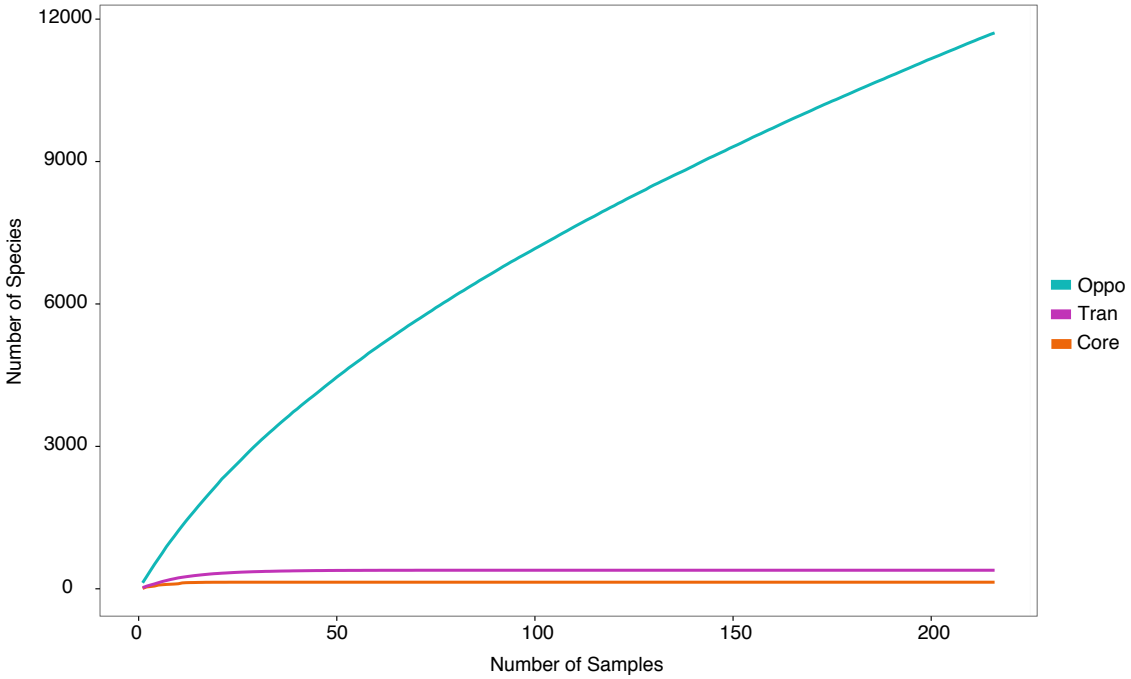


Figure B.3: Rarefaction curve aggregated over hosts and months for each assemblage.

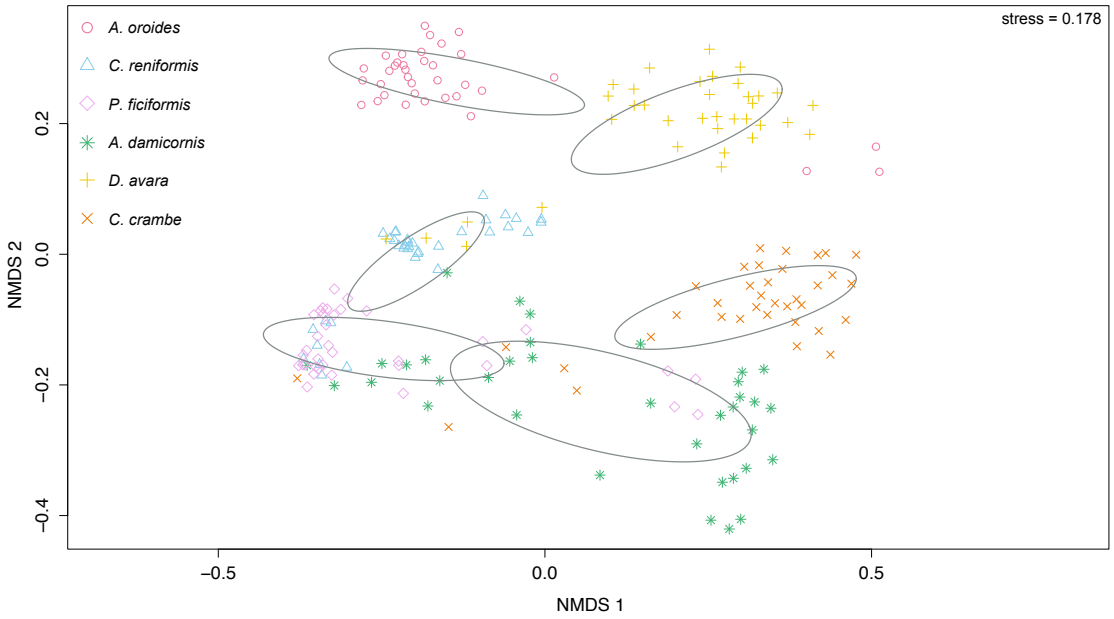


Figure B.4: Non-metric multidimensional scaling (NMDS) calculated on unweighted UniFrac distances of microbial species present among the 36 monthly samples from each host. Colours and shapes denote all monthly samples from a given host species. Host samples are surrounded by an ellipse showing the intraspecific variability across the time-series. *A. oroides*: Red circle, *C. reniformis*: Blue triangle, *P. ficiformis*: Purple diamond, *A. damicornis*: Green star, *D. avara*: Yellow cross (+), *C. crambe*: Orange cross (×)

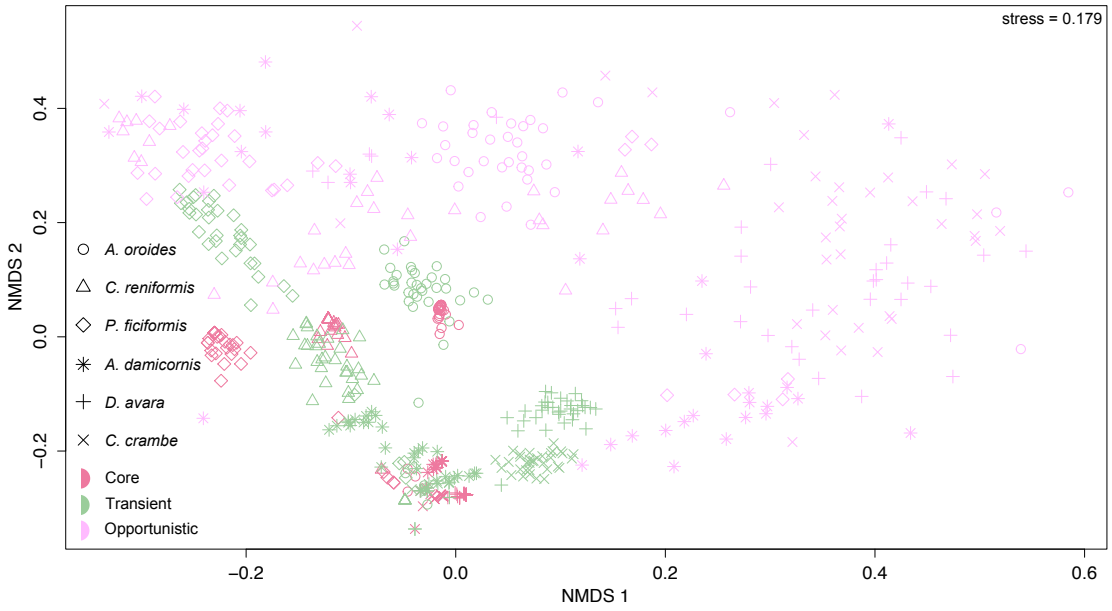


Figure B.5: Non-metric multidimensional scaling (NMDS) calculated on unweighted UniFrac distances of microbial species present among monthly samples from each host and assemblage. Colours denote all (36) monthly samples from a given host, and shapes denote all (216) monthly samples from a given assemblage. Circle, *C. reniformis*: triangle, *P. ficiformis*: diamond, *A. damicornis*: star, *D. avara*: cross (+), *C. crambe*: cross (×) and colours denote assemblages; purple: core microbiomes, green: transient assemblages, pink: opportunistic assemblages.

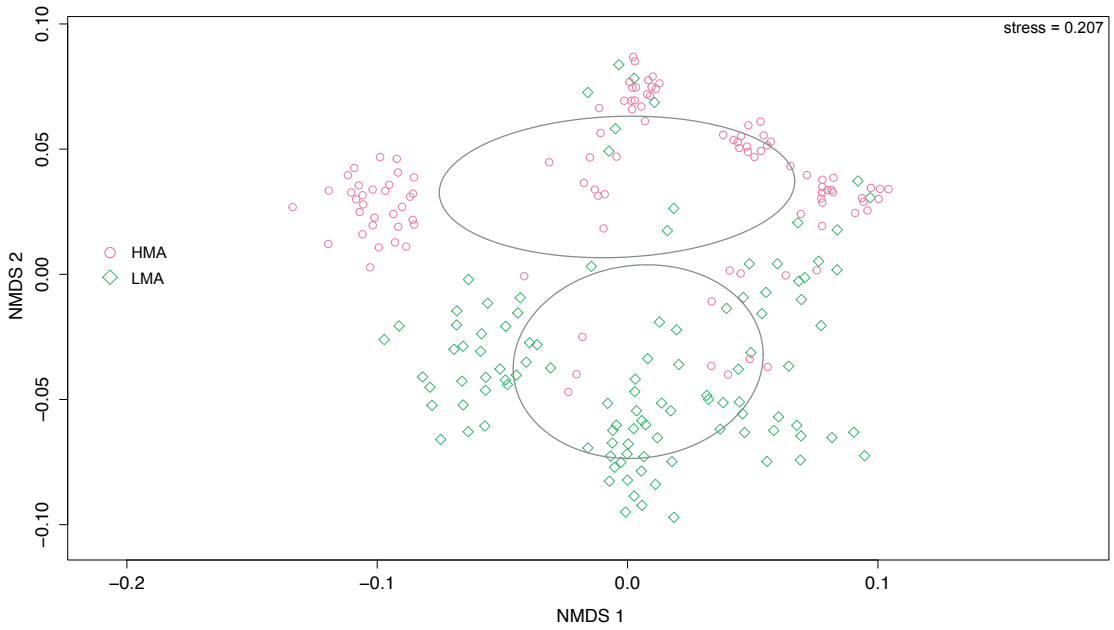


Figure B.6: Non-metric multidimensional scaling (NMDS) calculated on Jaccard distances of microbial species among monthly samples from each host classification (HMA/LMA). Red circles represent all (108) monthly samples from the three HMA hosts, while green triangles denote all (108) monthly samples belonging to the three LMA hosts.

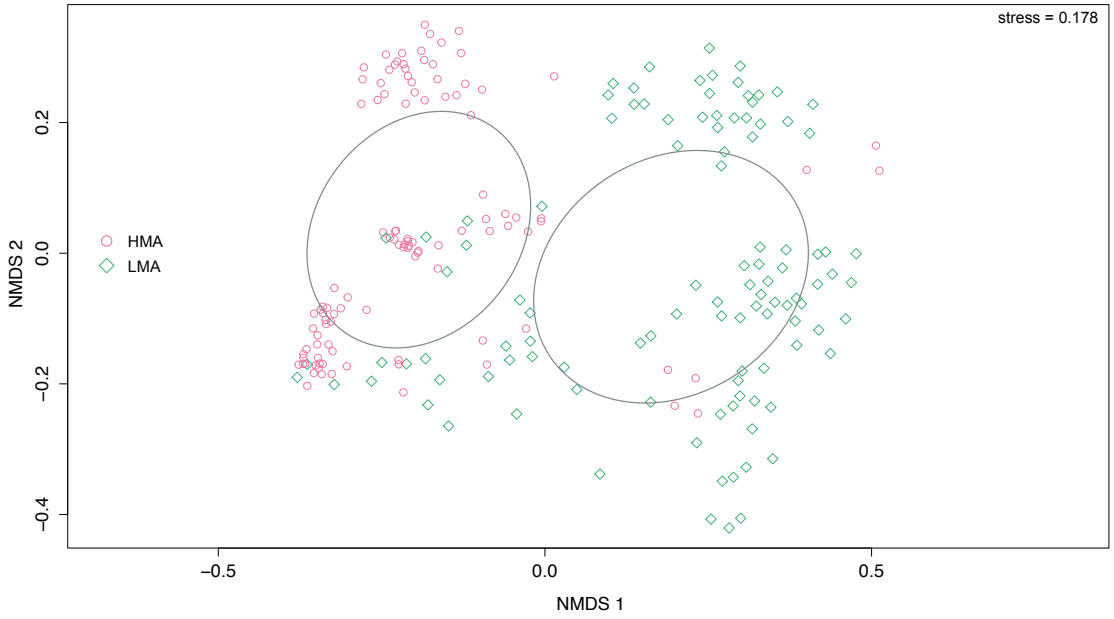


Figure B.7: Non-metric multidimensional scaling (NMDS) calculated on unweighted UniFrac distances of microbial species among monthly samples from each host classification (HMA/LMA). Red circles represent all (108) monthly samples from the three HMA hosts, while green triangles denote all (108) monthly samples belonging to the three LMA hosts.

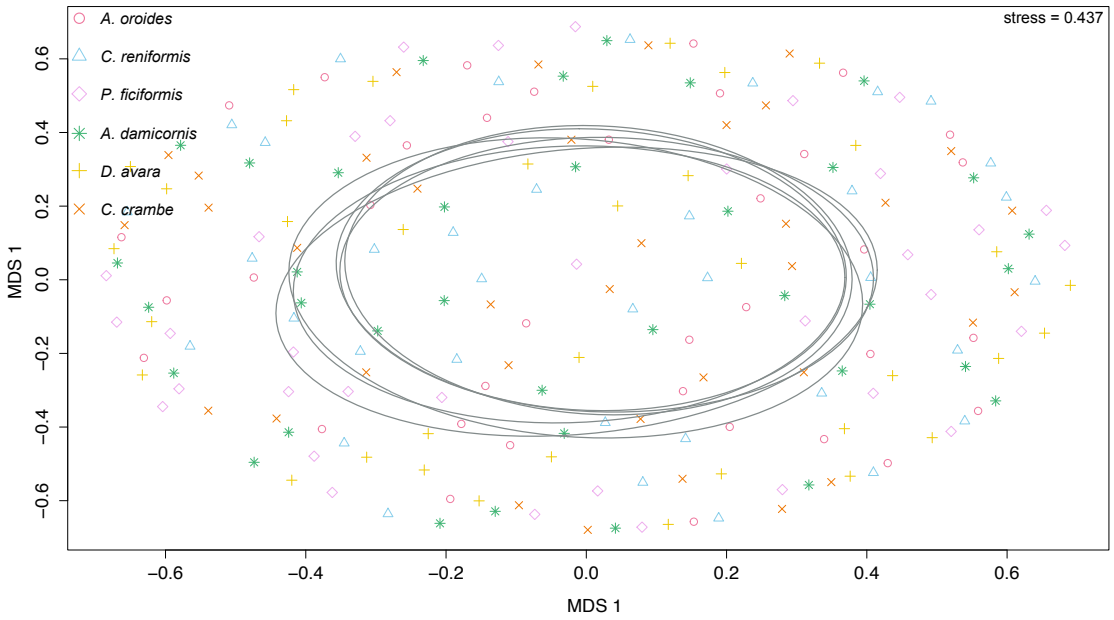


Figure B.8: Non-metric multidimensional scaling (NMDS) calculated on the average Jaccard distances of microbial species among random realisations. The average distance was calculated over 999 randomised realisations for each host. See *Null model* in the *Methods* for more details.

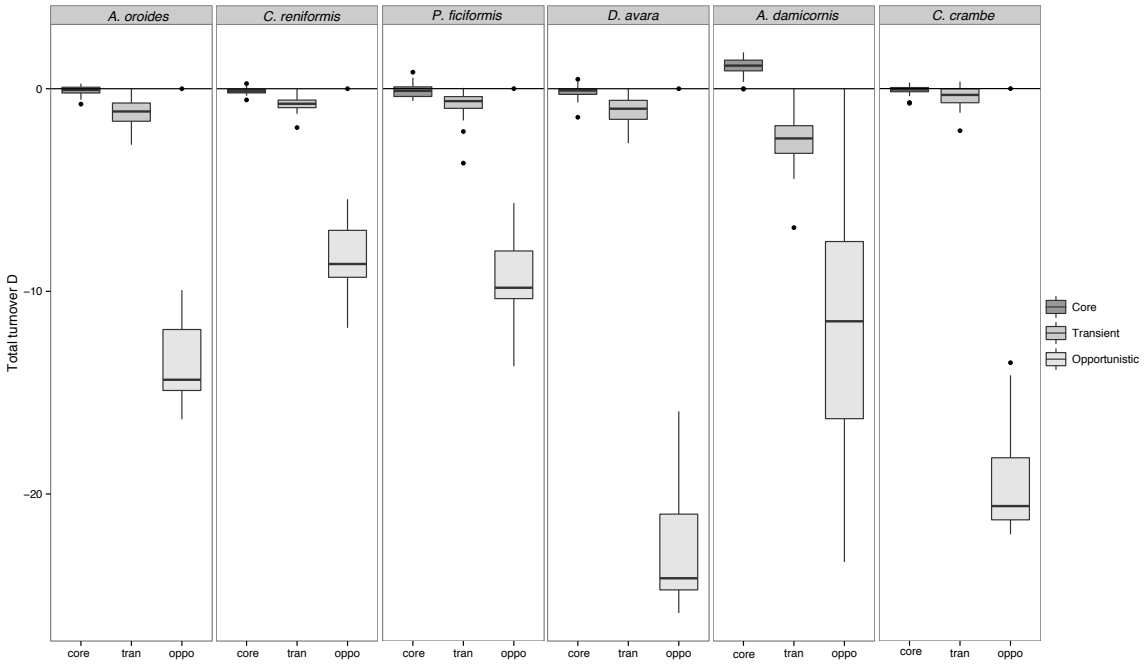


Figure B.9: Temporal total turnover (D) for for each assemblage and host.

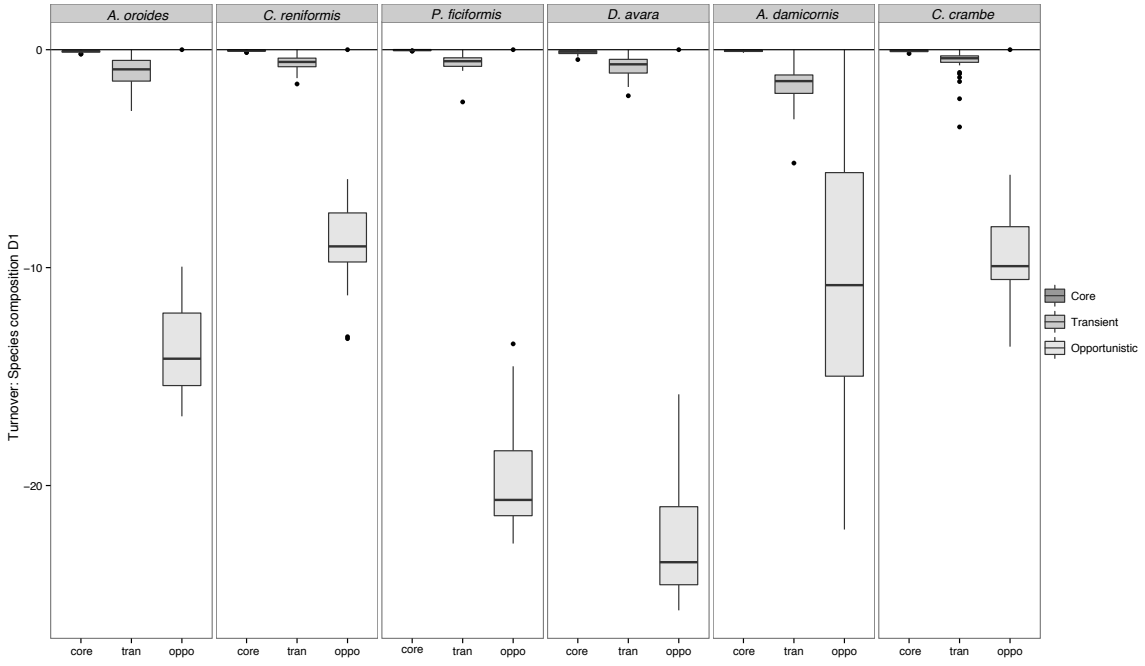


Figure B.10: Temporal turnover in terms of species composition (D_1) for for each assemblage and host.

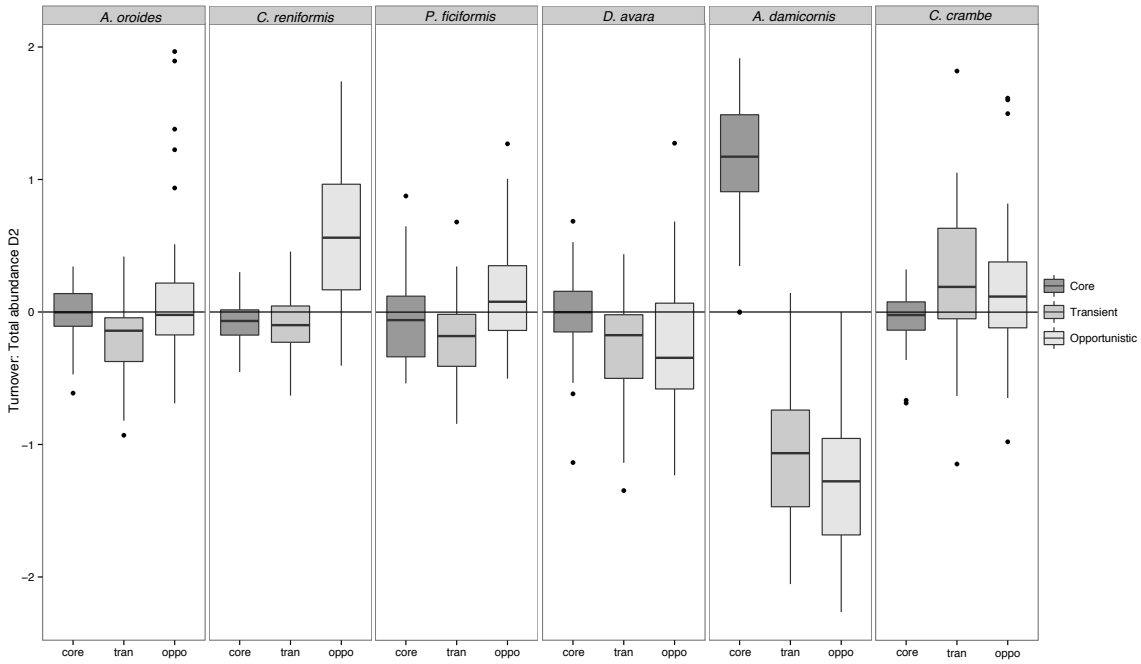


Figure B.11: Temporal turnover in terms of total abundance (D_2) for for each assemblage and host.

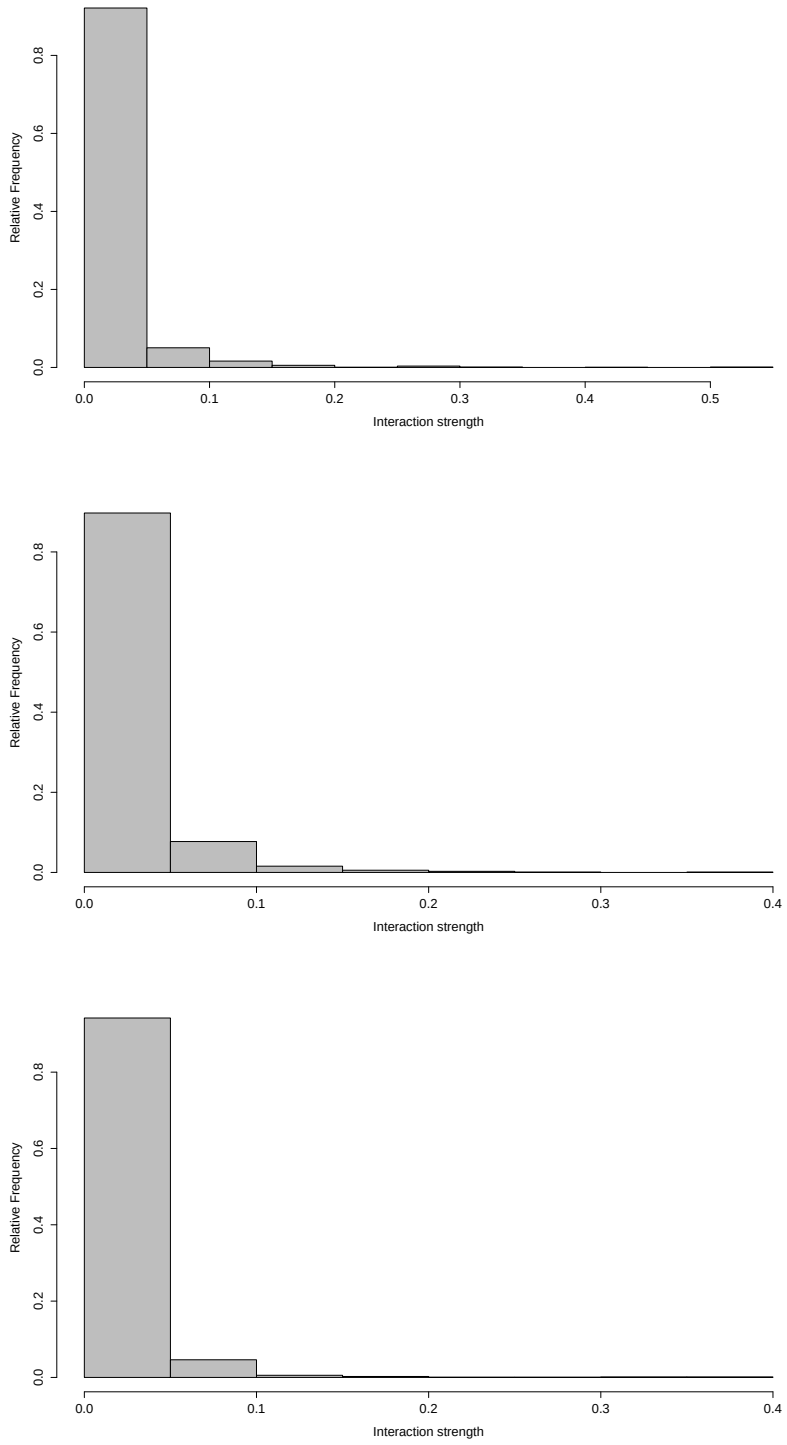
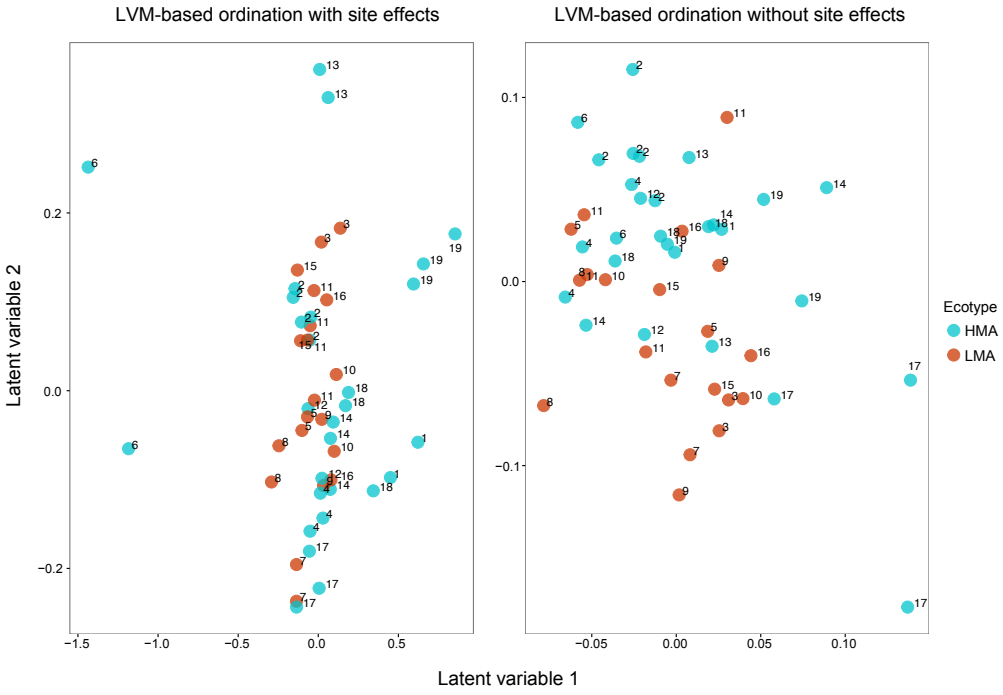
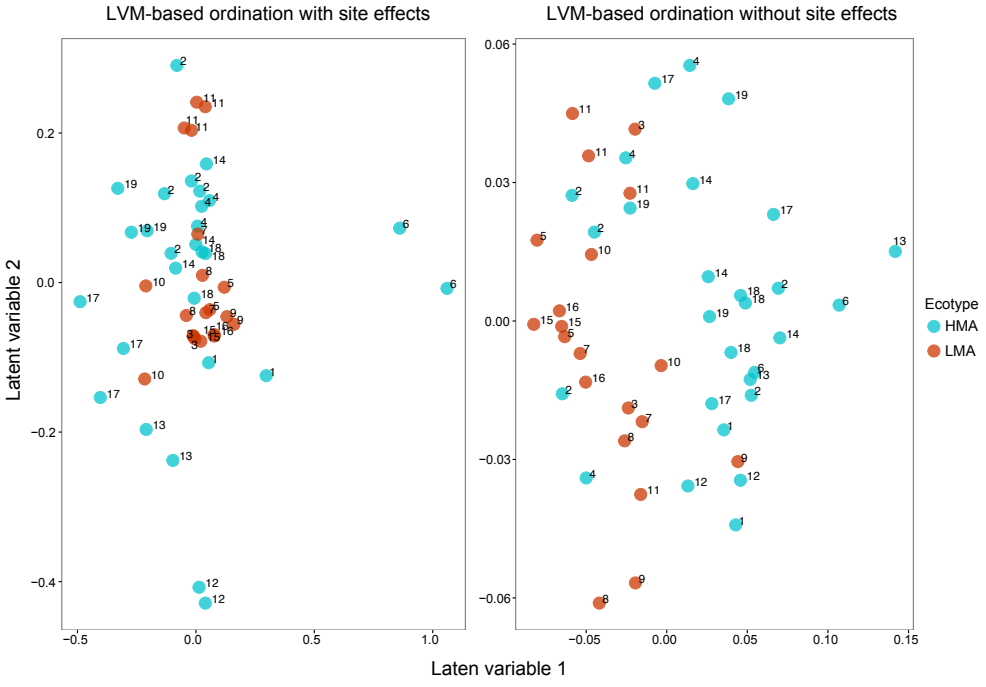


Figure B.12: Frequency histogram of interaction strengths for the core microbiome of *A. oroides* (top), *C. reniformis* (middle) and *P. ficiformis* (bottom). Interaction strength is calculated from the posterior distribution of the interaction coefficient α_{ij}. The distribution is skewed towards many weak and a few strong interactions

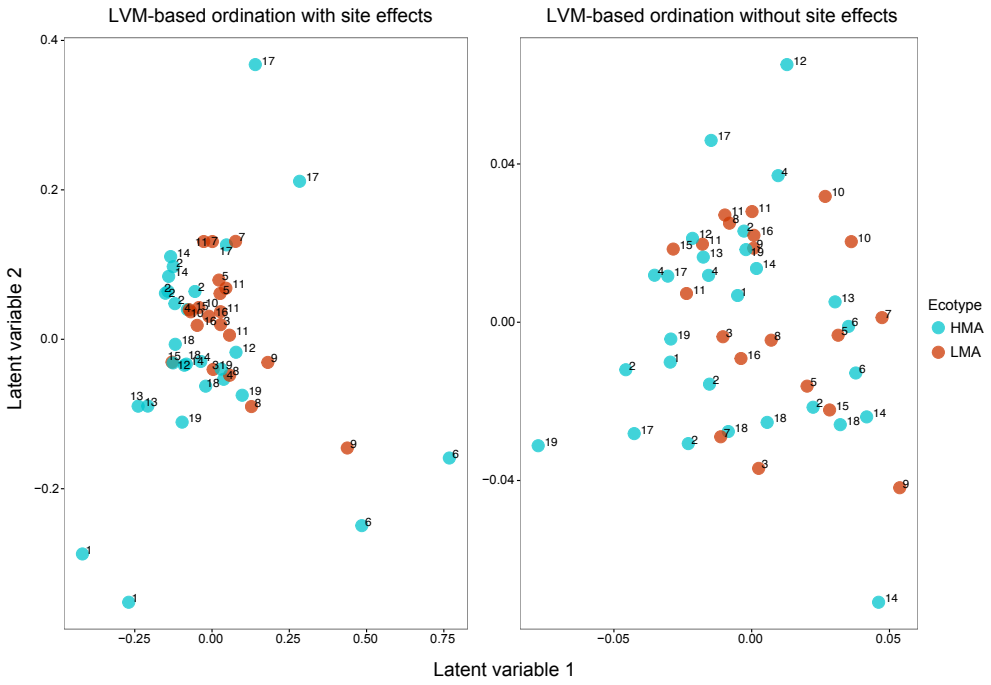
Chapter III


LVM-based ordination with site effects
LVM-based ordination without site effects
Latent variable 2
Latent variable 1
Ecotype
HMA
LMA

[a]

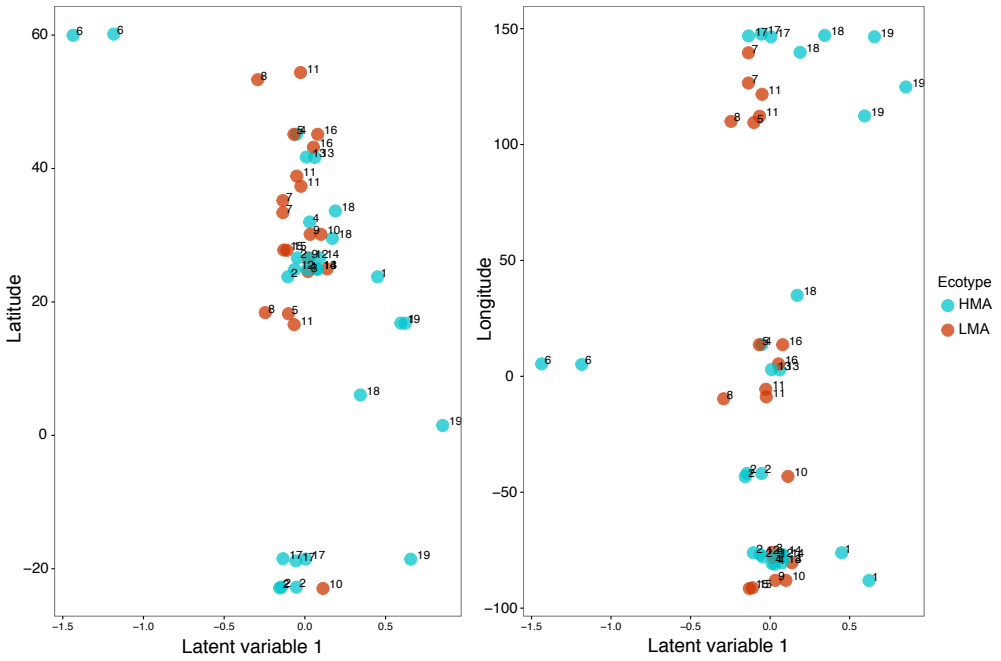

LVM-based ordination with site effects
LVM-based ordination without site effects
Latent variable 2
Laten variable 1
Ecotype
HMA
LMA

[b]

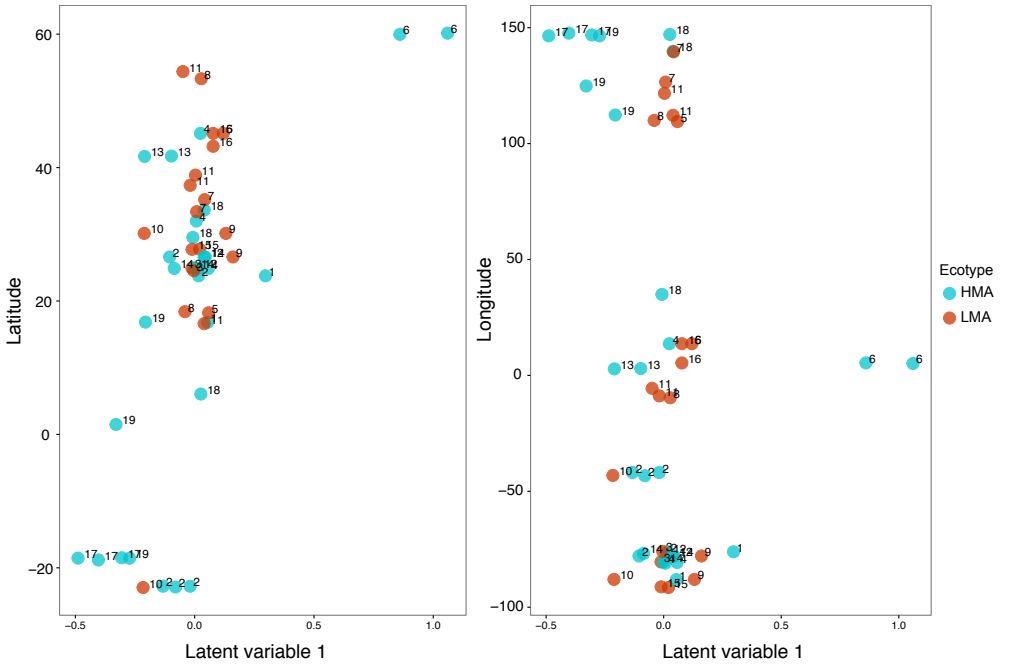


[c]

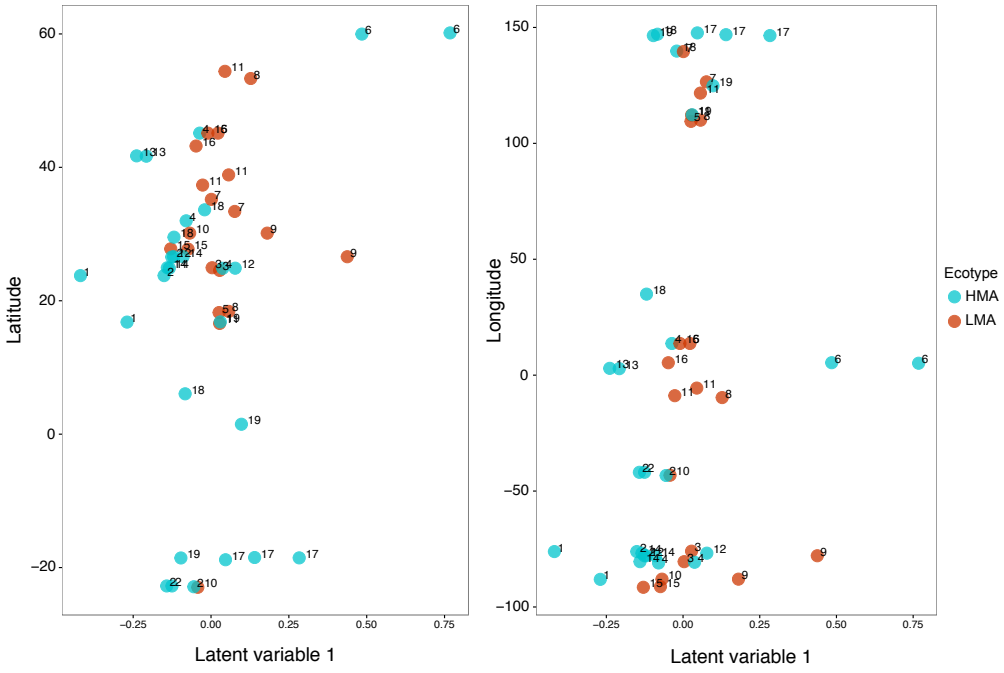
Figure C.1: Model-based unconstrained ordination with (left) and without (right) site effects included. Teal and red correspond to HMA and LMA hosts, respectively. Numbers correspond to host species (host IDs) in Table *3.1*. a represents OTUs at order (85%), b. at family (90%) and genus (95%) level, respectively.


Latitude
Latent variable 1
Longitude
Latent variable 1
Ecotype
HMA
LMA

[a]


Latitude
Latent variable 1
Longitude
Latent variable 1
Ecotype
HMA
LMA

[b]



[c]

Figure C.2: Model-based unconstrained ordination with site effects included. Latent variable 1 plotted against latitude and longitude. Teal and red correspond to HMA and LMA hosts, respectively. Numbers correspond to host species (host IDs) in Table *3.1*. a represents OTUs at order (85%), b. at family (90%) and genus (95%) level, respectively.

Chapter IV

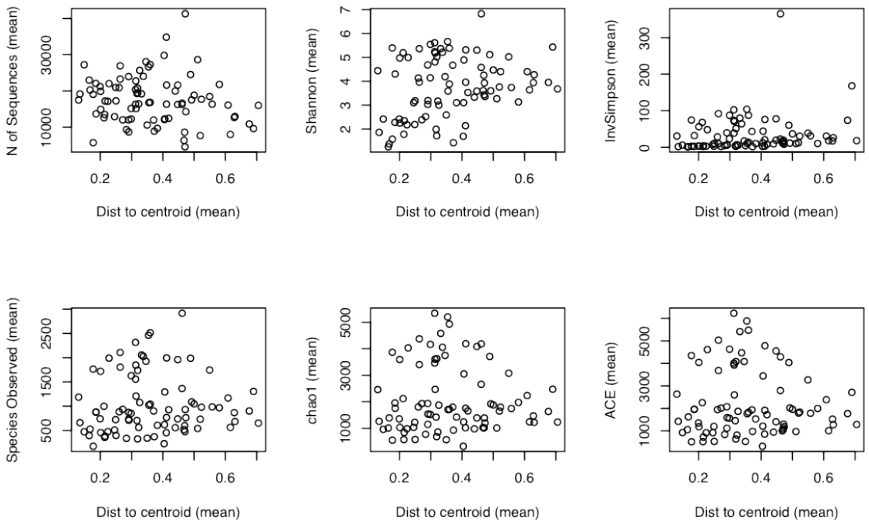


Figure D.1: Scatterplot of the intraspecific community dissimilarity (measured as distance of samples to group centroids) and various alpha-diversity measurements for 16S rRNA gene diversity of symbionts in individual sponge species

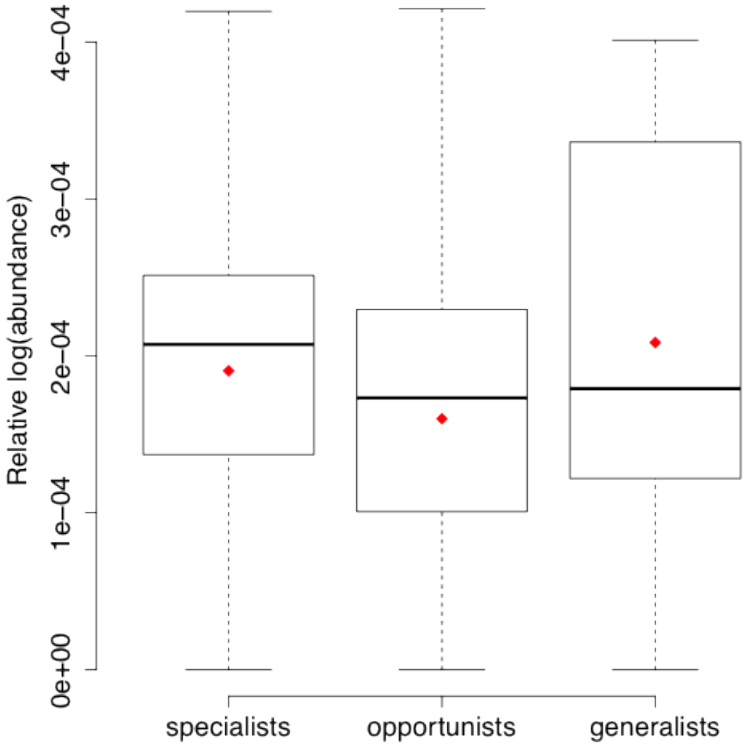


Figure D.2: Relative log abundances of OTUs across the entire sponge-bacteria network classified according to their connectivity pattern. Specialists are OTUs present in less than 5 different sponge species; opportunists are OTUs present in between 5 and 50 different sponge species; and generalists are OTUs present in more than 50 sponge species. Red diamonds show the mean value per group. In the boxes, central rectangles span the first quartile to the third quartile (the interquartile range). Segments inside rectangles show the median and ”whiskers” above and below the boxes show the locations of the minimum and maximum values

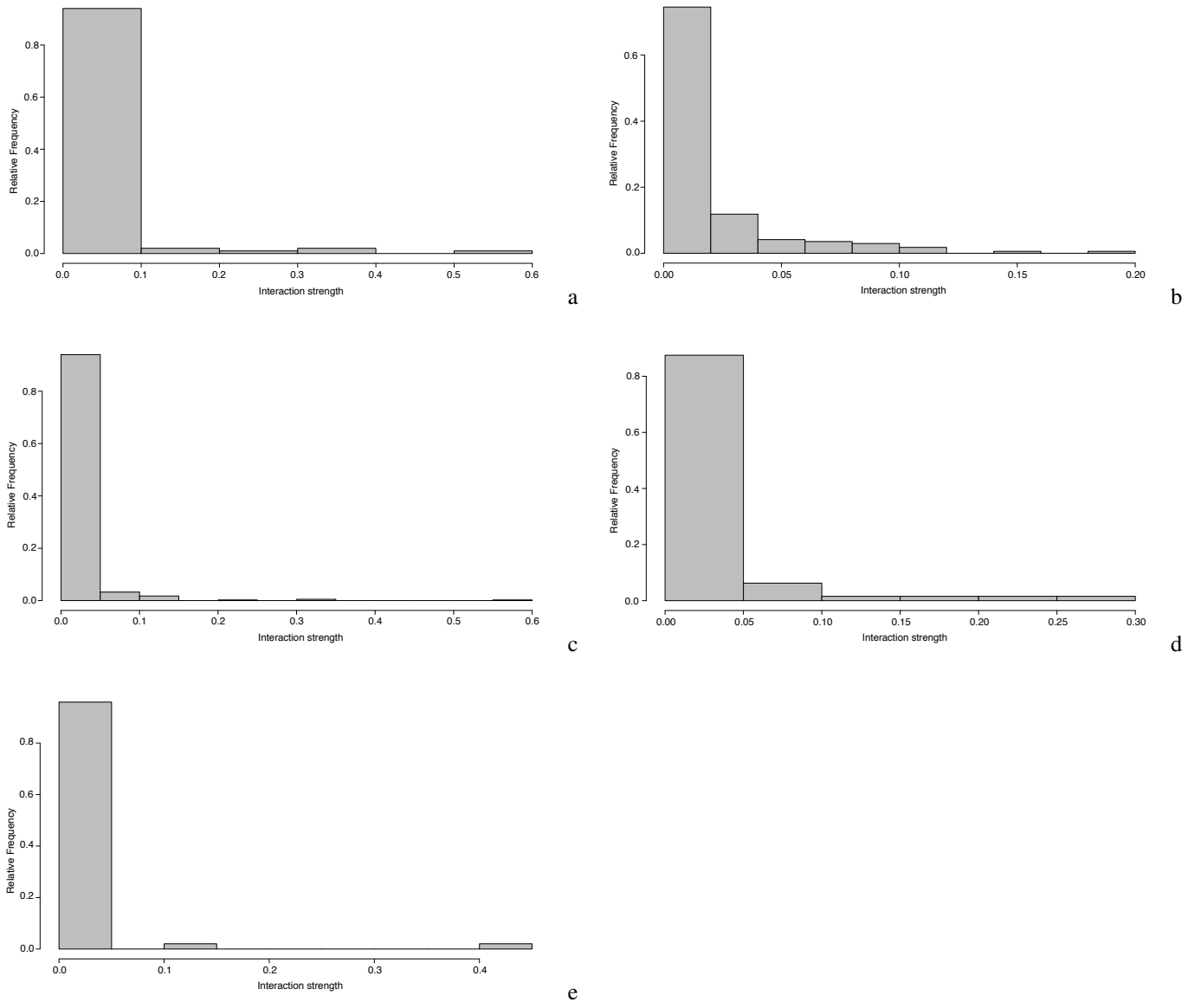
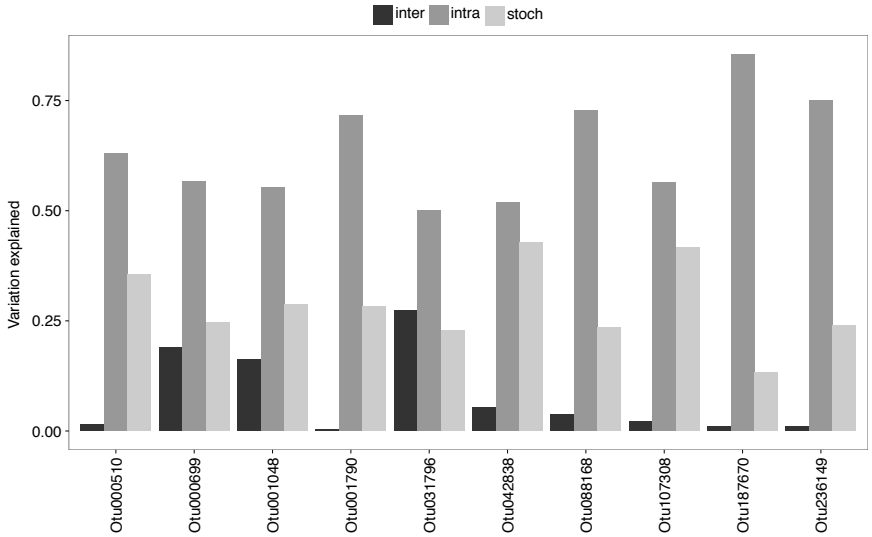
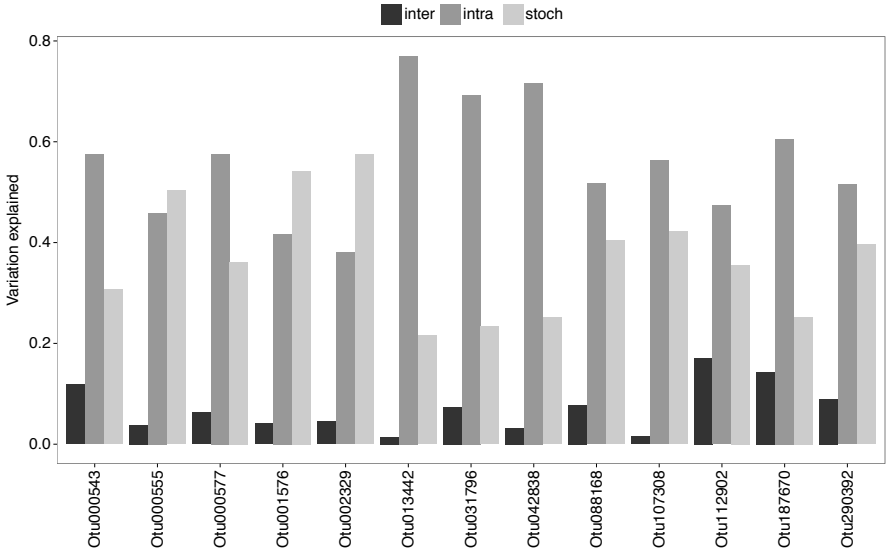


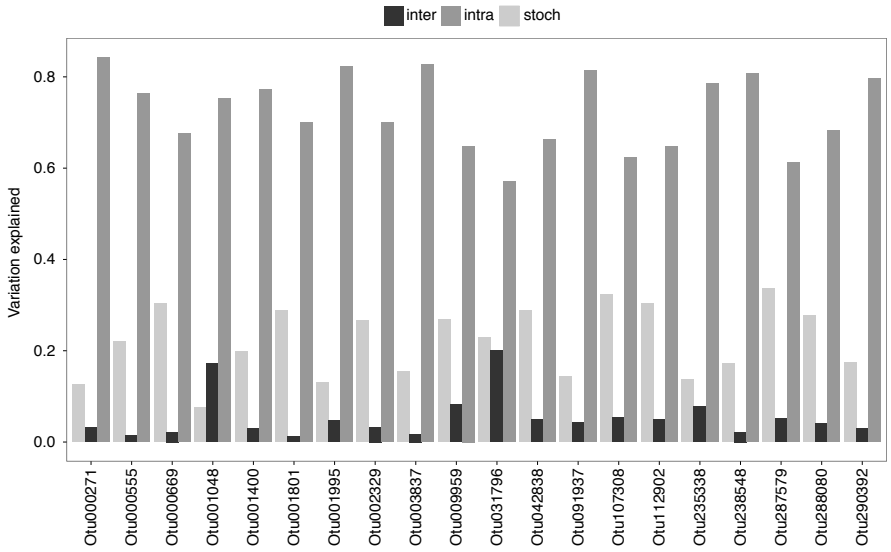
Figure D.3: Frequency histogram of interaction strengths for core microbiome of different sponge species. Interaction strength is calculated from the posterior distribution of the interaction coefficient α_{ij}. a. represent *S. fasciculata*, b. *I. variabilis*, c. *I. oros*, d. *C. delitrix*, e. *C. foliascens*. Note that across host species a skewed distribution is observed characterized by many weak and a few strong interactions


inter
intra
stoch
Variation explained
0.00
0.25
0.50
0.75
Otu000510
Otu000699
Otu001048
Otu001790
Otu031796
Otu042838
Otu088168
Otu107308
Otu187670
Otu236149

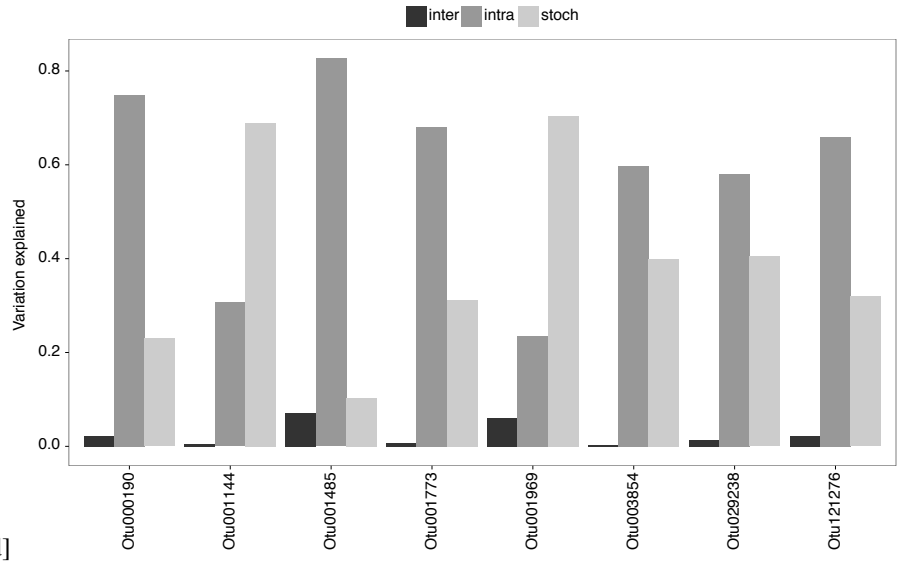
[a]


inter
intra
stoch
Variation explained
0.0
0.2
0.4
0.6
0.8
Otu000543
Otu000555
Otu000577
Otu001576
Otu002329
Otu013442
Otu031796
Otu042838
Otu088168
Otu107308
Otu112902
Otu187670
Otu290392

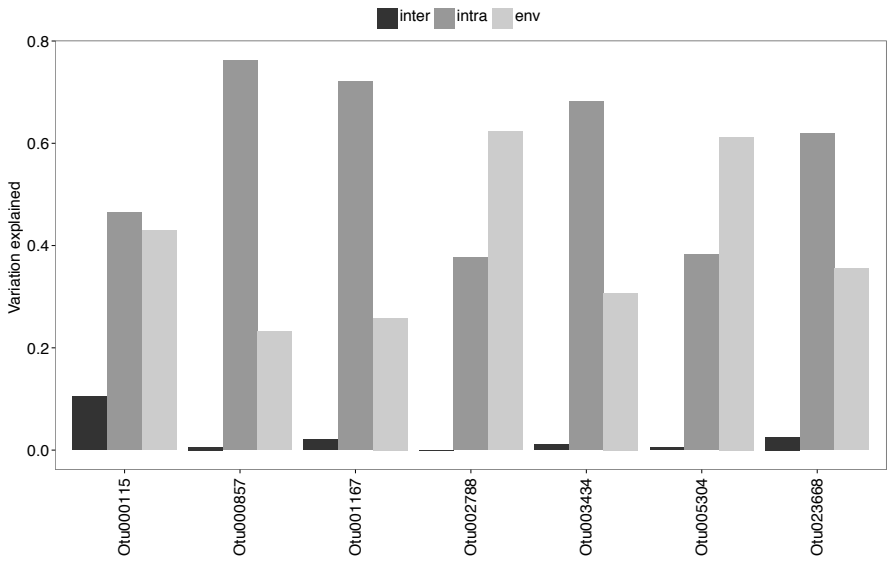
[b]


inter
intra
stoch
Variation explained
0.0
0.2
0.4
0.6
0.8
Otu000271
Otu000555
Otu000669
Otu001048
Otu001400
Otu001801
Otu001995
Otu002329
Otu003837
Otu009959
Otu031796
Otu042838
Otu091937
Otu107308
Otu112902
Otu235338
Otu238548
Otu287579
Otu288080
Otu290392

[c]

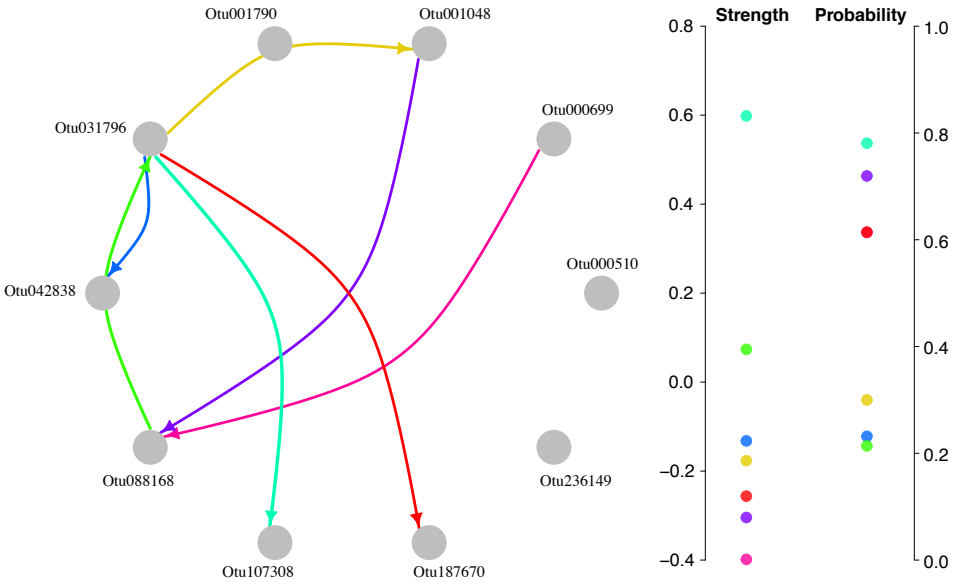

inter
intra
stoch
Variation explained
0.0
0.2
0.4
0.6
0.8
Otu000190
Otu001144
Otu001485
Otu001773
Otu001969
Otu003854
Otu029238
Otu121276

[d]

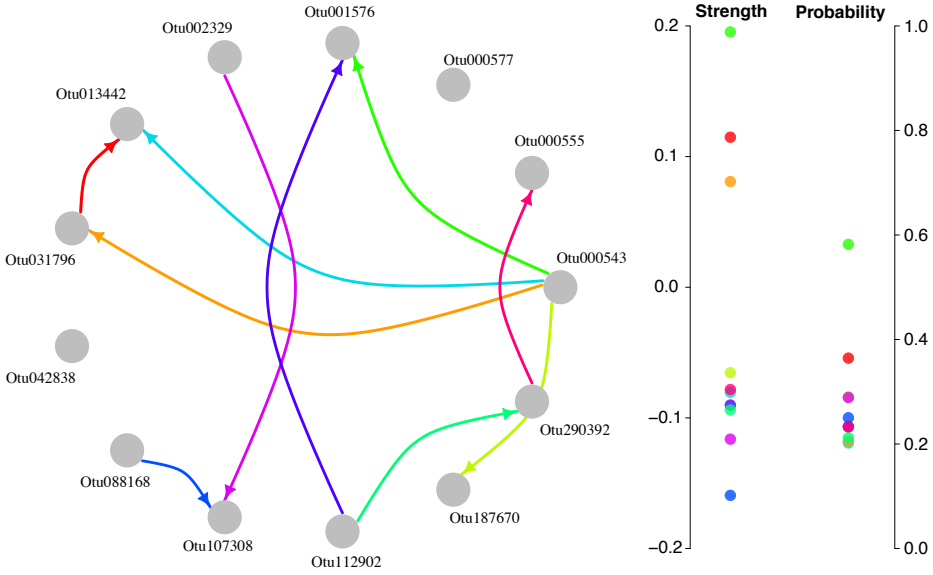


[e]

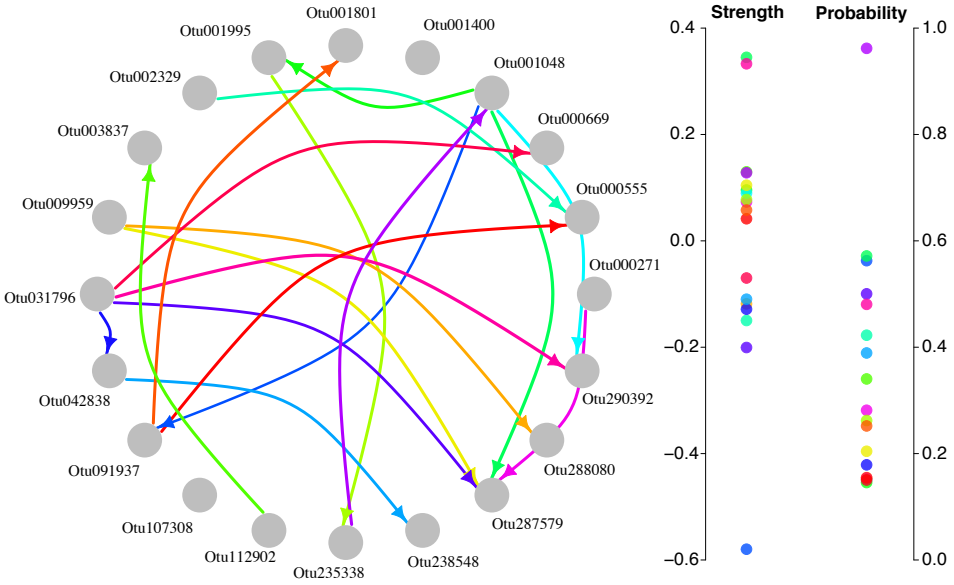
Figure D.3: Variation decomposition per OTU within each core microbiome. The y-axis shows the variation explained by interspecific interactions (black), interspecific interactions (i.e. density dependence) (dark grey) and stochasticity (light grey). a. represent *S. fasciculata*, b. *I. variabilis*, c. *I. oros*, d. *C. delitrix*, e. *C. foliascens*. The x-axis shows the different OTUs present in each core microbiome. Across hosts, density dependence explains the largest portion of the variation followed by stochasticity and interspecific interactions (for total values, see Figure 5.9)


Otu001790
Otu001048
Otu031796
Otu000699
Otu042838
Otu000510
Otu088168
Otu236149
Otu107308
Otu187670
Strength
Probability
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
1.0
0.8
0.6
0.4
0.2
0.0

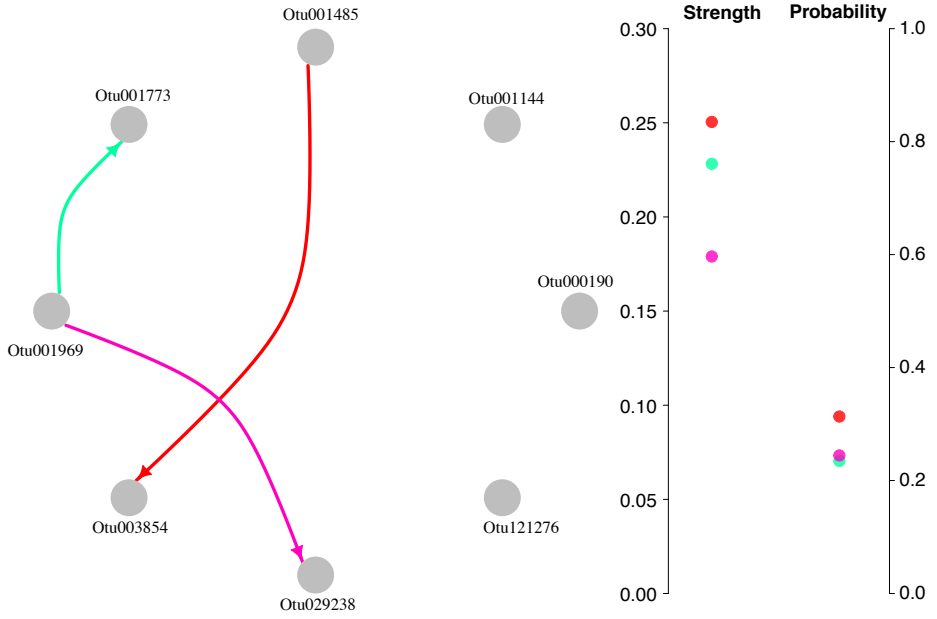
[a]


Otu002329
Otu001576
Otu000577
Otu013442
Otu000555
Otu031796
Otu000543
Otu042838
Otu290392
Otu088168
Otu187670
Otu107308
Otu112902
Strength
Probability
0.2
0.1
0.0
−0.1
−0.2
1.0
0.8
0.6
0.4
0.2
0.0

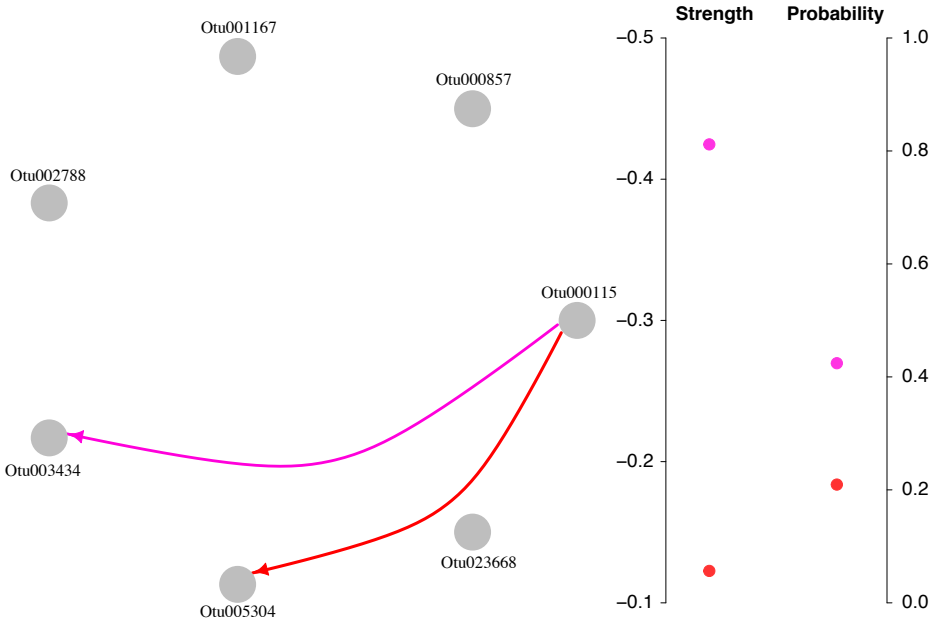
[b]


Otu001801
Otu001995
Otu001400
Otu002329
Otu001048
Otu003837
Otu000669
Otu009959
Otu000555
Otu000271
Otu031796
Otu042838
Otu290392
Otu091937
Otu288080
Otu107308
Otu287579
Otu112902
Otu238548
Otu235338
Strength
Probability
0.4
0.2
0.0
−0.2
−0.4
−0.6
1.0
0.8
0.6
0.4
0.2
0.0

[c]


Otu001485
Otu001773
Otu001144
Otu000190
Otu001969
Otu003854
Otu121276
Otu029238
Strength
Probability
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1.0
0.8
0.6
0.4
0.2
0.0

[d]



[e]

Figure D.3: Representative networks for each core microbiome. The links in these networks corresponds to the most credible ones. a. represent *S. fasciculata*, b. *I. variabilis*, c. *I. oros*, d. *C. delitrix*, e. *C. foliascens*. Each link's colour is mapped to a certain strength and probability displayed on the scales to the right. Note that the scale for the interaction strength ranges between negative and positive integers, while the scale for probability ranges from 0 to 1.

Chapter V

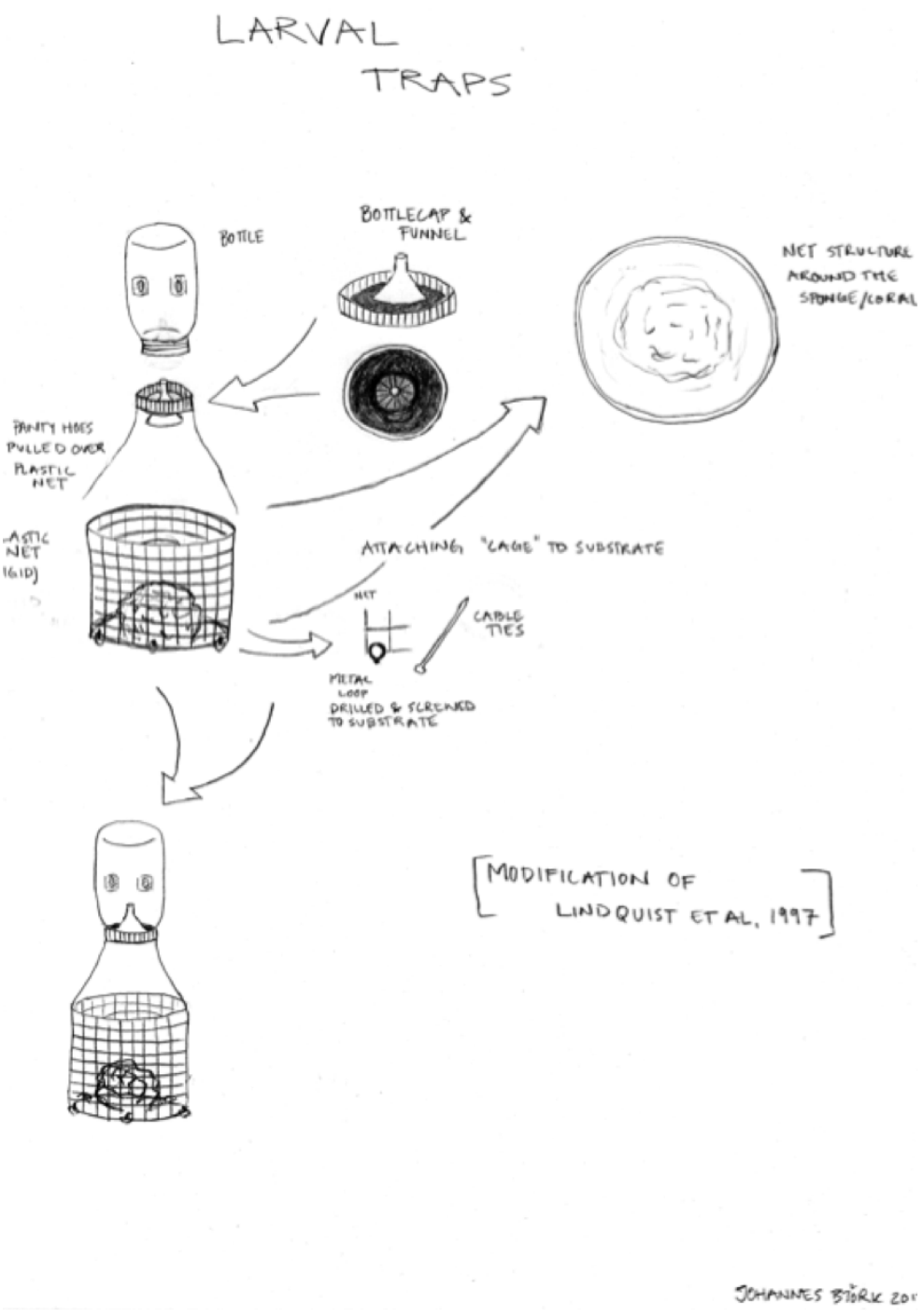


Figure E.1: Sketch of the constructed larvae trap that was used in the sampling protocol to capture dispersing larvae from adult hosts.

Table E.1: Sampled host species and their HMA-LMA classification. If classification does not exist at the species level, the classification is extrapolated from genus level.

Host species	Host classification	Reference
Acanthella acuta	LMA	Gloeckner et al. 2014
Agelas oroides	HMA	Wehrl 2006
Aplysina aerophoba	HMA	Erwin et al. 2015
Aplysina cavernicola	HMA	Erwin et al. 2015
Axinella damicornis	LMA	Erwin et al. 2015
Axinella polypoides	LMA	Wehrl 2006
Axinella verrucosa	LMA	Wehrl 2006
Cacospongia mollior	HMA	Hochmuth et al. 2010
Cacospongia scalaris	HMA	Hochmuth et al. 2010
Chondrosia reniformis	HMA	Gloeckner et al. 2014
Clathrina clathrus	LMA	Quévrain et al. 2014
Clathrina coriacea	LMA	Quévrain et al. 2014
Cliona celata	LMA	Blanquer et al. 2013
Cliona viridis	LMA	Blanquer et al. 2013
Corticium candelabrum	HMA	Caralt et al. 2007
Crambe crambe	LMA	Gloeckner et al. 2014
Crella pulvinar	LMA	Zhao-Ming Gao et al. 2015
Dysidea avara	LMA	Gloeckner et al. 2014
Dysidea fragilis	LMA	Gloeckner et al. 2014
Haliclona fulva	LMA	Sipkema et al. 2011
Haliclona mediterranea	LMA	Sipkema et al. 2011
Haliclona mucosa	LMA	Sipkema et al. 2011
Hemimycale columella	HMA	Blanquer et al. 2013
Ircinia fasciculate	HMA	Gloeckner et al. 2014
Ircinia oros	HMA	Gloeckner et al. 2014
Ircinia varabilis	HMA	Gloeckner et al. 2014
Oscarella lobularis	LMA	Gloeckner et al. 2014
Petrosia ficiformis	HMA	Gerçe et al. 2011
Phorbas fictitius	HMA	Dupont et al. 2013
Phorbas tenacior	HMA	Dupont et al. 2013
Raspaciona aculeata	LMA	Riesgo et al. 2009
Spirastrella cunctatrix	LMA	Erwin et al. 2015
Spongia agaricina	HMA	Noyer et al. 2014

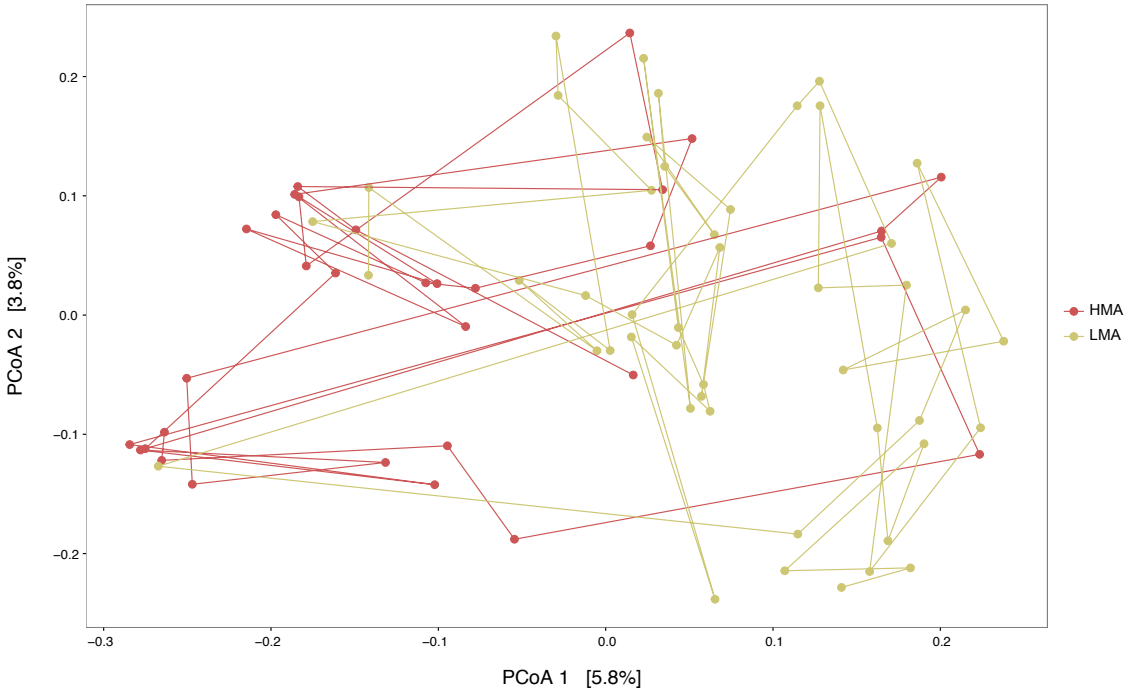


Figure E.2: Principal coordinates analysis of unweighted UniFrac distances across all adult host-associated microbial communities, including their larvae. Dots and connecting lines coloured in red represent HMA hosts, and khaki LMA hosts. Lines are included as a way of depicting each treatments dispersion.

Table E.2: Pairwise comparisons of UniFrac distances among modules within the starved partial symbiotic network. Each module is represented by its focal host

	A. aerophoba	*I. oros*	*C.viridis*	*C. crambe*	*D. avara*	*O. lobularis*	*H. columella*
A. aerophoba							
I. oros	F=0.977, P=0.468						
C.viridis	F=1.579, P=0.01	F=1.315, P=0.082					
C. crambe	F=0.897, P=0.62	F=0.892, P=0.621	F=1.017, P=0.416				
D. avara	F=1.173, P=0.172	F=1.110, P=0.284	F=1.120, P=0.251	F=0.934, P=0.573			
O. lobularis	F=1.587, P=0.014	F=1.069, P=0.36	F=1.094, P=0.281	F=0.978, P=0.497	F=1.025, P=0.404		
H. columella	F=1.495, P=0.049	F=1.360, P= 0.066	F=1.472, P= 0.021	F=1.073, P=0.326	F=1.321, P=0.087	F=1.271, P=0.114	

Notes: The analysis was done using AMOVA which is a nonparametric permutation method similar to ANOVA. Overall Test: F=3.670, P<0.001. There are no significant differences after correcting for multiple testing: Bonferoni correction (P<0.00238).

Table E.3: Pairwise comparisons of UniFrac distances among modules within the instant partial symbiotic network. Each module is represented by its focal host.

	A. aerophoba	*I. oros*	*C.viridis*	*C. crambe*	*D. avara*	*O. lobularis*	*H. columella*
A. aerophoba							
I. oros	F=0.736, P=0.828						
C.viridis	F=1.579, P<0.001	F=4.518, P<0.001					
C. crambe	F=3.758, P<0.001	F=3.427, P<0.001	F=7.237, P<0.001				
D. avara	F=2.519, P=0.002	F=2.285, P<0.001	F=4.060, P<0.001	F=3.458, P<0.001			
O. lobularis	F=2.702, P<0.001	F=2.434, P<0.001	F=5.615, P<0.001	F=3.776, P<0.001	F=1.860, P<0.001		
H. columella	F=5.728, P<0.001	F=4.005, P<0.001	F=5.755, P<0.001	F=4.608, P<0.001	F=3.863, P<0.001	F=3.555, P<0.001	

Notes: The analysis was done using AMOVA which is a nonparametric permutation method similar to ANOVA. Overall Test: F=3.700, P<0.001. All modules are significant different after multiple testing, except for the modules containing HMA host *A. aerophoba* and *I. oros*. Bonferoni correction (P<0.00238).

Table E.4: Network architecture for full overall and host-specific networks under instant and starved treatments.Number of hosts and OTUs in each network. Vulnerability and generality, interaction diversity and evenness, Specialisation (H_2), modularity and nestedness (weighted NODF). Note that the modularity is standardised to make it comparable across networks.

	Instant		Starved	
	Overall	Host-specific	Overall	Host-specific
N_{Host}	36	36	36	36
N_{OTU}	13697	13017	15440	15134
Vulnerability	24.24	23.94	30.53	30.30
Generality	1.19	1.20	1.16	1.16
Diversity	6.45	6.39	6.63	6.59
Evenness	0.49	0.49	0.50	0.50
Connectance	0.04	0.04	0.04	0.04
Specialisation	0.96	0.96	0.97	0.97
Modularity	0.95	0.95	0.96	0.96
Nestedness	1.37	1.28	1.42	1.40

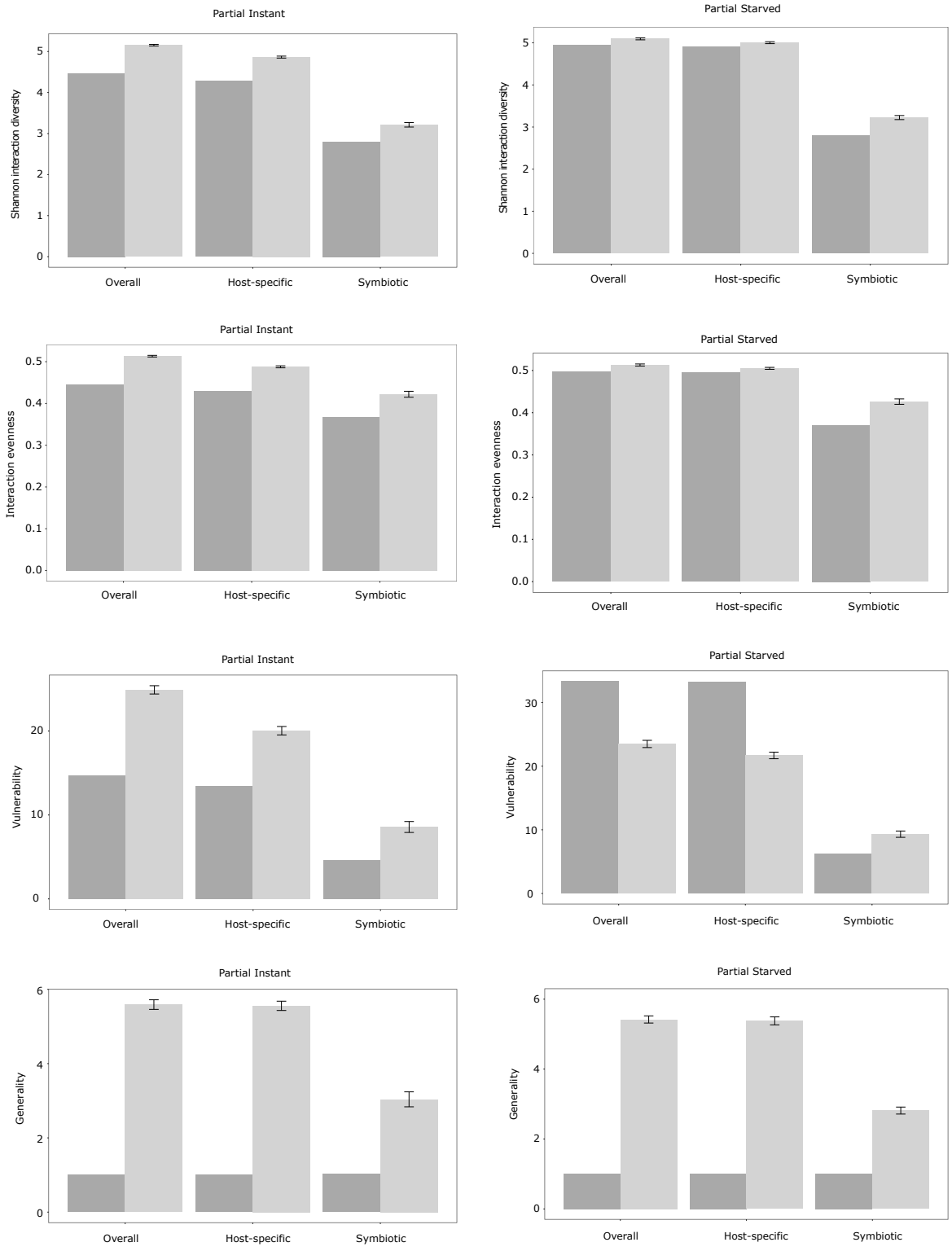


Figure E.3: Architectural properties of the sponge–microbe networks. Specifically for partial networks: overall, host-specific and symbiotic networks under both instant and starved treatment. Interaction diversity, interactions evenness, vulnerability and generality. Dark gray bars show the observed network architecture, while light gray bars display what is the expected architecture under the most conservative null model. Null expectations are plotted with mean±s.d. based on 100 randomisation. All facets of the observed network architecture deviate significantly from what is expected under the most conservative null model.

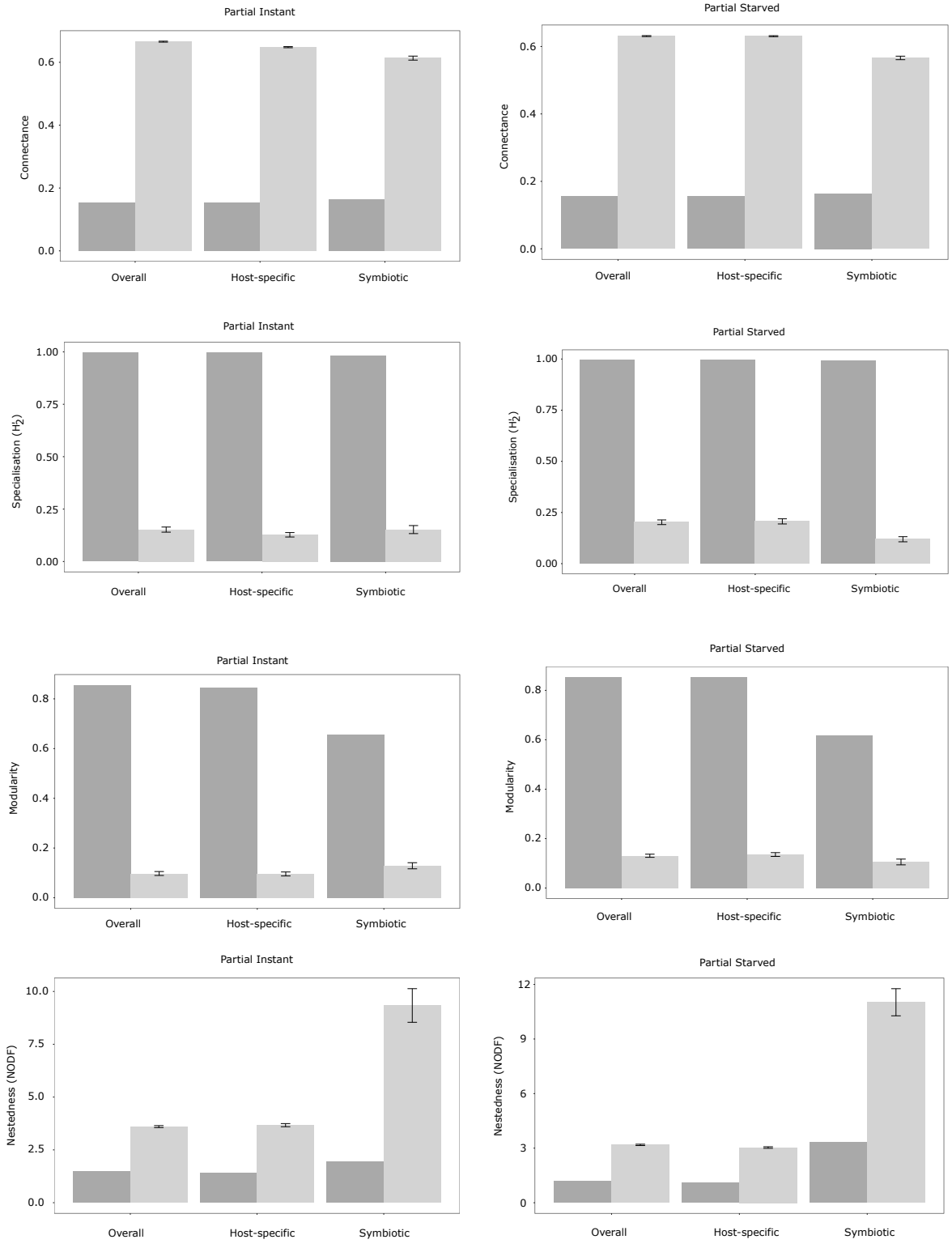


Figure E.4: Architectural properties of the sponge–microbe networks. Specifically for partial networks: overall, host-specific and symbiotic networks under both instant and starved treatments. Connectance, specialisation (H_2'), modularity and nestedness (weighted NODF). Dark gray bars show the observed network architecture, while light gray bars display what is the expected architecture under the most conservative null model. Null expectations are plotted with mean±s.d. based on 100 randomisation. All facets of the observed network architecture deviate significantly from what is expected under the most conservative null model.

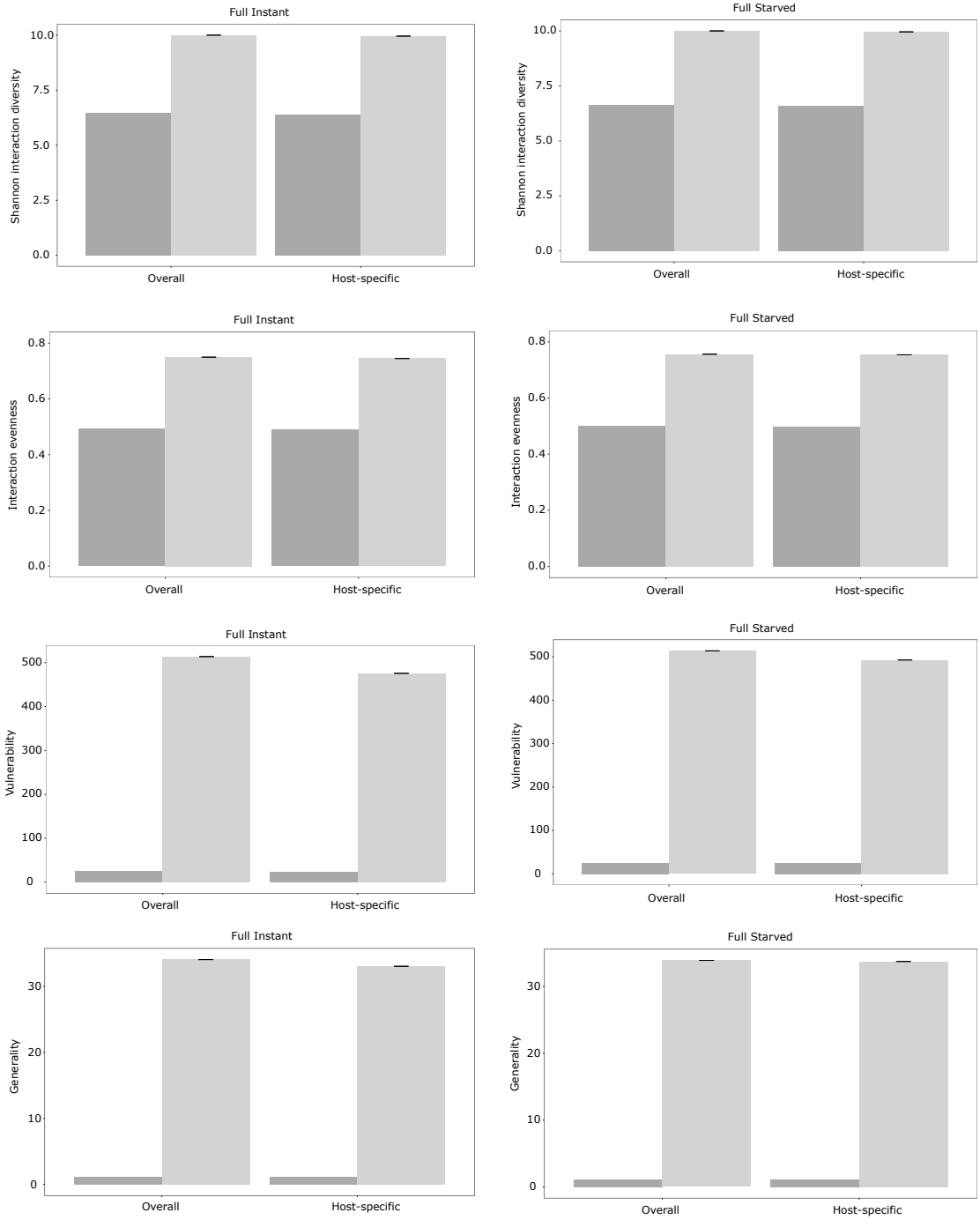


Figure E.5: Architectural properties of the sponge–microbe networks. Specifically for full networks: overall and host-specific networks under both instant and starved treatment. Interaction diversity, interactions evenness, vulnerability and generality. Dark gray bars show the observed network architecture, while light gray bars display what is the expected architecture under the least conservative null model. Null expectations are plotted with mean$\pm$s.d. based on 100 randomisation. All facets of the observed network architecture deviate significantly from what is expected under the most conservative null model.

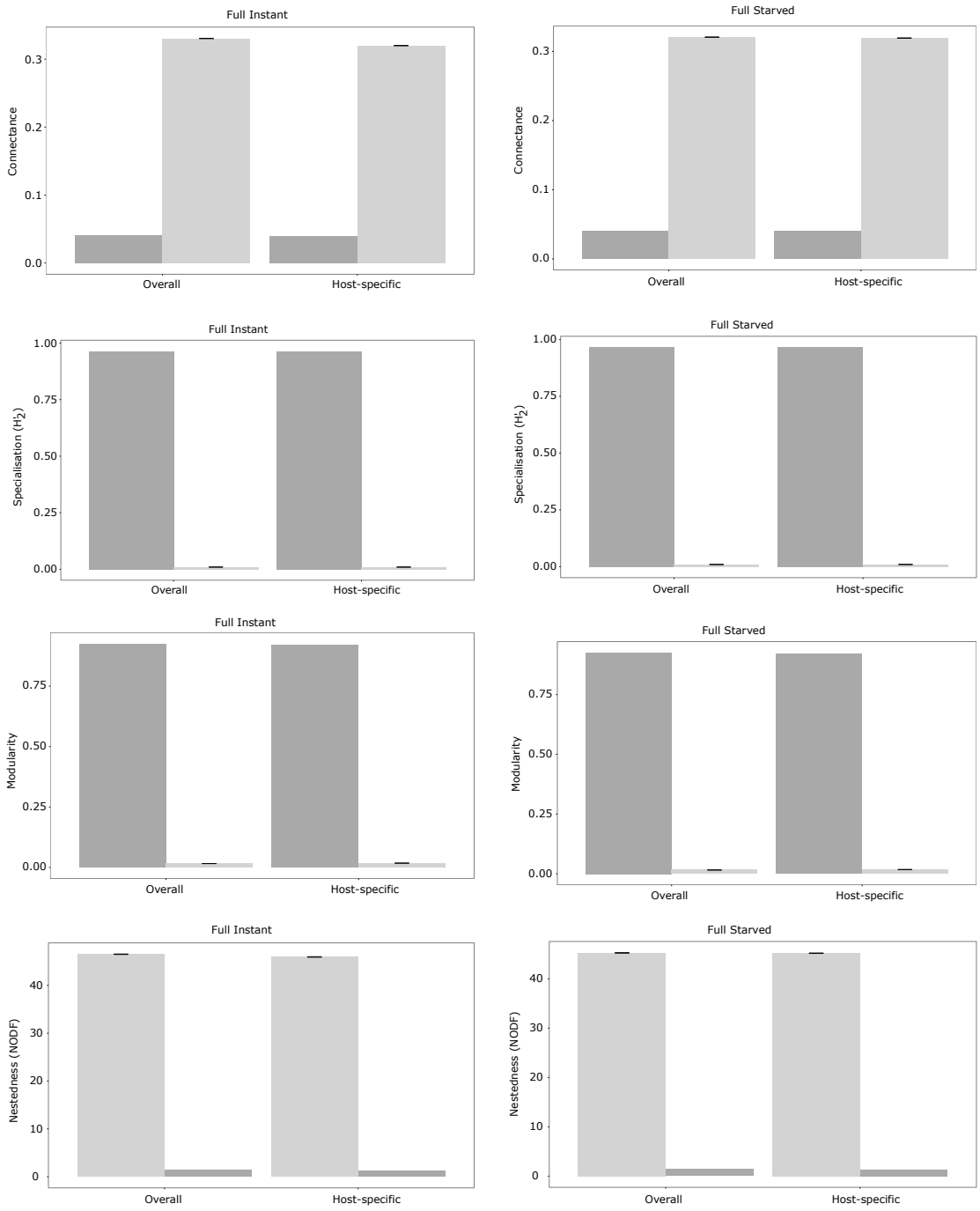


Figure E.6: Architectural properties of the sponge–microbe networks. Specifically for full networks: overall and host-specific networks under both instant and starved treatment. Connectance, specialisation (H_2'), modularity and nestedness (weighted NODF). Dark gray bars show the observed network architecture, while light gray bars display what is the expected architecture under the least conservative null model. Null expectations are plotted with mean±s.d. based on 100 randomisation. All facets of the observed network architecture deviate significantly from what is expected under the most conservative null model.

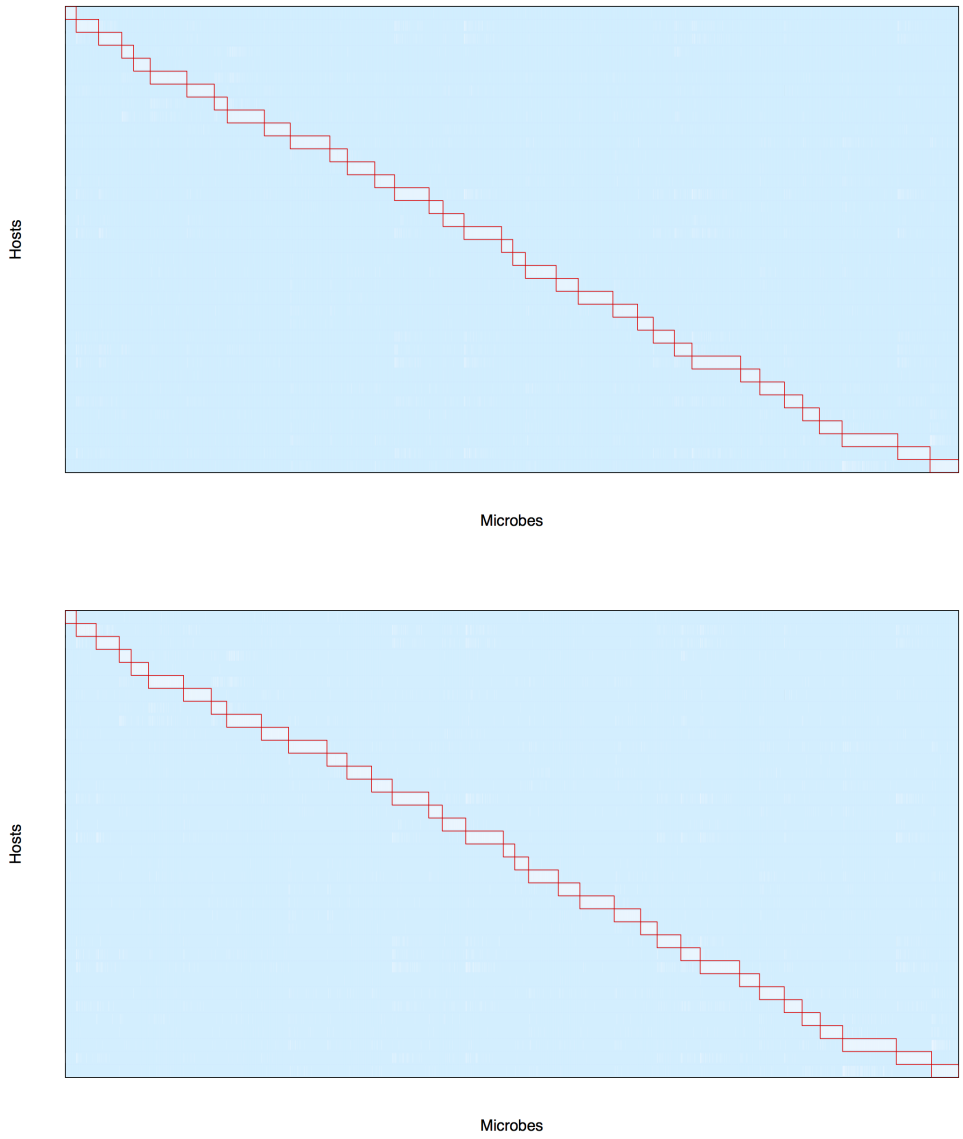


Figure E.7: The networks are organised in modules comprising nodes interacting more among themselves than with the rest of the network. Modules were detected using the LPAwb+ algorithm for weighted networks (Beckett 2016). Rows represent hosts and columns microbes. The figure corresponds to full networks: overall (top) and host-specific (bottom) network under instant treatment.

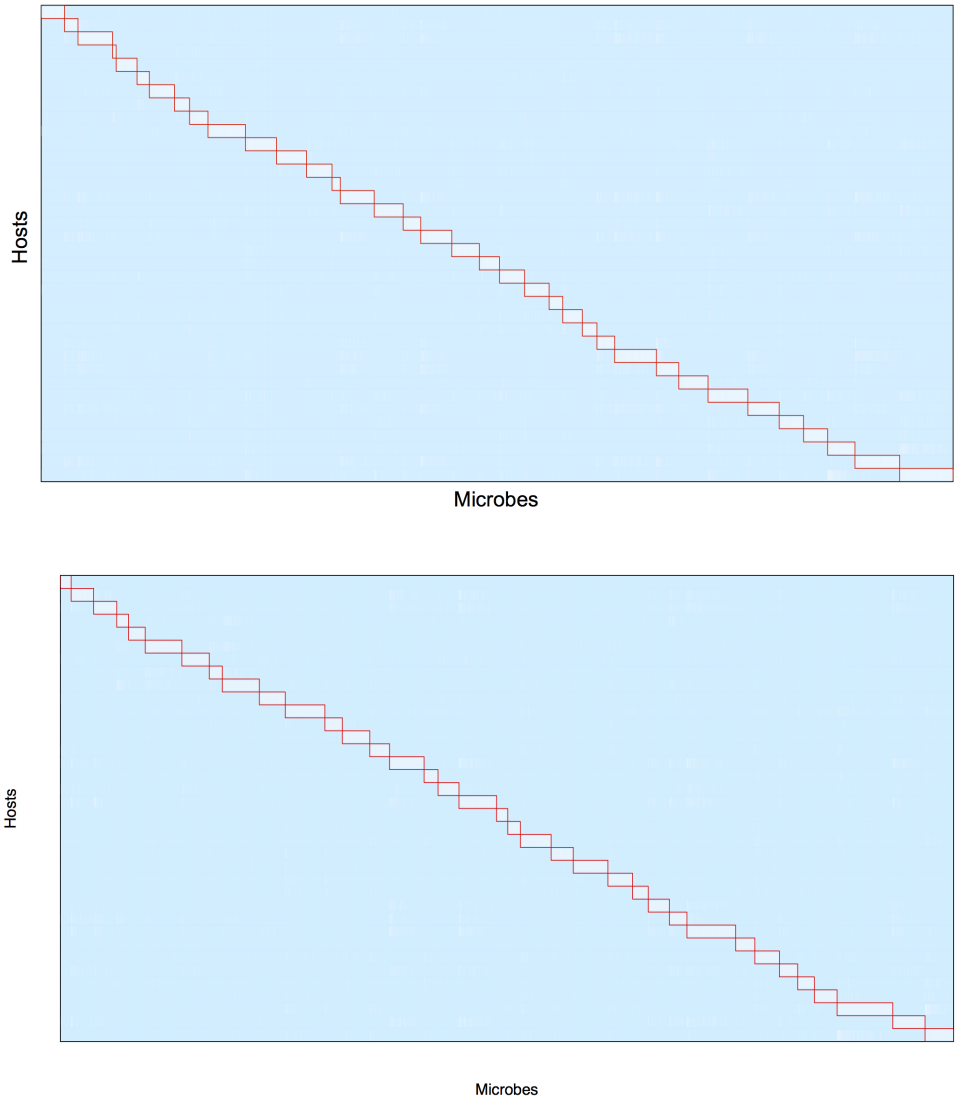


Figure E.8: The networks are organised in modules comprising nodes interacting more among themselves than with the rest of the network. Modules were detected using the LPAwb+ algorithm for weighted networks (Beckett 2016). Rows represent hosts and columns microbes. The figure corresponds to full networks: overall (top) and host-specific (bottom) network under starved treatment.

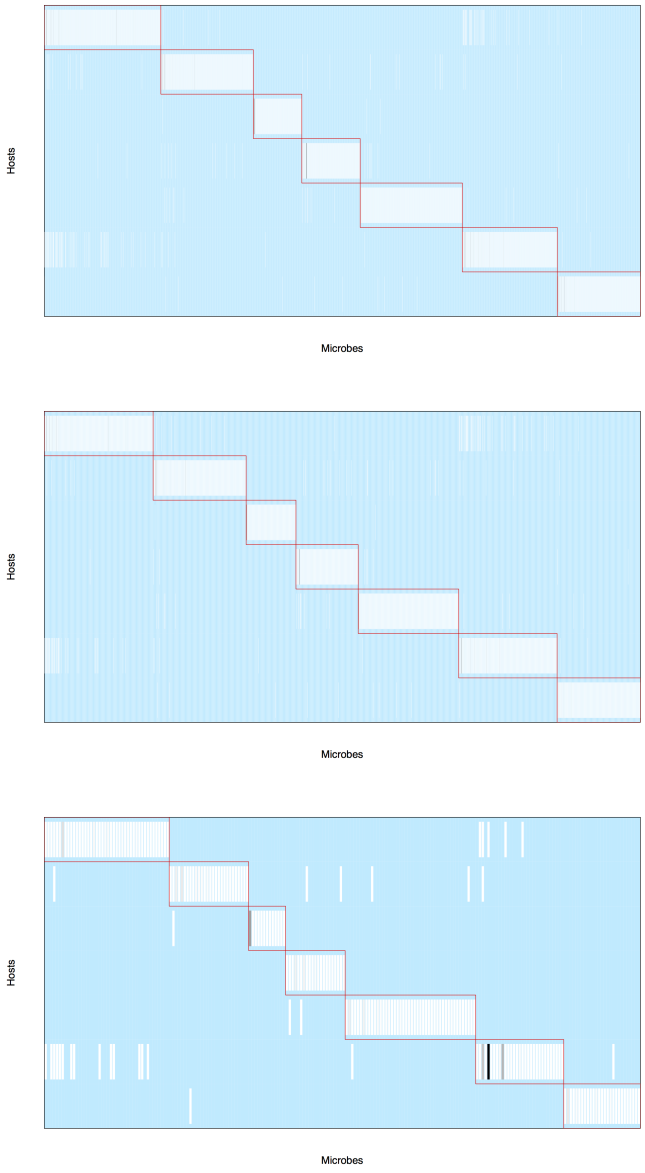


Figure E.9: The networks are organised in modules comprising nodes interacting more among themselves than with the rest of the network. Modules were detected using the LPAwb+ algorithm for weighted networks (Beckett 2016). Rows represent hosts and columns microbes. The figure corresponds to partial networks: overall (top), host-specific (middle) and symbiotic (bottom) network under instant treatment.

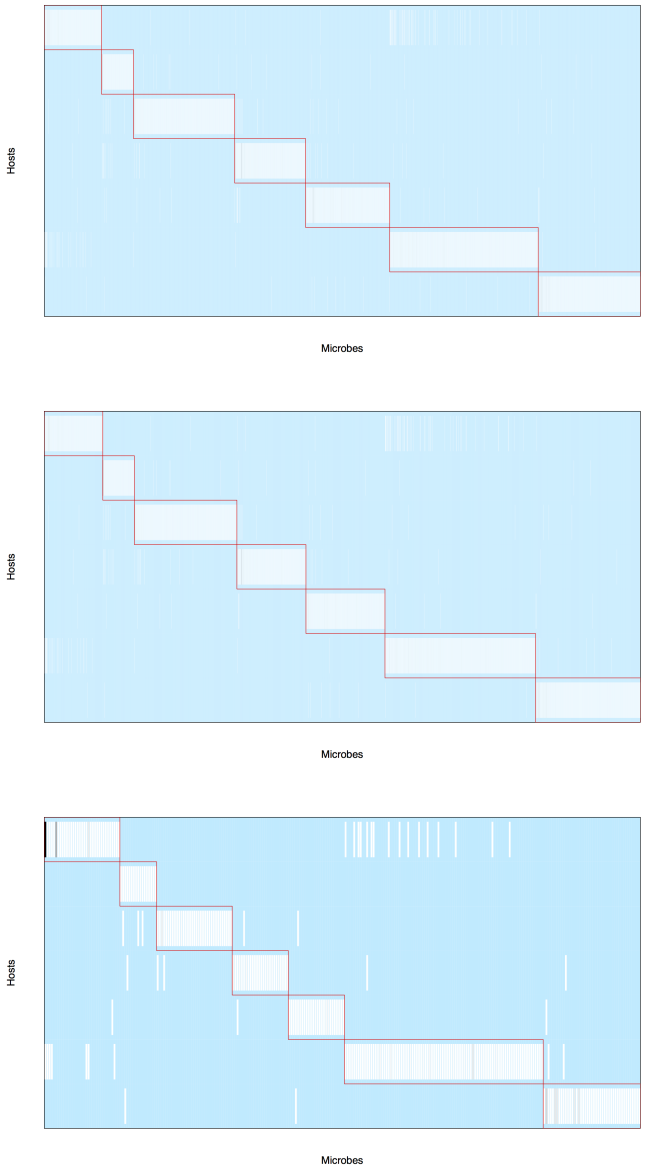


Figure E.10: The networks are organised in modules comprising nodes interacting more among themselves than with the rest of the network. Modules were detected using the LPAwb+ algorithm for weighted networks (Beckett 2016). Rows represent hosts and columns microbes. The figure corresponds to partial networks: overall (top), host-specific (middle) and symbiotic (bottom) networks under starved treatment.

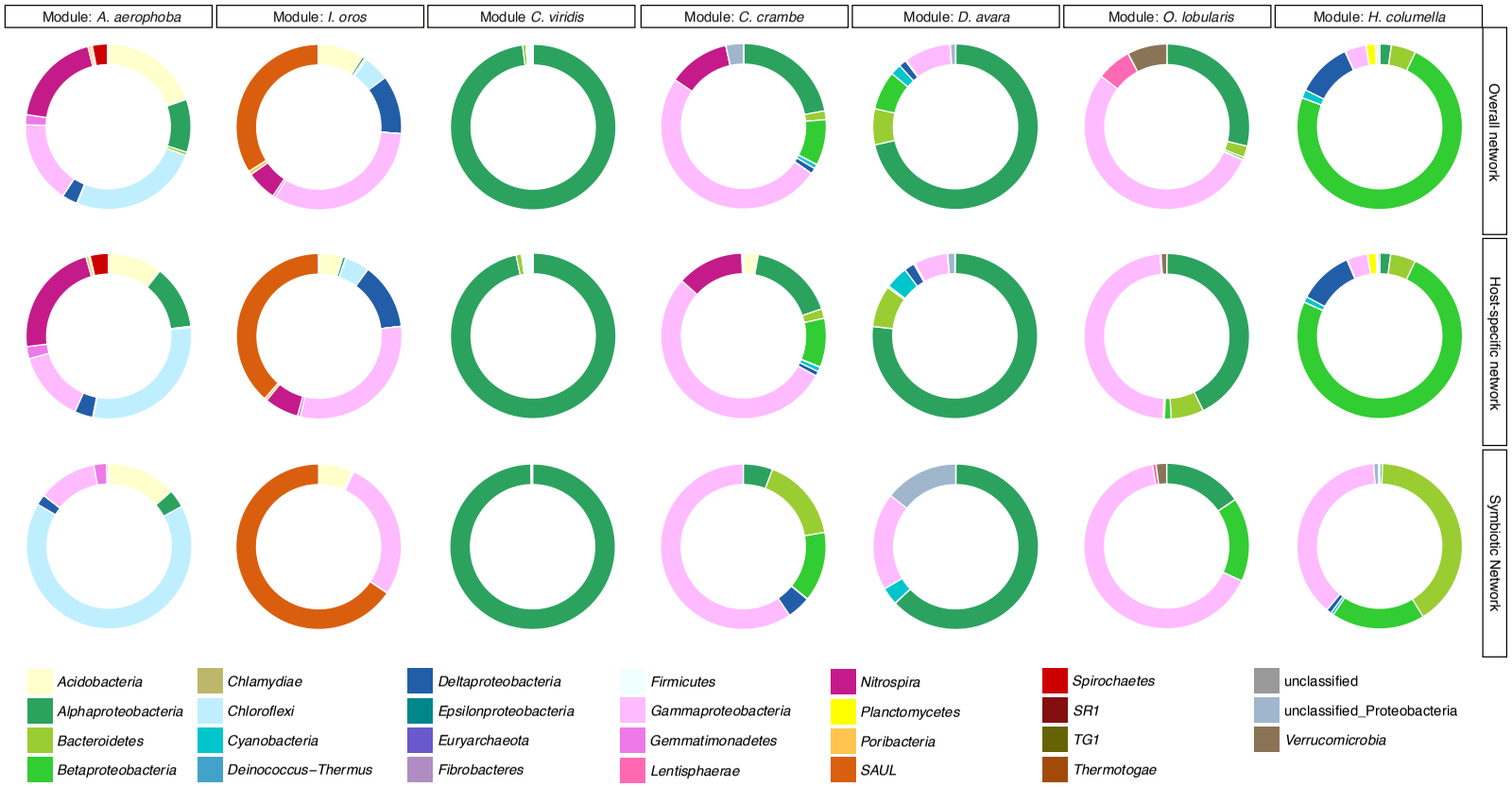


Figure E.11: Taxonomic profile for each module within the different partial networks under instant treatment. Colours represent different phyla and their width correspond to their relative abundance within each module. Rows represent network type; overall, host-specific, and symbiotic networks respectively, and each column corresponds to a given module within these networks.

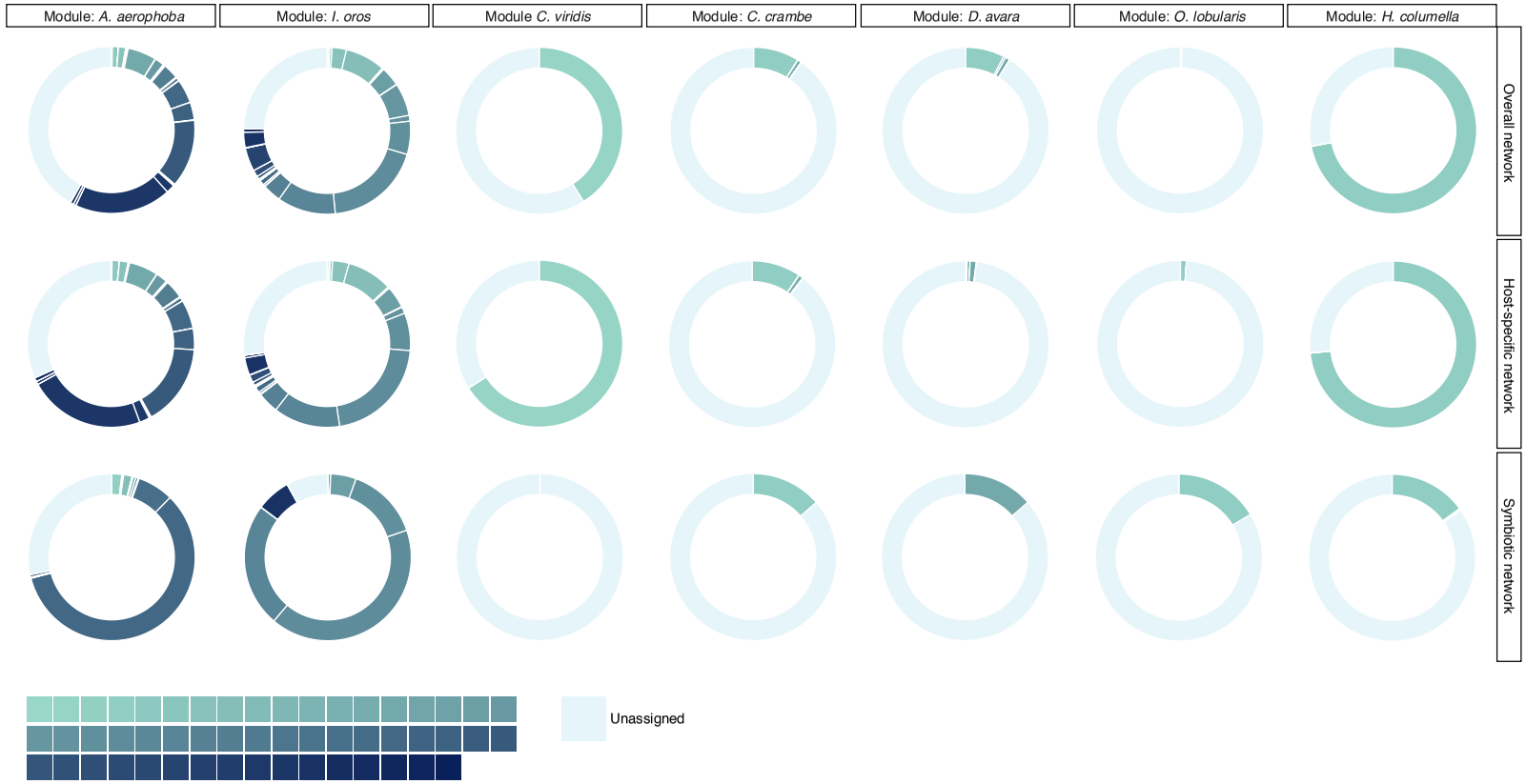


Figure E.12: 'Sponge-specific' profile for each module within the different partial networks under instant treatment. Colour shades represent different 'sponge-specific' clusters and their width correspond to their relative abundance within each module. The lightest shade represent the relative abundance of microbes not assigning to any 'sponge-specific' cluster. Rows represent network type; overall, host-specific, and symbiotic networks respectively, and each column corresponds to a given module within these networks.

Bibliography

1. Agrawal, A. F. & Lively, C. M. Modelling infection as a two-step process combining gene-for-gene and matching-allele genetics. *Proceedings. Biological sciences / The Royal Society / The Royal Society* **270,** 323–334 (2003).

2. Allesina, S. & Tang, S. Stability criteria for complex ecosystems. *Nature* **483,** 205–208 (2012).

3. Almaraz, P., Green, A. J., Aguilera, E., Rend??n, M. A. & Bustamante, J. Estimating partial observability and nonlinear climate effects on stochastic community dynamics of migratory waterfowl. *Journal of Animal Ecology* **81,** 1113–1125 (2012).

4. Almaraz, P. & Oro, D. Size-mediated non-trophic interactions and stochastic predation drive assembly and dynamics in a seabird community. *Ecology* **92,** 1948–1958 (2011).

5. Almeida-Neto, M., Guimaraes, P. R., Guimaraes, J., Loyola, R. D. & Werner, U. A consistent metric for nestedness analysis in ecological systems: reconciling. **117,** 1227–1239 (2008).

6. Anderson, S. A., Northcote, P. T. & Page, M. J. Spatial and temporal variability of the bacterial community in different chemotypes of the New Zealand marine sponge Mycale hentscheli. *FEMS Microbiology Ecology* **72,** 328–342 (2010).

7. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature May* **12,** 174–180 (2011).

8. Astorga, A., Oksanen, J., Luoto, M., Soininen, J., Virtanen, R. & Muotka, T. Distance decay of similarity in freshwater communities: Do macro- and microorganisms follow the same rules? *Global Ecology and Biogeography* **21,** 365–375 (2012).

9. Austin, M. P. Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecological Modelling* **157,** 101–118 (2002).

10. Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-Bacterial Mutualism in the Human Intestine. *Science* **307,** 1915–1920 (2005).

11. Backhed, F., Roswall, J., *et al.* Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host and Microbe* **17,** 690–703 (2015).

12. Ballesteros, E. Mediterranean Coralligenous Assemblages : a Synthesis of Present Knowledge. *Oceanography and Marine Biology: An Annual Review* **44,** 123–195 (2006).

13. Barberan, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *Isme J* **6,** 343–351 (2012).

14. Bascompte, J. Disentangling the Web of Life Biodiversity. *Science* **325,** 24–27 (2009).

15. Bascompte, J. & Jordano, P. *Mutualistic Networks* (Princeton University Press, 2014).

16. Bascompte, J. & Jordano, P. Plant-Animal Mutualistic Networks: The Architecture of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics* **38,** 567–593 (2007).

17. Bascompte, J., Jordano, P., Melián, C. J. & Olesen, J. M. The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 9383–9387 (2003).

18. Bassler, B. L. & Losick, R. Bacterially Speaking. *Cell* **125,** 237–246 (2006).

19. Bawa, K. S. Plant-Pollinator Interactions in Tropical Rain Forests. *Annual Review of Ecology and Systematics* **21,** 399–422 (1990).

20. Beckett, S. J. Improved community detection in weighted bipartite networks Subject Category : Subject Areas : Author for correspondence : (2016).

21. Bell, T. Experimental tests of the bacterial distance-decay relationship. *The ISME journal* **4,** 1357–65 (2010).

22. Bell, T., Ager, D., Song, J.-I., Newman, J. A., Thompson, I. P., Lilley, A. K. & van der Gast, C. J. Larger Islands House More Bacterial Taxa. *Science* **308,** 1884 (2005).

23. Bersier, L. F., Banasek-Richter, C. & Cattin, M.-F. Quantitative descriptors of food-web matrices. *Ecology* **83,** 2394–2407 (2002).

24. Bittleston, L. S., Pierce, N. E., Ellison, A. M. & Pringle, A. Convergence in Multispecies Interactions. *Trends in Ecology and Evolution* **31,** 269–280 (2016).

25. Björk, J. R., Díez-Vives, C., Coma, R., Ribes, M. & Montoya, J. M. Specificity and temporal dynamics of complex bacteria-sponge symbiotic interactions. *Ecology* **94,** 2781–2791 (2013).

26. Blüthgen, N., Fründ, J., Vazquez, D. P. & Menzel, F. What do interaction network metrics tell us about specialization and biological traits? *Ecology* **89,** 3387–3399 (2008).

27. Blüthgen, N., Menzel, F. & Blüthgen, N. Measuring specialization in species interaction networks. *BMC ecology* **6,** 9 (2006).

28. Blüthgen, N., Menzel, F., Hovestadt, T., Fiala, B. & Blüthgen, N. Specialization, Constraints, and Conflicting Interests in Mutualistic Networks. *Current Biology* **17,** 341–346 (2007).

29. Bogaert, D. *et al.* Variability and diversity of nasopharyngeal microbiota in children: A metagenomic analysis. *PLoS ONE* **6** (2011).

30. Boury-Esnault, N. & Rützler, K. Thesaurus of sponge morphology. *Smithsonian Contributions to Zoology,* 1–55 (1997).

31. Brouat, C. & Duplantier, J. M. Host habitat patchiness and the distance decay of similarity among gastrointestinal nematode communities in two species of Mastomys (southeastern Senegal). *Oecologia* **152,** 715–720 (2007).

32. Burgsdorf, I., Erwin, P. M., L??pez-Legentil, S., Cerrano, C., Haber, M., Frenk, S. & Steindler, L. Biogeography rather than association with cyanobacteria structures symbiotic microbial communities in the marine sponge Petrosia ficiformis. *Frontiers in Microbiology* **5,** 1–11 (2014).

33. Burmolle, M., Ren, D., Bjarnsholt, T. & Sorensen, S. J. Interactions in multispecies biofilms: Do they actually matter? *Trends in Microbiology* **22,** 84–91 (2014).

34. Cabreiro, F. & Gems, D. Worms need microbes too: Microbiota, health and aging in caenorhabditis elegans. *EMBO Molecular Medicine* **5,** 1300–1310 (2013).

35. Camerano, L. Desequilibrio dei viventi merc la reciproca distruzione–On the equilibrium of living beings by means of reciprocal destruction. *Atti della Reale Accademia delle Scienze di Torino* **15,** 393–414 (1880).

36. Caporaso, J. G., Kuczynski, J., *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7,** 335–336 (2010).

37. Caporaso, J. G., Lauber, C. L., Costello, E. K., *et al.* Moving pictures of the human microbiome. *Genome Biology* **12,** R50 (2011).

38. Caporaso, J. G., Lauber, C. L., Walters, W. a., *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal* **6,** 1621–1624 (2012).

39. Cardinale, B. J., Palmer, M. a. & Collins, S. L. Species diversity enhances ecosystem functioning through interspecific facilitation. *Nature* **415,** 426–9 (2002).

40. Chaffron, S., Rehrauer, H., Pernthaler, J. & Mering, C. A global network of coexisting microbes from\nenvironmental and whole-genome sequence data. *Genome Research* **2010,** 947–959 (2010).

41. Chase, J. M. Stochastic community assembly causes higher biodiversity in more productive environments. *Science (New York, N.Y.)* **328,** 1388–91 (2010).

42. Cho, I. & Blaser, M. The human microbiome: at the interface of health and disease. *Nature Reviews Genetics* **13,** 260–270 (2012).

43. Clark, J. S., Gelfand, A. E., Woodall, C. W. & Zhu, K. More than the sum of the parts: Forest Climate response from join specise distributions. *Ecological Applications* **24,** 990–999 (2014).

44. Cole, J. R. *et al.* Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Research* **42,** 633–642 (2014).

45. Coppari, M., Gori, A., Viladrich, N., Saponari, L., Canepa, A., Grinyó, J., Olariaga, A. & Rossi, S. The role of Mediterranean sponges in benthic–pelagic coupling processes: Aplysina aerophoba and Axinella polypoides case studies. *Journal of Experimental Marine Biology and Ecology* **477,** 57–68 (2016).

46. Costello, E. E. K., Lauber, C. C. L., Hamady, M., Fierer, N., Gordon, J. I. & Knight, R. Bacterial community variation in human body habitats across space and time. *Science* **326,** 8–11 (2009).

47. Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M. & Relman, D. a. The application of ecological theory toward an understanding of the human microbiome. *Science (New York, N.Y.)* **336,** 1255–62 (2012).

48. Cottingham, K. L., Brown, B. L. & Lennon, J. T. Biodiversity may regulate temporal variability of ecological systems. *Ecology Letters* **4,** 72–85 (2001).

49. Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: Networks, competition, and stability. *Science* **350,** 663–666 (2015).

50. Cram, J. A., Chow, C.-E. T., Sachdeva, R., Needham, D. M., Parada, A. E., Steele, J. A. & Fuhrman, J. A. Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *Isme J* **9,** 563–580 (2015).

51. Cressie, N., Calder, C. A., Clark, J. S., Hoef, J. M. V., Wikle, K., Clark, S., Jay, M., Wikle, K. & Carolina, N. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *HIERARCHICAL MODELS IN ECOLOGY* **19,** 553–570 (2009).

52. Crone, E. Contrasting effects of spatial heterogeneity and environmental stochasticity on population dynamics of a perennial wildflower. *Journal of Ecology* **XXX,** XXX (2016).

53. Darwin, R. C. in (ed M.A., F.R.S.) (1862).

54. De Wit, R. & Bouvier, T. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environmental Microbiology* **8,** 755–758 (2006).

55. Degnan, S. M. The surprisingly complex immune gene repertoire of a simple sponge, exemplified by the NLR genes: A capacity for specificity? *Developmental and Comparative Immunology* **48,** 269–274 (2015).

56. Delsuc, F., Metcalf, J. L., Wegener Parfrey, L., Song, S. J., Gonzlez, A. & Knight, R. Convergence of gut microbiomes in myrmecophagous mammals. *Molecular Ecology* **23,** 1301–1317 (2014).

57. Denwood, M. J. runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software* (in press).

58. Dethlefsen, L., McFall-Ngai, M. & Relman, D. A. An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* **449,** 811–818 (2007).

59. Donohue, I. *et al.* On the dimensionality of ecological stability. *Ecology Letters* **16,** 421–429 (2013).

60. Dormann, C. F., Fründ, J., Bluthgen, N. & Gruber, B. Indices, graphs and null models: analysing bipartite ecological networks. *The Open Ecology Journal* **2,** 7–24 (2009).

61. Dormann, C. F. & Strauss, R. A method for detecting modules in quantitative bipartite networks. *Methods in Ecology and Evolution* **5,** 90–98 (2014).

62. Dubilier, N., Bergin, C. & Lott, C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature reviews. Microbiology* **6,** 725–40 (2008).

63. Dunne, J. The Network Structure of Food Webs. *Ecological Networks. Linking structure to dynamics in food webs,* 386 (2006).

64. Dunne, J. A., Williams, R. J. & Martinez, N. D. Food-web structure and network theory: The role of connectance and size. *Pnas* **99,** 12917–12922 (2002).

65. Dupont, Y. L. & Olesen, J. M. Ecological modules and roles of species in heathland plant-insect flower visitor networks. *Journal of Animal Ecology* **78,** 346–353 (2009).

66. Easson, C. G. & Thacker, R. W. Phylogenetic signal in the community structure of host-specific microbiomes of tropical marine sponges. *Front Microbiol* **5,** 532 (2014).

67. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27,** 2194–2200 (2011).

68. Elli, M., Colombo, O. & Tagliabue, A. *Medical Hypotheses* **75,** 350–352 (2010).

69. Emmerson, M. C. & Raffaelli, D. Predator-prey body size, interaction strength and the stability of a real food web. *Journal of Animal Ecology* **73,** 399–409 (2004).

70. Emmerson, M. & Yearsley, J. M. Weak interactions, omnivory and emergent food-web properties. *Proceedings of the Royal Society B-Biological Sciences* **271,** 397–405 (2004).

71. Ereskovsky, A. & Tokina, D. Morphology and fine structure of the swimming larvae of Ircinia oros (Porifera, Demospongiae, Dictyoceratida). *Invertebrate Reproduction & Development* **45,** 137–150 (2004).

72. Erwin, P. M., Coma, R., López-Sendino, P., Serrano, E. & Ribes, M. Stable symbionts across the HMA-LMA dichotomy: low seasonal and interannual variation in sponge-associated bacteria from taxonomically diverse hosts. *FEMS Microbiology Ecology* **91,** fiv115 (2015).

73. Erwin, P. M., L??pez-Legentil, S., Gonz??lez-Pech, R. & Turon, X. A specific mix of generalists: Bacterial symbionts in Mediterranean Ircinia spp. *FEMS Microbiology Ecology* **79,** 619–637 (2012).

74. Erwin, P. M., Olson, J. B. & Thacker, R. W. Phylogenetic diversity, host-specificity and community profiling of sponge-associated bacteria in the northern Gulf of Mexico. *PLoS ONE* **6** (2011).

75. Estrada, E. Food webs robustness to biodiversity loss: The roles of connectance, expansibility and degree distribution. *Journal of Theoretical Biology* **244,** 296–307 (2007).

76. Estrela, S. & Brown, S. P. Metabolic and Demographic Feedbacks Shape the Emergent Spatial Structure and Function of Microbial Communities. *PLoS Computational Biology* **9** (2013).

77. Estrela, S., Whiteley, M. & Brown, S. P. The demographic determinants of human microbiome health. *Trends in Microbiology* **23,** 134–141 (2014).

78. Fagan William F. & Hurd L E. Hatch Density Variation of a Generalist Arthropod Predator : Population Consequenes and Community Impact. *Ecology* **75,** 2022–2032 (1994).

79. Faith, J. J. *et al.* The long-term stability of the human gut microbiota. *Science* **341,** 1237439 (2013).

80. Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N. S. & Thomas, T. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proceedings of the National Academy of Sciences* **109,** E1878–E1887 (2012).

81. Faraway, J. J. *Linear Models with R, Second Edition* (Taylor & Francis, 2014).

82. Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nature Reviews Microbiology* **10,** 538–550 (2012).

83. Fieseler, L., Horn, M., Wagner, M. & Hentschel, U. Discovery of the Novel Candidate Phylum " Poribacteria " in Marine Sponges Discovery of the Novel Candidate Phylum " Poribacteria " in Marine Sponges. *Applied and environmental microbiology* **70,** 3724–3732 (2004).

84. Finlay, B. J. Global dispersal of free-living microbial eukaryote species. *Science* **296,** 1061–1063 (2002).

85. Fisher, C. K. & Mehta, P. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS ONE* **9,** 1–10 (2014).

86. Flores, C. O., Meyer, J. R., Valverde, S., Farr, L. & Weitz, J. S. Statistical structure of host-phage interactions. *Proc Natl Acad Sci U S A* **108,** E288–97 (2011).

87. Flores, G. E., Caporaso, J. G., *et al.* Temporal variability is a personalized feature of the human microbiome. *Genome biology* **15,** 531 (2014).

88. Fortuna, M. A., Stouffer, D. B., Olesen, J. M., Jordano, P., Mouillot, D., Krasnov, B. R., Poulin, R. & Bascompte, J. Nestedness versus modularity in ecological networks: Two sides of the same coin? *Journal of Animal Ecology* **79,** 811–817 (2010).

89. Foster, J. A. & McVey Neufeld, K. A. Gut-brain axis: How the microbiome influences anxiety and depression. *Trends in Neurosciences* **36,** 305–312 (2013).

90. Franzosa, E. A., Huang, K., Meadow, J. F., Gevers, D., Lemon, K. P., Bohannan, B. J. M. & Huttenhower, C. Identifying personal microbiomes using metagenomic codes. *Proceedings of the National Academy of Sciences* **112,** E2930–E2938 (2015).

91. Fraune, S. & Bosch, T. C. G. Why bacteria matter in animal development and evolution. *BioEssays* **32,** 571–580 (2010).

92. Freeman, B. G. & Class Freeman, A. M. Rapid upslope shifts in New Guinean birds illustrate strong distributional responses of tropical montane species to global warming. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 4490–4 (2014).

93. Freeman, C. J., Easson, C. G. & Baker, D. M. Metabolic diversity and niche structure in sponges from the Miskito Cays, Honduras. *PeerJ* **2,** e695 (2014).

94. Freeman, C. J. & Thacker, R. W. Complex interactions between marine sponges and their symbiotic microbial communities. *Limnology and Oceanography* **56,** 1577–1586 (2011).

95. Freeman, C. J., Thacker, R. W., Baker, D. M. & Fogel, M. L. Quality or quantity: is nutrient transfer driven more by symbiont identity and productivity than by symbiont abundance? *The ISME journal* **7,** 1116–25 (2013).

96. Fromin, N., Hamelin, J., Tarnawski, S., Roesti, D., Forestier, N., Gillet, F., Aragno, M. & Rossi, P. Minireview Statistical analysis of denaturing gel electrophoresis (DGGE) fingerprinting patterns. *Environmental Microbiology* **4,** 634–643 (2002).

97. Frost, T. M. In situ measurements of clearance rates for the freshwater sponge Spongilla lucustris. *Limnology and Oceanography* **23,** 1034–1039 (1978).

98. Fuhrman, J. A. Microbial community structure and its functional implications. *Nature* **459,** 193–9 (2009).

99. Fuhrman, J. A., Steele, J. A., Hewson, I., Schwalbach, M. S., Brown, M. V., Green, J. L. & Brown, J. H. A latitudinal diversity gradient in planktonic marine bacteria. *Proceedings of the National Academy of Sciences* **105,** 7774–8 (2008).

100. Fukami, T. Historical contingency in community assembly : integrating niches, species pools, and priority effects. *Annual Review of Ecology Evolution and Systematics* **46,** 1–23 (2015).

101. Gaston, K. J. & Fuller, R. A. Commonness, population depletion and conservation biology. *Trends in Ecology and Evolution* **23,** 14–19 (2008).

102. Gil, R., Sabater-Muñoz, B., Latorre, A., Silva, F. J. & Moya, A. Extreme genome reduction in Buchnera spp.: toward the minimal genome needed for symbiotic life. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 4454–4458 (2002).

103. Gilbert, J., Steele, J., *et al.* Defining seasonal marine microbial community dynamics. *Isme J* **6,** 298–308 (2012).

104. Gilbert, J. A., Jansson, J. K. & Knight, R. The Earth Microbiome project: successes and aspirations. *BMC Biology* **12,** 69 (2014).

105. Giles, E. C., Kamke, J., Moitinho-Silva, L., Taylor, M. W., Hentschel, U., Ravasi, T. & Schmitt, S. *FEMS Microbiology Ecology* **83,** 232–241 (2013).

106. Gili, J. M. M. & Coma, R. Benthic suspension feeders in marine food webs. *Tree* **13,** 316–321 (1998).

107. Gloeckner, V., Lindquist, N., Schmitt, S. & Hentschel, U. Ectyoplasia ferox, an Experimentally Tractable Model for Vertical Microbial Transmission in Marine Sponges. *Microbial Ecology* **65,** 462–474 (2013).

108. Gloeckner, V., Wehrl, M., *et al.* The HMA-LMA Dichotomy Revisited: an Electron Microscopical Survey of 56 Sponge Species. *The Biological bulletin* **227,** 78–88 (2014).

109. Goeij, J. M. D., Oevelen, D. V., Vermeij, M. J. a., Osinga, R., Middelburg, J. J., Goeij, A. F. P. M. D. & Admiraal, W. Surviving in a Marine Desert: The Sponge Loop Retains Resources Within Coral Reefs. *Science* **342,** 108–110 (2013).

110. Gompertz, Benjamin. On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society B: Biological Sciences* **115,** 513–583 (1825).

111. Gotelli, N. J. *A Primer of Ecology* (Sinauer Associates, 1998).

112. Gotelli, N. J. Research frontiers in null model analysis. *Global Ecology & Biogeography* **10,** 337–343 (2001).

113. Gotelli, N. J. & McCabe, D. J. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. *Ecology* **83,** 2091–2096 (2002).

114. Green, J. & Bohannan, B. J. M. Spatial scaling of microbial biodiversity. *Trends in Ecology and Evolution* **21,** 501–507 (2006).

115. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* **72,** 5069–5072 (2006).

116. Guimarães, P. R., Rico-Gray, V., Oliveira, P. S., Izzo, T. J., dos Reis, S. F. & Thompson, J. N. Interaction Intimacy Affects Structure and Coevolutionary Dynamics in Mutualistic Networks. *Current Biology* **17,** 1797–1803 (2007).

117. Hacquard, S. *et al.* Microbiota and host nutrition across plant and animal kingdoms. *Cell Host and Microbe* **17,** 603–616 (2015).

118. Hanski, I. Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos* **38,** 210–221 (1982).

119. Hanski, I. Single-Species Metapopulation Dynamics: Concepts, Models and Observations. *Biological Journal of the Linnean Society* **42,** 17–38 (1991).

120. Harris, D. J. Running head: Species interactions in Markov networks Title: Estimating species interactions from observational data with Markov networks Author: David J. Harris: Population Biology; 1 Shields Avenue, Davis CA, 95616 (2015).

121. Hartman, A. L., Lough, D. M., Barupal, D. K., Fiehn, O., Fishbein, T., Zasloff, M. & Eisen, J. a. Human gut microbiome adopts an alternative state following small bowel transplantation. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 17187–17192 (2009).

122. Henderson, P. a. & Magurran, a. E. Direct evidence that density-dependent regulation underpins the temporal stability of abundant species in a diverse animal community. *Proceedings of the Royal Society B: Biological Sciences* **281,** 20141336–20141336 (2014).

123. Hentschel, U., Hopke, J., Horn, M., Anja, B., Wagner, M., Hacker, J., Bradley, S., Friedrich, A. B. & Moore, B. S. Molecular Evidence for a Uniform Microbial Community in Sponges from Different Oceans Molecular Evidence for a Uniform Microbial Community in Sponges from Different Oceans. *Applied and Environmental Microbiology* **68,** 4431–4440 (2002).

124. Hentschel, U., Piel, J., Degnan, S. M. & Taylor, M. W. Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology* **10,** 641–654 (2012).

125. Hillebrand, H., Watermann, F., Karez, R. & Berninger, U. G. Differences in species richness patterns between unicellular and multicellular organisms. *Oecologia* **126,** 114–124 (2001).

126. Hillerislambers, J., Adler, P. B., Harpole, W. S., Levine, J. M. & Mayfield, M. M. Rethinking Community Assembly through the Lens of Coexistence Theory. *Annu. Rev. Ecol. Evol. Syst* **43,** 227–48 (2012).

127. Honegger, R. in (eds Nicole, M. & Gianinazzi-Pearson, V.) 157–176 (Springer Netherlands, Dordrecht, 1996).

128. Hooper, J. N. A. & Soest, R. W. M. in (eds Hooper, J. N. A., Soest, R. W. M. & Willenz, P.) 1–7 (Springer US, Boston, MA, 2002).

129. Horner-Devine, M. C., Lage, M., Hughes, J. B. & Bohannan, B. J. M. A taxa-area relationship for bacteria. *Nature* **432,** 750–753 (2004).

130. Howe, H. E. & Smallwood, J. Ecology of seed dispersal. *Annual Review of Ecology and Systematics* **13,** 201–228 (1982).

131. Hui, F. K. C., Taskinen, S., Pledger, S., Foster, S. D. & Warton, D. I. Model-based approaches to unconstrained ordination. *Methods in Ecology and Evolution* **6,** 399–411 (2015).

132. Hui, F. K., Warton, D. I., Ormerod, J. T., Haapaniemi, V. & Taskinen, S. Variational Approximations for Generalized Linear Latent Variable Models. *Journal of Computational and Graphical Statistics* **8600,** – (2016).

133. Huse, S. M., Ye, Y., Zhou, Y. & Fodor, A. A. A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS ONE* **7,** 1–12 (2012).

134. Ims, R. A., Henden, J. A. & Killengreen, S. T. Collapsing population cycles. *Trends in Ecology and Evolution* **23,** 79–86 (2008).

135. Ings, T. C. *et al.* Ecological networks - Beyond food webs. *Journal of Animal Ecology* **78,** 253–269 (2009).

136. Ives, A. R., B, D., L, C. K. & R, C. S. Estimating Community Stability and Ecological Interactions From Time-Series Data. *Ecological Monographs* **73,** 301–330 (2003).

137. Ives, A. R. & Carpenter, S. R. Stability and diversity of ecosystems. *Science (New York, N.Y.)* **317,** 58–62 (2007).

138. Ives, A. R. & Helmus, M. R. Phylogenetic metrics of community similarity. *The American naturalist* **176,** E128–E142 (2010).

139. Iwamoto, T., Tani, K., Nakamura, K., Suzuki, Y., Kitagawa, M., Eguchi, M. & Nasu, M. Monitoring impact of in situ biostimulation treatment on groundwater bacterial community by DGGE. *FEMS Microbiology Ecology* **32,** 129–141 (2000).

140. James, G. A., Beaudette, L. & Costerton, J. W. Interspecies bacterial interactions in biofilms. *Journal of Industrial Microbiology* **15,** 257–262 (1995).

141. Janzen, D. H. *Janzen Cuándo hay coevolución.pdf* 1980.

142. Jiménez, E., Ribes, M., Jimenez, E., Ribes, M., Jiménez, E. & Ribes, M. Sponges as a source of dissolved inorganic nitrogen: Nitrification mediated by temperate sponges. *Limnology and Oceanography* **52,** 948–958 (2007).

143. Joppa, L. N., Bascompte, J., Montoya, J. M., Sol??, R. V., Sanderson, J. & Pimm, S. L. Reciprocal specialization in ecological networks. *Ecology Letters* **12,** 961–969 (2009).

144. Jordano, P., García, C., Godoy, J. a. & García-Castaño, J. L. Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 3278–3282 (2007).

145. Jordano, P. Coevolution in Multispecific Interactions among Free-Living Species. *Evolution: Education and Outreach* **3,** 40–46 (2010).

146. Jordano, P. Patterns of Mutualistic Interactions in Pollination and Seed Dispersal: Connectance, Dependence Asymmetries, and Coevolution. *American Naturalist* **129,** 657–677 (1987).

147. Jordano, P., Bascompte, J. & Olesen, M. J. Ivariant Properties in Coevolutionary Networks of Plant – Animal Interactions. *Ecology letters* **6,** 69–81 (2003).

148. Kamke, J., Rinke, C., Schwientek, P., Mavromatis, K., Ivanova, N., Sczyrba, A., Woyke, T. & Hentschel, U. The candidate phylum Poribacteria by single-cell genomics: New insights into phylogeny, cell-compartmentation, eukaryote-like repeat proteins, and other genomic features. *PLoS ONE* **9** (2014).

149. Kaparakis, M., Philpott, D. J. & Ferrero, R. L. Mammalian NLR proteins; discriminating foe from friend. *Immunology and cell biology* **85,** 495–502 (2007).

150. Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Jeffrey, I. Human nutrition, the gut microbiome, and immune system: envisioning the future. *Nature* **474,** 327–336 (2012).

151. Kery, M. & Royle, J. A. *Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1:Prelude and Static Models* (Elsevier Science, 2015).

152. Krasnov, B. R., Shenbrot, G. I., Mouillot, D., Khokhlova, I. S. & Poulin, R. Spatial variation in species diversity and composition of flea assemblages in small mammalian hosts: Geographical distance or faunal similarity? *Journal of Biogeography* **32,** 633–644 (2005).

153. Krause, A. E., Frank, K. a., Mason, D. M., Ulanowicz, R. E. & Taylor, W. W. Compartments revealed in food-web structure. *Nature* **426,** 282–285 (2003).

154. Lamont, B. B. Functional interactions within plants-the contribution of keystone and other species to biological diversity. *Biodiversity in Mediterranean ecosystems of Australia. Surrey Beatty, Chipping Norton, Australia,* 95–127 (1992).

155. Lee, O. O., Chiu, P. Y., Wong, Y. H., Pawlik, J. R. & Qian, P. Y. Evidence for vertical transmission of bacterial symbionts from adult to embryo in the Caribbean Sponge Svenzea zeai. *Applied and Environmental Microbiology* **75,** 6147–6156 (2009).

156. Lee, O. O., Wong, Y. H. & Qian, P. Y. Inter- and intraspecific variations of bacterial communities associated with marine sponges from San Juan Island, Washington. *Applied and Environmental Microbiology* **75,** 3513–3521 (2009).

157. Lee, W.-J. & Hase, K. Gut microbiota-generated metabolites in animal health and disease. *Nature chemical biology* **10,** 416–24 (2014).

158. Legendre, P. & Gallagher, E. D. Ecologically meaningful transformations for ordination of species data. *Oecologia* **129,** 271–280 (2001).

159. Lenski, R. E. & Levin, B. R. Constraints on the Coevolution of Bacteria and Virulent Phage: A Model, Some Experiments, and Predictions for Natural Communities. *The American Naturalist* **125,** 585 (1985).

160. Letten, A. D., Keith, D. A., Tozer, M. G. & Hui, F. K. C. Fine-scale hydrological niche differentiation through the lens of multi-species co-occurrence models. *Journal of Ecology* **103,** 1264–1275 (2015).

161. Ley, R., Turnbaugh, P., Klein, S. & Gordon, J. Microbial ecology: human gut microbes associated with obesity. *Nature* **444,** 1022–3 (2006).

162. Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science (New York, N.Y.)* **320,** 1647–1651 (2008).

163. Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science (New York, N.Y.)* **320,** 1647–1651 (2008).

164. Lindquist, N. Palatability of invertebrate larvae to corals and sea anemones. *Marine Biology* **126,** 745–755 (1996).

165. Lozupone, C., Stomabaugh, J., Gordon, J., Jansson, J. & Knight, R. Diversity , stability and resilience of the human gut microbiota. *Nature* **489,** 220–230 (2012).

166. Lozupone, C. & Knight, R. UniFrac : a New Phylogenetic Method for Comparing Microbial Communities UniFrac : a New Phylogenetic Method for Comparing Microbial Communities. *Applied and environmental microbiology* **71,** 8228–8235 (2005).

167. Lurgi, M., López, B. C. & Montoya, J. M. Climate change impacts on body size and food web structure on mountain ecosystems. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **367,** 3050–7 (2012).

168. Lurgi, M., Montoya, D. & Montoya, J. M. The effects of space and diversity of interaction types on the stability of complex ecological networks. *Theoretical Ecology* (2015).

169. Mackie, R. I., Sghir, A. & Gaskins, H. R. Developmental microbial ecology of the neonatal gastrointestinal tract. *American Journal of Clinical Nutrition* **69** (1999).

170. Magurran, A. E. & Henderson, P. A. Explaining the excess of rare species in natural species abundance distributions. *Nature* **422,** 714–716 (2003).

171. Maldonado, M. Intergenerational transmission of symbiotic bacteria in oviparous and viviparous demosponges, with emphasis on intracytoplasmically-compartmented bacterial types. *Journal of the Marine Biological Association of the UK* **87,** 1701–1713 (2007).

172. Maldonado, M., Ribes, M. & van Duyl, F. C. *Nutrient Fluxes Through Sponges. Biology, Budgets, and Ecological Implications* 113–182 (2012).

173. Maldonado, M. & Riesgo, A. Gametogenesis, embryogenesis, and larval features of the oviparous sponge Petrosia ficiformis (Haplosclerida, Demospongiae). *Marine Biology* **156,** 2181–2197 (2009).

174. Marino, S., Baxter, N. T., Huffnagle, G. B., Petrosino, J. F. & Schloss, P. D. Mathematical modeling of primary succession of murine intestinal microbiota. *Proc Natl Acad Sci U S A* **111,** 439–444 (2014).

175. Martiny, J. B. H., Eisen, J. A., Penn, K., Allison, S. D. & Horner-Devine, M. C. Drivers of bacterial beta-diversity depend on spatial scale. *Proceedings of the National Academy of Sciences* **108,** 7850–7854 (2011).

176. Martorell, C. & Freckleton, R. P. Testing the roles of competition, facilitation and stochasticity on community structure in a species-rich assemblage. *Journal of Ecology* **102,** 74–85 (2014).

177. May, R. M. C. *Stability and Complexity in Model Ecosystems* (Princeton University Press, 1973).

178. McCann, K. S. The diversity-stability debate. *Nature* **405,** 228–233 (2000).

179. McCann, K. S. *Food Webs (MPB-50)* (Princeton University Press, 2012).

180. McFall-Ngai, M. Are biologists in 'future shock'? Symbiosis integrates biology across domains. *Nature reviews. Microbiology* **6,** 789–792 (2008).

181. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences* **110,** 3229–3236 (2013).

182. McKellar, R. C. & Lu, X. *Modeling Microbial Responses in Food* (Taylor & Francis, 2003).

183. McNally, L., Viana, M. & Brown, S. P. Cooperative secretions facilitate host range expansion in bacteria. *Nature communications* **5,** 4594 (2014).

184. Mills, L. S., Soulé, M. E. & Doak, D. F. The Keystone-Species Concept in Ecology and Conservation. *Bioscience* **43,** 219–224 (1993).

185. Missaghi, B., Barkema, H., Madsen, K. & Ghosh, S. Perturbation of the Human Microbiome as a Contributor to Inflammatory Bowel Disease. *Pathogens* **3,** 510–527 (2014).

186. Mohamed, N. M., Rao, V., Hamann, M. T., Kelly, M. & Hill, R. T. Monitoring bacterial diversity of the marine sponge Ircinia strobilina upon transfer into aquaculture. *Applied and Environmental Microbiology* **74,** 4133–4143 (2008).

187. Moitinho-Silva, L., Bayer, K., Cannistraci, C. V., Giles, E. C., Ryu, T., Seridi, L., Ravasi, T. & Hentschel, U. Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea. *Molecular Ecology* **23,** 1348–1363 (2014).

188. Montoya, J. M., Pimm, S. L. & Sole, R. Ecological networks and their fragility. *Nature* **442,** 259–264 (2006).

189. Morales-Castilla, I., Matias, M. G., Gravel, D. & Ara??jo, M. B. Inferring biotic interactions from proxies. *Trends in Ecology and Evolution* **30,** 347–356 (2015).

190. Mougi, a. & Kondoh, M. Diversity of Interaction Types and Ecological Community Stability. *Science* **337,** 349–351 (2012).

191. Moya, A., Peretó, J., Gil, R. & Latorre, A. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nature reviews. Genetics* **9,** 218–29 (2008).

192. Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., Gonzalez, A., Fontana, L., Henrissat, B., Knight, R. & Gordon, J. I. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332,** 970–974 (2011).

193. Mutshinda, C. M., O'Hara, R. B. & Woiwod, I. P. A multispecies perspective on ecological impacts of climatic forcing. *Journal of Animal Ecology* **80,** 101–107 (2011).

194. Mutshinda, C. M., O'Hara, R. B. & Woiwod, I. P. What drives community dynamics? *Proceedings of the Royal Society B* **276,** 2923–2929 (2009).

195. Muyzer, G. & Smalla, K. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie van Leeuwenhoek* **73,** 127–41 (1998).

196. Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4,** 133–142 (2013).

197. Nemergut, D. R. *et al.* Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews* **77,** 342–356 (2013).

198. Newman, M. *Networks: an introduction* (OUP Oxford, 2010).

199. Nogueira, T., Rankin, D. J., Touchon, M., Taddei, F., Brown, S. P. & Rocha, E. P. C. Horizontal Gene Transfer of the Secretome Drives the Evolution of Bacterial Cooperation and Virulence. *Current Biology* **19,** 1683–1691 (2009).

200. Nyholm, S. V. & McFall-Ngai, M. J. The winnowing: Establishing the squid-vibrio symbiosis. *Nature Reviews Microbiology* **2,** 632–642 (2004).

201. Oh, J., Byrd, A. L., Deming, C., Conlan, S., Kong, H. H. & Segre, J. A. Biogeography and individuality shape function in the human skin metagenome. *Nat Rev* **514,** 59–64 (2014).

202. O'Hara, R. B. & Sillanpaa, M. J. A review of bayesian variable selection methods: What, how and which. *Bayesian Analysis* **4,** 85–118 (2009).

203. Olesen, J. M., Bascompte, J., Dupont, Y. L. & Jordano, P. The modularity of pollination networks. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 19891–19896 (2007).

204. Olesen, J. M., Bascompte, J., Elberling, H. & Jordano, P. Temporal Dynamics in a Pollination Network. *Ecology* **89,** 1573–1582 (2008).

205. Olesen, J. M., Eskildsen, L. I. & Venkatasamy, S. Invasion of pollination networks on oceanic islands: Importance of invader complexes and endemic super generalists. *Diversity and Distributions* **8,** 181–192 (2002).

206. Oliver, K. M., Russell, J. a., Moran, N. a. & Hunter, M. S. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 1803–7 (2003).

207. Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution* (2015).

208. Ovaskainen, O., Roy, D. B., Fox, R. & Anderson, B. J. Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution* (2015).

209. Paine, R. T. Food-web analysis through field measurement of per capita interaction strength. *Nature* **355,** 73–75 (1992).

210. Paine, R. T. A Note on Trophic Complexity and Community Stability. *The American Naturalist* **103,** 91–93 (1969).

211. Parfrey, L. W. & Knight, R. Spatial and temporal variability of the human microbiota. *Clin Microbiol Infect* **18,** 8–11 (2012).

212. Parrish, J. A. D. & Bazzaz, F. A. Competitive Interactions in Plant Communities of Different Successional Ages. *Ecology* **63,** 314–320 (1982).

213. Pascoal, C., Cássio, F., Nikolcheva, L. & Bärlocher, F. Realized Fungal Diversity Increases Functional Stability of Leaf Litter Decomposition Under Zinc Stress. *Microbial Ecology* **59,** 84–93 (2009).

214. Patefield, W. M. Algorithm AS 159 : An Efficient Method of Generating Random R C Tables with Given Row and Column Totals All use subject to JSTOR Terms and Conditions Algorithm Random R x C Tables with An Efficient Method of Generating Given Row and Column Totals. *Journal of the Royal Statistical Society. Series C* **30,** 91–97 (1981).

215. Pedros-Alio, C. Marine microbial diversity: can it be determined? *Trends in Microbiology* **14,** 257–263 (2006).

216. Peters, B. M., Jabra-Rizk, M. A., O'May, G. A., William Costerton, J. & Shirtliff, M. E. Polymicrobial interactions: Impact on pathogenesis and human disease. *Clinical Microbiology Reviews* **25,** 193–213 (2012).

217. Pimm, S. L. in, 1–11 (Springer Netherlands, Dordrecht, 1982).

218. Pimm, S. L. The Balance of Nature: What Is It and Why Care? *The Balance of Nature?: Ecological Issues in the Conservation of Species and Communities,* 434 (1991).

219. Pimm, S. L. The complexity and stability of ecosystems. *Nature* **307,** 321–326 (1984).

220. Pita, L., Turon, X., López-Legentil, S. & Erwin, P. M. Host rules: Spatial stability of bacterial communities associated with marine sponges (ircinia spp.) in the western mediterranean sea. *FEMS Microbiology Ecology* **86,** 268–276 (2013).

221. Pitta, D. W., Kumar, S., Vecchiarelli, B., Baker, L. D., Ferguson, J. D., Thomsen, N., Shirley, D. J. & Bittinger, K. Temporal dynamics in the ruminal microbiome of dairy cows during the transition period. *Journal of Animal Science* **92,** 4014–4022 (2014).

222. Plummer, M. *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling* 2003.

223. Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A. & Mccarthy, M. A. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution* **5,** 397–406 (2014).

224. Poppell, E., Weisz, J., Spicer, L., Massaro, A., Hill, A. & Hill, M. Sponge heterotrophic capacity and bacterial community structure in high- and low-microbial abundance sponges. *Marine Ecology* **35,** 414–424 (2014).

225. Poulin, R. The decay of similarity with geographical distance in parasite communities of vertebrate hosts. *Journal of Biogeography* **30,** 1609–1615 (2003).

226. Proctor, M., Yeo, P. & Lack, A. *The Natural History of Pollination (Collins New Naturalist Library, Book 83)* (HarperCollins Publishers, 2012).

227. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490,** 55–60 (2012).

228. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F. O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research* **41,** 590–596 (2013).

229. Raffaelli, D. G. & Hall, S. J. in (eds Polis, G. A. & Winemiller, K. O.) 185–191 (Springer US, Boston, MA, 1996).

230. Ramsey, M. M., Rumbaugh, K. P. & Whiteley, M. Metabolite cross-feeding enhances virulence in a model polymicrobial infection. *PLoS Pathogens* **7,** 1–8 (2011).

231. Ranjard, L. *et al.* Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nature communications* **4,** 1434 (2013).

232. Reche, I., Pulido-Villena, E., Morales-Baquero, R. & Casamayor, E. O. Does ecosystem size determine aquatic bacterial richness? *Ecology* **86,** 1715–1722 (2005).

233. Reiswig, H. M. Bacteria as food for temperate-water marine sponges. *Canadian Journal of Zoology* **53,** 582–589 (1975).

234. Reiswig, H. M. Partial Carbon and Energy Budgets of the Bacteriosponge Verohgia fistularis (Porifera: Demospongiae) in Barbados. *Marine Ecology* **2,** 273–293 (1981).

235. Relman, D. a. The human microbiome: ecosystem resilence and health. *Nutr Rev* **70,** 1–12 (2013).

236. Reveillaud, J., Maignien, L., Eren, M. a., Huber, J. a., Apprill, A., Sogin, M. L. & Vanreusel, A. Host-specificity among abundant and rare taxa in the sponge microbiome. *The ISME journal* **8,** 1198–1209 (2014).

237. Ribes, M., Jim??nez, E., Yahel, G., L??pez-Sendino, P., Diez, B., Massana, R., Sharp, J. H. & Coma, R. Functional convergence of microbes associated with temperate marine sponges. *Environmental Microbiology* **14,** 1224–1239 (2012).

238. Riesgo, A. & Maldonado, M. Differences in reproductive timing among sponges sharing habitat and thermal regime. *Invertebrate Biology* **127,** 357–367 (2008).

239. Robertson, S. J., Rubino, S. J., Geddes, K. & Philpott, D. J. Examining host-microbial interactions through the lens of NOD: From plants to mammals. *Seminars in Immunology* **24,** 9–16 (2012).

240. Robinson, C. J., Bohannan, B. J. M. & Young, V. B. From structure to function: the ecology of host-associated microbial communities. *Microbiology and molecular biology reviews : MMBR* **74,** 453–76 (2010).

241. Rooney, N. & McCann, K. S. Integrating food web diversity, structure and stability. *Trends in Ecology and Evolution* **27,** 40–45 (2012).

242. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nature Rev Immunology* **9,** 313–324 (2009).

243. Royle, J. A. & Dorazio, R. M. *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities* (Academic, 2008).

244. Ruan, Q., Dutta, D., Schwalbach, M. S., Steele, J. A., Fuhrman, J. A. & Sun, F. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* **22,** 2532–2538 (2006).

245. Ruby, E. G. Symbiotic conversations are revealed under genetic interrogation. *Nature Reviews Microbiology* **6,** 752–762 (2008).

246. Sauve, A. M. C., Fontaine, C. & Thébault, E. Structure-stability relationships in networks combining mutualistic and antagonistic interactions. *Oikos* **123,** 378–384 (2014).

247. Schauer, M., Balagué, V., Pedrós-Alió, C. & Massana, R. Seasonal changes in the taxonomic composition of bacterioplankton in a oligotrophic coastal system. *Aquat. Microb. Ecol.* **31,** 163–174 (2003).

248. Scher, J. U. & Abramson, S. B. The microbiome and rheumatoid arthritis. *Nature reviews. Rheumatology* **7,** 569–78 (2011).

249. Schläppy, M. L., Schöttner, S. I., Lavik, G., Kuypers, M. M. M., de Beer, D. & Hoffmann, F. Evidence of nitrification and denitrification in high and low microbial abundance sponges. *Marine Biology* **157,** 593–602 (2010).

250. Schloss, P. D. *et al.* Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75,** 7537–7541 (2009).

251. Schluter, J. & Foster, K. R. The Evolution of Mutualism in Gut Microbiota Via Host Epithelial Selection. *PLoS Biology* **10** (2012).

252. Schmitt, S., Angermeier, H., Schiller, R., Lindquist, N. & Hentschel, U. Molecular microbial diversity survey of sponge reproductive stages and mechanistic insights into vertical transmission of microbial symbionts. *Applied and Environmental Microbiology* **74,** 7694–7708 (2008).

253. Schmitt, S., Hentschel, U. & Taylor, M. W. Deep sequencing reveals diversity and community structure of complex microbiota in five Mediterranean sponges. *Hydrobiologia* **687,** 341–351 (2012).

254. Schmitt, S., Tsai, P., *et al.* Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *The ISME Journal* **6,** 564–576 (2012).

255. Schmitt, S., Wehrl, M., Bayer, K., Siegl, A. & Hen. *Marine sponges as models for commensal microbe-host interactions* 2007.

256. Schmitt, S., Weisz, J. B., Lindquist, N. & Hentschel, U. Vertical transmission of a phylogenetically complex microbial consortium in the viviparous sponge Ircinia felix. *Applied and Environmental Microbiology* **73,** 2067–2078 (2007).

257. Shade, A. & Handelsman, J. Beyond the Venn diagram: The hunt for a core microbiome. *Environmental Microbiology* **14,** 4–12 (2012).

258. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS Pandemic. *Cold Spring Harb Perspect Med* **1,** a006841 (2011).

259. Shimadzu, H., Dornelas, M. & Magurran, A. E. Measuring temporal turnover in ecological communities. *Methods in Ecology and Evolution* **6,** 1384–1394 (2015).

260. Simister, R. L., Deines, P., Botté, E. S., Webster, N. S. & Taylor, M. W. Sponge-specific clusters revisited: A comprehensive phylogeny of sponge-associated microorganisms. *Environmental Microbiology* **14,** 517–524 (2012).

261. Sipkema, D., de Caralt, S., Morillo, J. A., Al-Soud, W. A., S??rensen, S. J., Smidt, H. & Uriz, M. J. Similar sponge-associated bacteria can be acquired via both vertical and horizontal transmission. *Environmental Microbiology* **17,** 3807–3821 (2015).

262. Solé, R. V. & Montoya, J. M. Complexity and fragility in ecological networks. *Proceedings. Biological sciences / The Royal Society* **268,** 2039–45 (2001).

263. Song, S. J. *et al.* Cohabiting family members share microbiota with one another and with their dogs. *eLife* **2013,** 1–22 (2013).

264. Stacy, A., Everett, J., Jorth, P., Trivedi, U., Rumbaugh, K. P. & Whiteley, M. Bacterial fight-and-flight responses enhance virulence in a polymicrobial infection. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 7819–24 (2014).

265. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).

266. Steele, J. A. *et al.* Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *The ISME journal* **5,** 1414–25 (2011).

267. Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Rätsch, G., Pamer, E. G., Sander, C. & Xavier, J. B. Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota. *PLoS Computational Biology* **9,** 31–36 (2013).

268. Stenseth Nils Chr, Viljugrein Hildegunn, Takashi Saitoh, Hansen Thomas F, Kittilsen Marte O, Bølviken Erik & Glöckner Fredrik. Seasonality , density in Hokkaido cycles dependence , and voles population. *Pnas* **100,** 11478–11483 (2003).

269. Sunagawa, S. *et al.* Ocean plankton. Structure and function of the global ocean microbiome. *Science (New York, N.Y.)* **348,** 1261359 (2015).

270. Taylor, M. W., Thacker, R. W. & Hentschel, U. Evolutionary Insights from Sponges. *Science* **316,** 1854–1855 (2007).

271. Taylor, M. W., Schupp, P. J., Dahllof, I., Kjelleberg, S. & Steinberg, P. D. Host specificity in marine sponge-associated bacteria, and potential implications for marine microbial diversity. *Environmental Microbiology* **6,** 121–130 (2004).

272. Taylor, M. W., Schupp, P. J., De Nys, R., Kjelleberg, S. & Steinberg, P. D. Biogeography of bacteria associated with the marine sponge Cymbastela concentrica. *Environmental Microbiology* **7,** 419–433 (2005).

273. Taylor, M. W., Tsai, P., Simister, R. L., Deines, P., Botte, E., Ericson, G., Schmitt, S. & Webster, N. S. 'Sponge-specific' bacteria are widespread (but rare) in diverse marine environments. *The ISME Journal* **7,** 438–443 (2013).

274. Taylor, M. W., Radax, R., Stegar, D. & Wagner, M. Sponge-associated microorganisms: evolution, evology, and biotechnological potential. *Microbiology and Molecular Biology Reviews* **71,** 295–347 (2007).

275. Taylor, T. N., Remy, W., Hass, H. & Kerp, H. Mycological Society of America Fossil Arbuscular Mycorrhizae from the Early Devonian. *Mycologia* **87,** 560–573 (1995).

276. Thacker, R. W. & Freeman, C. J. *Sponge-Microbe Symbioses. Recent Advances and New Directions* 1st ed., 57–111 (Elsevier Ltd., 2012).

277. Thacker, R. & Starnes, S. Host specificity of the symbiotic cyanobacterium Oscillatoria spongeliae in marine sponges , Dysidea spp . *Marine Biology* **142,** 643–648 (2003).

278. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* **458,** 1018–1020 (2009).

279. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).

280. Thébault, E. & Fontaine, C. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* **329,** 853–856 (2010).

281. Thibaut, L. M. Diversity and stability of herbivorous fishes on coral reefs. *Ecology* **93,** 891–901 (2012).

282. Thomas, T. *et al.* Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *The ISME journal* **4,** 1557–1567 (2010).

283. Thompson, J. N. *Relentless evolution* (University of Chicago Press, 2013).

284. Thompson, J. N. *The coevolutionary process* (University of Chicago Press, 1994).

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

286. Thompson, J. N., Schwind, C., Guimarães, P. R. & Friberg, M. Diversification through multitrait evolution in a coevolving interaction. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 11487–92 (2013).

287. Thorson, J. T., Scheuerell, M. D., Shelton, A. O., See, K. E., Skaug, H. J. & Kristensen, K. Spatial factor analysis: A new tool for estimating joint species distributions and correlations in species range. *Methods in Ecology and Evolution* (2015).

288. Tkacz, A. & Poole, P. Role of root microbiota in plant productivity. *Journal of Experimental Botany* **66,** 2167–2175 (2015).

289. Toju, H., Guimarães, P. R., Olesen, J. M. & Thompson, J. N. Assembly of complex plant–fungus networks. *Nature Communications* **5,** 5273 (2014).

290. Trosvik, P. & de Muinck, E. J. Ecology of bacteria in the human gastrointestinal tract-identification of keystone and foundation taxa. *Microbiome* **3,** 44 (2015).

291. Tsai, K.-N., Lin, S.-H., Liu, W.-C. & Wang, D. Inferring microbial interaction network from microbiome data using RMN algorithm. *BMC systems biology* **9,** 54 (2015).

292. Tschöp, M. H., Hugenholtz, P. & Karp, C. L. Getting to the core of the gut microbiome. *Nature biotechnology* **27,** 344–346 (2009).

293. Tsuchida, T., Koga, R., Shibao, H., Matsumoto, T. & Fukatsu, T. Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, Acyrthosiphon pisum. *Molecular Ecology* **11,** 2123–2135 (2002).

294. Turnbaugh, P. J. & Gordon, J. I. The core gut microbiome, energy balance and obesity. *The Journal of Physiology* **587,** 4153–4158 (2009).

295. Turnbaugh, P. J., Hamady, M., *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).

296. Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-liggett, C., Knight, R. & Gordon, J. I. The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature* **449,** 804–810 (2007).

297. Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R. & Gordon, J. I. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–31 (2006).

298. Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R. & Gordon, J. I. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Science translational medicine* **1,** 6ra14 (2009).

299. Tylianakis, J. M., Tscharntke, T. & Lewis, O. T. Habitat modification alters the structure of tropical host-parasitoid food webs. *Nature* **445,** 202–205 (2007).

300. Ulrich, W. & Gotelli, N. Null Model Analysis of Species Nestedness Patterns. **88,** 1824–1831 (2007).

301. Uriz, M. J., Turon Xavier & Becerro Mikel A. Morphology and Ultrastructure of the Swimming Larvae of Crambe crambe (Demospongiae, Poecilosclerida). *Invertebrate Biology* **120,** 295–307 (2001).

302. Vacelet, J. & Donadey, C. Electron microscope study of the association between some sponges and bacteria. *Journal of Experimental Marine Biology and Ecology* **30,** 301–314 (1977).

303. Van Soest, R. W. M. *et al.* Global diversity of sponges (Porifera). *PLoS ONE* **7** (2012).

304. Vazquez, D. P. & Aizen, M. A. ASYMMETRIC SPECIALIZATION: A PERVASIVE FEATURE OF PLANT–POLLINATOR INTERACTIONS. **85,** 1251–1257 (2004).

305. Vazquez, D. P. & Aizen, M. A. Null Model Analyses of Specialization in Plant–Pollinator Interactions. *Ecology* **84,** 2493–2501 (2003).

306. Vazquez, D. P., Lomascolo, S. B., Maldonado, B. M., Chacoff, N. P., Jimena, D., Stevani, E. L. & Vitale, N. L. The strength of plant–pollinator interactions. *Ecology* **93,** 719–725 (2012).

307. Vazquez, D. P., Melian, C. J., Williams, N. M., Bluthgen, N., Krasnov, B. R. & Poulin, R. Species abundance and asymmetric interaction strength in ecological networks. *Oikos* **116,** 1120–1127 (2007).

308. Vazquez, D. P., Poulin, R., Krasnov, B. R. & Shenbrot, G. I. Species abundance and the distribution of specialization in host-parasite interaction networks. *Journal of Animal Ecology* **74,** 946–955 (2005).

309. W., N. J., Jordana, F. L. & M., M. R. Analysis of Artifacts Suggests DGGE Should Not Be Used For Quantitative Diversity Analysis. *Journal of Microbiological Methods* **92,** 256–263 (2013).

310. Wang, J., Shen, J., *et al.* Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *The ISME journal* **7,** 1310–1321 (2013).

311. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73,** 5261–5267 (2007).

312. Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C. & Hui, F. K. C. So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology and Evolution* **xx,** 1–14 (2015).

313. Warton, D. I., Wright, S. T. & Wang, Y. Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution* **3,** 89–101 (2012).

314. Webster, N. S., Luter, H. M., Soo, R. M., Bott??, E. S., Simister, R. L., Abdo, D. & Whalan, S. Same, same but different: Symbiotic bacterial associations in GBR sponges. *Frontiers in Microbiology* **3,** 1–11 (2012).

315. Webster, N. S. Sponge disease: A global threat? *Environmental Microbiology* **9,** 1363–1375 (2007).

316. Webster, N. S., Cobb, R. E., Soo, R., Anthony, S. L., Battershill, C. N., Whalan, S. & Evans-Illidge, E. Bacterial Community Dynamics in the Marine Sponge Rhopaloeides odorabile Under In Situ and Ex Situ Cultivation. *Marine Biotechnology* **13,** 296–304 (2011).

317. Webster, N. S., Negri, A. P., Munro, M. M. H. G. & Battershill, C. N. Diverse microbial communities inhabit Antarctic sponges. *Environmental Microbiology* **6,** 288–300 (2004).

318. Webster, N. S. & Taylor, M. W. Marine sponges and their microbial symbionts: Love and other relationships. *Environmental Microbiology* **14,** 335–346 (2012).

319. Webster, N. S., Taylor, M. W., Behnam, F., Lücker, S., Rattei, T., Whalan, S., Horn, M. & Wagner, M. Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environmental Microbiology* **12,** 2070–2082 (2010).

320. Webster, N. S., Webb, R. I., Ridd, M. J., Hill, R. T. & Negri, A. P. The effects of copper on the microbial community of a coral reef sponge. *Environmental Microbiology* **3,** 19–31 (2001).

321. Wehrl, M., Steinert, M. & Hentschel, U. Bacterial uptake by the marine sponge Aplysina aerophoba. *Microbial Ecology* **53,** 355–365 (2007).

322. Weisz, J. B., Lindquist, N. & Martens, C. S. Do associated microbial abundances impact marine demosponge pumping rates and tissue densities? *Oecologia* **155,** 367–376 (2008).

323. Weitz, J. S., Poisot, T., Meyer, J. R., Flores, C. O., Valverde, S., Sullivan, M. B. & Hochberg, M. E. Phage-bacteria infection networks. *Trends in Microbiology* **21,** 82–91 (2013).

324. Wiens, M., Korzhev, M., Perović-Ottstadt, S., Luthringer, B., Brandt, D., Klein, S. & Müller, W. E. G. Toll-like receptors are part of the innate immune defense system of sponges (Demospongiae: Porifera). *Molecular Biology and Evolution* **24,** 792–804 (2007).

325. Wilkinson, C. R. Microbial associations in sponges. III. Ultrastructure of the in situ associations in coral reef sponges. *Marine Biology* **49,** 177–185 (1978).

326. Wilkinson, C. R., Garrone, R. & Vacelet, J. Marine Sponges Discriminate between Food Bacteria and Bacterial Symbionts: Electron Microscope Radioautography and in situ Evidence. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **220,** 519–528 (1984).

327. Willson, M. F., Irvine, A. K. & Walsh, N. G. Vertebrate Dispersal Syndromes in Some Australian and New Zealand Plant Communities, with Geographic Comparisons. *Biotropica* **21,** 133–147 (1989).

328. Wootton, J. T. Estimates and tests of per capita interaction strength: Diet, abundance, and impact of intertidally foraging birds. *Ecological Monographs* **67,** 45–64 (1997).

329. Wootton, J. T. & Emmerson, M. Measurement of Interaction Strength in Nature. *Annual Review of Ecology, Evolution, and Systematics* **36,** 419–444 (2005).

330. Yang, J., Sun, J., Lee, O. O., Yim, H. W. & Qian, P. Y. Phylogenetic diversity and community structure of sponge-associated bacteria from mangroves of the Caribbean Sea. *Aquatic Microbial Ecology* **62,** 231–240 (2011).

331. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486,** 222–227 (2012).

332. Yin, Z., Zhu, M., Davidson, E. H., Bottjer, D. J., Zhao, F. & Tafforeau, P. Sponge grade body fossil with cellular resolution dating 60 Myr before the Cambrian. *Proceedings of the National Academy of Sciences of the United States of America* **112,** E1453–60 (2015).

333. Yuen, B., Bayes, J. M. & Degnan, S. M. The characterization of sponge nlrs provides insight into the origin and evolution of this innate immune gene family in animals. *Molecular Biology and Evolution* **31,** 106–120 (2014).

334. Zan, J. *et al.* A complex LuxR-LuxI type quorum sensing network in a roseobacterial marine sponge symbiont activates flagellar motility and inhibits biofilm formation. *Molecular Microbiology* **85,** 916–933 (2012).

335. Zhou, Y. *et al.* Biogeography of the ecosystems of the healthy human body. *Genome biology* **14,** R1 (2013).

336. Zinger, L., Boetius, A. & Ramette, A. Bacterial taxa-area and distance-decay relationships in marine environments. *Molecular Ecology* **23,** 954–964 (2014).

337. Zinger, L., Amaral-Zettler, L. A., *et al.* Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS ONE* **6,** 1–11 (2011).