Microbial Community Structure and Dynamics on Patchy Landscapes

by

Manoshi Sen Datta

B.S., University of California, Berkeley (2010)

Submitted to the Program in Computational and Systems Biology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2016

Author .....
Program in Computational and Systems Biology
May 20, 2016

Certified by .....
Jeff Gore
Associate Professor, Physics
Thesis Supervisor

Certified by .....
Otto X. Cordero
Assistant Professor, Civil & Environmental Engineering
Thesis Supervisor

Certified by .....
Martin F. Polz
Professor, Civil & Environmental Engineering
Thesis Supervisor

Accepted by .....
Christopher B. Burge
Director, Computational and Systems Biology Graduate Program

Microbial Community Structure and Dynamics on Patchy Landscapes

by

Manoshi Sen Datta

Submitted to the Program in Computational and Systems Biology
on May 20, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Microbes are tiny metabolic engines with large-scale effects on industry, the environment, and human health. Understanding how the micron-scale actions (and interactions) of individual microbes give rise to macro-scale consequences remains a major challenge in microbial ecology. However, for the most part, studies employ coarse-grained sampling schemes, which average over the heterogeneous microscopic structure of microbial communities. This has limited our ability to establish mechanistic links between dynamics occurring across these disparate spatial scales. However, such links are critical for (a) making sense of the tremendous extant microbial diversity on Earth, and (b) predicting how perturbations (e.g., global climate change) may influence microbial diversity and function.

In this thesis, I characterize the structure and dynamics of wild bacterial populations in the ocean at spatial scales of tens of microns. I then employ a simple, two-strain laboratory model system to link (cooperative) inter-species interactions at local scales to emergent properties at larger scales, focusing on spatially connected meta-communities undergoing range expansions into new territory. This work encompasses diverse environments (ranging from well-mixed communities in the laboratory to individual crustaceans) and approaches (including mathematical modeling, high-throughput sequencing, and traditional microbiological experiments).

Altogether, we find that the microscale environment inhabited by a microbe – that is, "what the neighborhood is like" and "who lives next to whom" – shapes the structure and dynamics of wild microbial populations at local scales. Moreover, these local interactions can drive patterns of biodiversity and function, even at spatial scales much larger than the length of an individual cell. Thus, our work represents a small step toward developing mechanistic theories for how microbes shape our planet's ecosystems.

Thesis Supervisor: Jeff Gore
Title: Associate Professor, Physics

Thesis Supervisor: Otto X. Cordero
Title: Assistant Professor, Civil & Environmental Engineering

Thesis Supervisor: Martin F. Polz
Title: Professor, Civil & Environmental Engineering

Acknowledgments

When I began my Ph.D. experience in June of 2010, I had no idea what I was getting into. I didn't anticipate how much I would learn or the struggles that getting a Ph.D. would entail. But, I also never imagined the community of talented, compassionate people who would make it all possible. I've never had the opportunity to express my gratitude so openly, and I find that I'm not very good at it. Regardless, just trust me – you all matter more than I can say.

Jeff Gore was (un)fortunate enough to have to put up with me from the very beginning. In the early years, he taught me how to "marinate in raw data" and patiently helped me correct my many missteps along the way. As a result, any good habits that I've developed as a scientist, I usually attribute to him. Even when my work took me far away from Jeff's core interests, his continued guidance made it easier for me to follow my passions and to take intellectual risks. Although I joined Jeff's lab on a whim in June 2010, it turned out to be one of the best decisions I've ever made.

Martin Polz played a pivotal role in shaping the direction of my Ph.D. work. After I finished my first paper, I knew that I wanted to work on more "natural" systems, but I didn't have the first idea as to where I should start. Martin took me on as a student, even though I didn't know anything about the ocean, environmental microbiology, or sequencing. His lab became the scientific playground through which I began to work on topics that I would never have broached otherwise. Martin has also been a great source of down-to-earth advice about how to place the results of my analyses in the context of the real world.

Finally, Otto Cordero allowed me to come into my own as a scientist. We started working together when he was just a postdoc with some pie-in-the-sky ideas for his lab. Early on, he allowed me to take ownership of some of those ideas, and together, we've managed to put some of them into practice through a truly synergistic working relationship. I've had so much fun. Otto has also made it a priority to help me advance my career, and I'm especially thankful for all of the advice and unique opportunities

that he's given me.

I'm also grateful to the members of my thesis committee over the years – Eric Alm, Penny Chisholm, and Daniel Segre – for their time and expert advice.

With three advisors come three wonderful groups of labmates who taught me new things, gave me advice that made my science better, and offered support when things weren't going well. It's a pleasure to work with such an intellectually engaged, fun-loving, and caring bunch. I could spend pages and pages talking about each one of my lab colleagues, but in the interest of brevity, I've limited myself to a few.

From the Gore lab, I owe special thanks to several colleagues, who have contributed immensely to my scientific education and emotional well-being: Eugene Yurtsev, Sherry Chao, Andrea Velenich, Tanya Artemova, Kevin Axelrod and Arolyn Conwill. Especially in the early days, these folks helped me adjust to the rigors of full-time research through valuable discussions, companionship, and comic relief. From the Polz lab, I'd especially like to thank Michael Cutler, for never failing to drop what he's doing to help me and for offering moral support. Kathryn Kauffmann has also been a source of many enjoyable late-night conversations; every time we talk, she tells me something new and cool about the ocean. Thanks also to Diana Chien, who was a firm friend, even before I joined the Polz lab, and with whom I've shared many Student Center pop gen study sessions, lengthy lunches, and long walks in search of bubble tea. Among my newest labmates in the Cordero lab, Tim Enke has also been a great friend and scientific collaborator.

My friends at MIT and beyond have also provided moral and intellectual support outside the lab. Thanks to my fellow CSB classmates – especially Leyla Isik, Mimi Xie, Thomas Gurry, Arshed Al-Obeidi, Chris Smillie, Anna Podgornaia, Vikram Agrawal, Christoph Engert, and Tracy Washington – for your friendship from the very beginning. I'd also be remiss if I didn't mention Martina Koeva and Pooja Upadhyaya, two lovely ladies whose support and encouragement has been invaluable to me over the years.

No matter how you slice it, this thesis would not have been possible without my parents, Supriyo and Anuradha Datta. They gave me the freedom to follow my own

path, while affording me every opportunity, educational and otherwise. They also taught me to value intellectual autonomy, as well as clarity and simplicity of thought. These are values that I try to emulate every day in my work, though I'm still working on it. I also owe thanks to my sister, Malika, for being a reliable source of obscure *Gilmore Girls* references.

Finally, thank you to Nate Cermak, who has been my constant companion, my scientific consultant, and my most ardent cheerleader. You're the best.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

By definition, microbes are invisible to the naked eye and are typically no more than a few microns in length. Yet, these miniscule organisms are the metabolic engines of our planet, collectively exerting tremendous control on processes at much larger spatial scales. These include (but are not limited to) the production of fermented foods (cheese or beer, for instance) [187], symbioses and dysbioses in the human body [30], and biogeochemical cycling in the ocean [89].

The classic mantra of microbial ecology ("Who is there and what are they doing?") suggests that we ought to be able to predict these emergent properties of microbial ecosystem function from the properties of the individual building blocks. Furthermore, with the advent of high-throughput sequencing technologies, it is now straightforward to characterize the basic building blocks of complex microbial consortia – that is, the diverse mixture of microbial cells, genes, and metabolisms that are present in a given ecosystem [62]. However, just as the extreme tensile strength of spider's silk cannot be explained simply by the mechanical properties of the component amino acids [49], simply knowing "who is there and what they're doing" is not enough.

Instead, much like spider's silk, microbial ecosystem function depends on interactions between the building blocks within a hierarchy of nested spatial scales [165, 105].

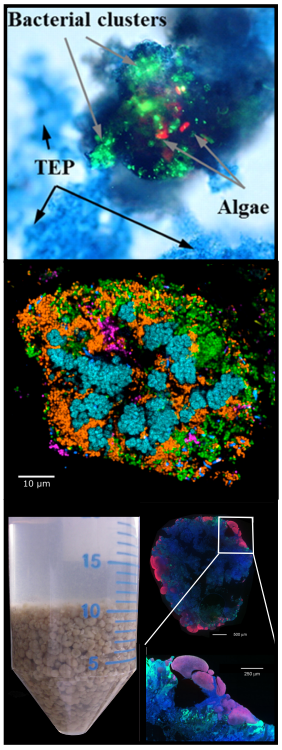
Individual microbes aggregate with others to form dense, diverse local communities (Table 1), where individual cells affect each other through a mixture of short-range (e.g., contact-mediated) and long-range (e.g., small molecule- mediated) interactions [105, 188]. At higher levels of organization, microbes can traverse a landscape of local communities, thereby forming a meta-community of local communities that are coupled by migration between them [113, 80]. Thus, linking the component building blocks of microbial populations to ecosystem-level properties will require (a) determining the factors that influence microbial population structure at these intermediate spatial scales, and (b) probing how emergent phenomena at these (and higher) levels of spatial organization stem from interactions between the building blocks.

In this chapter, I will summarize what we know about the structure and dynamics of wild microbial populations at the local community and meta-community scales. I will then discuss some implications of microbial meta-community structure for ecosystem biodiversity and function. Finally, I will outline the work in this thesis, in which we first characterize wild bacterial populations from the ocean at local community scales, and then use simple laboratory systems to probe the implications of coupling between local communities on meta-community structure and dynamics.

1.2 Microbial meta-communities on patchy landscapes

1.2.1 Local community patches

In the laboratory, microbes are often characterized in monoculture and in the context of a well-mixed test tube. In such an environment, inter-species interactions cannot occur, and all cells are evenly distributed throughout the environment. By contrast, in many natural environments, microbes of diverse taxonomic origins aggregate in local community "patches", either attached to surfaces (e.g., a food particle [177]) or to each other in multicellular flocs (Fig. 1-1). These highly localized community patches sometimes form because the underlying distribution of nutrients in the environment is patchy, with localized, nutrient-rich regions that cells can exploit in an



Environment	Structure	Community length scale	Reference(s)
Aquatic	Organic particles	1-1,000 µm	Kirchman (2010); Bar-Zeev, *et al*. (2012)
Aquatic	Live copepods	1,000-2,000 µm	Tang, *et al*. (2010)
Aquatic	Pink berries	500 µm-1 cm	Wilbanks, *et al*. (2014)
Terrestrial	Soil aggregates	2-2,000 µm	O'Donnell, *et al*. (2007); Franklin and Mills (2007); Raynaud and Nunan (2014)
Terrestrial	Leaf surface structures (e.g., trichomes)	60-100 µm	Esser, *et al*. (2015)
Animal	Colonic crypts	100-500 µm	Donaldson, *et al*. (2015)
Animal	Food particles	50-500 µm	Walker, *et al*. (2008); Van Wey, *et al*. (2011)
Animal	Dental plaques	10-100 µm	Welch, *et al.*, (2016) Valm, *et al.* (2012)
Industrial	Granular activated sludge bioreactors	500 µm-1 cm	Gonzales-Gil and Holliger (2014)

Figure 1-1: Examples of naturally occurring particulate microbial habitats. Besides the case of nutrient particles in aquatic environments, other types of particle structures can be found in a variety of other environments as well. These include (but are not limited to): colonic crypts in the human gut, trichomes and other surface structures on leaves, granules in activated sludge bioreactors, dental plaque, and many others [111, 98, 12, 185, 127, 60, 136, 52, 45, 180, 177, 68]. These particles represent discrete community units that can be individually sampled from the environment.

otherwise nutrient-poor landscape. However, even when the environment is nutrient-rich, many other mechanisms (including inter-species interactions, discussed in more detail below) can favor the formation of dense cellular clusters.

Interestingly, these local community patches are often on the order of 100 μm in diameter (Fig. 1-1) [157]. Thus, within these local patches, cell-to-cell distances are sufficiently short for cells to influence the physiological behaviors of other cells in the same patch through inter-species interactions. For instance, the biofilm matrix produced by one community member can restrict access to resources for other cells in the immediate neighborhood [46, 106]. Longer-range interactions – quorum sensing, antibiotic-mediated killing, and even electrical signals – can also occur within local community patches [105, 188, 131]. Together, these interactions between community members can lead to non-intuitive emergent phenomena at the local community scale that cannot be predicted by individual members [55, 51].

1.2.2 A meta-community of connected local patches

At higher levels of organization, local community patches are coupled by migration between them, thus forming a microbial meta-community [113, 80]. Indeed, in a given environment, there are typically many such patches, with coupling between them as a result of cellular dispersal from patch to patch (Table 1). In many environments, dispersal from these local community patches, especially surface-attached (biofilm) aggregates, is a common occurrence [113]. However, the extent to which dispersal occurs (and hence, the degree of coupling between local communities) may differ substantially. In some cases, dispersal occurs through passive mechanisms (e.g., shear flow in aquatic environments [158] or through the air in terrestrial environments [1]). In other cases, dispersal is active. For instance, in aquatic environments, some bacteria use chemotaxis to swim in a directed manner towards nutrient-rich patches, as well as other chemical stimuli [158, 159, 152]. Moreover, the mechanisms of dispersal from biofilms can be complex, involving a multitude of environmental cues, signal transduction pathways, and quorum sensing (a form of cell density-dependent signaling) [113]. Although many of these mechanisms have been studied in laboratory

settings, future work should be aimed at characterizing the mechanisms and rates of dispersal in complex, naturally occurring microbial communities.

1.2.3 A case study: Bacterial communities on marine particulate organic matter

One of the most well studied examples of local community patches are those formed by bacteria on organic particles in the marine environment. In oceans, lakes, and other aquatic environments, organic particles – ranging from decaying crustaceans to fish fecal pellets to polysaccharide gels – serve as nutrient-rich scaffolds for microbial communities (Fig. 1-1) [11]. Microbes from the surrounding water flock to these particles and assemble into dense multi-species consortia that consume and recycle particle resources before they sink out of zones of high productivity in the ocean [11, 152]. The assembly of the local communities on particles is shaped by the interplay between the cell behavior and ecological interactions. Traits such as swimming speed, chemotaxis, and surface attachment control the order of arrival of organisms to a particle, as well as their residence time [152, 190, 158]. At the same time, ecological interactions such as quorum sensing [70, 86], chemical antagonism [108], and exploitation of public goods [33] inhibit or facilitate growth. By modulating the abundance of particle-degrading bacteria and their exposure to particle surfaces, local interactions on POM can thus control the rates of particle degradation and biomass production, and consequently, the rates of carbon remineralization in the water column [11, 89].

1.3 Consequences of microbial meta-community structure for biodiversity and ecosystem function

The complex, spatially structured landscapes that microbes inhabit in nature can have a profound influence on population structure, dynamics, biodiversity, and function at larger spatial scales [80, 122]. How local migration affects meta-community population dynamics is the subject of metapopulation theory, a well-developed body of work in

the macro-ecological literature. I have summarized a small number of results below. However, for a comprehensive review of the field, see [80]).

1.3.1 Predicted consequences for biodiversity

Spatially extended populations, like those of a microbial meta-community, are predicted to support a higher level of biodiversity than a well-mixed, spatially uniform environment. In principle, this can occur through a variety of mechanisms. One of the most common examples is that of mutually antagonistic colonizers. In simple population dynamics models, one species excludes the other in any individual local community, and the winner can depend on something as simple as the order of arrival. However, these species will co-exist globally within the meta-community, that is, across many such patches [65]. Thus, the presence of spatially isolated local communities may allow for co-existence between competitors that would exclude each other in a well-mixed environment.

Dispersal between local communities is also predicted to support biodiversity at the meta-community scale. For instance, if a species goes extinct in a local community, it may be "rescued" by migrants from a neighboring community, which can re-seed the species [104, 80]. To our knowledge, this type of rescue effect has not been observed in wild microbial populations, but has been observed in some plant and animal populations [81, 80] and in simple laboratory ecosystems [25, 107].

On evolutionary timescales, spatially structured environments can also allow species with similar resource preferences to co-exist through life history trade-offs. In one of the most common of these trade-offs, referred to as a competition-dispersal trade-off, organisms are either specialized for exploiting local nutrient patches or, alternatively, for readily dispersing between patches [87, 190]. Thus, spatially structured populations can support many taxa that may not co-exist otherwise through fitness trade-offs on local versus global scales.

1.3.2 Predicted consequences for ecosystem function

Within local community aggregates, microbes in close physical proximity may have complementary metabolic repertoires that improve their combined functional productivity (compared to their productivity in isolation). Why certain taxa lose the ability to perform particular metabolic functions – thereby creating dependencies on other organisms – remains a key question in microbial evolution. However, in the most well known cases of metabolic complementarity, populations consume the metabolic waste products of others, often as electron donor or acceptors in anaerobic environments [114, 121].

Genomic studies suggest that metabolic complementation plays a crucial role in natural microbial communities. For example, Zelezniak, *et al.* showed computationally that there is a general trend for locally co-occurring populations to be enriched in metabolic complementarities, suggesting that interactions among micro-organisms are common and likely to emerge from pairing of incomplete or complementary metabolic pathways [191]. In a separate experimental study in methanogenic communities, Embree, *et al.* showed that amino acid auxotrophies create interdependencies between populations that control energy flux and contribute to community robustness [51]. These studies suggest that interactions through metabolic complementarities are common in nature and can have a large impact on community function.

1.4 Experimental approaches for studying microbial communities and meta-communities

Despite the importance of meta-community structure for microbial biodiversity and function, studies of wild microbial populations at local community and meta-community scales have been limited. Some of these limitations stem merely from historical precedent, while others stem from shortcomings in existing technologies. I have summarized some of the experimental approaches currently used to study microbial populations in nature at local community and meta-community scales.

1.4.1 Characterizing spatially isolated, naturally occurring local microbial communities

Individual microbes are small – no more than several microns in length. Therefore, the local communities that they form are also small (often roughly 100 μm in length) [60, 157, 21]. However, typical sampling methods target much larger spatial scales (for instance, a bucket of seawater or a scoop of soil), mostly for convenience. Sampling at "bucket scales" averages over variability across many local communities, making it impossible to identify the underlying drivers of local community structure in wild microbial populations. Nonetheless, some naturally occurring local microbial communities can be sampled individually (Fig. 1-1). Further work should be aimed at sampling and characterizing such structures with high replication in natural ecosystems.

In cases where local community aggregates have been sampled individually, several methods exist that allow us to visualize "who sits next to whom" in these aggregates. The most well known among these is fluorescence *in situ* hybridization (FISH), in which fluorescent probes bind to specific microbial sequences and are visualized via microscopy. Combining FISH-based techniques with many modes of microscopy, researchers have characterized the microscale spatial structure of microbial communities in a diverse range of ecosystems, including the oral microbiome [111], the mammalian intestine [48], in soil [15], and on marine snow [169]. Furthermore, FISH has been combined with mass spectrometry-based techniques, including nanoscale secondary ion mass spectrometry (NanoSIMS), to identify the metabolic roles of individual cells within complex microbial consortia, ranging from those living on symbiont-bearing coral polyps [183] to mouse intestines [14]. Recently, FISH was used with NanoSIMS to identify a syntrophic coupling based upon direct electron transfers between aggregates of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments [114]. Overall, using these techniques, we can characterize the physical structure of a microbial community at the microscale and with single-cell resolution.

1.4.2 Microfluidic approaches for studying microbial communities under defined physicochemical conditions

Biologically inspired microfluidic systems, including soil-on-a-chip [155], gut-on-a-chip [94, 16], and coral-on-a-chip [148], have been used successfully to study the microscale structure and dynamics of microbial communities. These systems provide precise control of the patch microhabitat (e.g., nutrient concentration, temperature, pH, and fluid flow), thereby producing defined patches for microbes to colonize. Furthermore, the colonization process can be visualized in real-time by coupling microfluidic devices with microscopy [155, 94, 16, 148]. For example, a recent study used a coral-on-a-chip system to visualize the dynamical process by which a coral polyp is infected by a known coral pathogen (*Vibrio coralliilyticus*) with a level of spatiotemporal resolution that could not be achieved in a natural marine ecosystem [148]. Altogether, microfluidic devices are a powerful addition to the microbial ecology toolbox, but to date, have not been extended to microbial communities as diverse as those found in nature.

1.4.3 Characterizing dispersal between patches in microbial populations

Biologically inspired microfluidic devices have also been used to characterize how microbes from the environment differ in their dispersal characteristics, which influences their ability to move from patch to patch [154]. This approach has been used to differentiate marine bacterial populations based upon their dispersal abilities, particularly rates of chemotaxis towards nutrient-rich substrates [190, 154]. Moreover, by pairing microfluidic devices with single-cell optical tracking, researchers can quantify the swimming trajectories of many single cells within a population. For instance, in one recent example, researchers used a microfluidic device in conjunction with time-lapse microscopy to track how individual bacteria from a natural marine bacterial assemblage formed clusters around decaying copepods, fecal pellets, and other natural marine detritus [152]. Thus, microfluidic devices can be used to characterize

variation in dispersal, even in complex microbial assemblages.

Linking the physiological characteristics of individual cells to dynamics at the level of the meta-community is often impossible in wild microbial populations. Thus, simplified laboratory systems can complement studies of microbial populations *in situ* by offering a high level of control over the species present, the growth environment, and migration parameters.

One commonly used experimental system is based upon a linear stepping-stone model (also known as the Levins metapopulation model) and simulates short-range dispersal between discrete, well-mixed subpopulations on a lattice [41, 93, 92, 38]. Briefly, a "meta-population" landscape is simulated on a one-dimensional lattice (one row of a 96-well culture plate, for example), where each well contains a well-mixed population of cells growing in identical growth media. After cells are inoculated into some portion of the wells, migration is simulated by pipetting cells from one well into another. Although these model systems are highly simplified representations of the natural environment, they have been used demonstrate possible effects of local dispersal on biodiversity [41, 93], host-phage co-evolution [92], and spatiotemporal signals before population collapse [38].

1.5 Goals of this thesis

In this thesis, we have characterized microbial populations within their local communities and meta-communities (described below), drawing on methods from traditional microbiology, high-throughput sequencing, statistics, and mathematical modeling. Note that I performed this work in collaboration with a number of co-authors, each of whom provided significant insights regarding the ideas and data described in this thesis.

In Chapter 2, **we considered how local microbial communities are structured on individual, naturally occurring patches**, using small crustaceans (*C. finmarchicus* copepods) as model nutrient patches. We found that copepod individuals harbor a common "core" set of bacterial taxa, but also have a significant

"flexible" component that differs from individual to individual. This structuring of local communities on individual copepod "patches" is partially controlled by physiological variability across individual hosts, but is also influenced by local inter-taxon interactions.

Next, in Chapter 3, **we explored the underlying community assembly dynamics that might give rise to local communities in natural environments**. To this end, we developed a semi-synthetic model system, which allowed us to study the dynamical process by which bacteria from natural seawater self-assemble into communities on defined particulate nutrient patches. We found that local interactions between particle-associated taxa drive rapid, reproducible ecological successions on these patches.

Finally, in Chapter 4, **we asked how cooperative interactions within local communities influence biodiversity at the level of the meta-community**. Using a simple laboratory ecosystem, we found that, even if cooperative alleles are weakly favored in an individual local community, they are strongly favored at the front of an expanding meta-community. This leads the overall prevalence of cooperative alleles within the meta-community to be substantially higher than what would be predicted in a well-mixed community without spatial structure.

Chapter 2

Inter-individual variability in the copepod microbiome

This work was done in collaboration with Amalia Aruda Almada, who made equal contributions to the manuscript.

Abstract

Copepods harbor diverse bacterial communities that play important roles in biogeochemical transformations in the marine environment. However, the factors that dictate the structure of copepod-associated bacterial communities have not been characterized, particularly at the level of copepod individuals. Here, we characterized the bacterial communities associated with nearly two hundred individual copepods from the species *Calanus finmarchicus*, an ecologically and biogeochemically important marine invertebrate in the North Atlantic, and we have examined how physiological changes associated with the transition from active growth to diapause may drive changes in their associated bacterial communities. Our findings demonstrated that while individual copepods share a common "core" microbiome, most bacterial taxa were patchily distributed across individual copepods. However, bacterial taxa were not simply randomly distributed, but instead, formed discrete clusters of taxa whose members were highly correlated in their distributions across individual copepods. The members of these clusters were typically strongly associated with common drivers, ranging from the presence of food in the gut to the abundance of other co-occurring bacterial clusters. Altogether, our findings reaffirm the hypothesis that copepods represent selective niches for bacterial communities and provide further insight into the potential factors that drive the composition of these communities.

2.1 Introduction

Calanoid copepods are major players in the marine ecosystem, serving both as lynchpins of aquatic food webs [172] and vehicles of vertical carbon transport [172, 90]. However, copepods also exert a more indirect influence, each acting as a selective, nutrient-rich habitat for tens of thousands of bacteria [153, 117, 163], which themselves initiate major chemical transformations [164, 145, 17, 173, 18]. While no more than a few millimeters in length, copepods are the most abundant metazoans in the ocean [179]. Thus, the bacterial communities that inhabit copepods, distinct from those "free-living" in ambient seawater [117, 43], likely exert a fundamental control on biogeochemical transformations in the marine environment. Nonetheless, what controls the composition of copepod-associated bacterial communities, including host physiology, remains poorly understood.

Living copepods present a dynamic habitat whose properties may depend strongly on host physiology [163]. When copepods feed, they amass food particles in their guts, thus presenting a nutrient-rich (albeit anoxic) environment compared to the surrounding seawater [161, 164]. Moreover, as copepods feed and defecate, they release nutrients into their surroundings, which bacteria can then exploit [119, 163]. Accordingly, bacteria are not uniformly distributed over the copepod surface, but instead, are often concentrated at the gut, mouth, and anus [26, 117]. The frequency with which copepods molt can also affect their associated bacterial communities by modulating the rate of community turnover [163]. Additionally, as copepods migrate vertically through the water column, their associated bacteria travel with them, further shaping the environment to which they are exposed [74, 163]. Furthermore, association with copepods is known to have dramatic impacts on the proliferation, virulence, and physiology of many bacterial species, including several pathogens. Metazoan hosts can often exert a strong selective force on associated microbiota, which is reflected in the frequent observation of species-specific and even host site-specific microbiomes.

Altogether, copepod physiology is likely to have a profound influence on their associated bacterial flora. However, given that the physiological state of a copepod

could vary drastically from individual to individual, the standard method of bulk sampling of copepods could lead to physiological averaging that would mask these effects. Individual sampling of copepods also offers the unique opportunity to study co-occurrence patterns among bacterial taxa within the communities, potentially allowing us to identify robust statistical associations between taxa driven by common niche preferences, interspecific dependencies, and other factors.

One of the most dramatic physiological transitions that copepods in the family Calanidae undergo is the transition from active growth to a diapausing (dormant) state. Each year, an enormous number of *Calanus finmarchicus* juveniles – tens of thousands per square meter in the North Atlantic [82, 67, 90] – enter diapause. This transition coincides with a number of physiological changes, including cessation of feeding [84, 85], delayed molting [116, 8], pH changes and accumulation of ammonium in the hemolymph [144, 147], differential gene expression [166, 9, 167], and lipid accumulation in an enlarged oil sac [84, 115]. Importantly, diapausing copepods migrate vertically from the surface to deep ocean basins, bringing with them vast amounts of carbon in the form of lipid stores that is eventually sequestered in the deep ocean [84, 90]. This process (the so-called "lipid pump") is a vital determinant of global carbon fluxes in the ocean [90]. Thus, the bacterial communities associated with diapausing copepods may further play a role in mediating carbon fluxes from the lipid pump, much like the microbe-mediated "biological pump" [89]. Nonetheless, to date, how copepod-associated bacterial communities change during a copepod's transition to diapause has not been characterized.

The goals of this study were (a) to characterize variability in copepod-associated bacterial communities across many individual copepods, and (b) to assess the potential drivers underlying this variability, including host physiology. To this end, we collected 189 individual copepods (*C. finmarchicus* copepodites, stage C5 from Trondheimfjord, Norway) on two separate dates during the early summer (June 6, 2012 and June 11, 2012). At this time of year, C5 copepods have typically just begun their descent from the surface ocean, and thus, are in the early stages of the transition from active growth to a diapausing (dormant) state (Methods; Appendix A). On

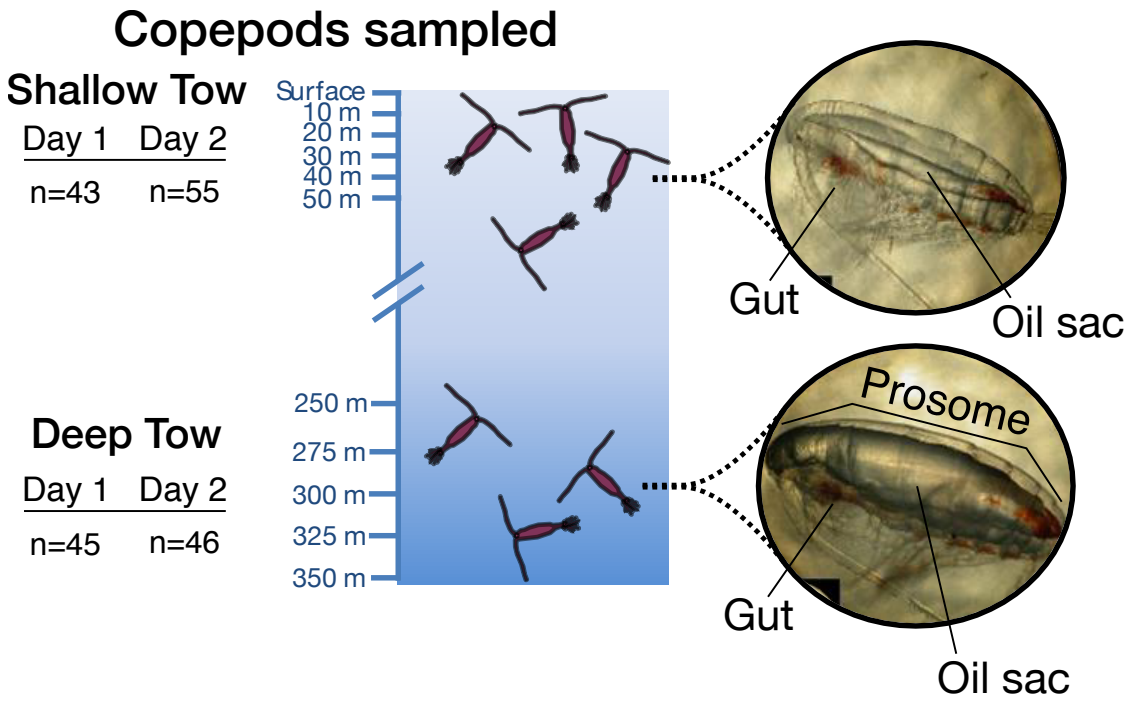


Figure 2-1: Sampling of individual copepods from the North Atlantic. Individual copepods were collected over two sampling dates at Tröllet Station near Trondheimfjord, Norway (189 individuals in total; *C. finmarchicus* C5 copepodids). To capture individuals in a wide range of physiological states, copepods were collected from surface (0-50 meter depths) and deep (250-350 meter depths) seawater. At the surface, copepods were largely in an active state, while at increased depths, copepods were often in diapause (dormant). Among copepod individuals, physiological differences were manifested by changes in morphology, including differences in prosome (body) volume, oil sac size, and the presence of food in the gut.

each of these dates, we sampled copepods inhabiting two depth strata (shallow, 0-50 meters, and deep, 250-340 meters below the surface), thereby allowing us to sample a broad physiological gradient between actively growing and diapausing states (Fig. 2-1). Accordingly, for each individual, we quantified several morphological characteristics that mark the physiological transition between active growth and diapause. We also characterized the composition of the copepod-associated bacterial community via 16S rRNA amplicon sequencing.

2.2 Results and discussion

2.2.1 Morphological and bacterial community variability across individual copepods

Copepods collected from shallow and deep seawater displayed the morphological hallmarks of active growth and diapause, respectively, although copepods at each depth exhibited substantial inter-individual variability. Consistent with previous studies [166, 9], deep-dwelling copepods had significantly larger body volumes (Fig. 2-2a, $p = 6 \times 10^{-19}$, two-sided Mann-Whitney U test) and fuller oil sacs (Fig. 2-2b, $p = 5 \times 10^{-16}$, two-sided Mann-Whitney U test) on average than their shallow-dwelling counterparts. Deep-dwelling copepods were also significantly less likely to have food in their guts (Fig. 2-2c, $p = 4 \times 10^{-8}$, Fisher's exact test) and, on average, harbored fewer bacterial cells (Fig. 2-2d, $p = 6 \times 10^{-5}$, two-sided Mann-Whitney U test) than those from shallow seawater. However, even copepods sampled from the same depth varied substantially in these morphological characteristics; for each characteristic, the variation (standard deviation) at a single depth was comparable to the difference in means between depths (Table S1). Thus, we hypothesized that this morphological variability between copepod individuals – potentially indicative of underlying physiological differences – might lead to variability in their associated bacterial communities.

Consistent with this hypothesis, individual copepods varied substantially in the bacterial communities they harbored. In analyzing community composition, we chose to focus on the 241 bacterial taxa whose mean abundance was above a defined threshold ($f_{min} = 2 \times 10^{-4}$) on both sampling dates (Materials and Methods; Fig. 2-3), which account for 90% of all copepod-associated sequenced reads. Among this taxon subset, over 85% (207 out of 241) were present on less than 90% copepod individuals, suggesting that these taxa were patchily distributed across individual copepods. This so-called "flexible" part of the copepod microbiome (Methods) was taxonomically diverse, with representatives from Gammaproteobacterial, Alphaproteobacterial, and

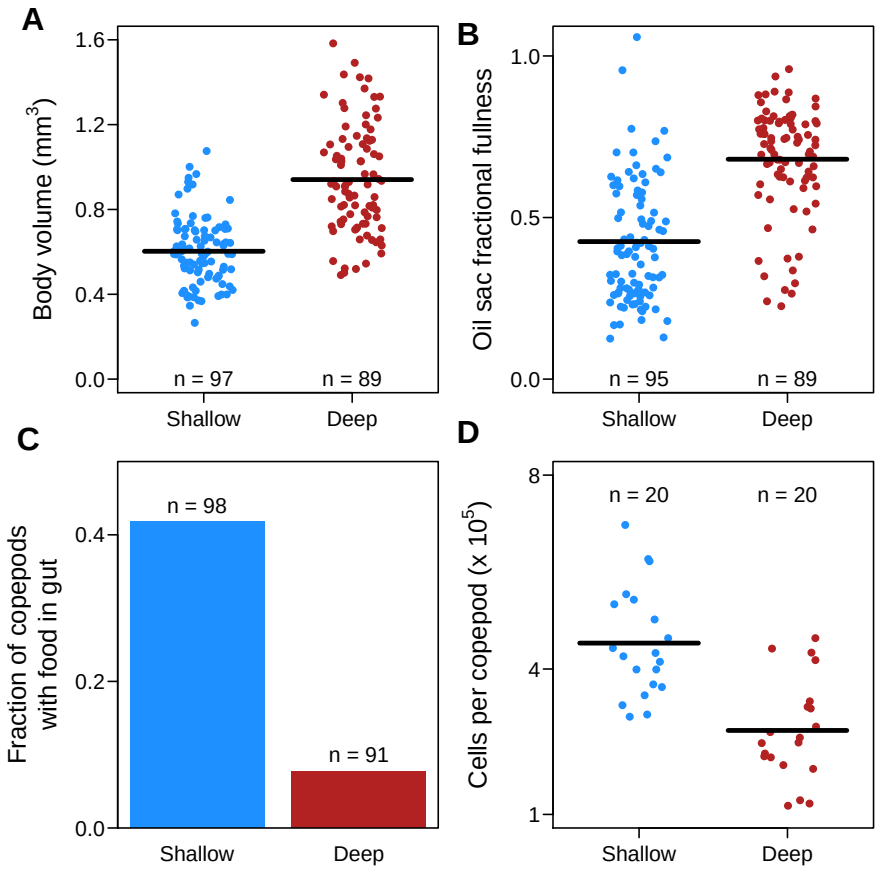


Figure 2-2: Quantification of morphological characteristics. Distribution of morphological characteristics for many individual copepods across shallow- and deep-dwelling populations. A roughly equal number of copepods (≈ 100) were sampled on two different sampling dates. (a) Prosome (body) volume. (b) Volume of oil sac. (c) Fraction of copepods with food present in the gut. (d) Number of bacterial cells per copepod. Procedures for quantifying each of these characteristics are described in Methods.

Flavobacteriial lineages.

Interestingly, Flavobacteriaceae, which are commonly found associated with copepods, were never detected on more than 90% of the copepod individuals, but represented 24 out of the 207 taxa in the flexible microbiome (Fig. 2-4). Thus, Flavobacteriaceae may not be persistent members of the *C. finmarchicus* microbiome. Similarly, previous studies have identified members of this clade as transient, potentially intestine-associated taxa [163, 118], although the biological drivers of their distribution have not been fully determined.

Despite significant variability in the bacterial communities associated with individual copepods, we also identified a set of "core" bacterial taxa, each of which was detected on more than 90% of the copepods sampled (Fig. 2-4). This core set comprised 34 taxa (out of the 241 above our abundance threshold, Fig. 2-3) and represented a range of diverse bacterial phyla, including Actinobacteria, Bacteroidetes, and Proteobacteria. However, 25 of the 34 core taxa were Proteobacteria, with 10 each being Gammaproteobacteria and Alphaproteobacteria. Of the gammaproteobacterial lineages, eight were Moraxellaceae or Vibrionaceae, both of which are bacterial families whose associations with calanoid copepods in the North Atlantic have been documented previously [43, 118]. Together, this core set of bacterial taxa comprised a highly variable fraction of the community from individual to individual (range: 1.5%-93%), but together were rarely the numerically dominant; in only 22 out of 189 copepod individuals did they account for more than 50% of the community.

How universal is this core microbiome? In previous studies, the composition of copepod-associated bacterial communities has been linked to host-specific environmental factors (e.g., copepod diet [163, 118], oxygen availability [18], or pH [164]), as well as external abiotic factors (e.g., salinity [47]). However, while one might expect this core microbiome to be species- and/or environment-specific as has been observed in other invertebrates [61], the existing literature suggests that the core microbiome we observed is not unique to *Calanus finmarchicus*. First, the bacterial taxa that we identified as most abundant in the *C. finmarchicus* core microbiome are also commonly observed in several other North Atlantic species of calanoid copepods (i.e., *Acartia*

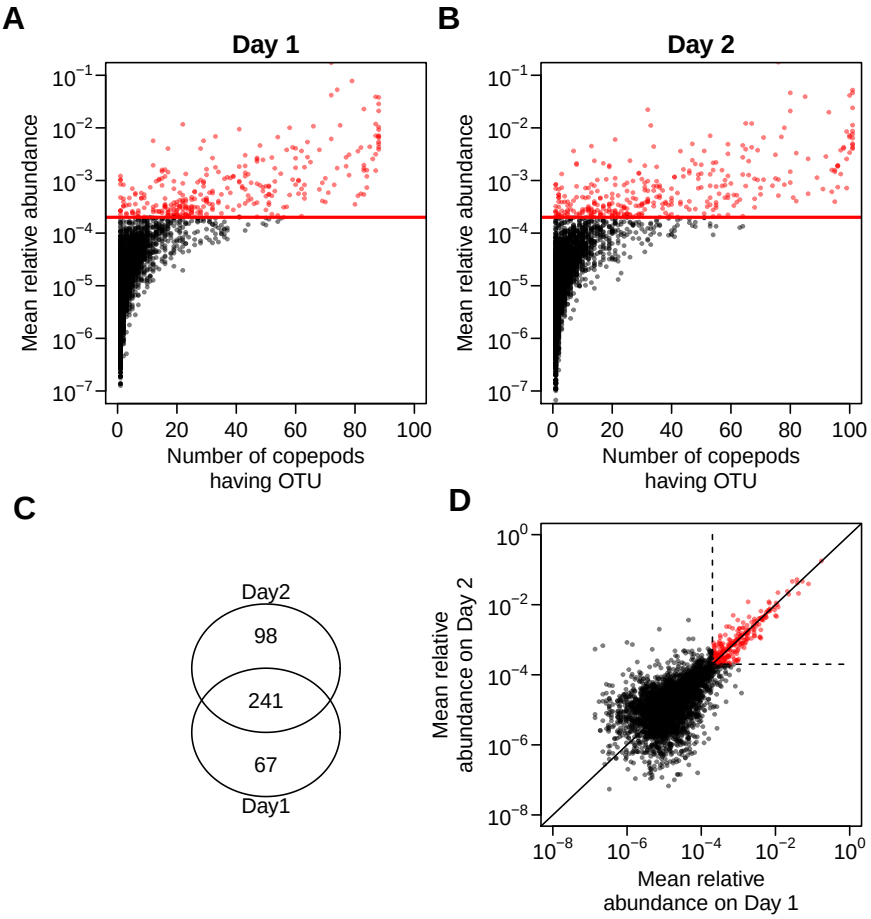


Figure 2-3: Defining OTU subset based on mean relative abundance threshold. (a) For the first sampling date (June 6, 2012), the mean relative abundance of each OTU versus the number of copepods in which that OTU was present at non-zero abundance. The red solid line indicates the mean relative abundance threshold that we imposed (2×10^{-4}). Points in red correspond to OTUs with a mean relative abundance above the threshold. (b) For the second sampling date (June 11, 2012), the mean relative abundance of each OTU versus the number of copepods in which that OTU was present at non-zero abundance. The red solid line indicates the mean relative abundance threshold that we imposed (2×10^{-4}). Points in red correspond to OTUs with a mean relative abundance above the threshold. (c) Venn diagram of the number of OTUs for each sampling date whose mean relative abundance was above the threshold. (d) OTU mean relative abundance on the first sampling date versus the second sampling date. Points in red correspond to OTUs with a mean relative abundance above the threshold on both sampling dates.

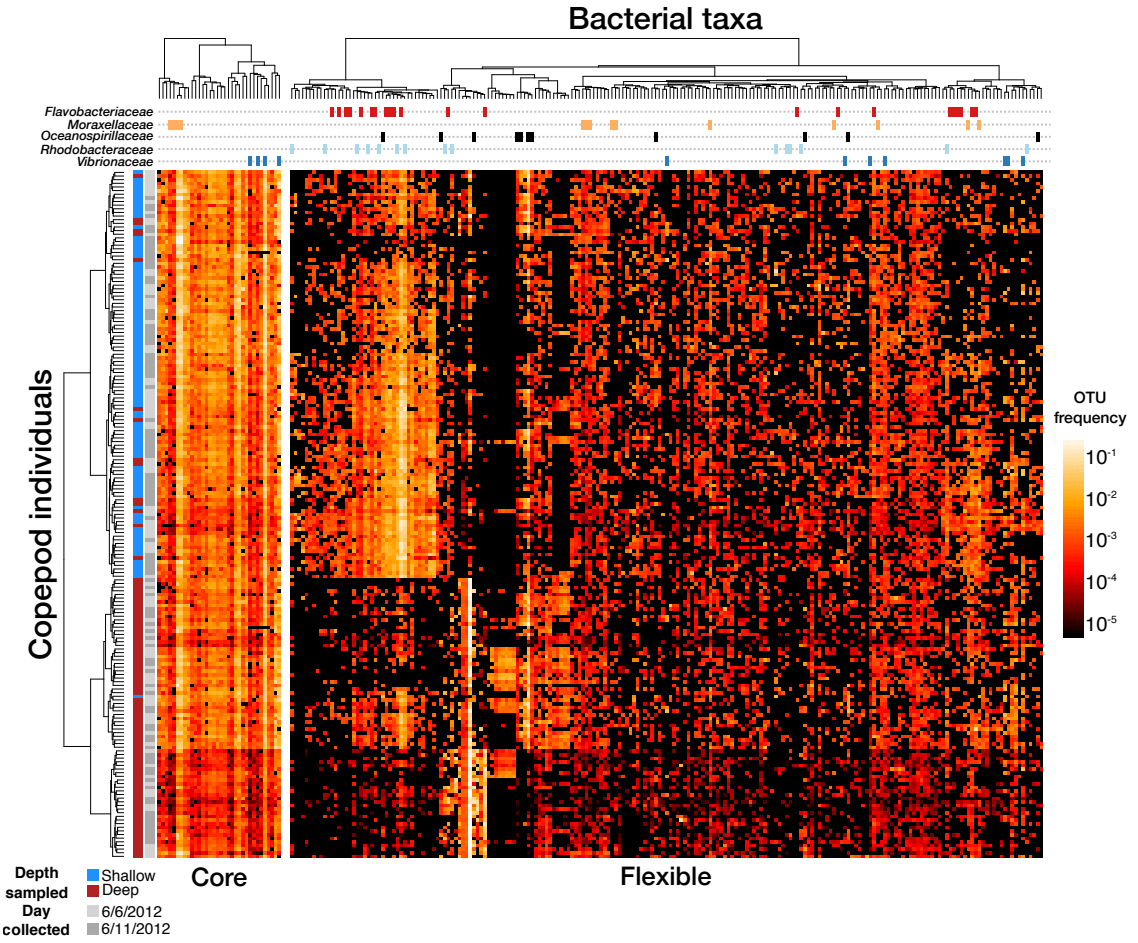


Figure 2-4: Physiologically diverse copepods share a common "core" microbiome, but also have a distinct flexible component. Compositional heat map of a subset of bacterial taxa (mean abundance $> 2 \times 10^{-4}$) present over nearly 200 copepod individuals. The left heat map shows the "core" microbiome (taxa present in >90% of copepod-associated communities). The right heat map shows the "flexible" microbiome (taxa present in <90% of the communities). In both heat maps, taxa are clustered by their correlation (distance metric: log-transformed Pearson correlation as estimated with SparCC; clustered with Ward's minimum variance method). Copepods are clustered by the overall similarity of their communities across core and flexible components (distance metric: Euclidean distance between log-transformed relative abundances; clustered with Ward's minimum variance method). The leftmost color-coded column indicates the depth at which an individual copepod was collected (blue, shallow; red, deep). The second from left color-coded column indicates the date on which the sample was collected (light gray, 6/6/2012; dark gray, 6/11/2012). The topmost lines indicate which taxa fall into one of the five most commonly observed taxonomic families.

sp., *Temora longicornis*, *Centropages* sp., and *Calanus helgolandis*) [43, 66, 149, 117]. The copepod species listed above span a wide range of feeding lifestyles, including herbivorous, omnivorous, and detrivorous lifestyles, suggesting that feeding alone may not dictate the composition of copepod-associated bacterial communities. Moreover, in a different aquatic environment (the Gulf of Maine), these same taxa are also observed to colonize multiple calanoid copepod species [118]. Together, this suggests that there may be a small core set of bacterial taxa that are well adapted to generic features of the niche provided by an individual copepod.

2.2.2 Identifying ecological factors shaping copepod-associated bacterial communities

Overall, we found that the vertical depth at which a copepod was sampled was a strong, but most likely indirect, predictor of the composition of its associated bacterial community. In particular, when individual copepods were clustered by the similarity (Spearman correlation) of their associated bacterial communities, they largely grouped by the vertical depth at which they were collected (Fig. 2-4). Furthermore, using PERMANOVA with copepod vertical sampling depth as the only predictor, we showed that depth also accounted for a significant portion of the variability dissimilarity between their bacterial communities (Table S2A, $R^2 = 0.32$, $p < 10^{-6}$, bivariate PERMANOVA with 1,000,000 permutations). Differences in bacterial composition between shallow and deep seawater could, in theory, account for variability in community composition between shallow- and deep-dwelling copepods, as copepods are colonized by bacteria from the ambient seawater [74] However, the bacterial communities from shallow-dwelling copepods were no more similar to shallow seawater than to deep seawater (and likewise for deep-dwelling copepods), suggesting that ambient seawater composition is not a direct differentiating factor (Fig. 2-5).

Instead, our data supported an alternative explanation: that host morphological characteristics, which co-vary with depth, account for depth's effects on copepod-associated bacterial communities. Indeed, while depth by itself yielded a PER-

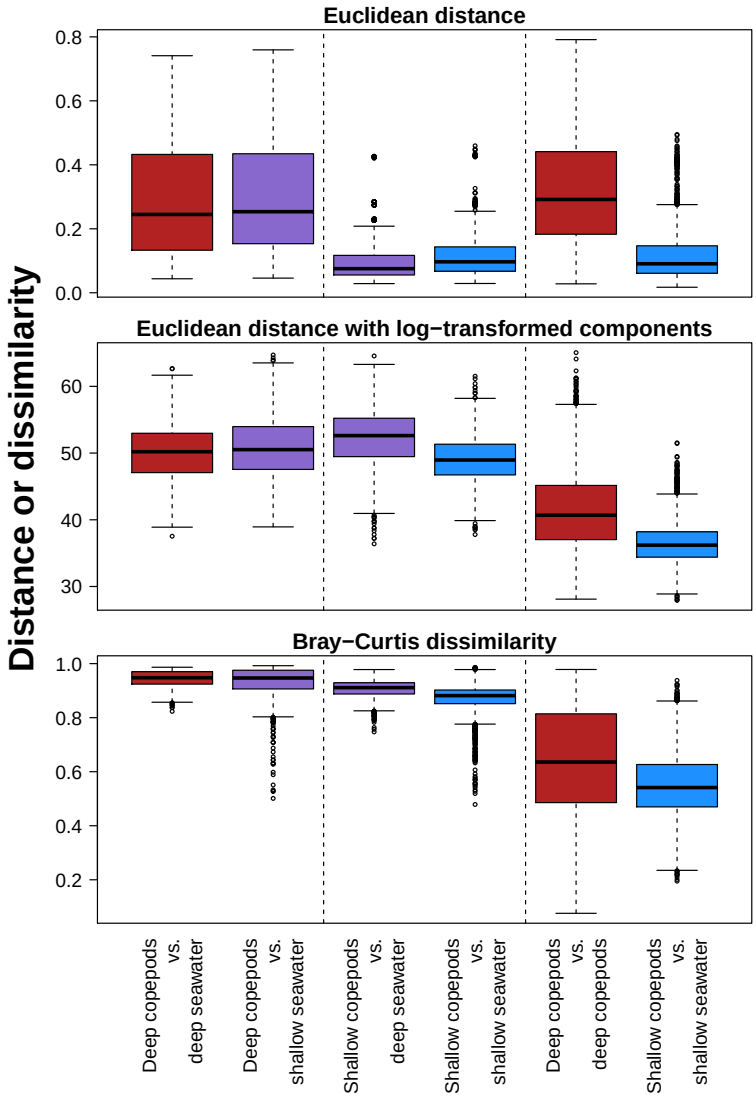


Figure 2-5: Similarity between copepods and the ambient seawater from which they were collected. The distance between copepod-associated community composition and seawater composition was assessed for each individual copepod (189 in total) with one of three distance/dissimilarity metrics. Each box-and-whisker plot describes the distribution of distances/dissimilarities for copepods of a particular type versus the ambient seawater type (deep or shallow) indicated.

MANOVA R^2 of 0.32, we found that, when controlling for copepod physiology via the morphological proxies that we measured (copepod vertical sampling depth, body volume, oil sac fullness, the presence of food in the gut), the R^2 for depth was reduced by 90%, though it remained statistically significant (Table S2B). Therefore, differences in individual-specific physiological characteristics shape the bacterial niche, and thus, they are likely to be important drivers of bacterial community composition. In particular, the specific habitat provided by an individual copepod, including its feeding history, may shape local selective pressures [161, 163], thereby fostering significant inter-individual variability in copepod-associated bacterial community composition. Moreover, on each copepod individual, bacteria exist at much higher cell densities than in the ambient seawater [163], suggesting that local interactions between bacterial taxa may also influence bacterial community composition.

To quantify the effects of the copepod host and co-occurring bacterial taxa on copepod-associated bacterial community composition, we took a two-step approach. First, we characterized the network of correlations between individual bacterial taxa, which allowed us to identify clusters of taxa whose distribution across individual copepods was strongly positively correlated. Second, using a multivariate linear regression approach (Methods; Appendix A), we identified putative associations linking these clusters to each other and to the morphological characteristics that we quantified for individual copepods. Importantly, this approach allowed us to quantify the effects of all variables independently, while controlling for the effects of the other variables.

Overall, we identified seven clusters of bacterial taxa, each of which contain taxa with strong correlations across individual copepods. To identify this structure, we used SparCC to estimate the Pearson correlation of log-transformed taxon abundances across all individual copepods, focusing on the 241 bacterial taxa whose mean abundance was above our pre-defined threshold (Fig. 2-3; Methods). Of these taxa, most (177 out of 241) were negligibly correlated with any of the other taxa. However, the remaining 64 taxa grouped into seven clusters (Fig. 2-6). Within clusters, pairs of taxa were typically highly positively correlated, while taxa from different clusters were often negatively correlated. The magnitudes of these correlations were consistent

across sampling dates (correlation=0.63, Fig. S4), suggesting that these associations are biologically reproducible. Notably, these clusters differed in their level of phylogenetic coherence: some clusters (e.g., Cluster 6, Fig. 2-6) only contained bacterial taxa from a single taxonomic class (Gammaproteobacteria), while others (e.g., Cluster 2, Fig. 2-6) contained representatives from several disparate classes (Actinobacteria, Alphaprotebacteria, and others).

Given these strongly correlated clusters of bacterial taxa, we next sought to identify the ecological factors that may influence their distributions across individual copepods. The compositional nature of the bacterial community data precluded the prediction of cluster relative abundances directly [2]. Instead, we applied a standard additive log-ratio transform, with which we quantified the log-transformed ratio of the abundance of each cluster relative to the abundance of the core microbiome for each copepod [2](Methods; Appendix A). We then used multivariate regression models to predict each cluster's log ratio-transformed abundance with copepod-specific morphological characteristics, as well as the abundances of other clusters (Fig. 2-7). Importantly, rather than identifying associations conditioned on a particular regression model, we used Bayesian model averaging to calculate the probability that a predictor had a non-zero effect over all models containing all possible subsets of the predictors [133]. Using this approach, we identified several significant associations, including those between bacterial clusters and copepod morphology, as well as those between pairs of bacterial clusters. These associations allowed us to identify possible ecological factors underlying fine-scale differentiation in copepod-associated bacterial communities at the level of individual copepod hosts.

2.2.3 Factors influencing bacterial cluster abundances within the copepod microbiome

We first identified a positive association between Cluster 2 – a cluster that was enriched among shallow-dwelling copepods compared to their deep-dwelling counterparts (Fig. 2-4) – and the presence of food in the copepod gut (Fig. 2-7). Given

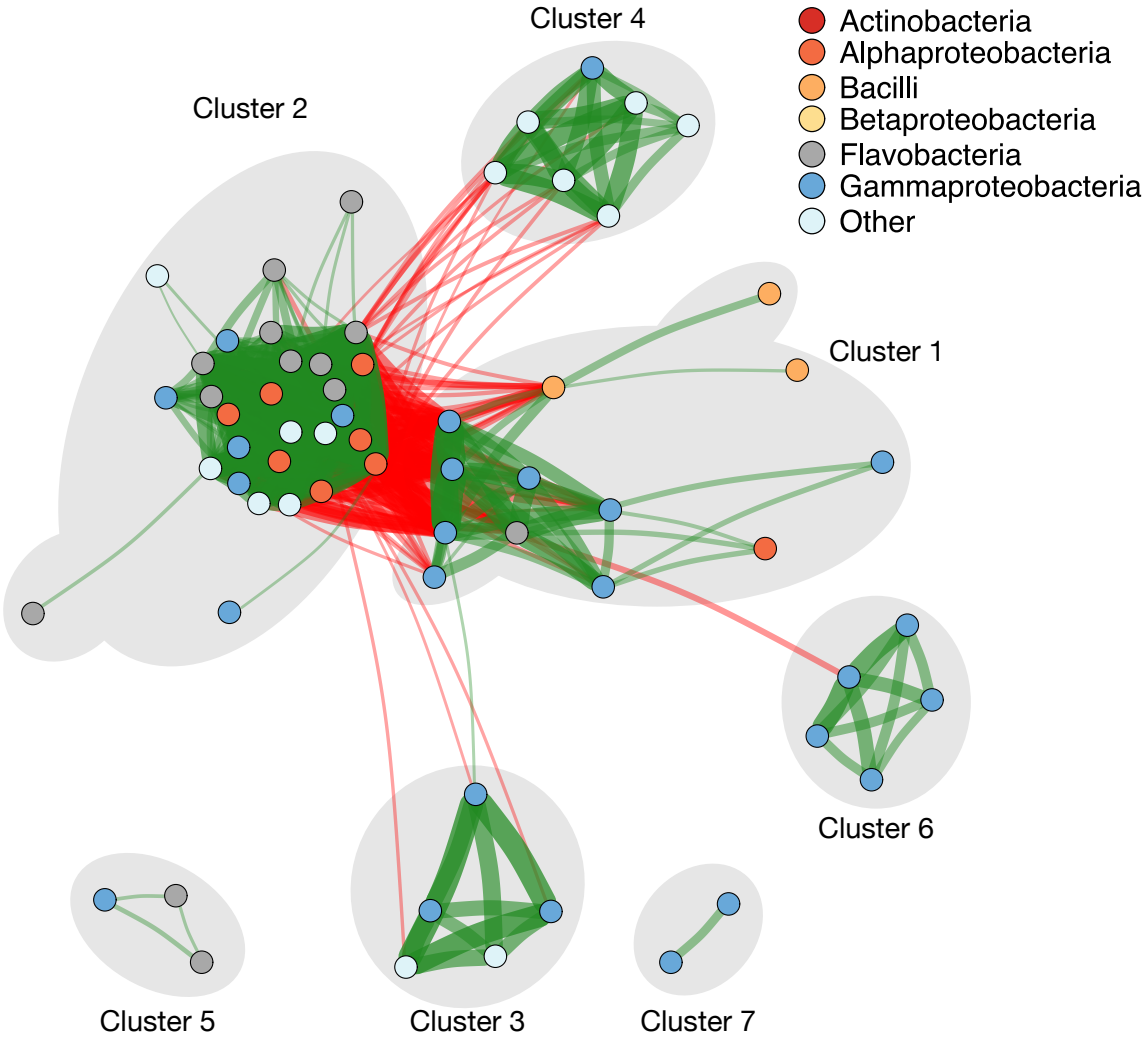


Figure 2-6: Across hundreds of copepod-associated communities, bacteria form a small number of highly positively correlated clusters. Nodes represent individual OTUs and are colored by taxonomy at the class level. Edges indicate positive (green) or negative (red) correlations between bacterial taxa as approximated by SparCC. The magnitude of the correlation is proportional to the width of the edge. Nodes for which all inter-taxon correlations are negligible ($|\rho| < 0.35$) are not shown. Nodes are arranged according to a Fruchterman-Reingold force-directed graph drawing algorithm [64]. The positions of some nodes have been adjusted for disambiguation of cluster designations. Clusters of highly positively correlated taxa are outlined in gray.

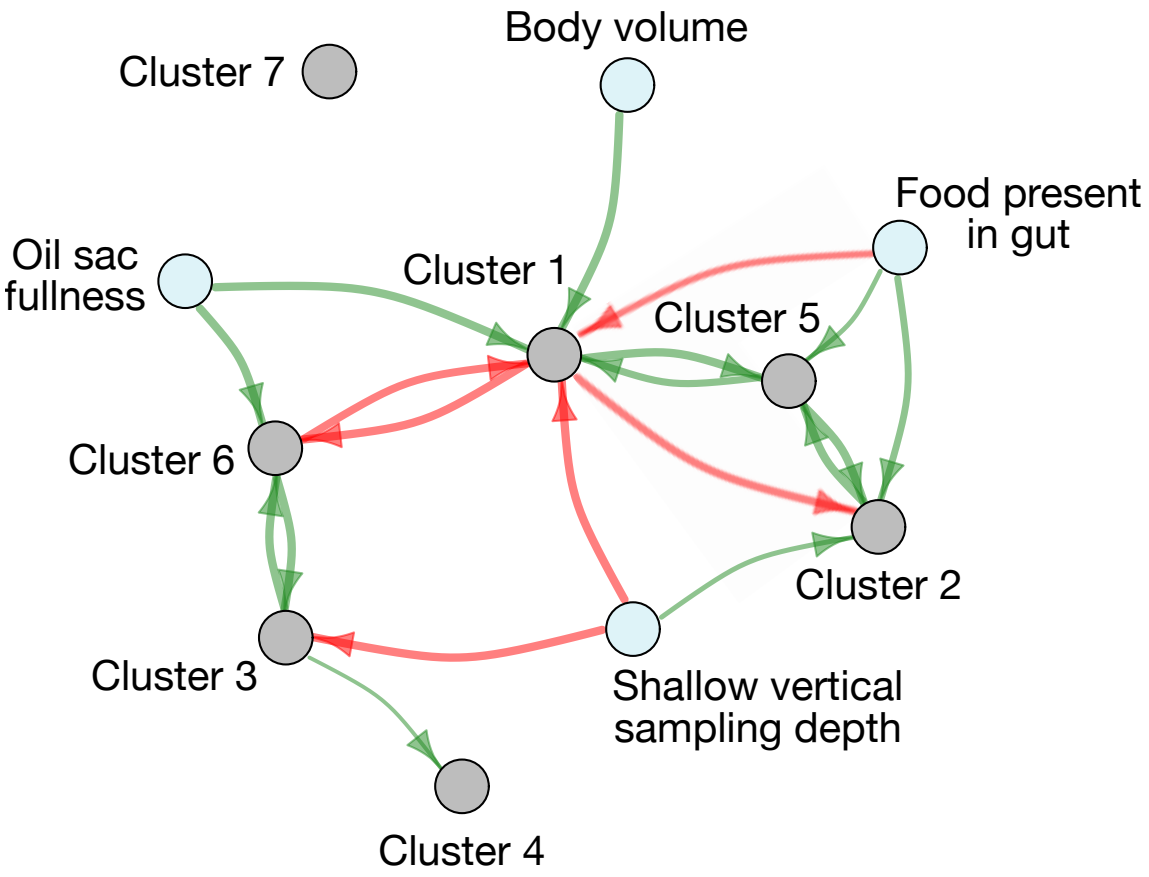


Figure 2-7: Cluster abundances are predicted by host morphology, as well as the abundances of other clusters. Nodes represent clusters of positively correlated bacterial taxa (gray) or copepod morphological characteristics (light blue). Directed edges indicate the association of a given predictor on the summed abundance of all bacterial taxa in a cluster, controlling for the effects of all other predictors (green: positive association; red: negative association). The width of the edge is proportional to the probability that a predictor has a non-zero effect on the cluster abundance, as determined by Bayesian model averaging. Edges for which the probability of a non-zero effect is less than 75% are not shown.

that 10 out of the 29 taxa in Cluster 2 were Flavobacteriacaeae, the association that we have identified supports a growing body of evidence indicating that Flavobacteriaceae (and perhaps other clades) are associated with the food a copepod ingests, rather than any intrinsic feature of a copepod.

Many mechanisms may account for the strong positive association between Cluster 2 abundance and food in the gut. However, one possible link is that Flavobacteriaceae are often highly abundant on diatoms [6] and other phytoplankton [24], which are standard sources of food for *C. finmarchicus*. Therefore, Flavobacteriaceae (and perhaps other Cluster 2 taxa) may reach high abundance within the copepod microbiome by "hitchhiking" into the gut on an ingested food particle. Alternatively, Cluster 2 taxa may typically reside at low levels in the copepod gut, but grow rapidly on the nutrients provided when copepods feed. Regardless of mechanism, the presence of food in the gut (through its effect on Cluster 2) is a significant differentiator of copepod microbiomes across vertical depths.

Among copepods collected at a single vertical depth, we also identified differences in their associated bacterial communities that could be linked to host-specific morphological variability. In particular, among deep-dwelling copepods, we identified two distinct subgroups based upon their associated bacterial communities (Fig. 2-4). One subgroup corresponded to those that harbored taxa from Clusters 3 and 6 at high relative abundance, while the other subgroup was comparatively depleted in taxa from these clusters. Interestingly, of the six taxa classified as Oceanospirillaceae across all clusters, four were found in Cluster 6 (comprising 80% of this cluster). This suggests that the Oceanospirillaceae present in our survey change in abundance as a correlated unit defined by Cluster 6.

Using our regression approach, we found that the abundance of Cluster 6 relative to the core microbiome was positively predicted by the fullness of a copepod's oil sac, a repository of nutrient-rich lipids amassed during the transition to diapause. To our knowledge, the effect of oil sac fullness on the copepod microbiome has not been explored previously, and indeed, would have been difficult to discern with standard bulk sampling methods. However, the mechanisms of this association are unclear.

One hypothesis is that the fullness of the oil sac may correspond to diet quality, since diatom blooms are typically a rich food source associated with lots of lipid storage. Thus, Oceanospirillaceae in Cluster 6 may be present at high abundance due to high food quality before the copepod entered diapause. Incidentally, the abundance of Cluster 3 was not predicted by oil sac fullness, but was well predicted by the abundance of Cluster 6, suggesting possible interactions between bacterial clusters.

Finally, we identified one cluster, Cluster 1, a hub in Fig. 2-7, which mediates differences in the diversity of bacterial communities harbored by shallow- and deep-dwelling copepods. Notably, this cluster was a defining feature of the communities associated with deep-dwelling copepods: on average, the relative abundance of Cluster 1 was significantly higher among deep-dwelling copepods (mean: 48%) than among shallow-dwelling copepods (mean: 2%) ($p = 4 \times 10^{-25}$, two-sided Mann-Whitney U test). While this cluster contained a diverse mixture of taxonomic classes (Fig. 2-6), it was dominated by a single taxon (seq1) that, alone, reached a maximum relative abundance of nearly 90% on individual copepods (Fig. 2-4). Surprisingly, seq1's closest taxonomic relatives are from the genus *Marinimicrobium*, a common genus of marine microorganisms whose association with brine shrimp has been documented previously [137], but whose associations with copepods have not been demonstrated. Thus, the bacterial communities associated with deep-dwelling copepods are dominated by taxa whose roles in the copepod microbiome have not, to our knowledge, been previously described.

The abundance of Cluster 1 strongly affects differences in overall community (Shannon) diversity between shallow- and deep-dwelling copepods. In particular, while some deep-dwelling copepods harbored bacterial communities that were as diverse as those of their shallow-dwelling counterparts, they were significantly less diverse on average (Fig. 2-8a, 2-8b). However, this difference was largely mediated by Cluster 1. Indeed, if we remove Cluster 1, the effective number of species (N_e) in the bacterial communities of deep-dwelling copepods increased dramatically (median $N_e = 18$ with Cluster 1 compared to $N_e = 53$ without Cluster 1), while the diversity of shallow-dwelling copepod communities did not change ($N_e = 65$ with

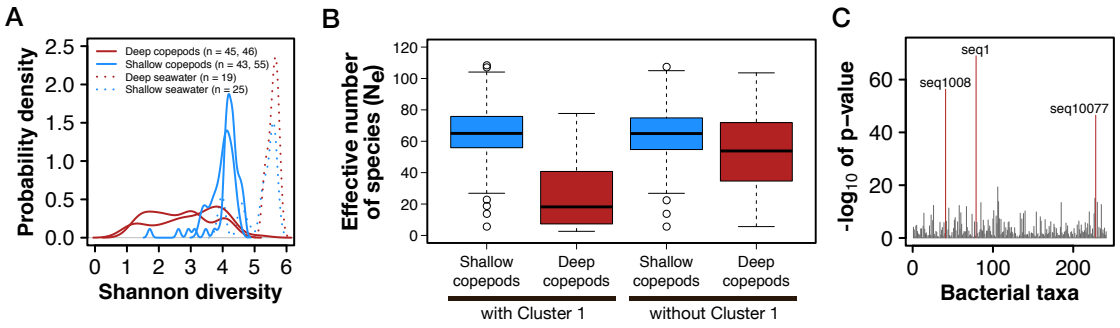


Figure 2-8: Differences in the diversity of copepod microbiomes are driven by a small number of taxa. (a) Kernel density plots of the Shannon diversity of bacterial communities associated with copepod individuals (solid lines) and the surrounding seawater (dotted lines). Samples were collected from the surface (blue lines) and at hundreds of meters in depth (red lines) on two different sampling dates (6/6/2012, light; 6/11/2012, dark). (b) The effective number of species ($N_e = e^{\text{Shannon diversity}}$) in bacterial communities associated with individual copepods. On the left, N_e was calculated based upon the abundances of all OTUs. On the right, counts associated with taxa from Cluster 1 were removed, and the remaining counts were renormalized before calculating N_e. (c) Manhattan plot of the p-value of the correlation of each individual bacterial taxon with Shannon diversity.

and without Cluster 1) (Fig. 2-8b). Moreover, this effect was specific to taxa associated with Cluster 1; nearly all taxa had statistically insignificant correlations with the community-wide Shannon diversity, with the exception of seq1, seq1008, and seq10077, all of which are members of Cluster 1 (Fig. 2-8c).

Despite the importance of Cluster 1 in determining the overall diversity of the community, the underlying drivers of its abundance remain unclear. Using our multivariate regression approach, we found that the abundance of Cluster 1 relative to the core microbiome was positively associated with copepod body volume and oil sac fullness and was negatively associated with shallow sampling depth and the presence of food in the gut, all of which suggest that Cluster 1 may become more abundant as copepods transition to diapause. However, Cluster 1's associations with Clusters 5 and 6 indicate the possible effects of inter-cluster interactions or, alternatively, common responses to other host-associated factors that were not measured in this study.

2.3 Conclusion

Individual copepods – even those of the same species and collected from the same habitat – vary considerably in the bacterial communities that they harbor. By leveraging the statistical power of our high-throughput sampling scheme, we have identified robust, statistically significant associations between certain bacterial taxa and specific physiological characteristics of their copepod hosts. Thus, we suggest that each copepod presents a distinct habitat modulated by physiological state, and that this presents a source of variability that cannot be observed with coarse-grained sampling techniques. We also identified both positive and negative associations between bacterial taxa, indicating that interactions between co-colonizing taxa may also shape the copepod microbiome.

A key step forward will be to characterize the mechanisms underlying the statistical associations identified in this study. In particular, we have identified several bacterial taxa that are positively associated with the presence of food in the gut. Are these taxa simply hitchhiking into the copepod on food particles (e.g., diatoms), or alternatively, are they simply "blooming" on the copepod in response to feeding-induced nutrient levels? We also identified a cluster of bacterial taxa that are positively associated with the fullness of the copepod's oil sac, but the origins of this association are unclear. Thus, future work should focus on identifying the source of copepod-bacterial associations, perhaps using *in vitro* experiments with bacterial isolates to complement studies in the field.

One of the most striking findings of this study was the prevalence of *Marinimicrobium* (in Cluster 1), which was endemic among deep-dwelling copepods up to 90% of these communities. While we showed that Cluster 1 abundance was associated with oil sac fullness, body volume, and the absence of food in the gut, as well as other bacterial clusters, it remains mechanistically unclear how these associations arise. Future work should focus on isolating and characterizing *Marinimicrobium* to understand how it comes to dominate the microbiomes of deep-dwelling copepods.

More broadly, there are many additional factors not measured here that may affect

how bacteria assemble functional communities on individual copepod microhabitats. For instance, copepod-associated bacterial communities may change over time as they adapt to the copepod habitat. Indeed, microbial communities associated with model marine chitin particulates have been found to undergo dramatic compositional and functional shifts in a matter of days, even in the absence of external perturbations [40]. This may be particularly relevant for diapausing copepods, since they can remain in the C5 stage (without molting or feeding) for several months, thereby allowing bacterial communities to develop over a longer period of time. Additionally, while we have focused on how copepods influence their bacterial inhabitants, bacteria may also influence their copepod hosts (for instance, by provoking significant transcriptional responses [4]). Ultimately, future work incorporating these factors will lead to a better understanding of copepod-bacterial interactions, and how these interactions affect the global marine ecosystem.

2.4 Methods

2.4.1 Sampling of individual copepods and seawater

Copepods were sampled at two different depths (shallow, 0-50 meters, and deep, 250-350 meters deep) on two separate dates during the early summer (June 6, 2012 and June 11, 2012), at which time, some individuals within the population had already descended from the surface ocean and entered into diapause. The surrounding seawater was also sampled at each depth on each of the sampling dates. Detailed sampling protocols are available in Appendix A.

2.4.2 Quantification of morphological characteristics

For all individual copepods, we quantified several morphological characteristics that often vary with host physiological state during the transition from active growth to diapause. These characteristics were the prosome (body) volume, oil sac volume, the fraction of the oil sac that was filled with lipids, and the presence or absence of food

in the gut. All of these characteristics were estimated from digital photographs. The presence or absence of food in the gut was assessed in a visually from these images. All others were quantified as described below.

Previous studies have shown that the true prosome volume of *C. finmarchicus* is well approximated by numerical integrations of lateral views of the prosome [115]. However, this method is computationally intensive and thus, does not scale well to large numbers of individual copepods. Instead, we approximated the shape of the prosome as an ellipsoid, allowing us to calculate the prosome volume as a function of its length and width. Notably, for a representative copepod test set, this ellipsoid approximation provided estimates that were consistent with the more complex numerical integration approach (data not shown). All measurements were calibrated with digital photographs of a stage micrometer taken repeatedly during sampling.

The volume and fractional fullness of the oil sac was estimated from digital photographs as described previously [115, 166]. Briefly, for each individual copepod, the oil sac was visually sectioned into cylindrical cross-sections whose volume could be quantified directly from the height and cross-sectional area. The total oil sac volume was estimated by summing over all cylindrical cross sections and adjusting for deviations from a perfectly cylindrical shape as previously described [115]. The fullness of the oil sac was estimated by normalizing the oil sac volume by an empirically defined apparent maximum capacity [115, 166].

2.4.3 Quantifying bacterial abundance on copepods and in seawater

The total number of bacteria present per copepod was quantified via fluorescence microscopy as previously described [162] with minor modifications.

Formalin-preserved copepods in 1.5-mL Eppendorf tubes were homogenized by (1) manually grinding the sample with a sterile plastic pestle (Axygen Scientific #PES15BSI) and (2) subsequently placing the sample in a water bath sonicator (Kendal #CD4820) for three minutes. Homogenized copepod samples were diluted

eight-fold with artificial seawater (sterile-filtered, 0.22 μm). Two replicate 2-mL aliquots of each diluted copepod sample were filtered onto a black polycarbonate 0.22-μm filter (EMD Millipore Isopore #GTBP02500) (vacuum pressure $<$ 10 mm Hg). Each filter was stained directly with a SYBR Gold solution (ThermoFisher Scientific #S-11494, 2X, diluted in TE buffer) [17, 19, 162] for 15 minutes, after which they were mounted in Citifluor Media (Ted Pella Inc. #NC9522768). Filters were imaged with a Zeiss Axiostar Plus microscope outfitted with epifluorescent illumination (excitation: 450-490 nm; emission: 515 nm). We first counted the average number of cells across fifty random view fields, with which we estimated the number of bacterial cells present per milliliter.

Blank samples – containing artificial seawater (0.22-μm filtered), but no copepods – were processed in the manner described above. For samples of seawater, five 1-mL aliquots were filtered onto a black polycarbonate filter (EMD Millipore Isopore #GTBP02500) before staining with SYBR Gold (ThermoFisher Scientific #S-11494).

2.4.4 DNA extraction from individual copepod and seawater samples

Total genomic DNA was extracted using a modification of a previously described protocol [194]. Briefly, samples were bead-beaten in 2-mL tubes with a mixture of 0.1-mm diameter silica, 1.4-mm diameter zirconium, and 4-mm diameter silica beads (OPS Diagnostics #PFMM 4000-100-28). After treatments with lysozyme and proteinase K, genomic DNA was extracted from samples in phenol:chloroform:isoamyl alcohol (25:24:1 v/v) solution. Subsequently, DNA was precipitated from samples overnight in isopropanol with GlycoBlue (1μL; Life Technologies #AM9516) as a co-precipitant. Precipitated DNA was resuspended in water (molecular biology grade) and stored at -20°C. All copepod and seawater samples were extracted in a random order to avoid bias, and blanks were processed simultaneously. Details of DNA extraction procedure are in Appendix A.

2.4.5 16S rRNA amplicon sequencing of copepod- and seawater-associated bacterial communities

Amplicon libraries (16S rRNA gene V4 hypervariable region) were prepared with modifications to a previously described protocol [130]. Details are described in Appendix A. To ensure that seawater and copepod samples were comparable, seawater samples were diluted 100-fold in water (molecular biology grade) such that identical amplification schemes could be used for all samples. Amplicon libraries of DNA extraction blanks, no-template control (NTC) samples, and defined mock communities (consisting of nine bacteria in known proportions) were prepared simultaneously.

After amplicon libraries were prepared, samples were multiplexed for sequencing. Copepod 18S rRNA amplicons were removed from this mixture via gel extraction (NucleoSpin Gel and PCR Clean-up kit, Macherey-Nagel #740609) before sequencing. Multiplexed, gel extracted samples were sequenced on an Illumina MiSeq (paired-end, 250+250) at the BioMicro Center (Massachusetts Institute of Technology, Cambridge, MA). Reads were merged and quality filtered with custom scripts calling USEARCH [50], mothur [146], and SmileTrain (https://github.com/almlab/SmileTrain).

2.4.6 Identifying operational taxonomic units (OTUs) via distribution based clustering (DBC)

Operational taxonomic units (OTUs) were identified using distribution-based clustering, an algorithm that uses both genetic distance and the distribution of sequences across samples to group reads into OTUs. This approach reduces the number of OTUs with redundant information and improves the power of many downstream analyses to describe biologically relevant trends [130].

A two-step process was used to identify OTUs. First, the DBC algorithm was applied to quality-filtered reads with a high abundance threshold (k_{fold} 10) to remove reads that likely arose through PCR or sequencing errors. This step generated 15,060 OTUs across all seawater and copepod samples. Second, the DBC algorithm was applied to this set of OTUs (95% identity threshold, k_{fold} 0). This step merged

OTUs with representative sequences that were >95% identical and were similarly distributed over samples. Altogether, this resulted in 9,642 OTUs across all seawater and copepod samples.

Taxonomic assignment of individual OTUs was performed with the RDP Classifier [182]. OTUs that were annotated as "Cryptomonadaceae", "Chlorophyta", "Bacillariophyta", "Streptophyta", or "Chloroplast" were removed prior to normalization, since these OTUs are likely non-bacterial in origin. After removal of these OTUs, 9,487 OTUs remained for subsequent analyses.

2.4.7 Identifying "abundant" OTU subset

From the 9,487 non-"Chloroplast" OTUs, we identified a subset of 241 OTUs that were used for subsequent analyses. First, for each of the two sampling dates (June 6, 2012 and June 11, 2012), we identified the list of OTUs whose mean relative abundance across all copepods sampled on that date was greater than 2×10^{-4}. These lists comprised roughly 300 OTUs for each date. Second, we found the intersection of these two OTU lists, identifying a set of 241 OTUs whose mean relative abundance across both sampling dates was above the threshold (Fig. 2-3). Together, these 241 OTUs corresponded to 89.9% of sequenced reads from all copepod samples.

We chose to identify our OTU subset of interest with a mean abundance threshold, since such a threshold allows for OTUs with a wide range of distributions across individual copepods (roughly 100 per sampling date). These include (1) OTUs present at a low, but detectable abundance across all copepods, as well as (2) OTUs present at high abundance on a single copepod. In practice, there are many possible thresholds (aside from 2×10^{-4}) and thresholding schemes (e.g., maximum or minimum relative abundance) that we could have used. However, the qualitative results that we have discussed here are robust to the choice of threshold (not shown).

2.4.8 Defining the "core" and "flexible" microbiome

Differences in sequencing depth precluded the use of an absolute cutoff in defining the core microbiome (e.g., requiring an OTU to be present in all sequenced copepods). Instead, we defined the "core" microbiome as OTUs that were present at non-zero relative abundance in at least 90% of copepod-associated bacterial communities. All OTUs that did not meet this criterion were considered a part of the "flexible" microbiome.

2.4.9 Using PERMANOVA to identify broad associations between metadata and bacterial communities

To assess sources of variation across copepod-associated bacterial communities, we carried out a permutational MANOVA (PERMANOVA) [7]. All analyses were performed with the *adonis* function from the R vegan package with a Euclidean distance metric and 10^6 permutations.

2.4.10 Identifying correlated clusters of taxa

Exact computation of correlations with compositional data – such as data obtained through 16S rRNA amplicon sequencing – can give spurious results [2]. As an alternative, we used SparCC, a procedure that allowed us to approximate the linear Pearson correlations between log-transformed components for compositional data [63]. We applied SparCC with the default parameters (-i 20, -x 10, -t 0.1) to compositional abundance data for the 241 OTUs whose mean abundance surpassed the defined threshold (Fig. 2-3). This procedure generated a 241×241 matrix of pairwise correlations between all OTUs.

To identify clusters of taxa where the within-cluster correlation was strongly positive, we used a three-step procedure. First, we binarized the correlation matrix in the following manner: taxon-taxon correlations (including negative correlations) with values below a given threshold (0.35) were set to 0, while all others were set to 1. Second, taxa were removed if (a) they were part of the core microbiome or (b) they lacked

non-zero correlations with other taxa after binarization. Third, strongly positively correlated clusters were identified using a community detection algorithm (walktrap community, details below). Using this approach, we identified seven clusters (Fig. 2-6) for subsequent analyses.

The community detection algorithm that we used was implemented in the walktrap.community function from the R igraph library (nsteps = 2). The walktrap algorithm has several advantages, including high quality of results for both test and real-world networks and low runtime [129], which made it suitable for use here.

It is also important to note that, of the 241 OTUs for which taxon-taxon correlations were estimated, the majority were excluded from the seven clusters identified, since they were not strongly correlated with any other taxa. As a result, only 64 OTUs were represented across the seven clusters. These 64 OTUs correspond to 68.0% of sequenced reads from all copepod samples (compared to 89.9% for all 241 OTUs), and thus, still represent a significant portion of the bacterial community.

2.4.11 Multivariate regressions

Our goal was to predict the abundances of individual bacterial clusters across individual copepods as a function of (a) the morphological characteristics of individual copepods and (b) the abundances of other bacterial taxa. To this end, we used a multivariate linear regression approach, which allowed us to quantify the effect of each explanatory variable on bacterial abundance, controlling for the effects of all other variables.

The problem of relating bacterial to their potential drivers is non-trivial for at least two reasons: (1) the large number of bacterial taxa and (2) the compositional nature of the data. Thus, our approach was to predict the abundances of each of the seven positively correlated clusters that we previously identified, controlling for the abundances of all other clusters. Additionally, we used an additive log-ratio transform [2] to remove the compositional sum constraint and to quantify the log-ratio absolute abundance of each cluster relative to the core microbiome. A more detailed description of our methods can be found in Appendix A.

Chapter 3

Rapid micro-scale successions on model marine particles

This chapter has been submitted for publication and is currently in revision.

Abstract

In the ocean, particles of organic matter harbor dense, diverse bacterial communities, which collectively digest and recycle essential nutrients. Traits like motility and exo-enzyme production are thought to allow individual taxa to colonize and exploit particle resources, but it remains unclear how community dynamics emerge as a function of these individual traits. By tracking the taxon and trait dynamics of bacteria attached to model marine particles, we demonstrate that particle-attached communities undergo rapid, reproducible successions driven by ecological interactions. Motile, particle-degrading taxa are selected for during early successional stages. However, this selective pressure is relaxed at later stages of colonization, as secondary consumers invade, which are unable to use the particle resource, but instead, rely on carbon from primary degraders to create a trophic chain that shifts community metabolism away from the particle substrate. These results demonstrate that microscale primary successions may control the dynamics of particle-attached bacteria in the ocean and that rapid shifts in community-wide metabolism could limit rates of marine particle degradation.

3.1 Introduction

Bacterial colonization of particulate organic matter (POM) in the ocean is a well-known example of microbial community assembly with important implications for global carbon cycling [3, 150]. On global scales, POM mediates the transfer of nearly two billion tons of carbon from the surface to the deep ocean [89]. However, at micrometer scales, marine particulates serve as spatially isolated, nutrient-rich microhabitats in an otherwise nutrient-poor environment [98]. Microbes from the surrounding seawater, representing a complex colonization pool of bacteria, archaea, eukaryotes, and viruses, attach to these particles, eventually forming dense multi-species communities [150]. Within these communities, local interactions between neighboring cells are predicted to play an important role in shaping community-level structure and function [96, 75]. These interactions include exploitation of public goods [33] (for instance, broadcasted degradation products of carbohydrate-active enzymes [151]), antagonistic interactions via antibiotics [108], and quorum sensing [70, 86]. Moreover, at regional scales, the efficiency with which bacteria move through a particle "landscape" via active or passive dispersal is likely to influence their ecological success [158, 190, 128, 1, 56, 156, 44]. How these processes combine to give rise to dynamics at the level of the community, particularly in the context of a diverse natural microbial assemblage, is still not well understood.

To enable studies of microbial community dynamics, we developed a model system inspired by bacterial colonization of POM in the ocean. We simulated POM with paramagnetic micro-particles (Fig. 3-1: median diameter 40.7 μm) made of chitin – a highly abundant biopolymer in the ocean [13]. We incubated these particles in a sample of coastal seawater, which contained a diverse microbial assemblage of nearly one million bacteria per milliliter, as well as myriad viruses and small eukaryotes [98]. Over nearly six days, bacteria from the surrounding seawater (Appendix B; Fig. 3-3) self-assembled into communities on the chitin particle microhabitats (Fig. 3-1; Fig. 3-2a, Fig. 3-2b). At discrete time intervals, we harvested pools of particles (roughly 1,000 per sample), thus allowing us to reconstruct the average community assem-

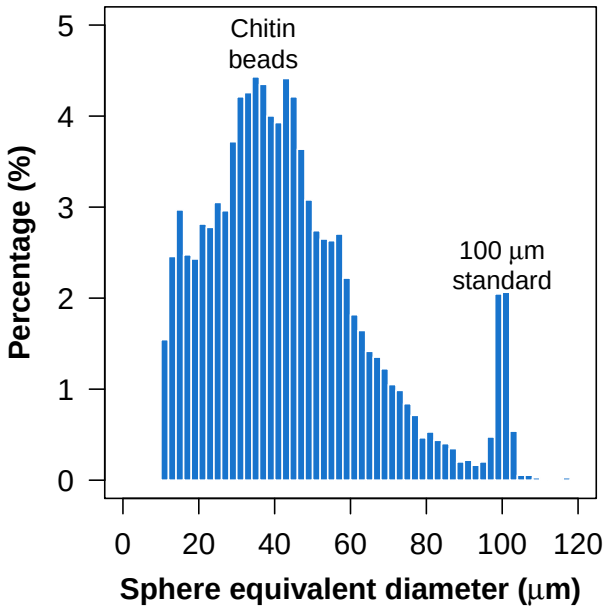


Figure 3-1: For many individual particles, the sphere equivalent diameter was measured with a Coulter counter (Beckman Coulter Multisizer 4, 560 μm aperture). Note that, for an irregularly shaped object, the sphere equivalent diameter refers to the diameter of a perfectly spherical object of identical volume. The chitin particles that we considered were roughly spherical, but irregularly shaped particles were also observed. As an internal standard, spherical polystyrene particles of a known diameter (100 μm, Thermo Scientific #4310A) were added to the sample before measurement.

bly dynamics occurring over many spatially distinct, but temporally synchronized, particles. To assess the reproducibility of these dynamics, we performed three replicates of the colonization process from a single, well-mixed seawater sample. Note that naturally-occurring POM varies widely in age, size, and chemical composition, making it challenging to discern the processes shaping their associated bacterial communities. In contrast, our model system offers a hybrid approach – maintaining high levels of microbial diversity, while reducing substrate heterogeneity – that allowed us to characterize complex community assembly dynamics in a defined environment with high temporal resolution.

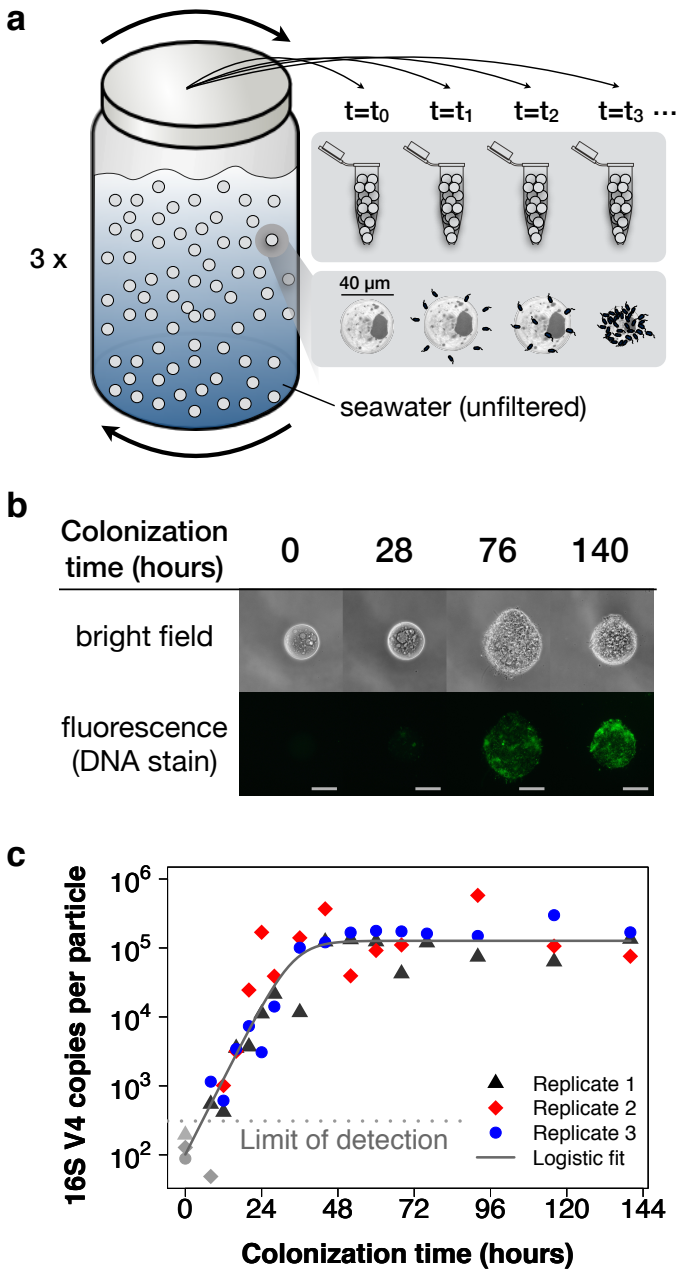


Figure 3-2: (A) Schematic of particle colonization procedure. (B) Particle-attached communities stained with SYBR Green I, a double-stranded DNA strain, and imaged with bright field (top) and fluorescence (bottom) microscopy (Appendix B). Scale bars represent 25 μm. Note that different particles are depicted for each time point. (C) Total 16S rRNA V4 copies per particle over time for three colonization replicates. Symbols in gray indicate measurements below the limit of detection of the assay. The gray line (-) indicates the fit to a logistic growth model.

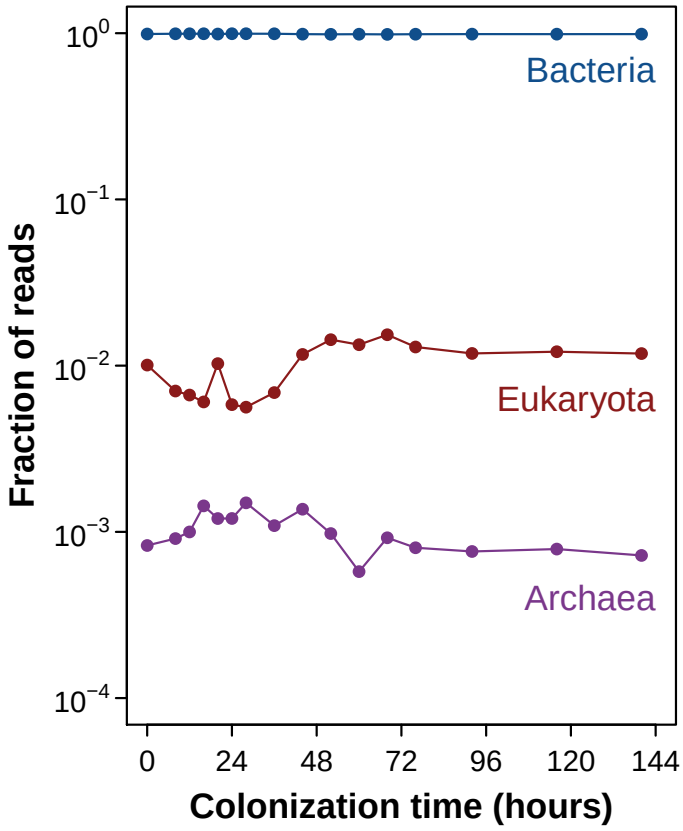


Figure 3-3: The fraction of read annotations corresponding to each of three taxonomic domains as a function of time. Metagenomic reads were annotated with MG-RAST (BLAT search against the M5NR database, e-value $< 10^{-5}$; length > 60 basepairs). Note that, in the standard MG-RAST pipeline, a single read may have more than one annotation. Thus, the number of read annotations is larger than the number of reads.

3.2 Results

3.2.1 Successions in particle-attached bacterial communities

Despite the extreme diversity of the surrounding microbial assemblage, we found that the overall growth dynamics of particle-attached bacterial communities were surprisingly simple. To characterize these combined growth dynamics, we quantified the number of copies of the V4 hypervariable region of the 16S rRNA gene – a rough proxy for the number of bacteria – present per particle on average over time. Across three colonization replicates, the dynamics were well described by a logistic growth model (Fig. 3-2c): bacterial communities underwent rapid exponential growth (doubling every 3.3 hours), eventually saturating at nearly 10^5 16S rRNA gene V4 copies per particle after only 40 hours of colonization.

Although the total abundance curve saturated early, the underlying colonization dynamics of individual taxa revealed a rapid ecological succession, with wholesale community turnover not only during exponential growth, but also long after the total bacterial abundance had saturated (Fig. 3-4a). In particular, many taxa experienced a sharp drop in absolute abundance, often by orders of magnitude, soon after reaching their peak absolute abundance levels (e.g. OTU 1 in Fig. 3-4c). As they dwindled, these taxa were replaced by others, which reached maximum levels that often matched (or exceeded) the earlier colonizers (Fig. 3-4b), but that declined in turn as still others supplanted them. In total, this dynamic process of community turnover brought 53 highly abundant taxa – present at $>1\%$ relative abundance in at least one time point – that each peaked in abundance at times ranging from 16-140 hours of incubation (Fig. 3-4a). While microbial successions are widely documented (e.g. in the human gut [99, 42], the soil [124], and the marine environment [168, 29]), such dramatic community turnover has not, to our knowledge, been observed on the spatial (microns) or temporal (hourly) scales documented here.

Given that these dynamics originated from the migration, growth, and interactions of many diverse bacteria, we predicted that chance events might give rise to divergent community dynamics, even from the nearly identical starting conditions of our three

colonization replicates [65]. However, across these replicates, individual taxon trajectories were highly reproducible (Fig. 3-4d; Fig. 3-5). For abundant taxa – present at >1% relative abundance at any time point and in any replicate – the median Spearman correlation between individual taxon trajectories from different replicates was greater than 0.8 (Fig. 3-4d). This high level of reproducibility indicates that technical variation across samples was minimal. However, it also suggests that the average process of community self-assembly is robust to ecological drift, particularly historical contingencies that can arise in a complex microbial milieu.

Importantly, while community turnover occurred continuously, we identified three discrete phases of colonization based upon changes in the community-wide diversity over time (Fig. 3-4e). In the first phase of colonization (t = 8-20 hours), the communities were at their most diverse (effective number of species, $N_{eff} \approx 180$ OTUs; Appendix B). The second phase was characterized by a significant decline in community-wide diversity, which reached a minimum ($N_{eff} \approx$ 20-30 OTUs) after 36-44 hours. However, in the third phase, the community-wide diversity rose again, eventually plateauing ($N_{eff} \approx$ 50-70 OTUs) after 72 hours. Notably, this non-monotonic trend in community diversity held for several diversity metrics (Fig. 3-6).

3.2.2 Mechanistic drivers of successional phases

What drives the community shifts that define these three phases of particle colonization? As is often true in plant communities, we hypothesized that temporal changes in the behavior and metabolism of particle-attached communities may shape the successional patterns that we observed [32, 57]. To test this hypothesis, we took two complementary approaches. First, we performed metagenomic sequencing of the time series to gain a holistic view of how the metabolic potential of the community changed with time. Second, we amassed and phenotypically characterized a collection of bacterial strains isolated from different phases of colonization. Using their 16S rRNA gene sequences, we mapped these isolates to the operational taxonomic units (OTUs) originally observed via 16S sequencing (Appendix B; Fig. 3-7). This allowed us to link the phenotypic traits of individual isolates to their taxon's colonization

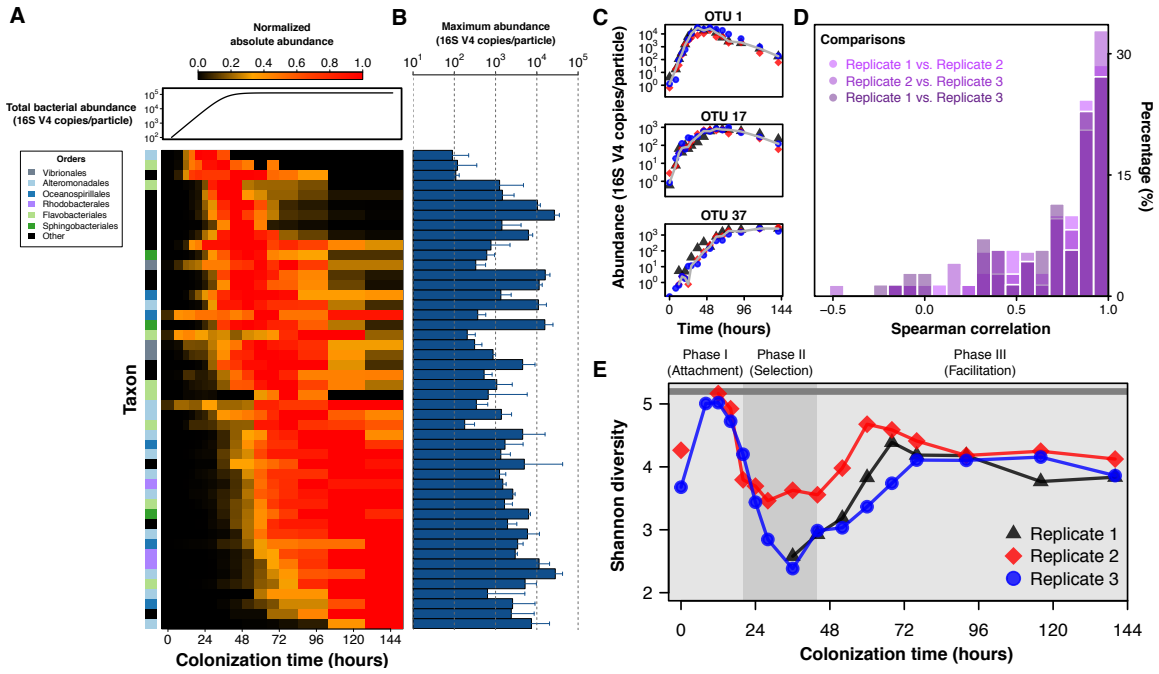


Figure 3-4: (A) Absolute abundance trajectories for individual taxa from a single colonization replicate (Replicate 2). Individual trajectories are normalized to the maximum value. Color bar indicates order-level taxonomic identities. Line plot above the heat map shows the logistic fit to the total bacterial abundance trajectory. (B) Maximum abundance per particle attained by each taxon. Error bars are standard deviations (n=3). (C) Absolute abundance trajectories of three representative taxa across colonization replicates. Gray lines indicate the median trajectories. (D) Histogram of cross-replicate correlations for individual taxa (Methods). (E) Shannon diversity over time for the three colonization replicates. Samples for which sequencing coverage was insufficient for the Shannon diversity to saturate have been omitted. The solid gray line indicates the initial Shannon diversity of the seawater.

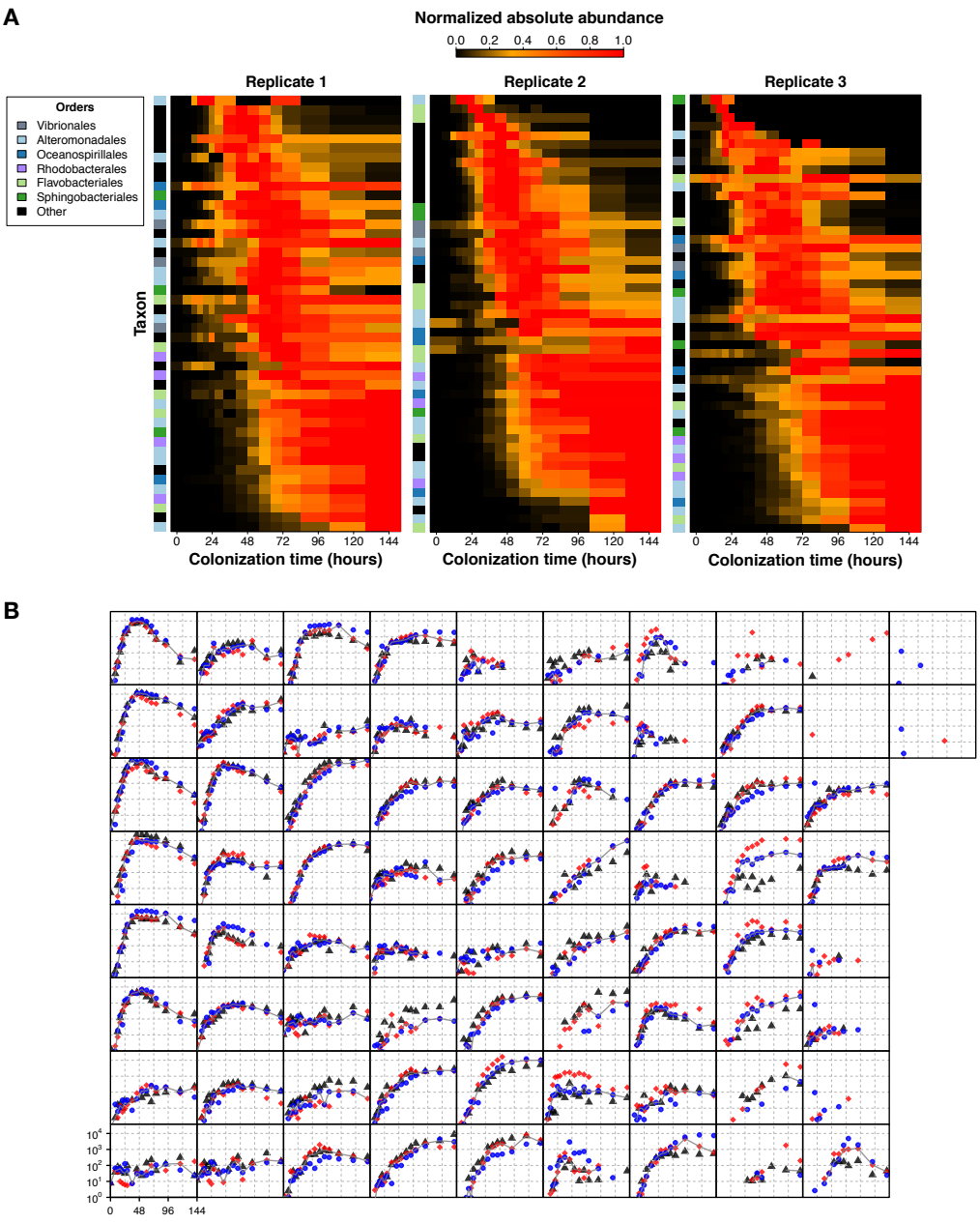


Figure 3-5: (A) Normalized smoothed absolute abundance trajectories for individual taxa in each of three colonization replicates. Each row corresponds to the dynamics of a single taxon (only taxa present at a relative abundance $> 1\%$ at any time point in a given replicate are shown). Trajectories were smoothed with a three-point running median filter and normalized by the maximum abundance attached by that taxon such that all values range from 0-1. Taxa are ordered by increasing center of mass of the trajectory. (B) Absolute abundance trajectories of all taxa present above the relative abundance threshold (1%) in any of the replicates. Grey lines indicate the median trajectory for a given taxon. All taxon trajectories are plotted with the same x- and y-axis limits.

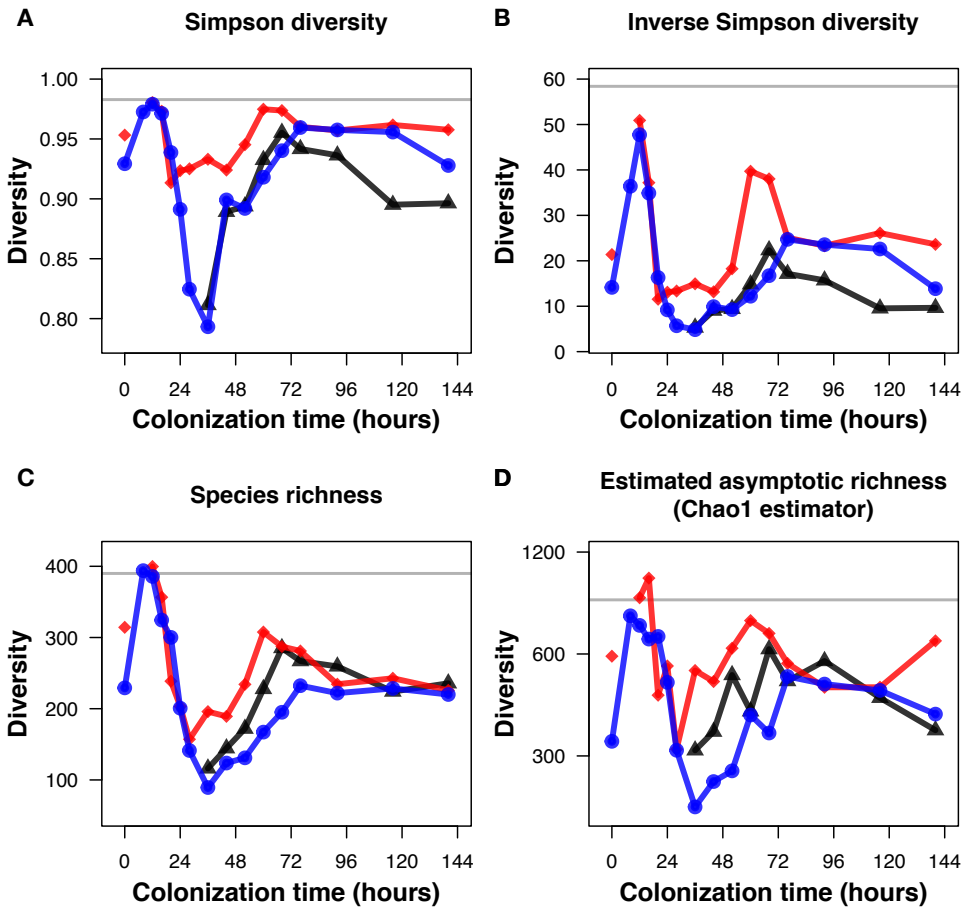
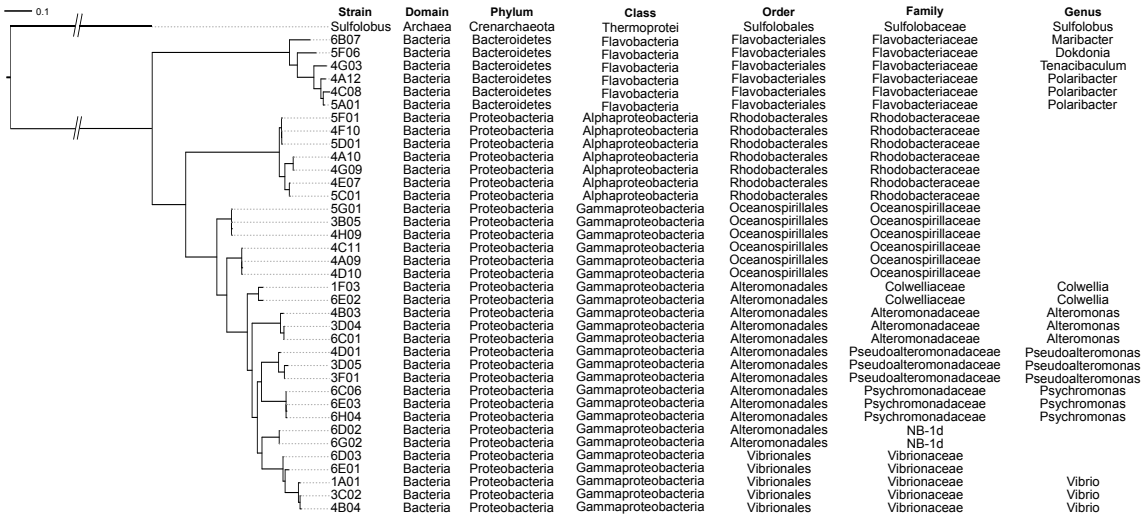


Figure 3-6: Particle-attached community diversity over time in each of three replicates, as calculated with several different diversity metrics. In all plots, the diversity of the initial seawater inoculum (t = 0 hours) is also indicated. In all cases, samples with less than 1500 sequenced reads were excluded. (A) Simpson diversity at each timepoint. (B) Inverse Simpson diversity at each timepoint. (C) Species richness at each timepoint. To account for differences in sampling effort, each community was subsampled to 1,500 reads before calculating species richness. Thus, only the relative species richness is meaningful here. Qualitative differences in species richness are robust to subsample size. (D) Estimated asymptotic species richness over time. Asymptotic estimators extrapolate the number of species that would be observed, given infinite sampling effort. In particular, the Chao1 estimator is a robust estimator of the minimum asymptotic species richness.



Strain	Domain	Phylum	Class	Order	Family	Genus
Sulfolobus	Archaea	Crenarchaeota	Thermoprotei	Sulfolobales	Sulfolobaceae	Sulfolobus
6B07	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Maribacter
5F06	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Dokdonia
4G03	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Tenacibaculum
4A12	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Polaribacter
4C08	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Polaribacter
5A01	Bacteria	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Polaribacter
5F01	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	
4F10	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	
5D01	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	
4A10	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	
4G09	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	
4E07	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	
5C01	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	
5G01	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae	
3B05	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae	
4H09	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae	
4C11	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae	
4A09	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae	
4D10	Bacteria	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae	
1F03	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Colwelliaceae	Colwellia
6E02	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Colwelliaceae	Colwellia
4B03	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	Alteromonas
3D04	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	Alteromonas
6C01	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	Alteromonas
4D01	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Pseudoalteromonadaceae	Pseudoalteromonas
3D05	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Pseudoalteromonadaceae	Pseudoalteromonas
3F01	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Pseudoalteromonadaceae	Pseudoalteromonas
6C06	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Psychromonadaceae	Psychromonas
6E03	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Psychromonadaceae	Psychromonas
6H04	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	Psychromonadaceae	Psychromonas
6D02	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	NB-1d	
6G02	Bacteria	Proteobacteria	Gammaproteobacteria	Alteromonadales	NB-1d	
6D03	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	
6E01	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	
1A01	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Vibrio
3C02	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Vibrio
4B04	Bacteria	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Vibrio

Figure 3-7: Isolate full-length 16S sequences (roughly 1,400 basepairs) were aligned against the Silva reference database. From this alignment, a maximum-likelihood tree was generated with PhyML (substitution model: GTR+gamma). Taxonomic classifications for strains (identified with the Ribosomal Protein Database) are indicated in the table. Classifications with <80% confidence as determined by RDP are not shown.

dynamics.

Overall, our data suggest that the three phases of colonization were governed by distinct ecological processes: (i) Phase I, attachment, (ii) Phase II, selection, and (iii) Phase III, replacement by secondary consumers. In Phase I, particle-attached bacterial communities were as diverse as the seawater from which the colonizers originated (Fig. 3-4e; Fig. 3-6), despite low total bacterial abundance (Fig. 3-2c). Moreover, the frequencies of gene families associated with chitin metabolism (e.g., GH18 family chitinases) were low (Fig. 3-8a). This suggests that, at early stages of colonization, the composition of particle-attached communities is not determined by growth on the particle substrate, but instead, may be governed by particle attachment ability. In general, particle attachment is a complex trait influenced by bacteria-particle encounter rates, chemotaxis, biofilm production, and the expression of chitin binding proteins. Nonetheless, given the diversity of taxa able to colonize particles in Phase I, this suggests that particle attachment is a weak selective filter.

By contrast, the dramatic decline in community-wide diversity that defined Phase

II (t = 20-44 hours) was likely driven by strong ecological selection for chitin metabolism and rapid dispersal ability. In particular, gene families associated with chitin metabolism peaked in relative abundance in Phase II (Fig. 3-8a), while nearly all others remained constant in time (Fig. 3-9). These gene families were associated with multiple stages of chitin metabolism, including extracellular chitin degradation (with GH18 family chitinases), chitin-specific substrate attachment (via chitin-binding proteins), chemotaxis towards chitin monomers, and catabolism of chitin oligomers (Fig. 3-8a). Indeed, among taxa with isolate representatives, four out of the five taxa that were highly abundant in Phase II could grow in culture with chitin as the sole carbon source (Fig. 3-8b; Fig. 3-10a). Interestingly, all four of these chitin-metabolizing taxa were also able to (i) broadcast extracellular chitinases into the surrounding environment (Fig. 3-8b; Fig. 3-10e) and (ii) to consume two common chitinase degradation products, chitin monomers (N-acetylglucosamine or GlcNAc) and dimers (N',N'-diacetylchitobiose or $(GlcNAc)_2$) (Fig. 3-8b; Fig. 3-10b, Fig. 3-10c). Moreover, all taxa that gained prominence in Phase II were motile under laboratory conditions (Fig. 3-8b; Fig. 3-10d), highlighting that rapid dispersal via active swimming may influence colonization order.

As particle-attached communities entered Phase III (t = 44-140 hours), the community-wide diversity rose again from its Phase II minimum as, simultaneously, the ecological selection for chitin metabolism and rapid dispersal that defined Phase II was relaxed. Community wide, the relative levels of gene families associated with chitin metabolism and chemotaxis towards chitin degradation products declined in Phase III, sometimes by orders of magnitude (e.g. GH18 chitinases, Fig. 3-8a). Similarly, among taxa with isolate representatives that reached prominence in Phase III, none were motile under laboratory conditions (Fig. 3-8b), and the majority (8 out of 11) were unable to grow in culture with chitin as the sole carbon source (Fig. 3-8b, Fig. 3-10a). Incidentally, the minority that could metabolize chitin did not broadcast extracellular chitinases, nor could they typically consume chitin monomers or dimers (Fig. 3-8b; Fig. 3-10b, Fig. 3-10c), suggesting a non-canonical mechanism of chitin metabolism compared to their Phase II counterparts (Appendix B). Altogether, despite their widespread inabil-

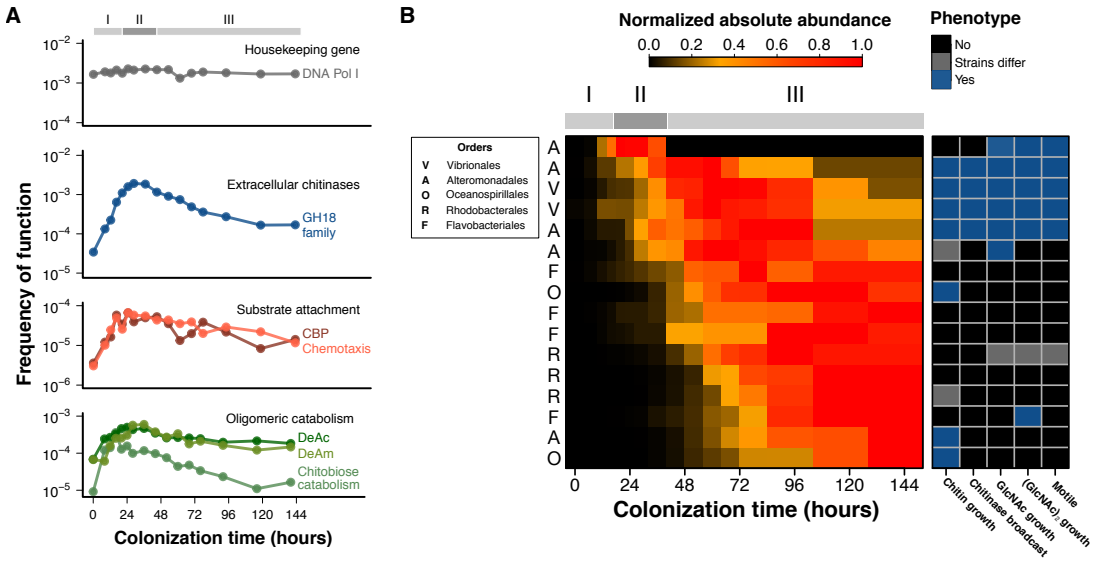


Figure 3-8: In both subpanels, phases of colonization (I, II, and III) are indicated with the gray color bar. (A) The fraction of read annotations mapped to a given functional category over time. Abbreviations: "DNA Pol I" = DNA polymerase I (EC 2.7.7.7); "GH18 family" = Glycoside Hydrolase Family 18; "CBP" = chitin-binding protein (Auxiliary Activity Family 10); "Chemotaxis" = N-acetylglucosamine regulated methyl-accepting chemotaxis protein; "DeAc" = N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25); "DeAm" = Glucosamine-6-phosphate deaminase (EC 3.5.99.6); "Chitobiose catabolism" = $(GlcNAc)_2$ Catabolic Operon (SEED Subsystem). (B) Left heat map: Absolute abundance trajectories of isolated taxa. Leftmost letter identifiers show order-level taxonomic identities. Right heat map: whether isolates do (blue) or do not (black) display a functional trait (assays described in Methods; Appendix B). Gray: within-taxon isolates differ in their phenotype. The number of isolates surveyed per taxon ranged from 1-3.

ity to consume chitin, the primary particle resource, Phase III-dominant taxa often grew to levels that rivaled those from Phase II (Fig. 3-4b). Thus, Phase III marked a community-wide shift in metabolism away from chitin towards other nutrient sources.

Given their inability to metabolize chitin directly, we hypothesized that Phase III-dominant taxa instead consumed nutrient byproducts produced by chitin metabolizers. To test this hypothesis, we co-cultured isolates from two Phase III-dominant taxa (from distinct bacterial phyla, Proteobacteria and Bacteroidetes) with each of six chitin metabolizing isolates (representing three orders within Gammaproteobacteria). Of the six chitin metabolizers, three could broadcast extracellular chitinases, while

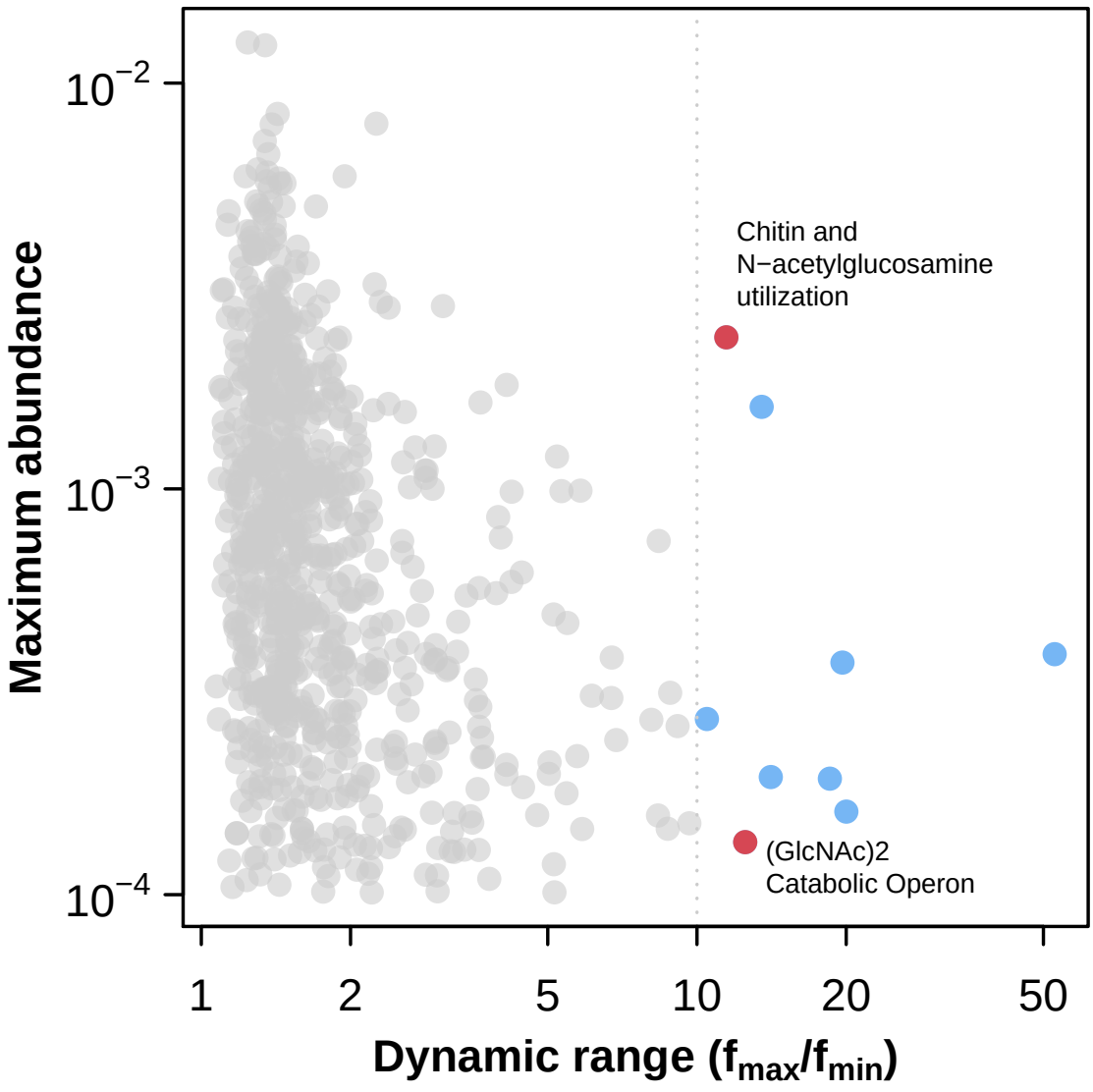


Figure 3-9: The maximum relative abundance reached versus the overall dynamic range for functional categories. Each point corresponds to a single functional category as annotated with MG-RAST (SEED Subsystem Level 3, similar to a KEGG pathway). Only functional categories that reached a maximum relative abundance $> 10^{-4}$ are plotted. Functional categories for which the dynamic range (maximum abundance divided by minimum abundance) was less than 10 are plotted in gray. Those with a dynamic range greater than 10 are: $(GlcNac)_2$ Catabolic Operon, Alkanesulfonates Utilization, Atlg48360, Chitin and N-acetylglucosamine utilization, Conjugative transfer, Conjugative transposon, Bacteroidales, Ectoine biosynthesis and regulation, Heme biosynthesis orphans, Phage tail proteins 2.

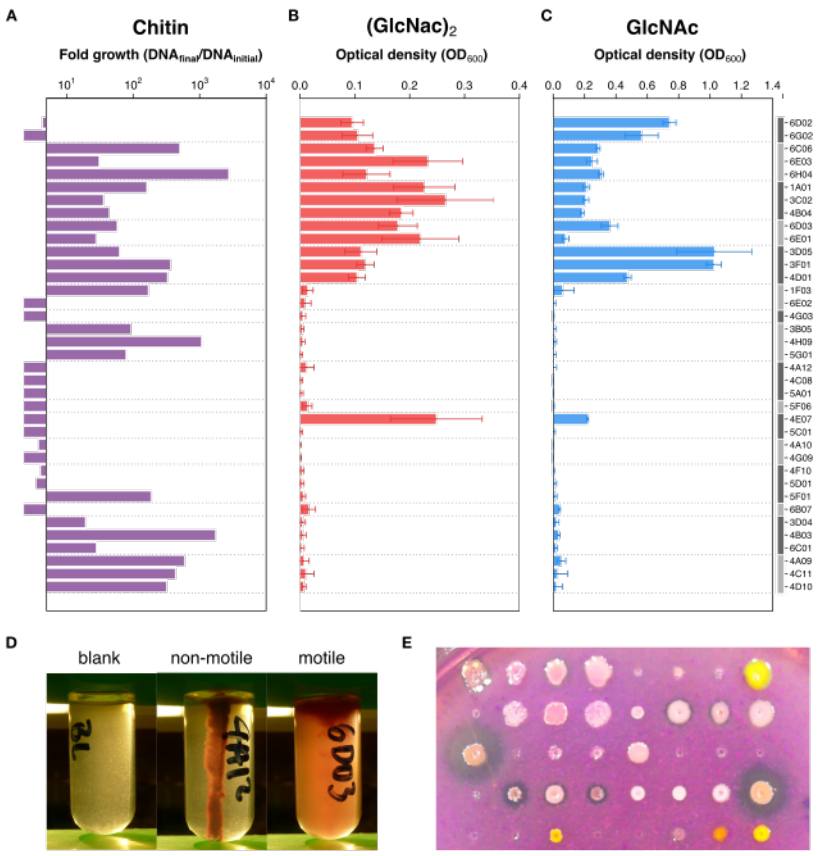


Figure 3-10: Phenotypes of isolate representatives of OTUs from 16S sequencing. Isolate names are indicated on the right and are grouped by the OTUs to which they correspond (light/dark gray bars). Here, OTU order corresponds to ordering in Fig. 3-8b. (A) Isolate fold growth with chitin particles (1,000 particles/mL) as the sole carbon source. Culture growth was assayed by quantifying the amount of DNA present in the sample, both initially (t = 0 days) and after growth (t = 7 days). Fold growth is defined as [DNA]t=7/[DNA]t=0. Isolates for which fold growth < 5 (falling to the left of the y axis) were deemed to have no significant growth. (B) Isolate growth with N-acetylglucosamine (GlcNAc) (0.5% w/v) as the sole carbon source. Total culture yield was assayed after 48 hours via optical density (OD600). Error bars are standard deviations (n=4). (C) Isolate growth with N',N'-diacetylchitobiose ((GlcNAc)$_2$) (0.1% w/v) as the sole carbon source. Total culture yield was assayed after 48 hours via optical density (OD600). Error bars are standard deviations (n=4). (D) Example results from agar stab assay of motility after 7 days of incubation. Left image is an un-inoculated tube (no evidence of growth). Middle image is for a non-motile strain (growth along stab line, no growth elsewhere). Right image is for a motile strain (growth throughout tube). (E) Example from chitinase secretion assay. Each spot corresponds to the growth for an individual isolate. Zones of clearing around a colony, indicating secretion of extracellular chitinases, can be assessed visually.

three did not (Fig. 3-8b; Fig. 3-11), suggesting potential differences in their ability to sustain a non-chitin-metabolizing subpopulation [88]. Altogether, we found that isolates of Phase III-dominant taxa grew robustly on chitin particles in 10 out of 12 co-cultured pairs, despite their inability to grow in monoculture. Indeed, the enhancement of their growth was often quite dramatic; in some cases, isolates grew 1,000-fold (roughly 10 doublings) over seven days in co-culture, with little or no growth in monoculture. Interestingly, the degree of growth enhancement did not depend on whether the chitin metabolizing co-culture partner could broadcast extracellular chitinases (Fig. 3-11).

How do chitin metabolizers facilitate the growth of Phase III-dominant taxa? Previous studies have documented "cheater" strains – specialized for consumption of GlcNAc and $(GlcNAc)_2$ – that do not produce chitinases themselves, but can exploit chitin-degrading taxa by scavenging for their degradation products [13, 138, 46]. However, the taxa that dominated Phase III were unlikely to be canonical cheaters. In particular, gene families involved in GlcNAc and $(GlcNAc)_2$ catabolism decreased in relative abundance from Phase II to Phase III (Fig. 3-8a), suggesting that Phase III was not enriched in taxa that specialized in the consumption of these products. Similarly, only 1 out of the 11 isolated taxa that were highly abundant in Phase III was able to grow in culture with GlcNAc or $(GlcNAc)_2$ as the sole carbon source (Fig. 3-8b; Fig. 3-10b, Fig. 3-10c). However, in the same minimal medium, these isolates could grow on many other carbon sources (Fig. 3-12), indicating that growth deficits stemmed from a lack of a suitable carbon source, rather than auxotrophies or missing co-factors. More broadly, this implies that chitin metabolizers facilitate the invasion of Phase III-dominant taxa by providing them with alternative carbon sources. Possible sources include, but are not limited to, cell debris, biofilm-associated exopolysaccharides, or small metabolic byproducts (e.g. organic acids).

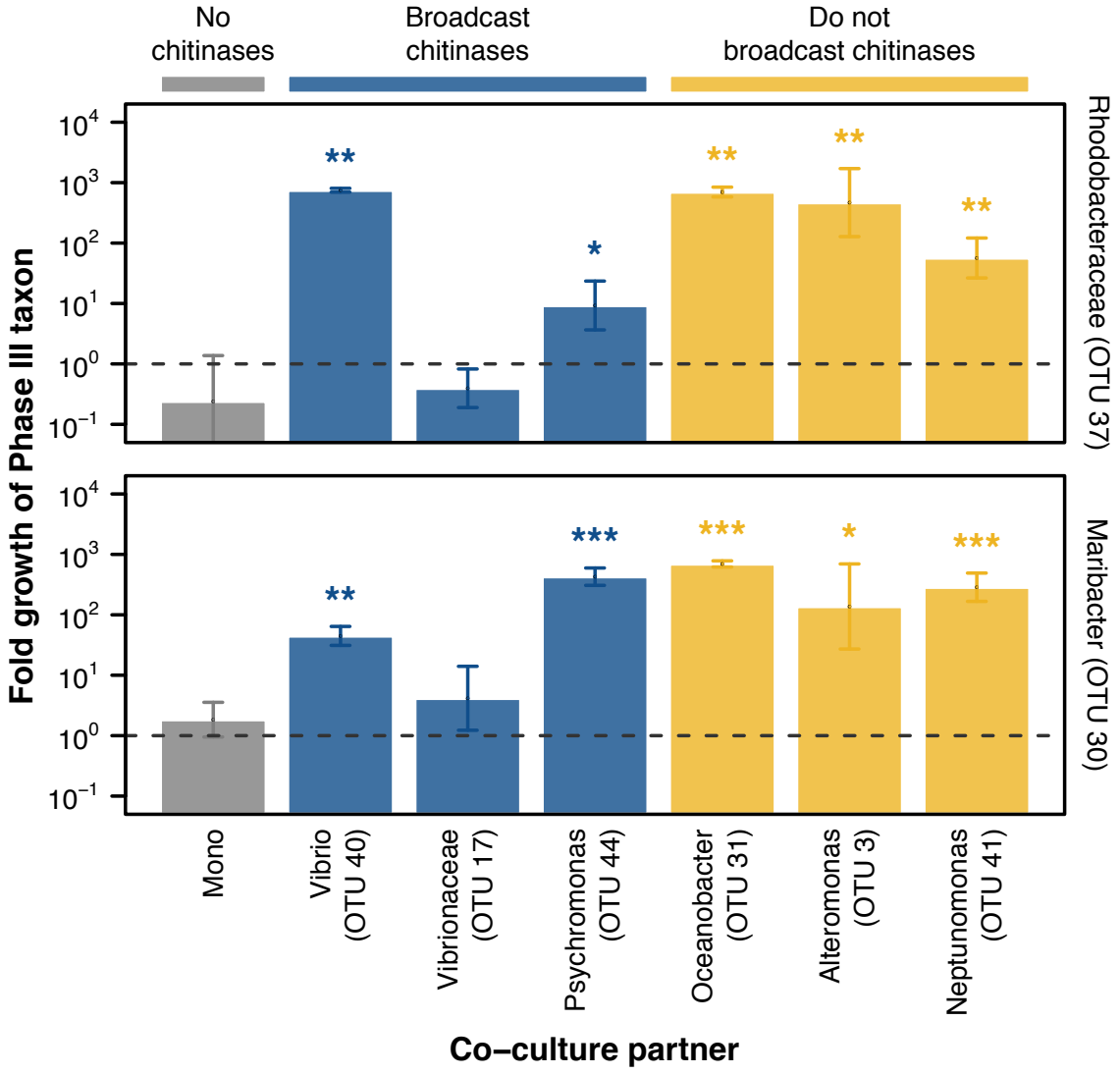


Figure 3-11: Fold growth of two Phase III-dominant taxa (OTUs 37 and 30) in monoculture ("Mono", gray) and in co-culture with chitin-degrading partners ("OTU X", blue/yellow). Blue bars: partners that broadcast extracellular chitinases. Yellow bars: partners that do not broadcast extracellular chitinases. Strains and their co-culture partners were characterized taxonomically with the Ribosomal Protein Database (RDP) classifier. The lowest level of classification with >80% confidence is indicated for each strain. Asterisks: when fold growth in co-culture is significantly different than in monoculture (two-tailed t-test; * = $0.01 < p < 0.05$; ** = $0.001 < p < 0.01$; *** = $p < 0.001$). Error bars are standard deviations over three biological replicates.

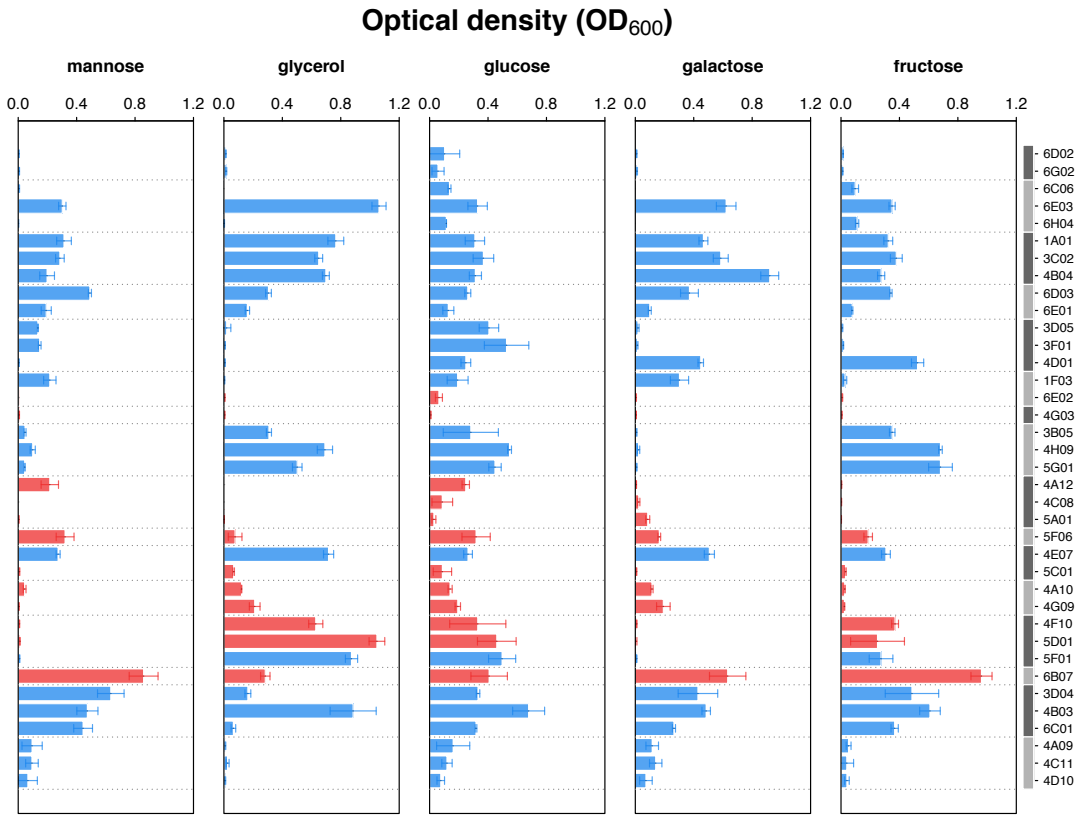


Figure 3-12: Individual isolates were grown with mannose, glycerol, glucose, galactose, or fructose as the sole carbon source (0.5% w/v). Total yield of cultures was assayed after 48 hours via optical density (OD600). Isolate names are indicated on the right. Isolates are grouped by the OTUs to which they correspond (light/dark gray bars), and OTU order corresponds to ordering in Fig. 3-8b. Data for strains that could not grow on chitin, GlcNAc, or $(GlcNAc)_2$ are indicated with red bars. Error bars are standard deviations (n=4).

3.3 Discussion

Overall, we have demonstrated that bacterial communities colonizing nutrient-rich microhabitats undergo successional dynamics driven by two factors – dispersal limitation and facilitative interactions – that, together, drive primary successions at the scale of tens of microns. Together, our results suggest that the existing theory of successions that has been developed for plants and animals may be applied to complex natural microbial communities, thereby providing a basis for linking microbial community structure to their population dynamics and activity.

Our work also illustrates that micro-scale ecological dynamics may have important consequences for global ecosystem processes. In particular, the rapid successional transition from primary particle degraders (in Phase II) to secondary consumers (in Phase III) that we observed in our system suggests that the bacteria commonly found on naturally occurring particles may not be the primary particle degraders. Instead, most particle-attached bacteria may be secondary consumers that recycle waste products from primary degraders. These secondary consumers could increase the biomass yield of the particle-attached community, while decreasing particle degradation rate, as they compete with primary degraders for essential resources like space or oxygen. Therefore, the timescale of this transition could influence the balance between organic matter consumption and biomass buildup in the ocean, potentially a key factor shaping particle remineralization rates in the ocean. Further work should be aimed at understanding the impact of particle-attached community dynamics from microscopic to global scales.

3.4 Methods

3.4.1 Sampling of seawater

Coastal ocean surface water samples were collected on 7 October 2013 from a sampling site located near Northeastern University's Marine Science Center (Canoe Beach, Nahant, Massachusetts, USA; 42°25'11.5"N, 70°54'26.0"W). At the time of sample

collection (roughly 3 PM UTC), the water temperature was 16.5°C, while the ambient air temperature was 18.1°C. Salinity was measured to be 29.7 ppt using a handheld refractometer (VWR #89370-226).

3.4.2 Colonization of chitin particles in seawater

Two milliliters of chitin magnetic beads (New England Biolabs #E8036L; roughly 2.5 x 105 beads/mL) stored in 20% ethanol were washed three times with 50 mL of artificial seawater. Beads were resuspended in 100 mL of artificial seawater, resulting in a bead stock at 5,000 beads/mL. In each of three 1-L screw-cap high-density polyethylene (HDPE) bottles, 16 mL of the bead stock were added to 800 mL of unfiltered seawater, yielding a final bead concentration of 100 beads/mL.

Bottles were rotated end-over-end at 4 rotations per minute on a homemade bottle rotator under ambient lighting and temperature conditions. At each time point and for each bottle sample, 50 mL of the seawater/bead mixture (5,000 beads total) were transferred into a 50-mL conical tube (Corning Life Sciences #352070). Beads were separated from the surrounding seawater with a neodymium magnet (McMaster-Carr #5862K38). The seawater supernatant was transferred back into the original bottle. The beads that remained were then gently washed three times with 50 mL of artificial seawater before being resuspended in 5 mL of artificial seawater. For subsequent analyses (extraction of genomic DNA, isolate collection, and bead imaging), 1 mL of washed beads (containing roughly 1,000 beads in total) was transferred into each of five 1.5-mL Eppendorf tubes. At a subset of time points, 1-mL samples were also collected to characterize the surrounding seawater pool.

3.4.3 Quantification of total particle-attached bacteria over time

As described above, 1,000 beads were prepared from each bottle at each time point. Total bacterial DNA was quantified for each of these samples using a quantitative PCR (qPCR) assay (Appendix B). Briefly, each sample was amplified in a qPCR reaction,

allowing a Ct value (the number of cycles required for the PCR amplification curve to cross a threshold) to be calculated for all samples. The number of 16S V4 copies present in a sample was calculated from the Ct by using a standard curve.

3.4.4 Illumina 16S library preparation

Genomic DNA was extracted from all samples with the MasterPureâĎć DNA Purification Kit (Epicentre #MCD85201) (with modifications described in Appendix B). Amplicon libraries (16S rRNA gene V4 hypervariable region) were prepared according to the method described in Preheim, et al [130]. Samples were sequenced on an Illumina MiSeq (PE 250+250) at the BioMicro Center (Massachusetts Institute of Technology, Cambridge, MA). Reads were merged and quality filtered with custom scripts, and were clustered into operational taxonomic units (OTUs) (97% identity cutoff) with UCLUST and USEARCH (http://www.drive5.com/usearch/). The relative abundance of each OTU was calculated by normalizing per-OTU read counts by the total number of reads in the sample. The absolute abundance of each OTU at each time point was calculated by multiplying the OTU's relative abundance at a given time point by the total amount of particle-attached bacteria at that time point.

3.4.5 Plotting absolute abundance trajectories

Only taxa present at a relative abundance greater than 1% at any time point are shown. Plotted values are the medians over three colonization replicates. Data was smoothed with a three-point running median filter and normalized by the maximum.

3.4.6 Cross-replicate correlations

For all taxa in a taxon subset – present at >1% relative abundance in any replicate and at any time point – we calculated the Spearman correlation between trajectories in two colonization replicates (e.g. Replicate 1 vs. Replicate 2). We repeated this process for all pairs of replicates.

3.4.7 Metagenomic sequencing of particle-attached communities

For a single replicate time series (Replicate 2), metagenomic libraries were prepared for all time points with the Illumina Nextera XT DNA Sample Preparation Kit (Illumina # FC-131-1024) and Illumina Nextera XT DNA Sample Preparation Index Kit (Illumina # FC-131-1001). Sequencing libraries were normalized before sequencing with a modified protocol (Appendix B). All samples were sequenced on an Illumina MiSeq (PE 250+250) at the Genomic Diversity Center (ETH Zürich, Zürich, Switzerland).

3.4.8 Functional annotation of metagenomic reads

Before annotation, reads were quality filtered using the standard quality-control pipeline in MG-RAST (http://metagenomics.anl.gov/). Briefly, reads are pre-processed by using SolexaQA to trim low-quality regions. Subsequently, artificial duplicate reads were identified (using a k-mer approach) and removed. Remaining sequences were screened for matches to model organisms (e.g. human, mouse, cow, fly) and also removed.

Quality-filtered reads were assigned to functional categories using two methods. In the first method (used for "DNA Pol I", "Chemotaxis", "Chitobiose catabolism", "DeAc", and "DeAm"), reads were annotated in MG-RAST with SEED Subsystems, manually curated, hierarchical functional categorization system. Annotation transfer cutoffs were the defaults (e-value $<$ 10-5; minimal alignment length = 60 basepairs), although, qualitatively, results were consistent over a wide range of values. Note that MG-RAST allows a single read to be assigned to multiple functional categories (i.e., # of read annotations $\geq$ # of reads). Therefore, within each sample, counts were normalized to the number of annotations, rather than to the number of reads.

The second method (used for "GH18 family" and "CBP") involved annotation with manually curated HMMs provided by the dbCAN database (http://csbl.bmb.uga.edu/dbCAN/). The database of metagenomic read sequences was searched with profile HMMs for

GH18 (chitinases) and AA10 (chitin-binding proteins) using hmmsearch (HMMER 3.1b1, May 2013, http://hmmer.janelia.org/).

3.4.9 Culturing isolates from particle samples

As described above, tubes containing 1,000 beads were prepared from each incubation bottle at each time point. At a subset of these time points (t = 8, 24, 52, 76, 92, and 140 hours), tubes were sonicated in a bath sonicator (Cole-Parmer #8891, now discontinued) on "low" for five cycles (30 seconds on, 30 seconds off). Beads were separated from the supernatant with a neodymium magnet (McMaster-Carr #5862K38). The supernatant was then divided and plated at three different dilutions (1:104, 1:105, 1:106) on each of two types of plates: (1) Marine Broth 2216 (Difco #279110), or (2) Tibbles-Rawling minimal media with 0.2% N-acetylglucosamine. Both plate types were prepared with 1.5% agar (BD #214010). Plates were incubated at room temperature for 7 days to allow both slow-growing and fast-growing strains to become visible.

Following growth on plates, roughly 50 colonies were picked as representatives from each time point. To purify strains for subsequent analyses, colonies were re-streaked three times onto fresh plates containing the medium from which they were first picked. Stocks of strain isolates were prepared by first growing strains for two days at room temperature in Marine Broth 2216 liquid medium (Difco #279110), and then mixing the saturated culture and 80% glycerol in equal volumes in a cryovial. All stocks were stored at -80°C. For taxonomic classification of isolates, the 16S rRNA gene of each strain was sequenced via Sanger sequencing (Appendix B). This information was used to map isolates to OTUs identified via culture-independent methods (Appendix B).

3.4.10 Growth experiments with isolates

Media was prepared by supplementing Tibbles-Rawling minimal media with the desired carbon source. Strains were pre-grown to saturation for 48 hours in Marine Broth 2216 medium prepared according to the manufacturerâĂŹs instructions (Ma-

rine Broth 2216, dehydrated; Difco #279110). When strains were to be grown on chitin, strains were pre-grown in Marine Broth 2216 medium supplemented with 1,000 chitin beads/mL.

Isolate growth was assessed on each of three carbon sources: chitin resin (New England Biolabs #S6651L; added at 103 beads/mL), N-acetylglucosamine (Sigma-Aldrich #A3286-100G; 0.5% w/v), and N,N-diacetylchitobiose (D1523-10MG; 0.1% w/v). For most cases, culture growth was quantified by measuring OD600 with a spectrophotometer (Tecan Infinite F500). However, growth on chitin beads proved difficult to measure accurately with standard optical density measurements. Instead, culture growth was estimated based upon the change in total DNA content over time. At discrete time points, 500 μL of the culture was transferred into a 1.5-mL tube. Total genomic DNA was extracted from each of these samples using a MasterPureTM DNA Purification Kit (Epicentre #MCD85201; as described above). Total double-stranded DNA content was quantified for each sample with a Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies #P7589).

3.4.11 Chitinase broadcasting assay

A plate-based chitin clearing assay was used to assess chitinase broadcasting ability. Bacterial cultures were grown in Marine Broth 2216 until saturation. A small volume of these saturated cultures (typically 5 μL) was spotted onto a chitin clearing assay plate containing colloidal chitin stained with Remazol Brilliant Violet 5R. Colloidal chitin was prepared and stained as previously described [135], using an aqueous solution of 1.5% w/v sodium dichromate (Sigma #398063-100G) and 1.5% w/v potassium sodium tartrate (Sigma #217255-100G) as a mordant. After cultures were spotted onto plates, plates were incubated at room temperature for 5 days before imaging.

3.4.12 Motility assay

A standard agar stab assay was used to assess the potential for motility among isolates. Motility test agar medium was prepared with Marine Broth 2216 (Difco

#279110), Bacto Agar (BD #214010) (0.25% w/v), and 2,3,5-triphenyltetrazolium chloride (TTC) solution (Sigma-Aldrich #17779-10X10ML-F) and autoclaved. Media was aliquoted in autoclaved glass tubes (VWR #47729-576) with plastic closures (Cole-Parmer #EW-04500-01) (5 mL of media per tube) and allowed to cool to room temperature. Using an inoculating needle (Thomas #TL0000), a stab inoculation was made from a single colony for each strain into a media-filled glass tube. Tubes were incubated at room temperature for 7 days before cultures were analyzed. Evidence of motility was assessed visually. If growth occurred only along the stab line, strains were considered non-motile under these conditions; otherwise, strains were deemed motile. All results were confirmed via microscopy with liquid cultures in Marine Broth 2216 medium.

3.4.13 Isolate co-culture experiments

To characterize interactions between community members, we performed co-culture experiments with pairs of strain isolates. In each case, a chitin-degrading strain was mixed with a non-chitin-degrading strain, with the degrader in large excess (degrader:non-degrader $\approx$ 90:10). Combinations of strains, as well as monocultures of each strain, were grown in Tibbles-Rawling minimal media with chitin beads (1,000 beads/mL) at a total starting cell density of 106 cells/mL. Cultures were grown for 7 days at room temperature and rotated end-over-end at 6 rotations per minute. For each culture, samples were harvested at the beginning (t = 0 days) and end (t = 7 days) of the growth period and frozen at -80°C for subsequent analyses.

Quantifying the total number of cells in each sample is experimentally challenging. Therefore, the total amount of genomic DNA present in each sample was measured. First, total genomic DNA was extracted from each of these samples using a MasterPure DNA Purification Kit (Epicentre #MCD85201; modifications in Appendix B). Then, total double-stranded DNA content was quantified for each sample with a Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies #P7589).

To estimate the relative abundance of each strain within the co-cultures, amplicon libraries (16S rRNA gene V4 hypervariable region) were prepared according to a

previously described protocol [130]. Samples were sequenced on an Illumina MiSeq (paired-end, 250-basepair reads) at the BioMicro Center (Massachusetts Institute of Technology, Cambridge, MA). The absolute abundance of each strain in each sample was calculated by multiplying the strain's relative abundance in a given sample by the total amount of genomic DNA present in that sample.

Chapter 4

Range expansion promotes cooperation in an experimental microbial metapopulation

This chapter was published in a similar form in Datta, *et al.*, 2013 [41]

Abstract

Natural populations throughout the tree of life undergo range expansions in response to changes in the environment. Recent theoretical work suggests that range expansions can have a strong effect on evolution, even leading to the fixation of deleterious alleles that would normally be outcompeted in the absence of migration. However, little is known about how range expansions might influence alleles under frequency- or density-dependent selection. Moreover, there is very little experimental evidence to complement existing theory, since expanding populations are difficult to study in the natural environment. In this study, we have used a yeast experimental system to explore the effect of range expansions on the maintenance of cooperative behaviors, which commonly display frequency- and density-dependent selection and are widespread in nature. We found that range expansions favor the maintenance of cooperation in two ways: (1) through the enrichment of cooperators at the front of the expanding population, and (2) by allowing cooperators to "outrun" an invading wave of defectors. In this system, cooperation is enhanced through the coupling of population ecology and evolutionary dynamics in expanding populations, thus providing experimental evidence for a novel mechanism through which cooperative behaviors could be maintained in nature.

4.1 Introduction

Natural populations often increase the geographical area that they occupy through population growth and dispersal into new territory [53, 71, 175, 120]. Such events, termed range expansions, occur repeatedly during the life history of a species in response to changes in the environment (for instance, forest fires or seasonal changes in climate), as well as changes in phenotype that allow the species to colonize regions that were previously inaccessible to them [53, 181]. Despite their widespread occurrence, the effect of range expansions on genetic diversity has only begun to be explored. Recent theoretical analyses suggest that genetic drift on the low-density front of an expanding population can lead to the fixation of neutral (or even deleterious) alleles in a way that mimics positive selection [54, 101]. This effect, known as "allele surfing," is a stochastic effect that has been demonstrated experimentally for neutral alleles in expanding bacterial colonies [77] and was recently suggested to underlie the global patterns of human phenotypic variation observed today [110, 10, 125].

Here, we chose to study the maintenance of cooperative alleles in the context of populations undergoing range expansions. By definition, a "cooperator" provides a benefit to other members of the population at a cost to itself [126]. Examples of cooperation are ubiquitous in the wild, ranging from siderophore production in bacteria to the formation of communities in human populations [126, 186, 184, 188]. However, given the appearance of "defectors" that exploit the benefit provided by the cooperators without incurring the cost, explaining the origin and maintenance of cooperation in nature remains a key challenge in evolutionary biology [126, 184, 39, 28]. Several studies (both experimental and theoretical) have suggested that spatial structuring may play a key role in the maintenance of cooperation in nature [73, 35, 123, 189]. Moreover, previous studies have showed that population dynamics can play a key role in maintaining cooperative alleles [31, 36]. Thus, we hypothesized that the population dynamics of expanding populations might support the maintenance of cooperation in nature.

More broadly, cooperative alleles also comprise some of the most well known ex-

amples of alleles for which selection is not constant, but instead depends upon instantaneous allele frequencies or population densities [69, 174, 192, 140, 103, 132, 141]. While range expansions are predicted to have a strong effect on the evolution of neutral, deleterious, and beneficial alleles, alleles under frequency- or density-dependent selection have not been considered previously. However, since range expansions often create spatial heterogeneity in both population density and allele frequency [53, 77], we hypothesized that, in contrast to stochastic allele surfing events, there might be a deterministic coupling between population and evolutionary dynamics that would be broadly generalizable to alleles under frequency- or density-dependent selection.

To probe how range expansions shape the maintenance of cooperation, we used a well-characterized budding yeast model system. In this system, a cooperator strain (SUC2) catalyzes the enzymatic conversion of sucrose into a consumable energy source (glucose and fructose) [39, 69, 178, 72]. Although the cooperators retain some preferential access to the sugars that they produce, roughly 99% of this resource is lost to neighboring cells [39, 69, 178]. In our system, this behavior facilitates the growth of an obligate defector strain (Δsuc2) that cannot degrade sucrose, but can consume the sugars produced by neighboring cells [39, 69, 178, 72]. This system shows negative frequency-dependent selection, which leads to coexistence between the cooperator and defector genotypes in our system [39, 69, 178, 72, 109, 76] [for a comprehensive study of the eco-evolutionary feedback inherent in these cooperator-defector dynamics, see (35)]. The cooperator strain also displays an Allee effect (where the per capita growth rate of the population is maximal at intermediate population densities). This type of growth is predicted to have a strong effect on the dynamics of expanding populations [34, 37, 139, 171].

We used an experimental system (similar to that in [92]) based upon a linear stepping-stone model (also known as the Levins metapopulation model) [95, 80, 122]. This theoretical framework, which is well established in the evolutionary and ecological literature, simulates short-range dispersal through nearest-neighbor migration between discrete, well-mixed subpopulations on a lattice (Fig. 4-1A) [95, 80]. Following this scheme, we allowed populations to expand in a habitat of twelve wells (a

single row of a 96-well plate), where each well contained a well-mixed population of cells growing in identical growth media (Fig. 4-1B). To start the experiment, a portion of the wells were populated with cells, while the remainder were unpopulated to allow for expansion. Each day, a fraction m/2 of the contents of each well was transferred into the nearest neighbor wells. Next, the entire population was diluted by a fixed dilution factor (600 in Figures 2-4), which represents a death process in the growth dynamics. Subsequently, the cells grew overnight. The process was repeated for nine days. We note that previous experimental studies of range expansions have focused on expanding colonies on agar plates, which are continuous in both time and space and dominated by stochastic sampling events on the front [101, 77]. However, our experimental setup offers a higher level of control over migration and growth parameters, better approximates the patchiness of the natural environment, and allows us to study deterministic effects that are unlikely to manifest themselves in expanding colonies.

4.2 Results

4.2.1 Cooperators move as a traveling wave

We first explored the spatio-temporal dynamics of pure cooperators undergoing a range expansion. In agreement with theoretical predictions [122], we found that an expanding population of cooperators moves as a traveling wave. Over time, the population adopted a characteristic profile in space consisting of a high-density bulk region and a low-density front (Fig. 4-2A, Fig. 4-2B). The shape of the spatial profile was time-invariant and well fit by the solution to a simple model of range expansions with an Allee effect (see Appendix C and [122] for a more detailed treatment). The functional form of the solution has three free parameters: 1) $\rho_{\max}$, the density in the bulk population, 2) w, the width of the traveling wave, and 3) X_m, the midpoint of the wave (Fig. 4-2C). We also found that the wave profile traveled over time with a constant velocity ($v_{coop} = 0.65 \pm 0.01$ wells/day), which we measured by fitting the

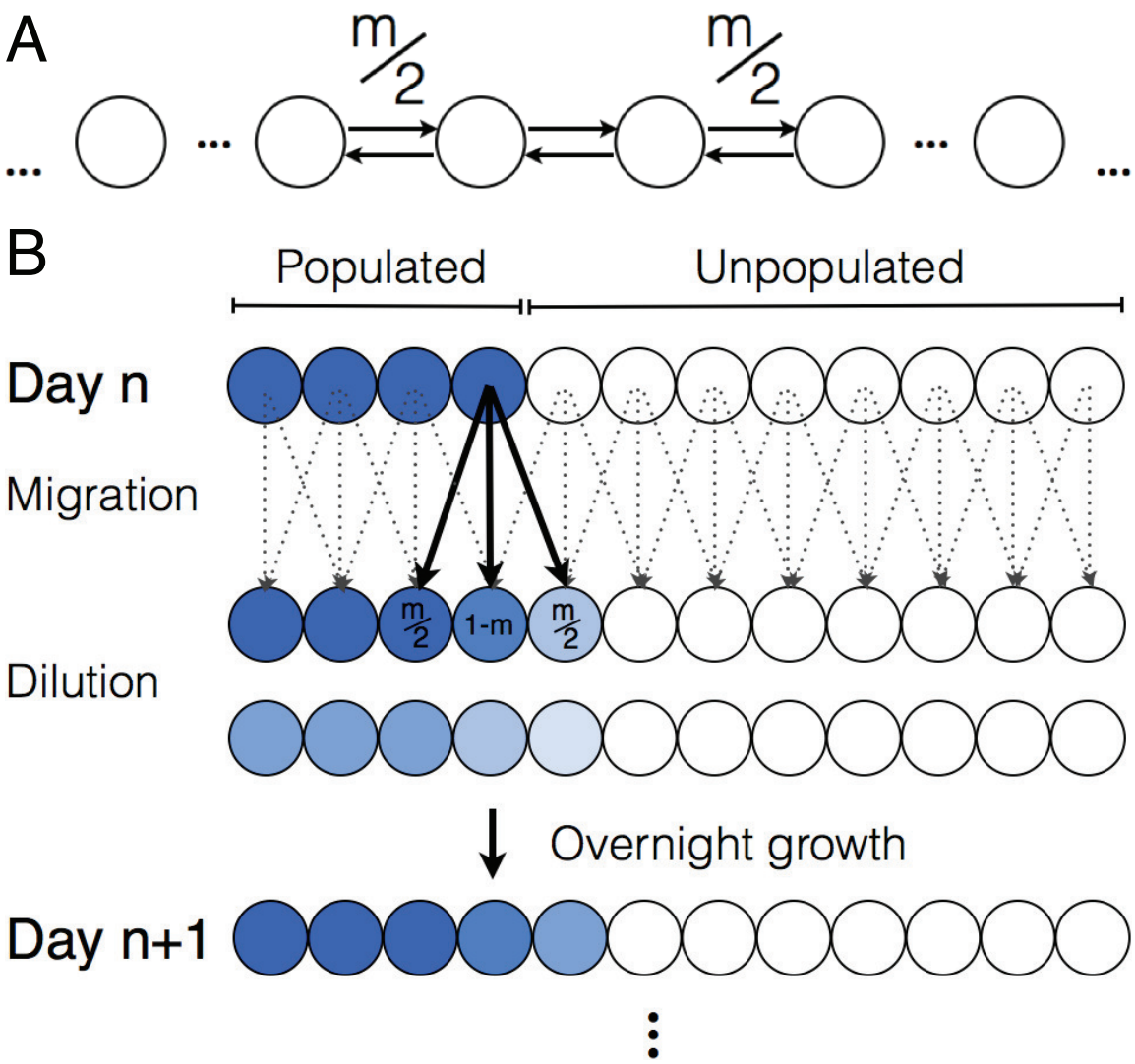


Figure 4-1: (A) A schematic of a linear stepping-stone (or Levins metapopulation) model [95, 80, 122]. The habitat consists of a linear array of subpopulations, each of which is either empty or contains a well-mixed population of individuals. Subpopulations are coupled to each other by nearest-neighbor migration with a fraction m leaving for the two neighboring wells at each time step. (B) Our daily protocol of growth, migration, and death follows the assumptions of the stepping-stone model. Each day, a fraction m/2 of the population from any individual well is migrated into each of the neighboring wells. Boundary conditions are reflective, meaning that wells on the edge receive a contribution m/2 from their only neighbor and 1 – m/2 from the corresponding well. Subsequently, the entire population is diluted by a dilution factor (ranging from 200-1000) that is fixed for the duration of the experiment. The entire population is then grown for 22 hours at 30°C. The process is repeated for several days as the population expands into the unpopulated region. For all experiments, the environmental conditions are spatially homogeneous – all wells (including unpopulated wells) are seeded with identical growth medium (YPD + CSM-His + 2% sucrose + 0.4X histidine).

time trajectory of the wave midpoint (X_m) to a line (Fig. 4-2D). The high quality of these fits ($R^2 = 0.99$ for fit shown here) suggests that the velocity is indeed constant over time. Given the migration scheme, the maximum wave velocity (v_{max}) is 1 well/day, since one additional well is populated each day. However, as illustrated above, the velocity can certainly be less than v_{max}, as it is an emergent property of the expanding population. To our knowledge, this is the first experimental measurement of traveling waves in a population with an Allee effect [139].

4.2.2 Mixed cooperator-defector waves are enriched in cooperators at the front

Next, we considered the range expansion of a mixed population of cooperators and defectors. Similar to the expanding cooperator population, the mixed population also moved as a traveling wave with a constant profile and velocity ($v_{mixed} = 0.49 \pm 0.02$ wells/day) (Fig. 4-3A, Fig. 4-4). Interestingly, we observed that mixed waves of cooperators and defectors traveled more slowly than waves of pure cooperators ($v_{mixed} < v_{coop}$), even though the bulk density of the mixed population was higher than that of the cooperators alone [109] (Fig. 4-2A, Fig. 4-3A, Fig. 4-5A). This result suggests that, in general, intraspecies interactions can have strong and complex effects on the growth and spread of populations.

In addition, we observed significant spatial heterogeneity in allele frequencies within the mixed population wave. Under well-mixed conditions, the cooperator and defector alleles are mutually invasible in our system [69], resulting in stable coexistence between the two alleles at a low frequency of cooperators (Fig. 4-3B, Fig. 4-5B). In agreement with the well-mixed prediction, the high-density bulk population maintained a stable equilibrium frequency of cooperators of about 15%. However, we observed that the frequency of cooperators was significantly larger on the low-density front of the expanding population wave, reaching a frequency that was three times higher than that found in the bulk (roughly 45%) (Fig. 4-3C, Fig. 4-3D). It is also interesting to note that, despite the decline in total population density at the front

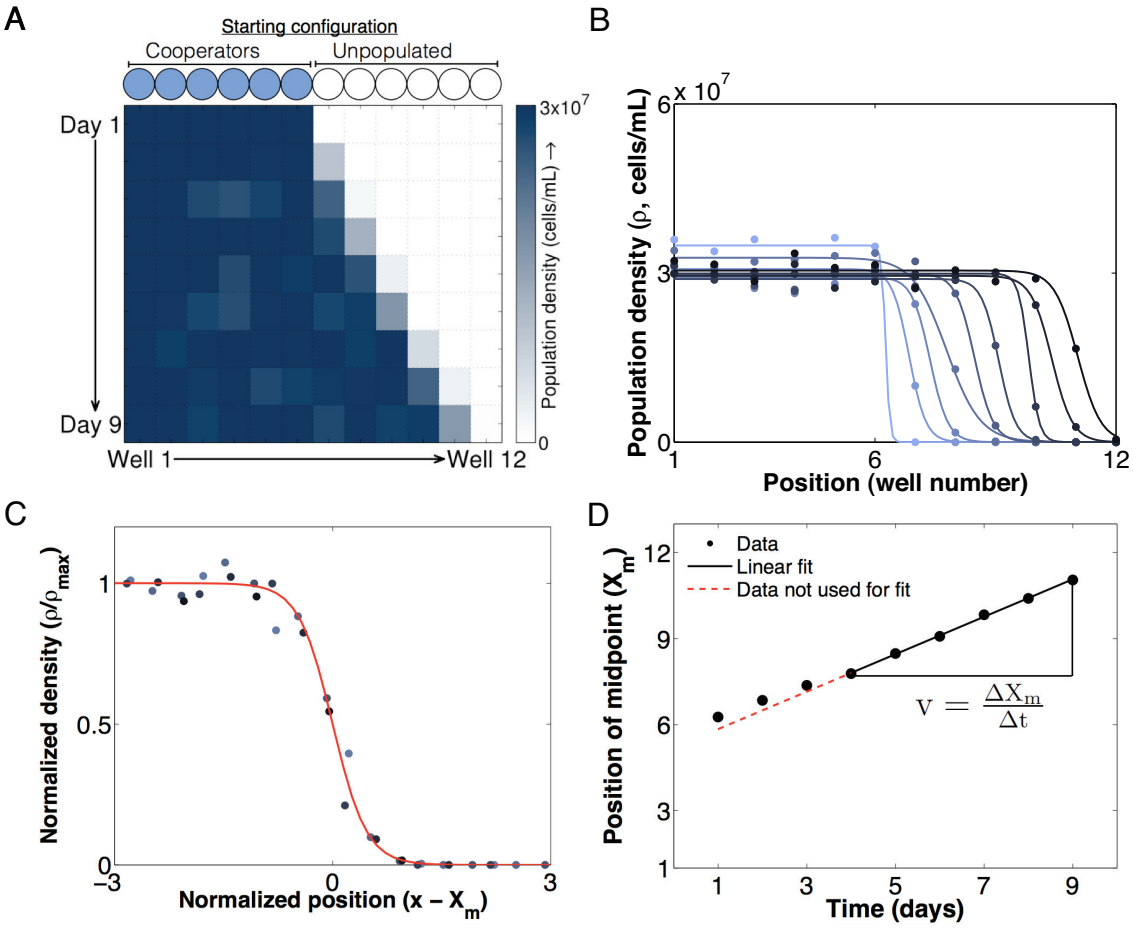


Figure 4-2: (A) The experimentally observed one-dimensional expansion of a pure cooperator population over nine days (m = 0.5 and dilution factor = 600). (B) Density profiles over time for expanding populations of pure cooperators (darker colors indicate later timepoints). Circles are measurements of population density at a particular time and spatial position. Lines are fits of individual profiles to the hyperbolic tangent function derived in Appendix C. (C) An overlay of the density profiles from the last six days of the experiment (darker circles indicate later timepoints). Each profile is normalized to the maximum density found in the bulk population ($\rho_{\max}$) and shifted by its midpoint position (X_m). The first three days are not included, as the expanding population had not yet reached a steady-state profile. The red line shows a theoretical fit to the hyperbolic tangent function derived from a standard reaction-diffusion model of expanding populations (discussed in the Appendix C). (D) We measure the velocity of the traveling wave by plotting the position of the density profile midpoint (X_m) vs. time and then finding the slope of the line. As in (C), the first three days are not included.

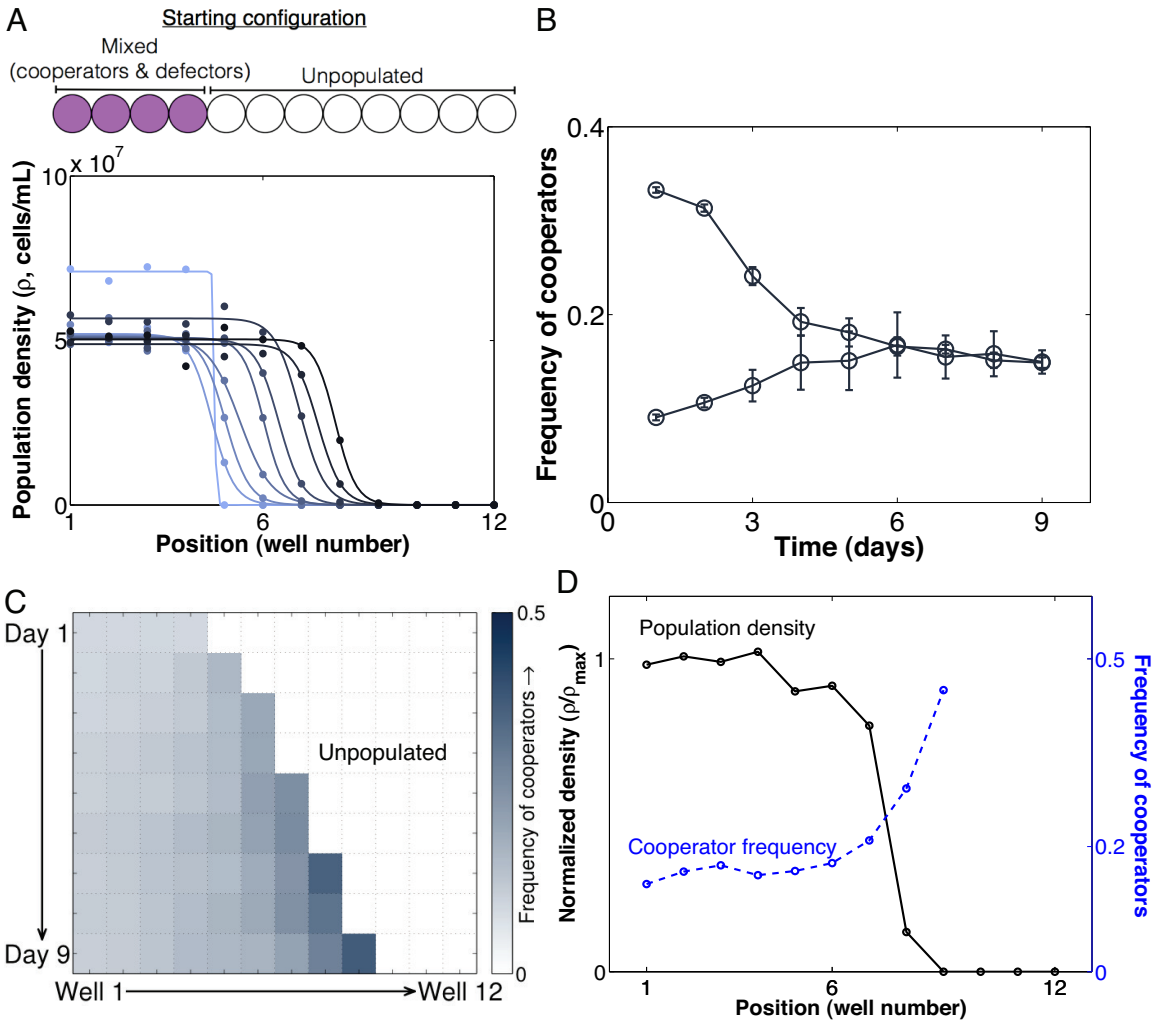


Figure 4-3: A) Population density profiles for the one-dimensional expansion of a mixed cooperator-defector population (m = 0.5 and dilution factor = 600) at its equilibrium cooperator frequency. Darker colors indicate later timepoints. Circles are density measurements from a particular day of the experiment. Lines are fits of individual profiles to the hyperbolic tangent function derived in Appendix C. (B) In a well-mixed population with no dispersal, the cooperator and defector alleles coexist at an intermediate frequency (in this case, roughly 15%). (C) The frequency of cooperators over time in an expanding mixed cooperator-defector population. The frequency of cooperators is strongly enriched on the front of the expanding population wave (45%), while the bulk population remains at the equilibrium frequency predicted in Fig. 4-3B. (D) An overlay of the density profile (in black) with the cooperator frequency profile (in blue) from Day 9 of the expansion of the mixed cooperator-defector population. The density profile is normalized to the maximum density found in the bulk population ($\rho_{\max}$). Cooperators are enriched at the low-density front of the expanding population.

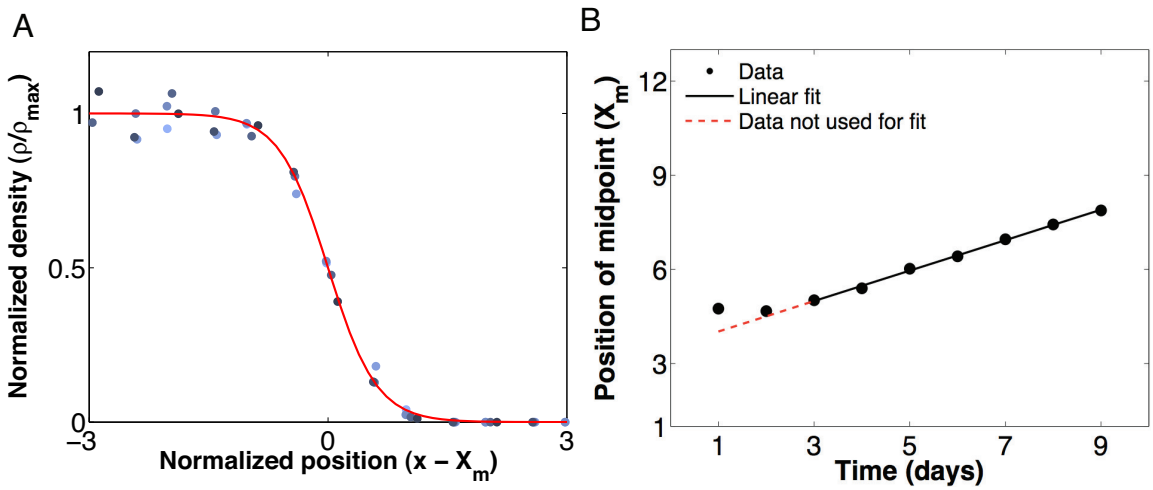


Figure 4-4: (A) An overlay of the density profiles from the last seven days of a one-dimensional expansion of a mixed cooperator-defector population (m = 0.5) and dilution factor = 600) at its equilibrium cooperator frequency. Each profile is normalized to the maximum density found in the bulk population ($\rho_{\max}$) and shifted by its midpoint position (X_m). The red line shows a fit to the hyperbolic tangent function derived in Appendix C. (B) Similar to the pure cooperator wave, we can measure the velocity of the mixed cooperator-defector wave by plotting the position of the density profile midpoint (X_m) vs. time and then finding the slope of the line. As in (A), data from the first two days are not included in the fit, since the expanding population has not reached a steady-state density profile.

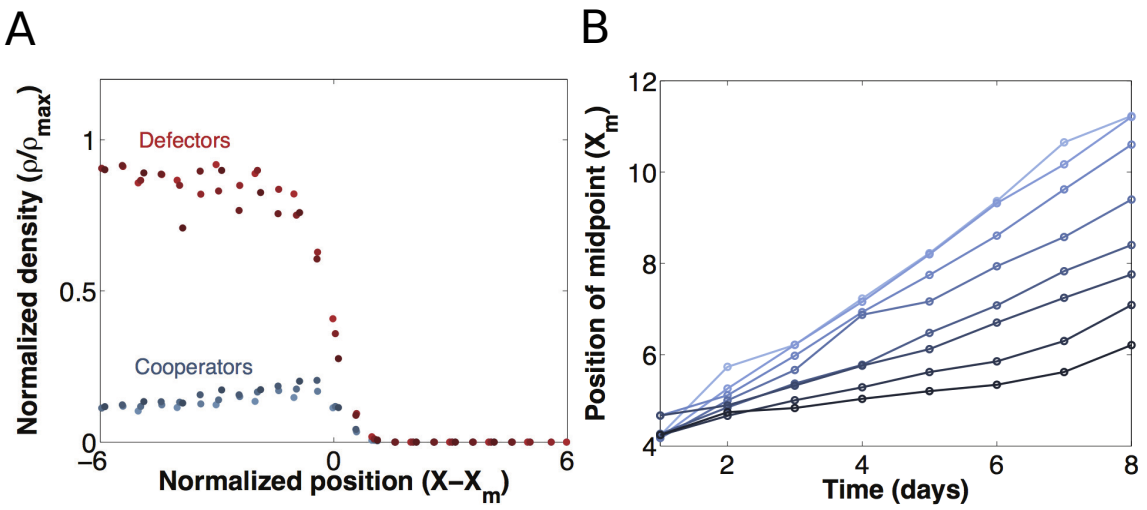


Figure 4-5: (A) An overlay of the cooperator and defector population density profiles within the mixed cooperator-defector wave. Data is shown from the last four days of the experiment (m = 0.5 and dilution factor = 600), where darker circles indicate later timepoints. Each profile is normalized to the maximum density found in the bulk population ($\rho_{\max}$) and shifted by its midpoint position (X_m). Both cooperators and defectors adopt a time-invariant spatial profile, and the cooperator density peaks at a position near the front of the mixed cooperator-defector wave. (B) The midpoint of expanding cooperator waves (X_m) plotted over time for a range of dilution factors between 200-1000 (m = 0.5). Darker colors indicate higher dilution factors. All X_m trajectories are roughly linear in time with slopes that decrease monotonically with increasing dilution factors.

of the expanding population, the density of cooperators actually reaches a peak near the front before declining to zero at the tip of the wave (Fig. 4-6A).

The cooperator cells are enriched at the front of the traveling wave due to their ability to outcompete defectors at low cell densities. At high densities, defector cells can outcompete the cooperators, as they can take advantage of the sugar produced by the cooperators without having to pay the metabolic "cost" of production. However, at low densities, defectors in the population cannot rely on a high density of cooperator cells to provide the sugar needed for growth [69, 142, 27]. Cooperators are able to outcompete defector cells under these conditions because they (selfishly) retain a small fraction of the sugars that they produce [69]. Thus, density-dependent growth dynamics, which allow for stable coexistence between the two alleles [69, 142, 27], also leads to strong deterministic enrichment of the cooperative allele on the low-density front of expanding population waves.

4.2.3 Cooperators can outrun an invading wave of defectors

Next, we considered the process by which defectors invade a spatially extended population of cooperators. As noted previously, in the absence of migration, defectors can invade a population of cooperators over time (Fig. 4-3B). We also observed the invasion of defectors into a spatially extended region of cooperators (Fig. 4-7A, Fig. 4-7B). This invasion takes place via a genetic wave with a time-invariant frequency profile (Fig. 4-7C) and a constant velocity (Fig. 4-7D) that we termed the "invasion velocity" ($v_{invasion} = 0.55 \pm 0.03$ wells/day). Thus, we used our experimental system to characterize the genetic wave of invasion for a competing allele. The notion of a genetic wave was first introduced by Fisher [59] and Kolmogorov [100], but thus far, experimental characterizations of genetic waves in spatially extended populations have been scarce (however, see examples in [125, 112]). Notably, we found that the invasion velocity was less than the velocity of migrating cooperators ($v_{invasion} < v_{coop}$). We were intrigued by this comparison, since it suggested that, if a population of cooperators continued to migrate as it was invaded, the two populations might never completely mix. To explore this idea further, we repeated our measurements of v_{coop}, $v_{invasion}$,

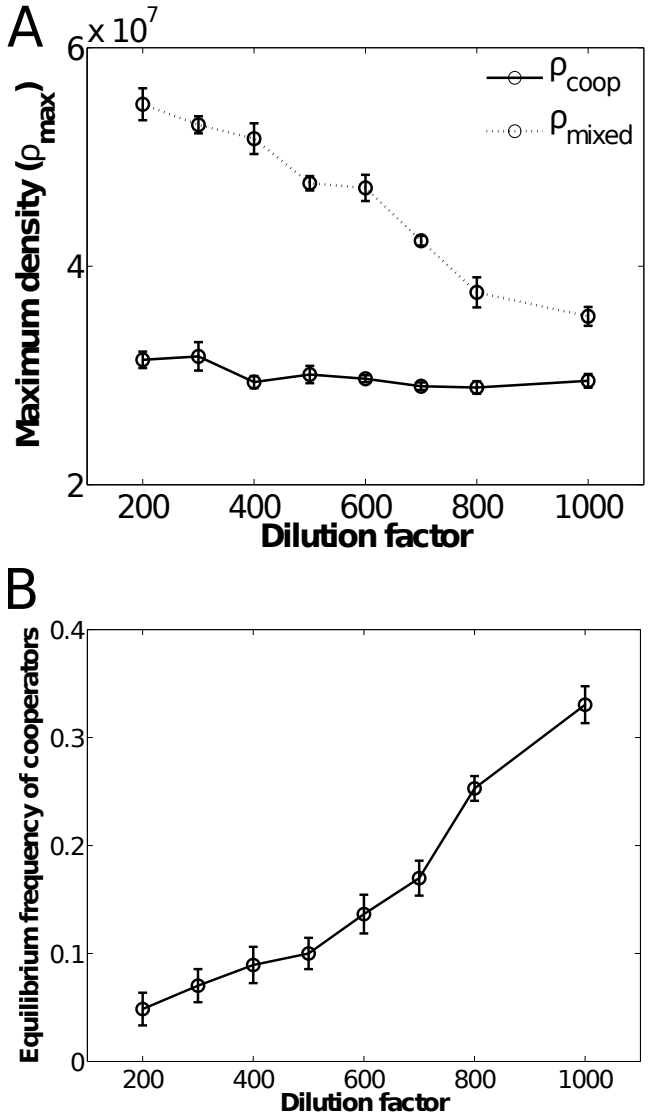


Figure 4-6: (A) The maximum population density (found in the bulk population) over a range of dilution factors between 200-1000. Values of $\rho_{\max}$ were estimated based upon fits of individual density profiles to the hyperbolic tangent function derived in Appendix C. Error bars indicate standard errors in $\rho_{\max}$ from the last five days of the experiment. It is important to note that histidine is limited in the growth media (see Methods). Since the cooperator strain is a histidine auxotroph (and the defector strain is not), these conditions limit the growth of the cooperator strain without strongly affecting the defector strain. Error bars indicate the SEM for measurements with n = 6. (B) The equilibrium frequency of cooperators that would be reached in a well-mixed population undergoing serial growth and dilution by a fixed dilution factor. Estimates of the equilibrium frequency were obtained by averaging the cooperator frequency in the leftmost two wells (those least affected by expansion dynamics on the front) over the last five days of the experiment. Error bars indicate the SEM for measurements with n = 6.

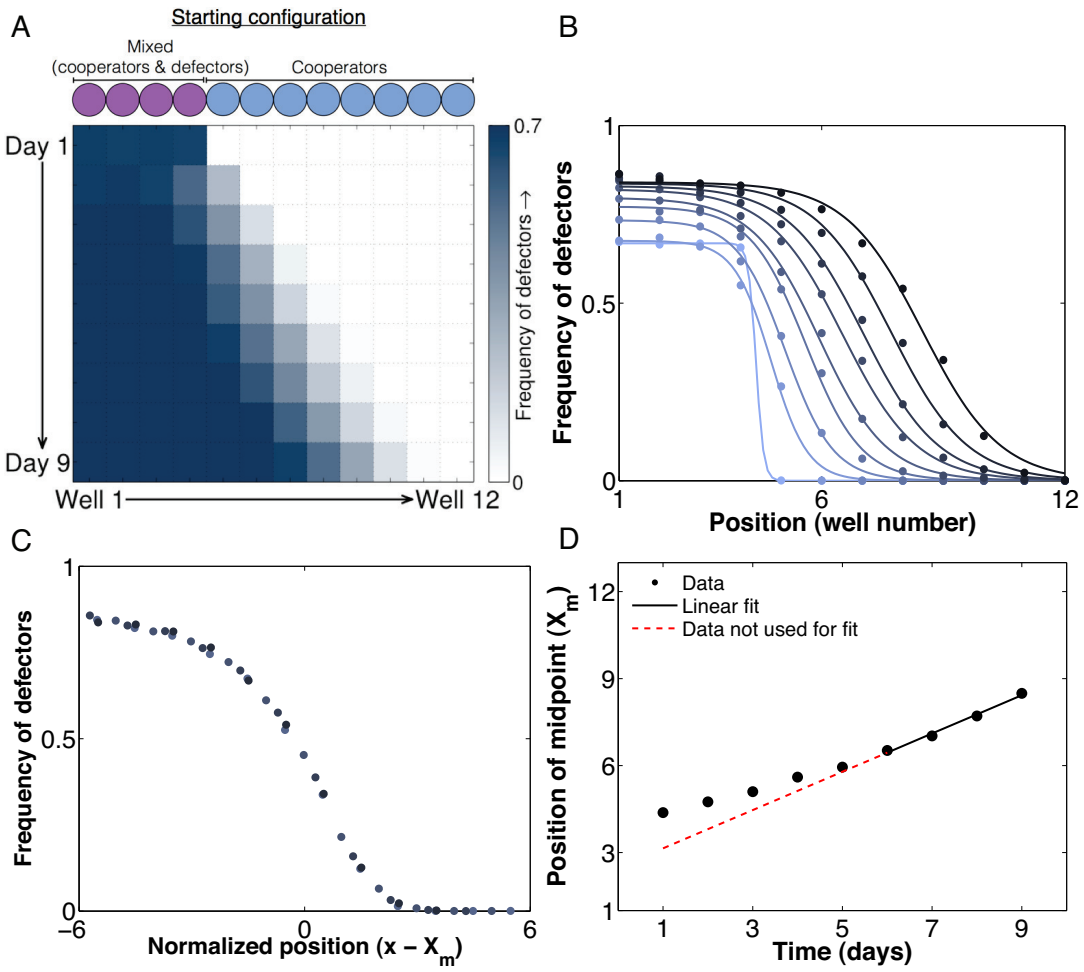


Figure 4-7: Note that, unlike in the previous figures, the frequency of *defectors*, not cooperators, is plotted here. (A) The experimentally observed one-dimensional invasion of defectors into a spatially extended population of cooperators (m = 0.5 and dilution factor = 600). Since defectors (Δsuc2) cannot grow in the medium used here without cooperators present, the invasion begins from a mixed cooperator-defector population near its equilibrium genetic composition. (B) Frequency profiles over time for an invading population of defectors (darker colors indicate later timepoints). Circles are measurements of the frequency of defectors at a particular time and spatial position. Lines are fits of individual profiles to the hyperbolic tangent function derived in Appendix C. It is important to note that this fit is based upon a theoretical prediction, but is simply used as an approximation to the observed sigmoidal shape to infer the wave midpoint (X_m). (C) An overlay of the frequency profiles from the last four days of the experiment (darker circles indicate later timepoints). Each profile is shifted by its midpoint position (X_m). We note that no red theoretical fit line is shown here because, to our knowledge, there is no theoretical prediction for the shape of the genetic wave profile. (D) We can measure the velocity of the invading genetic wave by using the same procedure that we used for the population density wave (Fig. 4-2D). The first six days of the experiment are excluded from the fit, as the population had not yet reached a steady-state frequency profile and density.

and v_{mixed} over a range of dilution factors from 200 to 1000 (Fig. 4-8A). The dilution factor, which modulates the daily death rate, allowed us to vary the dynamics of the system by changing the effective "severity" of the environment. Using this experimental control knob, we hoped to probe a broader range of dynamics in the system.

As we increased the dilution factor, we found that each of the velocities we measured (v_{coop}, $v_{invasion}$, and v_{mixed}) decreased monotonically (Fig. 4-8A, Fig. 4-6B for v_{coop}). For population waves (e.g., v_{coop} and v_{mixed}), this result can be explained by the fact that newly populated wells on the front of the wave cannot grow to measurable densities at higher dilution factors, thereby impeding the spread of the population wave. For the genetic wave ($v_{invasion}$), this monotonic decrease is consistent with the dependence predicted by the Fisher-Kolmogorov model (see [34, 37] and Appendix C). Interestingly, we also found that the mixed cooperator-defector wave traveled more slowly than both the pure cooperator wave and the defector invasion wave over the entire range of dilution factors that we probed (Fig. 4-8A). This result draws comparison to theoretical predictions suggesting that opportunistic pathogens can slow the migration of biological species [83], and it indicates that interactions between cells can have a significant influence on population expansions.

Most surprisingly, although v_{coop} and $v_{invasion}$ showed similar qualitative trends, a comparison of the two velocities immediately revealed two distinct regimes (Fig. 4-8A). At high dilution factors (above 700), we observed that $v_{invasion}$ was greater than v_{coop}, while $v_{invasion}$ was less than v_{coop} at low dilution factors (below 700). These relationships between v_{coop} and $v_{invasion}$ are surprising, given that under well-mixed conditions, cooperators have a growth advantage at low densities (high dilution factors), while defectors are favored at high densities (low dilution factors). However, it is important to note that, while these two expansion processes are both emergent properties of the well-mixed dynamics in individual wells, they occur under significantly different environmental conditions. In particular, cooperators spread by colonizing previously unoccupied regions; thus, at the front of the wave, population densities (and therefore, the concentration of glucose available for consumption) are low. On

the other hand, defectors spread into wells that are already occupied by cooperators, in which population densities are high and growth conditions are favorable (glucose concentration is high). Thus, our analysis shows that the velocities of range expansions are strongly coupled to the environment and can produce relationships that are not immediately apparent from dynamics observed in a well-mixed context. Interestingly, a simple phenomenological model of yeast growth in sucrose that we have used previously [69, 142, 27] predicts our experimentally observed crossing of velocities as a function of dilution factor (Fig. 4-10).

From an ecological perspective, the relationship between v_{coop} and $v_{invasion}$ has several interesting implications. At high dilution factors ($v_{invasion} < v_{coop}$), cooperators and defectors will eventually become completely mixed in space, after which the two alleles will spread together into new territories (Fig. 4-9). In contrast, at low dilution factors ($v_{invasion} > v_{coop}$), cooperators can "outrun" an invading wave of defectors (Fig. 4-8B). In this case, a sufficiently large region of pure cooperators expands over time, even in the presence of an invasive allele (Fig. 4-8C, Appendix C). Broadly, the ability of cooperators to outrun an invading defector wave provides a second mechanism through which range expansions could help to maintain cooperation in nature.

4.3 Discussion

We have presented an experimental study of the effect of range expansions on the maintenance of cooperation. Using yeast populations in the laboratory to study range expansions gave us direct control over all migration parameters and experimental conditions. As a result, we were able to perform a high-resolution analysis of an expanding population of cooperating alleles in a manner that could not easily be replicated in the natural environment.

Using this approach, we found that range expansions favor the maintenance of cooperation in two ways. First, cooperation is strongly enriched on the front of expanding populations compared to the bulk, even without spatial heterogeneity in the

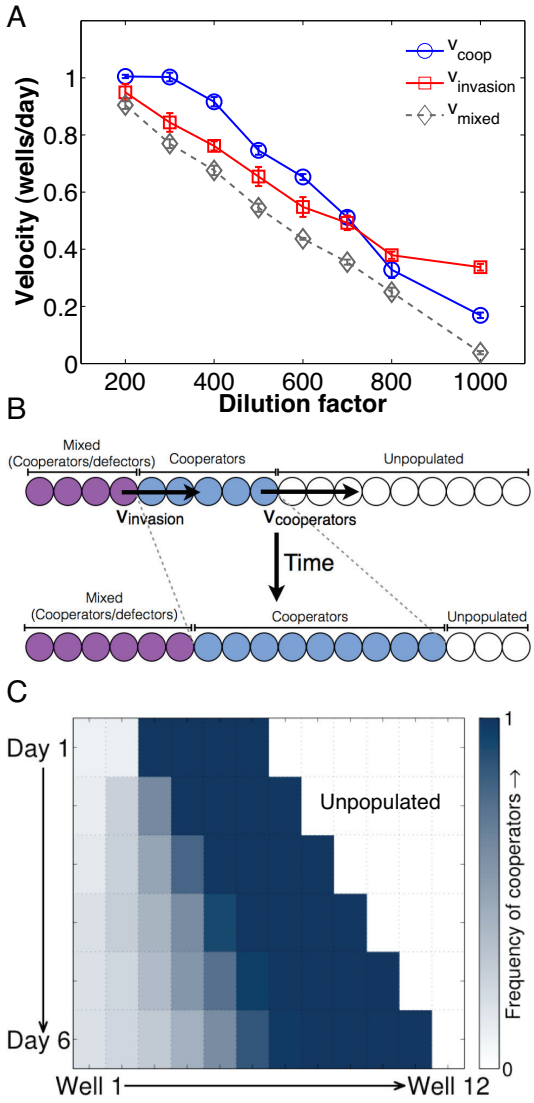


Figure 4-8: (A) Measurement of the velocities of pure cooperators (v_{coop}), invading defectors ($v_{invasion}$), and the mixed cooperator-defector wave (v_{mixed}) over a range of dilution factors indicates two regimes: (1) at high dilution factors, defectors invade more quickly than the cooperators can escape ($v_{invasion} > v_{coop}$), and (2) at low dilution factors, cooperators can "outrun" the invasion ($v_{invasion} < v_{coop}$). Error bars for the velocities are standard errors in the slope of the X_m vs. time plot. Asterisks indicate the magnitude of the p-value for the difference between v_{coop} and $v_{invasion}$ at a particular dilution factor (* denotes $p < 0.05$; ** denotes $p < 0.01$; *** denotes $p < 0.001$). (B) A schematic depicting the case in which cooperators can outrun defectors, in which the region occupied by the cooperators increases over time, even as the defectors invade. (C) Experimental observation of cooperators outrunning an invading wave of defectors ($m = 0.5$ and dilution factor $= 400$). At this dilution factor, $\Delta v = v_{coop} - v_{invasion} \approx 0.2$ wells/day. Over five days, the "headstart" region containing pure cooperators increases from four to nearly five wells, consistent with the observed Δv.

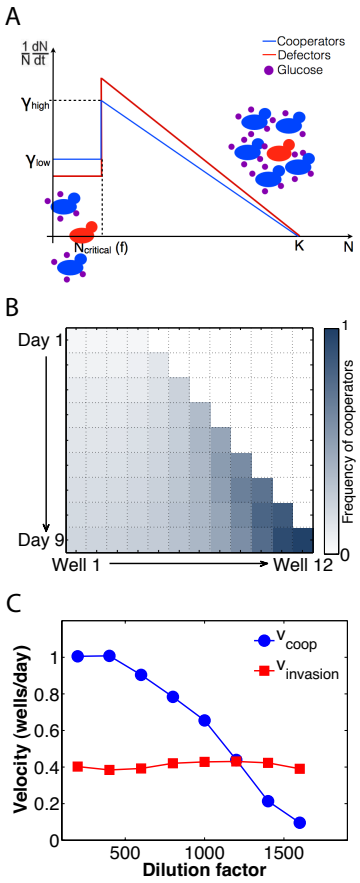


Figure 4-9: (A) Formulation of the model used to simulate growth dynamics in individual wells in discrete simulations. The schematic depicts the per-capita growth rate of cooperators and defectors as a function of cell density. Yeast growth is modeled as logistic with a carrying capacity ($K \approx 10^8$) cells/mL. Low and high density growth phases are delineated by a critical cell density, ($N_{critical} \approx 3 \times 10^5$) cells/mL. Since $N_{critical}$ is over two orders of magnitude lower than K, $\gamma_{c,low}$ and $\gamma_{d,low}$ are assumed to be approximately constant (no logistic decline). However, above $N_{critical}$, the growth rates are assumed to decrease from maximum values (K, $\gamma_{c,high}$ and $\gamma_{d,high}$) according to the logistic equation. (B) Discrete simulation of the frequency of cooperators as a function of time and space. Parameter values are as described in Appendix C. As we found experimentally, the discrete model predicts that the cooperator allele is enriched at the front of expanding populations. (C) Discrete simulation of the frequency of cooperators as a function of time and space. Parameter values are as described in Appendix C. As we found experimentally, the discrete model predicts that the cooperator allele is enriched at the front of expanding populations. (C) Discrete simulation of the velocity of cooperators (v_{coop}) and the velocity of defector invasion ($v_{invasion}$) as a function of dilution factor. Parameter values are as described in Appendix C. As we demonstrated experimentally, the discrete model predicts two growth regimes, in which $v_{invasion} < v_{coop}$ at low dilution factors and $v_{invasion} > v_{coop}$ at high dilution factors.

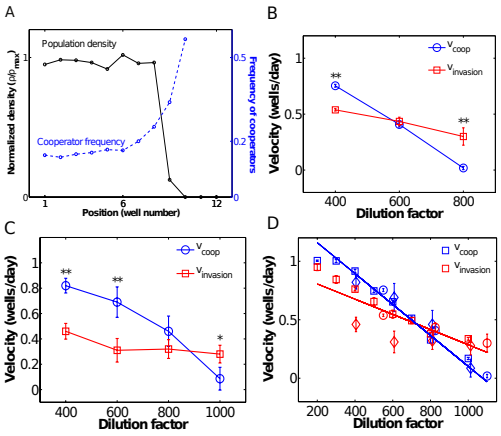


Figure 4-10: (A) An overlay of the density profile (in black) with the cooperator frequency profile (in blue) from Day 9 of the expansion of a mixed cooperator-defector population (m = 0.5 and dilution factor = 400). The density profile is normalized to the maximum density found in the bulk population ($\rho_{\max}$). Consistent with data shown in the main text, the cooperative allele is strongly enriched at the front of the expanding population. (B) Replicate measurements of the velocities of pure cooperators (v_{coop}) and invading defectors ($v_{invasion}$) over a range of dilution factors. Asterisks indicate the magnitude of the p-value for the difference between v_{coop} and $v_{invasion}$ at a particular dilution factor (* denotes $p < 0.05$; ** denotes $p < 0.01$; *** denotes $p < 0.001$). Error bars for velocities are standard errors in the slope of the X_m vs. time plot. (C) Additional replicate measurements of the velocities of pure cooperators (v_{coop}) and invading defectors ($v_{invasion}$) over a range of dilution factors. Asterisks are the same as in (B). In all three experimental replicates (including that shown in the main text), we found that there were two regimes, where at high dilution factors, defectors invade more quickly than the cooperators can escape ($v_{invasion} > v_{coop}$), and at low dilution factors, cooperators can "outrun" the invasion ($v_{invasion} < v_{coop}$). (D) A comparison of the velocity of expanding cooperators (v_{coop}) and the velocity of invading defectors ($v_{invasion}$) as a function of dilution factor. Data is compiled from three independent experiments. To account for experiment-to-experiment variability, all v_{coop} data was normalized by their x intercept (that is, the extrapolated dilution factor where $v_{coop} = 0$) (normalization factors were determined empirically to be 1.4 and 1.01 for data shown in (b) and (c), respectively). Dilution factors for the corresponding invasion velocities were rescaled by the same factor. We then compared the slopes of v_{coop} and $v_{invasion}$ as a function of dilution factor by using a multiple linear regression model (analogous to that discussed in Methods). Although there is a great deal of variation between experiments, the crossing of v_{coop} and $v_{invasion}$ is still statistically significant ($p = 4 \times 10^{-4}$). We note that the exact correspondence between velocity and dilution factor was highly variable between experiments. In simulations, we found that this difference could be explained by a 2% decrease in the growth rates of the two strains, which could easily arise through small variations in the media formulation or growth conditions. As a result, experiment-to-experiment variation primarily captures small variations in media formulation, differences in the temperature of the incubator or the room on a given week, or intrinsic differences between single colonies.

environmental conditions. Given that range expansions are common in natural populations, they could aid in the maintenance of cooperation in nature. Moreover, the mechanism of enrichment can be applied more generally to any allele favored at low densities. Second, we demonstrated that cooperators can "outrun" an invading wave of defectors under certain conditions. Since most natural populations are spatially extended, outrunning provides a plausible mechanism through which cooperation could be maintained in spatially extended populations.

To our knowledge, the two mechanisms that we have demonstrated here are different from those described previously. These novel mechanisms do not invoke kin selection or any form of reciprocity, both of which are deterministic effects that have been shown to lead to stable coexistence between cooperators and defectors under well-mixed conditions (12, 25). Stochastic effects due to population bottlenecks (i.e. allele surfing [53] and Simpson's paradox [31]) are also unlikely to play a large role in our experiments, given the relatively large population size, rates of dispersal between wells, and an Allee effect, which limits population growth below a critical density. Moreover, the mechanisms that we described do not rely on spatial heterogeneity in environmental conditions generated by deteriorating conditions [192, 193] or habitat destruction [193, 23]. Thus, our experiments suggest two distinct deterministic mechanisms in which cooperation could be maintained in a spatially homogeneous environment.

We also believe that our results are broadly generalizable, as many examples of obligate cheating behaviors like those in our system exist in nature. Indeed, several recent studies in microbes isolated from the environment (see [33] for an example) indicate that complete gene loss – which leads to obligate cheating – is a common mechanism through which defectors can arise in natural populations. However, we do note that facultative cheating has been observed in experimental populations [160, 143], which we have not studied here.

Given that cooperation is enriched at the expanding wave front and migrating cooperators can outrun an invading defector wave, we might expect cooperators to be able to "split" – that is, spatially separate themselves – from a mixed population

of cooperators and defectors (see Fig. 4-10 for schematic). This splitting effect would allow a cooperator population effectively to purify itself of defectors in a way that could not be achieved in a non-migrating population. However, recent theoretical predictions suggest that these two features are necessary, but not sufficient for splitting to occur [102]. Indeed, in line with this prediction, we found that the cooperator and defector populations moved at the same velocity within the mixed population wave (e.g., Fig. 4-6), even though the velocities of each population individually were significantly different. This finding suggests that splitting does not occur in our system, even at dilution factors where both cooperation enrichment and outrunning are observed.

In part, the absence of splitting can be attributed to the overall lowered velocity of the mixed cooperator-defector wave compared to the pure cooperators ($v_{mixed} < v_{coop}$ over all dilution factors tested). This suggests that cooperators are "slowed down" by their interactions with defectors, allowing the two alleles to move together in the mixed wave in a way that maintains allelic diversity in expanding populations. A second corollary of this slowing down effect is that, given that both v_{mixed} and v_{coop} decrease monotonically with the severity of the environment, v_{mixed} will reach zero while v_{coop} is still positive. As such, our results suggest that a population of pure cooperators may be able to expand into harsh environments that would be inaccessible to the mixed population. Thus, our experiments provide support for an additional mechanism, which had previously been described theoretically [192], through which cooperation could be aided via ecological constraints.

Overall, we suggest that the coupling of ecological and evolutionary effects drives the spatio-temporal dynamics of cooperation in our system. However, given the generality of our analyses, these ideas can easily be extended to the general case of alleles under frequency- or density-dependent selection, which are a widespread feature of many natural ecosystems. Thus, our findings support the growing, but still underappreciated, notion that eco-evolutionary feedback may dictate the growth and survival of natural populations [69, 142, 27].

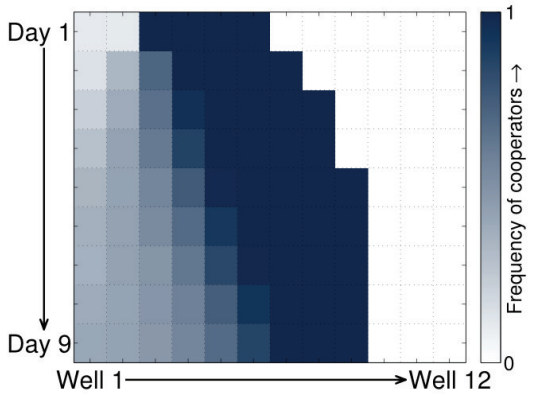


Figure 4-11: Experimental observation of defectors invading a spatially extended population (m = 0.5 and dilution factor = 800). Under these conditions, $\Delta v = v_{coop} - v_{invasion} \approx -0.1$ wells/day. Over nine days, the "headstart" region containing pure cooperators decreases from four to three wells, suggesting that the cooperators and defectors will eventually become completely mixed and spread together under these conditions.

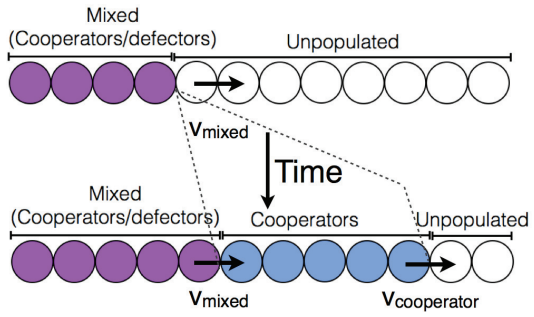


Figure 4-12: In theory, it is possible for cooperators to "split" from an expanding mixed population of cooperators and defectors, after which the cooperators travel as a pure wave ahead of the mixed wave. We did not observe this effect experimentally, even under conditions where (1) the cooperator allele was enriched on the front and (2) the cooperators could "outrun" the defectors. Our result is consistent with theoretical predictions suggesting that these two features are necessary, but not sufficient conditions for splitting to occur.

4.4 Methods

4.4.1 Strains

All strains are identical to those used in Gore, et al. (2009) [69]. Strains were derived from the haploid BY4741 strain of Saccharomyces cerevisiae (mating type a, EUROSCARF). The cooperator strain has a wildtype SUC2 gene, a mutated HIS3 (*his*3Δ1), and has yellow fluorescent protein (YFP) expressed constitutively from the ADH1 promoter (inserted using plasmid pRS401 with a MET17 marker). The defector strain lacks the SUC2 gene (EUROSCARF Y02321, SUC2::kanMX4), has the wildtype HIS3 gene, and has the fluorescent protein tdTomato expressed constitutively from the PGK1 promoter (inserted using plasmid pRS301 containing a HIS3 marker).

4.4.2 Experimental protocols

All experiments were performed in 200 μL batch culture in BD Falcon 96-well Microtest plates. All cultures were grown at 30°C in synthetic media (YNB and CSM-His) supplemented with 2% sucrose and 0.4X (8 μg/mL) histidine. It is important to note that, since the cooperator strain is a histidine auxotroph, the histidine concentration can be used to tune the âĂIJcostâĂİ of cooperation by preferentially limiting the growth of the cooperators [69].

Cultures were shaken continuously at 800 r.p.m. during growth. To avoid evaporation and cross-contamination between wells, plates were covered for the duration of the experiment with Parafilm Laboratory Film. Each day, the frequency of cooperators and defectors was measured via flow cytometry (BD FACS LSR II HTS)(Fig. 4-13). The population density was estimated with flow cytometry and benchmarked with optical density measurements at 620 nm with Thermo Scientific Multiskan FC microplate spectrophotometer. Cultures underwent a migration step and a dilution step each day following 22 hours of overnight growth. During the migration step, a portion m/2 of the cells from each well was transferred into each of the two neigh-

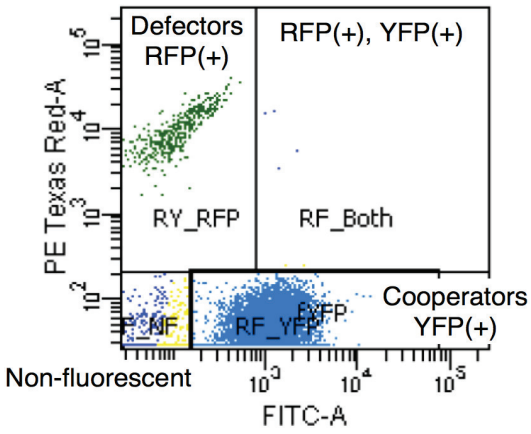


Figure 4-13: The cooperator strain is labeled with YFP that is expressed constitutively from the ADH1 promoter, while the defector strain is labeled with tdTomato expressed constitutively from the PGK1 promoter. We distinguish the two strains with a Becton Dickinson (BD) LSR II HTS flow cytometer with an excitation laser at 488 nm. An emission filter at 530/30 nm detects YFP fluorescence, while a filter at 575/26 nm detects RFP fluorescence. The plot above is from a sample from Day 6 of the expansion of a mixed population of cooperators and defectors. The two strains are distinguished based upon RFP fluorescence and separated with the gates shown. This separation identifies 782 defectors out of a total of 17,983 cells, yielding an estimate of $f = 0.96$ as the frequency of cooperators. A small number of cells (181) were non-fluorescent (gated in the bottom right), and seven counts were deemed both RFP- and YFP-positive, which indicated that multiple cells were detected simultaneously.

boring wells on a new 96-well plate. The remaining 1-m cells were transferred to the corresponding well on the new plate. We used m = 0.5 in all experiments. Boundary conditions were reflective, meaning that wells on the edge received a contribution m/2 from their single neighbor, with the remainder coming from the corresponding well.

Serial dilutions were performed each day with a fixed dilution factor ranging from 200 to 1000. The data presented in Fig. 4-2- 4-7 were diluted by a factor of 600 each day, 4-8C by a factor of 400, and Fig. 4-8A, 4-8B, and 4-6B- 4-5 with dilution factors of 200, 300, 400, 500, 600, 800, and 1000.

It is important to note that all data shown in the main text was obtained from a single nine-day run of the experiment. In three independent experiments, we found that the dilution factor at which particular velocities were observed was highly variable (±20%). However, our core conclusions were robust to this variation in each experiment (see Fig. 4-12 for examples).

Chapter 5

Conclusions

5.1 Summary of findings

In many environments, microbes inhabit spatially structured meta-communities, consisting of many local community aggregates coupled by dispersal between them. How are microbial communities structured within these local communities, and how do microbes self-assemble into these structures? How might inter-species interactions within local communities affect population structure at the level of the meta-community? In the work described in this thesis, we sought to address these questions, using both naturally occurring marine microbial communities and simple laboratory communities as model systems. In particular, this research led to the findings described below.

5.1.1 Local communities formed on individual copepods are shaped by host physiological variability and inter-species interactions

In Chapter 2, we characterized the local bacterial communities inhabiting nearly two hundred individual millimeter-scale crustaceans (*C. finmarchicus* copepods), each of which can be viewed as a habitat "scaffold" with a distinct microenvironment for bacteria to colonize. Although all surveyed copepods shared a common "core" set of bac-

terial taxa, most bacterial taxa were patchily distributed across individual copepods. We found that the distributions of many bacterial taxa were driven by physiological differences between copepod hosts. Thus, fine-scale variability in local community scaffolds may influence biodiversity in bacterial communities. Furthermore, we inferred strong positive and negative associations between groups of bacterial taxa, supporting the role of inter-taxon interactions shaping wild bacterial populations in the ocean.

5.1.2 Bacteria undergo rapid, reproducible succession during local community assembly on model marine particles

Given the inherent physicochemical and life history variability in naturally occurring community "scaffolds", it is difficult to quantify the dynamics of microbial community assembly on these scaffolds. To circumvent this variability, we used a semi-synthetic model system – immersing defined synthetic particle scaffolds into a naturally occurring microbial assemblage from coastal seawater. Using this system, we found that bacteria undergo complex successional dynamics on model particles, with rapid turnover of community members on hourly timescales. These successions were driven by substrate utilization and inter-taxon interactions, suggesting that the local microenvironment shapes bacterial community dynamics. Additionally, the dramatic turnover we observed suggests that local bacterial communities may undergo dynamics on timescales much faster than most traditional sampling schemes.

5.1.3 Range expansion in spatially extended meta-communities promotes cooperation compared to well-mixed environments

In Chapter 4, we used a simple laboratory model system – based upon the classic Levins metapopulation model – to characterize how cooperative alleles spread through an expanding microbial meta-community. Using this approach, we identi-

fied two mechanisms through which cooperative alleles may be strongly favored at meta-community scales, despite being only weakly favored within any individual local community. Both mechanisms are driven by the coupling of ecological and evolutionary effects, which drive the spatiotemporal dynamics of cooperation in our system. Thus, our findings support the growing, but still underappreciated, notion that eco-evolutionary feedback may dictate growth and survival in natural populations. More broadly, our findings also show that interactions within local communities can give rise to non-intuitive allele frequency distributions in a spatially structured environment.

5.2 Future directions

5.2.1 Exploring the microbial milieu beyond bacteria

In characterizing natural microbial communities, the work described in this thesis has focused exclusively on bacteria. However, naturally occurring microbial assemblages – including seawater, our focal ecosystem – contains a much wider range of microbes. These include myriad archaea, viruses, and small eukaryotes, all of which are important players in global ecosystems. Thus, developing a complete picture of local microbial communities requires us to move beyond a purely bacteria-centric view.

5.2.2 Alternative stable states in microbial communities

A natural next step is to characterize the statistical properties of individual, micron-scale local communities in high-throughput. This will allow us to identify robust patterns of taxon co-occurrence. Furthermore, we will able to assess how how inherently stochastic factors (for instance, priority effects) might cause communities to diverge to alternative stable states.

To interpret results in light of underlying ecological interactions, we need to control for variability in patch composition and historical contingencies explicitly. In the ideal case, each patch would be identical in physiochemical composition and life history, but this type of controlled, highly replicated patch structure does not exist in most

natural ecosystems. Therefore, synthetic or semi-synthetic laboratory systems, in which patch properties are tightly controlled, can complement studies of naturally occurring microbial communities.

The model system that we described in Chapter 3 offers a unique opportunity to broach these questions. By controlling the patch size and composition, as well as the pool of potential colonizers, this approach allows us to analyze many individual communities as discrete entities, each of which is a self-organized replicate from the same pool of colonizers. This model system offers a new way to broach the question of "who tends to co-occur with whom" at scales of 10-100 microns, a first step towards reconstructing interactions between taxa in a complex community. By comparing the microbial communities associated with many individual particles, we hope to identify robust statistical associations between taxa across many replicate communities, to probe the space of possible communities, and potentially identify alternative stable states.

5.2.3 Linking dynamics occurring on a single patch to emergent dynamics at meta-community, ecosystem, and global scales

Ultimately, our objective is to develop predictive models that can link individual microbial activity to metapopulation, ecosystem, and global function.

In Chapter 4, we sought to bridge the gap between the local community and meta-community scales with a simple laboratory system. While this laboratory system is an oversimplied abstraction of reality, asking these same questions in a natural setting is extremely challenging. Typically, the mechanisms and rates of dispersal are not known, and the diversity of the surrounding microbial milieu (and interactions between them) complicates analyses. Thus, addressing this problem will require a combination of field studies, biophysical modeling, and controlled experiments in the lab (see methods described in Chapter 1).

Moving beyond the meta-community scale to even higher levels of organization

presents an additional challenge. For various parts of the ocean, several long-term time series have been collected, characterizing the same body of water at "bucket scales" over a period of months or years. Collecting complementary microscale time series at these same sites could be very informative in identifying relationships between dynamics at different spatial scales.

Appendix A

Chapter 2: Supplementary Methods and Data

A.1 Reagent preparation

DNA extraction buffer	
Tris buffer (pH 8)	0.1 M
Sodium EDTA (pH 8)	0.1 M
Phosphate buffer (pH 8)	0.1 M
NaCl	1.5 M
CTAB	0.5% w/v

A.2 Details of sampling individual copepods and seawater

A.2.1 Collection of individual copepods

To collect copepods at two depths (0-50 meters and 250-350 meters), we used a double trip close-open-close system with a 75-centimeter diameter ring and a 150-μm conical mesh net. Once the net was recovered, the contents of the cod end were poured into 10-gallon buckets filled with cold, sterile-filtered (0.22 μm) seawater. Before use,

buckets were wiped with 10% bleach and thoroughly rinsed with Milli-Q water and sterile-filtered (0.22 μm) seawater.

Once back on shore, copepod samples were stored in the dark at 4ÂžC for processing. Live copepods were periodically sieved (500 μm) from the containers, rinsed with cold, sterile-filtered seawater, and placed into an ice-chilled Petri dish. From the Petri dish, they were captured individually using an ethanol-sterilized (95%) wide-bore glass Pasteur pipette, mounted on an autoclaved glass slide, and photographed alive with a Canon EOS-20D camera attached to a Zeiss Stemi 2000C stereomicroscope. After imaging, copepod samples collected for analysis of the associated bacterial community were based into microcentrifuge tubes containing 750 μL of RNALater solution (Sigma-Aldrich #R0901) and frozen (-20°C) until processing. Copepod samples collected for bacterial abundance counts were preserved in 1% formalin and stored at 4°C until processing.

A.2.2 Collection of seawater

To characterize the seawater pool from which copepods were colonized, seawater was sampled at 10-meter intervals from 0-50 meter depths and at 25-meter intervals from 250-350 meters. Water samples were collected in 5-L Niskin bottles. Temperature and salinity measurements were obtained with a SeaBird CTD system. Seawater samples bacterial abundance counts were preserved in 1% formalin. Seawater samples for sequencing were filtered (1 L) onto Sterivex filters (0.22 μm) and stored at -80°C.

A.3 Details of DNA extraction procedure

DNA was extracted according to the following procedure:

1. For seawater samples, Sterivex filters were removed from their casing and cut with a sterile razor blade in a sterile Petri dish.

2. Cut Sterivex filters and individual copepod samples were individually transferred into 2-mL screw-cap tubes with ethanol-sterilized forceps. Each tube

contained a mixture of molecular biology grade 0.1 mm silica, 1.4 mm zirconium, and 4 mm silica beads (OPS Diagnostics PFMM 4000-100-28).

3. To each tube, DNA extraction buffer (400 μL; reagent description in Appendix A) and sodium dodecyl sulfate (100 μL; 10% w/v) were added.

4. Samples were bead-beaten at top speed on a vortexer using a Vortex-Genie2 adapter (Mo-Bio).

5. Lysozyme (20 μL; 10% w/v) was added to each sample, upon which samples were incubated for 30 minutes at 37°C.

6. Proteinase K (20 μL; 10 mg/mL) was added to each sample, upon which samples were incubated for 30 minutes at room temperature.

7. To each tube, phenol:chloroform:isoamyl alcohol (500 μL; 25:24:1 v/v; pH 8) was added, upon which tubes were (a) slowly inverted on a rotating tube holder for 10 minutes and (b) centrifuged at 12,000 x g for 5 minutes.

8. From each tube, the supernatant was extracted with 400 μL of chloroform, and then inverted and centrifuged, as previously described.

9. For each sample, genomic DNA was precipitated at room temperature overnight with 100% isopropanol (0.6 x sample volume) and GlycoBlue (Life Technologies #AM9561) as a co-precipitant.

10. To wash precipitated DNA, DNA was centrifuged at 13,000 x g for 30 minutes and rinsed twice with ethanol (70% v/v).

11. Genomic DNA was resuspended in water (molecular biology grade; 30 μL) and stored at -20°C.

A.4 Details of 16S rRNA V4 amplicon sequencing.

Libraries for 16S rRNA paired-end sequencing were prepared using a previously described protocol [130] with the following modifications:

- Addition of BSA. Copepod DNA has been found to contain inhibitory substances that preclude PCR amplification under standard conditions [117, 22]. Thus, BSA was added to all PCR reactions at a final concentration of 0.4 μg/μL.

- Pre-amplification. Low quantities of DNA template in individual copepod samples prevented proper amplification with the PE16S_V4_U515_F and PE16S_V4_E786_R primers [130], which target the V4 region of the 16S rRNA gene and contain a second-step priming site. Consequently, the copepod samples were pre-amplified with U515F (5'-GTGCCAGCMGCCGCGGTAA-3') and E786R (5'-GGACTACHVGGGTWTCTAAT-3') primers, each of which are shorter primers that lack the second-step priming site. Pre-amplification was performed for 15 cycles to enrich for 16S rRNA V4 template. During the subsequent amplification step, the samples were amplified for 10 cycles with the PE16S_V4_U515_F and PE16S_V4_E786_R primers [130].

- Q5 polymerase. All reactions were performed with the Q5 High-Fidelity DNA Polymerase (New England Biolabs #M0491L), a polymerase whose fidelity is purportedly higher than that of Phusion polymerase.

More specifically, the amplification steps were performed according to the protocols described below. At each PCR step, amplifications were performed in quadruplicate 25-μL reactions. Replicate amplifications were pooled and cleaned using Agencourt AMPure XP-PCR beads (Beckman Coulter #A63880) before proceeding to the next step. Quantitative PCR reactions were performed before each PCR step to optimize the number of amplification cycles. Selected PCR products were also visualized on 2% TBE (tris-borate-EDTA) agarose gels to confirm the presence of the correct amplicon at each step.

A.4.1 Pre-amplification

Primers

515F (GTGCCAGCMGCCGCGGTAA)

806R (GGACTACHVGGGTWTCTAAT)

PCR reaction setup

	Volume (1 reaction)
ddH2O	11.75 μL
5X Q5 Buffer	5 μL
dNTPs (10 mM)	0.5 μL
515F (3 μM)	2.5 μL
806R (3 μM)	2.5 μL
BSA (20 μg/μL)	0.5 μL
Template	2 μL
Q5 polymerase	0.25 μL
Total	25 μL

Cycling conditions

Step	Temperature	Time
Initial denaturation	98°C	30 seconds
Amplification (15 cycles)	98°C	10 seconds
	50°C	60 seconds
	72°C	90 seconds
Final extension	72°C	10 minutes

A.4.2 Amplification

Primers

PE16S_V4_U515_F (ACACGACGCTCTTCCGATCTYRYRGTGCCAGCMGCCGCGGTAA)
PE16S_V4_E786_R (CGGCATTCCTGCTGAACCGCTCTTCCGATCTGGACTACHVGGGTWTCTAAT)

PCR reaction setup

	Volume (1 reaction)
ddH2O	11.75 μL
5X Q5 Buffer	5 μL
dNTPs (10 mM)	0.5 μL
PE16S_V4_U515_F (3 μM)	2.5 μL
PE16S_V4_E786_R (3 μM)	2.5 μL
BSA (20 μg/μL)	0.5 μL
Template	2 μL
Q5 polymerase	0.25 μL
Total	25 μL

Cycling conditions

Step	Temperature	Time
Initial denaturation	98°C	30 seconds
Amplification (10 cycles)	98°C	30 seconds
	52°C	30 seconds
	72°C	30 seconds
Final extension	72°C	10 minutes

A.4.3 Adapter addition

Primers

PE-III-PCR-F: (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT)
PE-IV-PCR-XXX: (CAAGCAGAAGACGGCATACGAGATXXXXXXXXXXCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT)

PCR reaction setup

	Volume (1 reaction)
ddH2O	8.65 μL
5X Q5 Buffer	5 μL
dNTPs (10 mM)	0.5 μL
PE-III-PCR-F (3 μM)	3.3 μL
PE-IV-PCR-XXX (3 μM)	3.3 μL
Template	4 μL
Q5 polymerase	0.25 μL
Total	25 μL

Cycling conditions

Step	Temperature	Time
Initial denaturation	98°C	30 seconds
Amplification (10 cycles)	98°C	30 seconds
	83°C	30 seconds
	72°C	30 seconds
Final extension	72°C	10 minutes

A.5 Details of multivariate linear regression analyses.

Here, we have outlined the procedure that we used to identify drivers of bacterial abundance across individual copepods.

1. *Estimate relative abundances of each OTU.* Normalizing the number of counts for a given OTU by the total counts in the sample overestimates the number of zeros for rare OTUs. This can lead to artifacts driven by variability in sequencing depth across samples, rather than true biological variability. Moreover, for cases in which data must be log-transformed, zeros are problematic. As an alternative, we used a Bayesian framework (detailed in [63]) to estimate the true relative abundances from observed count data. Briefly, it has been shown that the posterior joint distribution of fractions is equal to the Dirichlet distribution, assuming a uniform prior and unbiased sampling of reads during sequencing. Therefore, we estimated OTU fractions in the following manner:

$$p(\bar{x}|\bar{N}) = \mathrm{Dir}(\bar{N} + 1)$$

where $\bar{x}$ is a vector of the true fractions, and $\bar{N}$ is the vector of observed counts. The authors note that this approach implicitly assumes that all OTUs are present in a sample, even if they were not detected, and that this assumption may not be reasonable for rare OTUs. However, our analyses deal primarily with OTUs whose mean abundance was comparatively high.

2. *Renormalize relative abundances to subset of 241 abundant OTUs.* Previously, we had identified a subset of 241 OTUs whose mean relative abundance was above a defined threshold in both sampling dates. Given that this was the subset that we considered in quantifying inter-taxon correlations, we re-calculated their relative abundances by normalizing to the total number of counts assigned to these OTUs.

3. *Calculate the relative abundance for each cluster, the core microbiome, and those not in clusters or the core.* Each of the 241 OTUs could be assigned uniquely to the following nine categories: clusters 1-7, the core microbiome, or not in any cluster or the core. For each of these categories, the total relative abundance was simply the summed relative abundances of all OTUs assigned to that category.

4. *Calculate the additive log ratio-transformed abundances for each category.* For each category, we performed the additive log-ratio transform described in Materials and Methods. To perform this transform, we used the alr function from the R compositions library [176].

5. *For each cluster, estimate the probability that a given explanatory variable has a non-zero effect on the log ratio-transformed cluster abundance.* For each cluster, we fit a multivariate regression model to the log ratio-transformed cluster abundances. This model was of the following form:

$$\begin{aligned}\log\left(\frac{f_{\text{cluster i}}}{f_{\text{core}}}\right) =& \beta_d(\text{vertical sampling depth}) + \beta_b(\text{body volume}) + \\ & \beta_o(\text{oil sac fullness}) + \beta_f(\text{food in gut}) + \\ & \sum_{j\neq i}\beta_j \log\frac{f_{\text{cluster j}}}{f_{\text{core}}} + \beta_n \log\frac{f_{\text{not in cluster or core}}}{f_{\text{core}}}\end{aligned} \tag{A.1}$$

To account for model uncertainty, we used Bayesian model averaging. This allowed us to calculate the probability that a given explanatory variable has a non-zero effect on the cluster abundance across all possible regression models formulated from combinations of explanatory variables [134]. To perform this calculation, we used the bicreg function from the R BMA library.

We also considered alternative model formulations that included interactions between the copepod vertical sampling depth and copepod morphological characteristics (Fig. S8). These interaction terms added flexibility to the model

by allowing for effects of explanatory variables to differ between shallow- and deep-sampled copepod populations.

A.6 Supplementary Tables

	Prosome volume (mm^3)	Oil sac fractional fullness	Cells per copepod individual
Copepods assayed			
shallow	97	95	20
deep	89	89	20
Mean			
shallow	0.94	0.43	4.5×10^5
deep	0.60	0.68	2.7×10^5
Standard deviation			
shallow	0.16	0.21	1.1×10^5
deep	0.25	0.17	1.0×10^5

Table A.1: Summary statistics for copepod morphological characteristics. Details of how morphological characteristics were assayed for individual copepods are described in Materials and Methods and Appendix A.

A

Predictors	Df	Sum of squares	F statistic	R^2	p-value
Depth	1	5.58	111.87	0.319	0.000001
Residuals	182	11.9		0.681	
Total	183	17.48			

B

Predictors	Df	Sum of squares	F statistic	R^2	p-value
Body volume	1	6.81	136.50	0.390	0.000001
Food in gut	1	0.80	15.96	0.046	0.000009
Oil sac fractional fullness	1	0.32	6.49	0.019	0.003329
Day collected	1	0.21	4.14	0.012	0.020445
Depth	1	0.46	9.19	0.026	0.000424
Residuals	178	8.88		0.508	
Total	183	17.48			

Table A.2: PERMANOVA summary tables. In each table, "Df" means "degrees of freedom". (A) PERMANOVA summary statistics when copepod vertical sampling depth ("Depth") is used as the sole predictor of bacterial community composition. (B) PERMANOVA summary statistics when copepod morphological characteristics (e.g., "Body volume"), as well as copepod vertical sampling depth, are used as predictors of bacterial community composition.

Appendix B

Chapter 3: Supplementary Methods and Discussion

B.1 Supplementary Methods

B.1.1 Preparation of common reagents

Artificial seawater

A mixture of 40 g of sea salts (Sigma #S9883-1KG) and 1 L of Milli-Q deionized water was prepared. This mixture was filtered through a 0.22-μm filter using vacuum filtration (Corning Life Sciences #CLS430517). Note that artificial seawater is a mixture of salts that does not contain sources of carbon or nitrogen.

Minimal medium (no carbon source)

Minimal medium, containing sources of nitrogen, phosphorus, and sulfur, but not carbon, was prepared with a protocol adapted from Tibbles and Rawling [170].

Required reagents

- NaCl (Sigma-Aldrich #S3014)
- $MgSO_4 - 7H_2O$ (Sigma-Aldrich #63138)
- $MgCl_2 - 6H_2O$ (Sigma-Aldrich #M2393)

- $CaCl_2 - 2H_2O$ (Sigma-Aldrich #C7902)
- NH_4Cl (Sigma-Aldrich #A9434)
- Tris (1M, pH 8.0) (ThermoFisher Scientific #AM9855G)
- Disodium EDTA (0.5 M, pH 8.0) (ThermoFisher Scientific #AM9260G)
- K_2HPO_4 (Sigma-Aldrich #P3786)
- KH_2PO_4 (Sigma-Aldrich #P5655)
- $FeSO_4 - 7H_2O$ (Sigma-Aldrich #F8633)
- $Na_2MoO_4 - 2H_2O$ (Sigma-Aldrich #331058)
- Vitamin solution, 1000X (as previously described [58])
- Trace metals solution, 1000X (as previously described [170])

For 1 liter of minimal medium:

1. Prepare "Part I" (2X) solution.

   (a) Add the following components to deionized water (final volume of 450 mL).

Component	Amount
NaCl	51.9 g
$MgSO_4 - 7H_2O$	6 g
$MgCl_2 - 7H_2O$	4 g
$CaCl_2 - 2H_2O$	0.24 g
Tris (1M, pH 8.0)	50 mL
Na_2EDTA(0.5M)	5.4 mL
NH_4Cl	20 mL

   (b) Adjust pH of solution to 7.8.

   (c) Add deionized water to a final volume of 500 mL.

   (d) Autoclave solution to sterilize.

2. Prepare "Part II" (2X) solution.

   (a) Add the following components to deionized water (final volume of 450 mL).

Component	Amount
K_2HPO_4	1.6 g
KH_2PO_4	0.4 g

(b) Adjust pH of solution to 8.0.

(c) Add deionized water to a final volume of 500 mL.

(d) Autoclave solution to sterilize.

Note: For long-term storage, keep Part I and Part II in separate containers.

3. Once cooled, combine Parts I and II with the following additives to produce 1 L of media:

Component	Amount
Part I (2X)	250 mL
Part II (2X)	250 mL
$FeSO_4$ solution (1000X)	1 mL
Na_2MoO_4 solution	1 mL
Vitamins (1000X)	1 mL
Trace metals (1000X)	1 mL
Carbon source solution	?
Deionized water	to 1000 mL

B.1.2 Genomic DNA extractions

Before DNA extractions were performed, all samples were frozen at -80°C for purposes of long-term storage and cell lysis. Total genomic DNA was extracted from samples using a MasterPure DNA Purification Kit (Epicentre #MCD85201). The protocol for "fluid samples" was used with the following modifications:

- Reagent volumes were scaled up to accommodate a 500 μL sample.
- Before isopropanol was added, glycogen (Fermentas #R0551) was added to each sample (final concentration of 0.5 μg/μL) to increase yield.
- Following the addition of isopropanol, samples were stored at -20°C overnight to increase DNA yield.

- DNA pellets were resuspended in 50 μL of autoclaved Milli-Q deionized water, rather than in Tris-EDTA buffer.

B.1.3 Quantification of total particle-attached bacteria

To quantify the total number of 16S V4 copies per particle over time, we performed quantitative PCR (qPCR) assays for each sample as described below.

Setup for qPCR reactions

Primers

- 515F (GTGCCAGCMGCCGCGGTAA)
- 806R (GGACTACHVGGGTWTCTAAT)

PCR reaction setup

	Volume (1 reaction)
ddH_2O	9.4 μL
5X HF Buffer	4 μL
dNTPs (10 mM)	0.4 μL
515F (10 μM)	2 μL
806R (10 μM)	2 μL
Template	2 μL
Phusion	0.2 μL
SYBR Green I (200X)	0.1 μL

"Phusion" = Phusion High-Fidelity Polymerase (New England Biolabs #M0530L). SYBR Green I working stock (at 200X) was prepared by diluting the original concentrated stock provided by the manufacturer (10,000X) with DMSO.

Cycling conditions

Step	Temperature	Time
Initial denaturation	98°C	30 seconds
Amplification (35 cycles)	98°C	10 seconds
	50°C	60 seconds
	72°C	90 seconds
Final extension	72°C	10 minutes

Quantitative PCR reactions were performed with a Bio-Rad CFX96 Real-Time PCR Detection System.

Preparation of standard curve A sample of 16S V4 amplicons (amplified from a mixture of equal volumes of genomic DNA from all timepoints) was used to prepare the standard curve. Total double-stranded DNA content was measured for this sample using a Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies #P7589). Standards were prepared with serial 10-fold dilutions (8 in total) from this sample. The concentration for each of these standards was assumed to be consistent with the dilution. Quantitative PCR was performed in triplicate for each of these samples (with the protocol described above), thereby allowing a Ct value to be estimated for each standard.

In accordance with theory, a plot of the Ct value vs. log(DNA concentration) was found to be linear over a wide range of DNA concentrations, saturating at very low concentrations. Linear regression was used to obtain the equation for the best-fit line of the non-saturating data points in log-linear space. This equation was used to calculate the amount of DNA in each sample.

Limit of detection

- Identify the most concentrated standard whose Ct is indistinguishable from the no-template controls.
- Calculate the standard deviation of the Ct
- $\text{LOD} = \text{C}_{\text{t,neg}} + 2 * \text{SD}(\text{C}_{\text{t,neg}})$

B.1.4 Imaging of particle-attached communities

Samples were fixed for imaging by adding formalin at 1% by volume, and then storing the samples for 24 hours at 4°C. After this incubation period, the formalin/artificial seawater mixture was removed from particles, using a neodymium magnet (McMaster-Carr #5862K38) to separate particles from supernatant. Particles were re-suspended in PBS and stored at 4°C.

Before imaging, samples were stained with a double-stranded DNA stain (SYBR Green I Nucleic Acid Gel Stain; Life Technologies #S-7563). A working stock of the DNA stain (at 200X) was prepared in DMSO from the concentrated stock solution (10,000X). This stock was added at 0.5% v/v (final concentration was 1X in solution) to each bead sample. Beads were imaged with a Zeiss epifluorescence microscope at 40X magnification. Excitation and emission spectra of SYBR Green I are published by the manufacturers.

B.1.5 Calculating the effective number of species (N_{eff}) in a community

The effective number of species (N_{eff}) was calculated based upon the Shannon diversity of the community as described in Jost, 2006 [91] and indicated below.

$$\text{Shannon diversity} = D = -\sum p_i \ln p_i$$

$$N_{eff} = \exp D$$

B.1.6 Metagenomic sequencing of particle-attached communities

The amount of DNA present was measured with a Qubit dsDNA HS (High Sensitivity) Assay Kit (ThermoFisher Scientific #Q32851).

For library preparation, 1 ng of each sample was used. Libraries were prepared with the Illumina Nextera XT DNA Sample Preparation Kit (Illumina # FC-131-

1024) and Illumina Nextera XT DNA Sample Preparation Index Kit (Illumina # FC-131-1001). Since our libraries resulted in a wide range of fragment sizes and were at low concentrations overall, the Nextera XT library normalization protocol was not used. Instead, a double-sided size selection was performed with Agencourt AMPure XP beads to select for DNA fragments between 300-700 basepairs. The concentration of fragments between 300-700 basepairs was assessed with an Agilent 2100 Bioanalyzer, after which samples were pooled.

B.1.7 Sanger sequencing of 16S rRNA for isolates

To quantify the total number of 16S V4 copies per particle over time, we performed quantitative PCR (qPCR) assays for each sample as described below.

Setup for qPCR reactions

Primers

- 27F (AGAGTTTGATCMTGGCTCAG)
- 1492R (TACGGYTACCTTGTTACGACTT)

PCR reaction setup

	Volume (1 reaction)
ddH_2O	9.4 μL
5X HF Buffer	4 μL
dNTPs (10 mM)	0.4 μL
515F (10 μM)	2 μL
806R (10 μM)	2 μL
Template	2 μL
Phusion	0.2 μL

"Phusion" refers to Phusion High-Fidelity Polymerase (New England Biolabs #M0530L). Instead of using genomic DNA as the template, 2 μL of a saturated culture (grown

in Marine Broth 2216; Difco #279110) diluted 1:1000 in deionized water was added directly to the PCR reaction tube.

Cycling conditions

Step	Temperature	Time
Initial denaturation	98°C	30 seconds
Amplification (35 cycles)	98°C	30 seconds
	50°C	30 seconds
	72°C	90 seconds
Final extension	72°C	10 minutes

Following the PCR reaction, isopropanol precipitation was used to purify the PCR products. Two Sanger sequencing reactions were performed for each of these purified PCR products, using either the forward (27F) or reverse (1492R) PCR primer as the sequencing primer (GENEWIZ, Inc.). The sequences obtained from the two reactions were merged to obtain, in the majority of cases, the fully sequenced PCR construct.

B.1.8 Mapping isolates to OTU sequences

Isolates were mapped at 100% identity to the OTU V4 sequences identified during 16S sequencing. Note that, while isolates mapped to a particular OTU are identical in the 16S V4 hypervariable region, they are often not identical in other regions of the 16S rRNA gene.

B.2 Supplementary Discussion

B.2.1 Bacteria are the dominant particle colonizers at early stages of colonization

Seawater contains a diverse range of microbial life, including not just bacteria, but also viruses, phytoplankton, and zooplankton, among others. However, in this work,

we have largely focused on the bacterial dynamics of particle colonization for the following reasons:

- **Bacteria are typically the first colonizers of naturally occurring particles and dominate at early stages.** It has been observed that, during the early phases of particle decomposition, bacterial abundance increases rapidly. Subsequently, as bacterial biofilms form and particles aggregate, microscopic bacterivores and other eukaryotes increase in abundance within particle-attached communities. This pattern has been observed for naturally occurring particles [20], as well as in simplified laboratory systems [97].

- **Empirically, bacteria were the dominant particle colonizers on the timescale of our experiment (Fig. 3-3)**.

Nonetheless, it is possible that non-bacterial microbes influence bacterial colonization dynamics. This includes turnover due to viral predation or eukaryotic grazers, or interkingdom signaling between phytoplankton and bacteria [5]. This will undoubtedly be an interesting subject of future work.

B.2.2 Alternative modes of chitin degradation

In this study, we observed a small subset of taxa that were:

- Able to grow on chitin as the sole source of carbon
- Do not secrete chitinases extracellularly
- Could not grow on GlcNAc or $(GlcNAc)_2$ as the sole source of carbon

In the canonical mode of chitin degradation, bacteria secrete chitinases into the environment. These chitinases are active extracellularly, thus allowing bacteria to degrade insoluble chitin polymers with which they do not have direct contact.

Given that the taxon subset of interest does not secrete chitinases extracellularly, they do not follow this canonical strategy. However, an alternative mode of chitin degradation has been documented in which bacteria do not release chitinases

extracellularly, but instead, tether them to their surfaces. In such cases, bacteria are assumed only to degrade chitin with which they have made surface contact. It is hypothesized that these taxa are slower to degrade insoluble chitin, but are also less prone to invasion by cheaters. We have not demonstrated for this taxon subset that the chitinases are indeed surface-tethered. However, there may be interesting mechanisms (and resulting ecological dynamics) that warrant further exploration.

Even without releasing chitinases extracellularly, we might still expect taxa with surface-tethered chitinases to be able to grow on common chitinase degradation products (GlcNAc and $(GlcNAc)_2$). However, the isolates that we collected in this study did not. We have not identified the mechanism that allows for this particular growth pattern. However, we have listed some hypothetical mechanisms below that could be tested experimentally:

- **The chitinases produced by these taxa generate larger (or different) enzymatic degradation products.** Thus, these strains are not genetically equipped to import or metabolize GlcNAc or $(GlcNAc)_2$, but may be able to metabolize other degradation products.

- **These strains consume chitin degradation products, but only in the presence of chitin.** To our knowledge, such a mechanism has not been documented. However, it is known that chitin degradation itself is dependent on both the genetic and environmental context. Thus, consumption of chitin degradation products may be similarly regulated.

- **These strains do not produce chitinases, but instead produce chitin deacetylases.** Such strains could produce acetate (and glucosamine) as enzymatic byproducts, rather than GlcNAc and $(GlcNAc)_2$. We would predict that such strains would grow well on acetate and/or glucosamine as the sole source of carbon. Notably, chitin deacetylases have been found in marine bacteria, fungi, and insects, but have largely been studied in fungal species.

Appendix C

Chapter 4: Supplementary Methods and Discussion

C.1 Supplementary Text

C.1.1 Analytical results regarding population density waves

Although empirical studies of expanding populations are few and far between, the ecological theory of range expansions has a long history. Spatial models of range expansions based upon reaction-diffusion equations were first discussed by Fisher [59] and Kolmogorov [100] in the late 1930s, and a good summary of the results are found in [122]. Surprisingly, the wave front profiles that we observed empirically were well approximated by these continuous time, continuous space models, even though our experiments were discrete in time and space. Here, we summarize the formulation of a model of range expansions and some useful results. In accordance with our experiments, we only discuss one-dimensional expansions here.

The dynamics of range expansions are dictated by both the growth of the population and the dispersal of individuals into unpopulated territory. For short-range, isotropic dispersal, the process can be modeled with a diffusion term. Thus, the model

can be formulated as the reaction-diffusion equation shown below:

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} + G_c(c)c \tag{C.1}$$

where $c(x,t)$ is the population density at position x and time t, D is the effective diffusion coefficient for population dispersal, and G_c is the per capita growth rate of the population.

In principle, $G_c(c)$, could be any one of a number of functions, depending upon the growth dynamics of the specific population in question. In our cooperatively growing yeast populations, we found that $G_c(c)$ depends non-monotonically on the population density ($c(x,t)$) [37]. In general, a habitat has a carrying capacity (K) due to resource limitation, and populations usually grow at a reduced or negative rate close to this upper bound on the density. Populations displaying cooperative behaviors also tend to grow slowly or not at all at low densities, since interactions between individuals are limited. Thus, unlike the standard logistic model, the per capita growth rate is maximized at an intermediate population density. This non-monotonic dependence of the per capita growth rate on the population density is known as the Allee effect [122, 34, 37]. The most common model of growth with an Allee effect assumes the following form for the per capita growth rate:

$$G_c(c) = g_c(K - c)(c - c^*) \tag{C.2}$$

where K is the carrying capacity, c^* is the critical population density, and g_c modulates the overall magnitude of the per capita growth rate [122, 34]. The strong Allee effect describes the case in which $c^* > 0$, while a weak Allee effect occurs when $-\frac{K}{2} < c^* < 0$ [34].

This reaction-diffusion equation admits traveling wave solutions with a time-invariant density profile that moves at a constant velocity. Although nonlinear partial differential equations of this type are often difficult to solve analytically, exact solutions for the velocity and shape of the wave profile are known exactly. The expression for the velocity is shown below:

$$v = \begin{cases} \sqrt{\frac{Dg_c}{2}(K - 2c^*)}, & \text{if } c^* \geq -\frac{K}{2} \\ 2\sqrt{Dg_c K|c^*|}, & \text{if } c^* < -\frac{K}{2} \end{cases} \tag{C.3}$$

Previous work has shown that our experimental system demonstrates a strong Allee effect ($c^* > 0$) and that K and c^* approach each other with increasing dilution factor [37]. Thus, the analytical expression shown above predicts that the traveling wave velocity should decrease with increasing dilution factor, in line with what we observed experimentally.

The shape of the time-invariant wave profile is given by

$$c(\xi) = \frac{\mathrm{K}}{1 + e^{\sqrt{\frac{g_c}{2D}}\mathrm{K}\xi}} = \frac{\mathrm{K}}{2}\left[1 - \tanh\frac{1}{2}\sqrt{\frac{g_c}{2D}}\mathrm{K}\xi\right] = \frac{\rho_{\max}}{2}\left[1 - \tanh\left(\frac{x - \mathrm{X_m}}{w}\right)\right] \tag{C.4}$$

where $\xi = x - vt$, and we set $K = \rho_{\max}$, $\xi = x - \mathrm{X_m}$, and $w = 2K\sqrt{\frac{2D}{g_c}}$ to connect this analysis to our experimental results.

The population density wave profiles that we observed experimentally were well fit by this functional form. Thus, by fitting profiles from each day to this function, we obtained estimates of $\mathrm{X_m(t)}$ and $\rho_{\max}(\mathrm{t})$.

C.1.2 Analytical results regarding genetic waves

The spreading of cooperator and defector alleles can also be modeled with a reaction-diffusion equation like the one shown below:

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2} + G_f(f)f \tag{C.5}$$

where $f(x,t)$ is the frequency of the defector allele at position x and time t, D is the effective diffusion coefficient for population dispersal, and G_f is the relative growth rate of defectors, which is a function of f to model frequency-dependent dynamics.

Frequency-dependent selection is most often modeled with the following function:

$$G_f(f) = g_f(1-f)(f^*-f) \tag{C.6}$$

where $g_f \geq 0$ is the strength of selection and f^* is the equilibrium frequency of defectors in a well-mixed population.

Using this model and the results from [100], the velocity of defectors invading a spatially extended population of cooperators can be shown to be

$$\mathrm{v}_{\mathrm{invasion}} = 2\sqrt{Dg_f f^*} \tag{C.7}$$

As shown previously, the equilibrium frequency of defectors decreases with increasing dilution factor. Thus, if g_f remains constant, the above result predicts that $\mathrm{v}_{\mathrm{invasion}}$ should decrease monotonically with the dilution factor.

C.1.3 Discussion of outrunning

In the main text, we suggest that a "sufficiently large" leading region of cooperators is required for outrunning to occur. To develop this idea further, we can consider the relative movement of the invasion wave (consisting of a mixture of cooperators and defectors in our experiments) and the pure cooperator wave, both of which have finite widths because organisms are discrete entities [78]. If the separation between the two waves is such that the invasion wave front ends before the cooperator wave front begins, then outrunning will occur if and only if $\mathrm{v}_{\mathrm{coop}} > \mathrm{v}_{\mathrm{invasion}}$.

Interestingly, the notion of outrunning could be extended to a wide range of systems outside of the case in which cooperators outrun an invading wave of defectors. A particular example with frequency-dependent selection is considered in the Supplementary Appendix of [79].

Bibliography

[1] Rachel I Adams, Marzia Miletto, John W Taylor, and Thomas D Bruns. Dispersal in microbes: fungi in indoor air are dominated by outdoor air and show dispersal limitation at short distances. *The ISME Journal*, 7(7):1262–1273, July 2013.

[2] J. Aitchison. *The statistical analysis of compositional data*. Monographs on statistics and applied probability. Chapman and Hall, London ; New York, 1986.

[3] Alice L Alldredge and Mary W Silver. Characteristics, dynamics and significance of marine snow. *Progress in Oceanography*, 20(1):41–82, January 1988.

[4] Amalia A. Almada and Ann M. Tarrant. Vibrio elicits targeted transcriptional responses from copepod hosts. *FEMS Microbiology Ecology*, page fiw072, April 2016.

[5] S A Amin, L R Hmelo, H M van Tol, B P Durham, L T Carlson, K R Heal, R L Morales, C T Berthiaume, M S Parker, B Djunaedi, A E Ingalls, M R Parsek, M A Moran, and E V Armbrust. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature*, 522(7554):98–101, May 2015.

[6] S. A. Amin, M. S. Parker, and E. V. Armbrust. Interactions between Diatoms and Bacteria. *Microbiology and Molecular Biology Reviews*, 76(3):667–684, September 2012.

[7] Marti J. Anderson. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1):32–46, February 2001.

[8] E. G. Arashkevich, K.S. Tande, A. F. Pasternak, and B. Ellertsen. Seasonal moulting patterns and the generation cycle of Calanus finmarchicus in the NE Norwegian Sea, as inferred from gnathobase structures, and the size of gonads and oil sacs. *Marine Biology*, 146(1):119–132, December 2004.

[9] Amalia M. Aruda, Mark F. Baumgartner, Adam M. Reitzel, and Ann M. Tarrant. Heat shock protein expression during stress and diapause in the marine copepod Calanus finmarchicus. *Journal of Insect Physiology*, 57(5):665–675, May 2011.

[10] Q. D. Atkinson. Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science*, 332(6027):346–349, April 2011.

[11] Farooq Azam and Francesca Malfatti. Microbial structuring of marine ecosystems. *Nature Reviews Microbiology*, 5(10):782–791, October 2007.

[12] Edo Bar-Zeev, Ilana Berman-Frank, Olga Girshevitz, and Tom Berman. Revised paradigm of aquatic biofilm formation facilitated by microgel transparent exopolymer particles. *Proceedings of the National Academy of Sciences of the United States of America*, 109(23):9119–9124, June 2012.

[13] Sara Beier and Stefan Bertilsson. Bacterial chitin degradation-mechanisms and ecophysiological strategies. *Frontiers in microbiology*, 4:149–149, 2013.

[14] D. Berry, B. Stecher, A. Schintlmeister, J. Reichert, S. Brugiroux, B. Wild, W. Wanek, A. Richter, I. Rauch, T. Decker, A. Loy, and M. Wagner. Host-compound foraging by intestinal microbiota revealed by single-cell stable isotope probing. *Proceedings of the National Academy of Sciences*, 110(12):4720–4725, March 2013.

[15] J. Bertaux, U. Gloger, M. Schmid, A. Hartmann, and S. Scheu. Routine fluorescence in situ hybridization in soil. *Journal of Microbiological Methods*, 69(3):451–460, June 2007.

[16] Sangeeta N. Bhatia and Donald E. Ingber. Microfluidic organs-on-chips. *Nature Biotechnology*, 32(8):760–772, August 2014.

[17] Samantha L Bickel and Kam W Tang. Microbial decomposition of proteins and lipids in copepod versus rotifer carcasses. *Marine Biology*, 157(7):1613–1624, April 2010.

[18] Samantha L Bickel and Kam W Tang. Zooplankton-associated and free-living bacteria in the York River, Chesapeake Bay: comparison of seasonal variations and controlling factors. *Hydrobiologia*, 722(1):305–318, 2014.

[19] Samantha L Bickel, Kam W Tang, and Hans-Peter Grossart. Ciliate epibionts associated with crustacean zooplankton in german lakes: distribution, motility, and bacterivory. *Frontiers in microbiology*, 3:243, 2012.

[20] B A Biddanda and L R Pomeroy. *Microbial aggregation and degradation of phytoplankton-derived detritus in seawater. I. Microbial succession.* Marine ecology progress series. Oldendorf, 1988.

[21] James Q Boedicker, Katie Brenner, and Douglas B Weibel. Spatial Structure of Microbes in Nature and the Biophysics of Cell-Cell Communication. In *The Physical Basis of Bacterial Quorum Communication*, pages 53–81. Springer New York, New York, NY, 2015.

[22] Petra Brandt, Gunnar Gerdts, Maarten Boersma, Karen H Wiltshire, and Antje Wichels. Comparison of different DNA-extraction techniques to investigate the bacterial community of marine copepods. *Helgoland Marine Research*, 64(4):331–342, 2010.

[23] Michael A. Brockhurst, Angus Buckling, and Andy Gardner. Cooperation Peaks at Intermediate Disturbance. *Current Biology*, 17(9):761–765, May 2007.

[24] Alison Buchan, Gary R. LeCleir, Christopher A. Gulvik, and JosÃľ M. GonzÃąlez. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nature Reviews Microbiology*, 12(10):686–698, August 2014.

[25] Tormod Vaaland Burkey. Metapopulation Extinction in Fragmented Landscapes: Using Bacteria and Protozoa Communities as Model Ecosystems. *The American Naturalist*, 150(5):568–591, November 1997.

[26] Kevin R. Carman and Fred C. Dobbs. Epibiotic microorganisms on copepods and other marine crustaceans. *Microscopy Research and Technique*, 37(2):116–135, April 1997.

[27] Hasan Celiker and Jeff Gore. Competition between species can stabilize public-goods cooperation within a species. *Molecular Systems Biology*, 8, November 2012.

[28] Hasan Celiker and Jeff Gore. Cellular cooperation: insights from microbes. *Trends in Cell Biology*, 23(1):9–15, January 2013.

[29] Ashvini Chauhan, Jennifer Cherrier, and Henry N Williams. Impact of sideways and bottom-up control factors on bacterial community succession over a tidal cycle. *Proceedings of the National Academy of Sciences*, 106(11):4301–4306, March 2009.

[30] Ilseung Cho and Martin J Blaser. The human microbiome: at the interface of health and disease. *Nature Reviews: Genetics*, 13(4):260–270, April 2012.

[31] J. S. Chuang, O. Rivoire, and S. Leibler. Simpson's Paradox in a Synthetic Microbial System. *Science*, 323(5911):272–275, January 2009.

[32] J H Connell and R O Slatyer. Mechanisms of succession in natural communities and their role in community stability and organization. *American naturalist*, 111(982):1119, 1977.

[33] Otto X Cordero, Laure-Anne Ventouras, Edward F DeLong, and Martin F Polz. Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proceedings of the National Academy of Sciences*, 109(49):20059–20064, December 2012.

[34] Franck Courchamp, Tim Clutton-Brock, and Bryan Grenfell. Inverse density dependence and the Allee effect. *Trends in Ecology & Evolution*, 14(10):405–410, October 1999.

[35] R. Craig Maclean and C. Brandon. Stable public goods cooperation and dynamic social interactions in yeast. *Journal of Evolutionary Biology*, 21(6):1836–1843, November 2008.

[36] Jonas Cremer, Anna Melbinger, and Erwin Frey. Evolutionary and population dynamics: A coupled approach. *Physical Review E*, 84(5), November 2011.

[37] L. Dai, D. Vorselen, K. S. Korolev, and J. Gore. Generic Indicators for Loss of Resilience Before a Tipping Point Leading to Population Collapse. *Science*, 336(6085):1175–1177, June 2012.

[38] Lei Dai, Kirill S Korolev, and Jeff Gore. Slower recovery in space before collapse of connected populations. *Nature*, 496(7445):355–358, 2013.

[39] James A. Damore and Jeff Gore. Understanding microbial cooperation. *Journal of Theoretical Biology*, 299:31–41, April 2012.

[40] Manoshi S. Datta, Elzbieta Sliwerska, Jeff Gore, Martin F. Polz, and Otto X Cordero. Microbial interactions lead to rapid ecological succession on model marine particles. *in revision*, 2016.

[41] Manoshi Sen Datta, Kirill S Korolev, Ivana Cvijovic, Carmel Dudley, and Jeff Gore. Range expansion promotes cooperation in an experimental microbial metapopulation. *Proceedings of the National Academy of Sciences*, 110(18):7354–7359, 2013.

[42] Lawrence A David, Ana Weil, Edward T Ryan, Stephen B Calderwood, Jason B Harris, Fahima Chowdhury, Yasmin Begum, Firdausi Qadri, Regina C LaRocque, and Peter J Turnbaugh. Gut microbial succession follows acute secretory diarrhea in humans. *mBio*, 6(3):e00381–15, 2015.

[43] D De Corte, I Lekunberri, E Sintes, JAL Garcia, S Gonzales, and G J Herndl. Linkage between copepods and bacteria in the North Atlantic Ocean. *Aquatic Microbial Ecology*, 72(3):215–225, June 2014.

[44] Francisco Dini-Andreote, Michele de Cassia Pereira e Silva, Xavier Triado-Margarit, Emilio O. Casamayor, Jan Dirk van Elsas, and Joana Falcao Salles. Dynamics of bacterial community succession in a salt marsh chronosequence: evidences for temporal niche partitioning. *The ISME journal*, 8(10):1989–2001, October 2014.

[45] Gregory P Donaldson, S Melanie Lee, and Sarkis K Mazmanian. Gut biogeography of the bacterial microbiota. *Nature Reviews Microbiology*, 14(1):20–32, January 2016.

[46] Knut Drescher, Carey D Nadell, Howard A Stone, Ned S Wingreen, and Bonnie L Bassler. Solutions to the public goods dilemma in bacterial biofilms. *Current biology : CB*, 24(1):50–55, January 2014.

[47] C Dziallas, H P Grossart, and K W Tang. Distinct communities of free-living and copepod-associated microorganisms along a salinity gradient in Godthabsfjord, West Greenland. *Arctic*, 45(4):471–480, 2013.

[48] Kristen A Earle, Gabriel Billings, Michael Sigal, Joshua S Lichtman, Gunnar C Hansson, Joshua E Elias, Manuel R Amieva, Kerwyn Casey Huang, and Justin L Sonnenburg. Quantitative Imaging of Gut Microbiota Spatial Organization. *Cell host & microbe*, 18(4):478–488, September 2015.

[49] Davoud Ebrahimi, Olena Tokareva, Nae Gyune Rim, Joyce Y. Wong, David L. Kaplan, and Markus J. Buehler. Silk-Its Mysteries, How It Is Made, and How It Is Used. *ACS Biomaterials Science & Engineering*, 1(10):864–876, October 2015.

[50] Robert C Edgar. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)*, 26(19):2460–2461, October 2010.

[51] Mallory Embree, Joanne K Liu, Mahmoud M Al-Bassam, and Karsten Zengler. Networks of energetic and metabolic interactions define dynamics in microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*, 112(50):15450–15455, November 2015.

[52] Daniel S Esser, Johan H J Leveau, Katrin M Meyer, and Kerstin Wiegand. Spatial scales of interactions among bacteria and between bacteria and the leaf surface. *FEMS Microbiology Ecology*, 91(3):–, March 2015.

[53] Laurent Excoffier, Matthieu Foll, and Remy J. Petit. Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40(1):481–501, December 2009.

[54] Laurent Excoffier and Nicolas Ray. Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, 23(7):347–351, July 2008.

[55] K Faust and J Raes. Microbial interactions: from networks to models. *Nature Reviews Microbiology*, 10(8):538–550, 2012.

[56] Scott Ferrenberg, Sean P O'Neill, Joseph E Knelman, Bryan Todd, Sam Duggan, Daniel Bradley, Taylor Robinson, Steven K Schmidt, Alan R Townsend, Mark W Williams, Cory C Cleveland, Brett A Melbourne, Lin Jiang, and Diana R Nemergut. Changes in assembly processes in soil bacterial communities following a wildfire disturbance. *The ISME Journal*, 7(6):1102–1111, June 2013.

[57] Noah Fierer, Diana Nemergut, Rob Knight, and Joseph M Craine. Changes through time: integrating microorganisms into the study of succession. *Research in microbiology*, 161(8):635–642, October 2010.

[58] Kai Finster, Yuichi Tanimoto, and Friedhelm Bak. Fermentation of methanethiol and dimethylsulfide by a newly isolated methanogenic bacterium. *Archives of microbiology*, 157(5):425–430, 1992.

[59] R. A. Fisher. The wave of advance of advantageous genes. *Annals of Eugenics*, 7(4):355–369, June 1937.

[60] Rima B Franklin and Aaron L Mills, editors. *The Spatial Distribution of Microbes in the Environment*. Springer Science & Business Media, Dordrecht, September 2007.

[61] Sören Franzenburg, Jonas Walter, Sven Künzel, Jun Wang, John F Baines, Thomas CG Bosch, and Sebastian Fraune. Distinct antimicrobial peptide expression determines host species-specific bacterial associations. *Proceedings of the National Academy of Sciences*, 110(39):E3730–E3738, 2013.

[62] Eric A Franzosa, Tiffany Hsu, Alexandra Sirota-Madi, Afrah Shafquat, Galeb Abu-Ali, Xochitl C Morgan, and Curtis Huttenhower. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nature Reviews Microbiology*, 13(6):360–372, June 2015.

[63] Jonathan Friedman and Eric J Alm. Inferring Correlation Networks from Genomic Survey Data. *PLoS Computational Biology (PLOSCB) 8(9)*, 8(9):e1002687, 2012.

[64] Thomas M. J. Fruchterman and Edward M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164, November 1991.

[65] T Fukami. Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annual Review of Ecology*, 2015.

[66] G Gerdts, P Brandt, K Kreisel, M Boersma, K L Schoo, and A Wichels. The microbiome of North Sea copepods. *Helgoland Marine Research*, 67(4):757–773, June 2013.

[67] A. Gislason and T. Silva. Abundance, composition, and development of zooplankton in the Subarctic Iceland Sea in 2006, 2007, and 2008. *ICES Journal of Marine Science*, 69(7):1263–1276, September 2012.

[68] Graciela Gonzalez-Gil and Christof Holliger. Aerobic Granules: Microbial Landscape and Architecture, Stages, and Practical Implications. *Applied and Environmental Microbiology*, 80(11):3433–3441, June 2014.

[69] Jeff Gore, Hyun Youk, and Alexander van Oudenaarden. Snowdrift game dynamics and facultative cheating in yeast. *Nature*, 459(7244):253–256, May 2009.

[70] Lone Gram, Hans-Peter Grossart, Andrea Schlingloff, and Thomas Kiørboe. Possible quorum sensing in marine snow bacteria: production of acylated homoserine lactones by Roseobacter strains isolated from marine snow. *Applied and environmental microbiology*, 68(8):4111–4116, August 2002.

[71] Michael E. Gray, Thomas W. Sappington, Nicholas J. Miller, Joachim Moeser, and Martin O. Bohn. Adaptation and Invasiveness of Western Corn Rootworm: Intensifying Research on a Worsening Pest. *Annual Review of Entomology*, 54(1):303–321, January 2009.

[72] D. Greig and M. Travisano. The Prisoner's Dilemma and polymorphism in yeast SUC genes. *Proceedings of the Royal Society B: Biological Sciences*, 271(Suppl_3):S25–S26, February 2004.

[73] Ashleigh S. Griffin, Stuart A. West, and Angus Buckling. Cooperation and competition in pathogenic bacteria. *Nature*, 430(7003):1024–1027, August 2004.

[74] Hans-Peter Grossart, Claudia Dziallas, Franziska Leunert, and Kam W Tang. Bacteria dispersal by hitchhiking on zooplankton. *Proceedings of the National Academy of Sciences of the United States of America*, 107(26):11959–11964, June 2010.

[75] Hans-Peter Grossart, Thomas Kiørboe, Kam Tang, and Helle Ploug. Bacterial colonization of particles: growth and interactions. *Applied and environmental microbiology*, 69(6):3500–3509, June 2003.

[76] John H. Koschwanez, Kevin R. Foster, and Andrew W. Murray. Sucrose Utilization in Budding Yeast as a Model for the Origin of Undifferentiated Multicellularity. *PLoS Biology*, 9(8):e1001122, August 2011.

[77] O. Hallatschek, P. Hersen, S. Ramanathan, and D. R. Nelson. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50):19926–19930, December 2007.

[78] Oskar Hallatschek and KS Korolev. Fisher waves in the strong noise limit. *Physical Review Letters*, 103(10):108103, 2009.

[79] Oskar Hallatschek and David R Nelson. Life at the front of an expanding population. *Evolution*, 64(1):193–206, 2010.

[80] Ilkka Hanski. *Metapopulation Ecology*. OUP Oxford, March 1999.

[81] Ilkka Hanski. Metapopulations of animals in highly fragmented landscapes and population viability analysis. *Population Viability Analysis. University of Chicago Press, Chicago, IL*, pages 86–108, 2002.

[82] M Heath, P Boyle, A Gislason, W Gurney, S Hay, E Head, S Holmes, A Ingvarsdottir, S Jonasdottir, and P Lindeque. Comparative ecology of over-wintering Calanus finmarchicus in the northern North Atlantic, and implications for life-cycle patterns. *ICES Journal of Marine Science*, 61(4):698–708, June 2004.

[83] Frank M. Hilker, Mark A. Lewis, Hiromi Seno, Michel Langlais, and Horst Malchow. Pathogens can Slow Down or Reverse Invasion Fronts of their Hosts. *Biological Invasions*, 7(5):817–832, September 2005.

[84] Hans-Jurgen Hirche. Overwintering of Calanus finmarchicus and Calanus helgolandicus. *Marine Ecology Progress Series*, 11:281–290, March 1983.

[85] Hans-Jurgen Hirche. Spatial distribution of digestive enzyme activities of *Calanus finmarchicus* and *C.hyperboreus* in Fram Strait/Greenland Sea. *Journal of Plankton Research*, 11(3):431–443, 1989.

[86] Laura R Hmelo, Tracy J Mincer, and Benjamin A S Van Mooy. Possible influence of bacterial quorum sensing on the hydrolysis of sinking particulate organic carbon in marine environments. *Environmental Microbiology Reports*, 3(6):682–688, December 2011.

[87] Marcel Holyoak, Mathew A Leibold, and Robert D Holt. *Metacommunities: spatial dynamics and ecological communities*. University of Chicago Press, 2005.

[88] Nina Jagmann, Katharina Styp von Rekowski, and Bodo Philipp. Interactions of bacteria with different mechanisms for chitin degradation result in the formation of a mixed-species biofilm. *FEMS Microbiology Letters*, 326(1):69–75, January 2012.

[89] Nianzhi Jiao, Gerhard J Herndl, Dennis A Hansell, Ronald Benner, Gerhard Kattner, Steven W Wilhelm, David L Kirchman, Markus G Weinbauer, Tingwei Luo, Feng Chen, and Farooq Azam. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nature Reviews Microbiology*, 8(8):593–599, August 2010.

[90] Sigrun Huld Jonasdottir, Andre W Visser, Katherine Richardson, and Michael R Heath. Seasonal copepod lipid pump promotes carbon sequestration in the deep North Atlantic. *Proceedings of the National Academy of Sciences of the United States of America*, page 201512110, September 2015.

[91] L Jost. Entropy and diversity. *Oikos*, 113(2):363–375, 2006.

[92] Benjamin Kerr, Claudia Neuhauser, Brendan J. M. Bohannan, and Antony M. Dean. Local migration promotes competitive restraint in a host-pathogen 'tragedy of the commons'. *Nature*, 442(7098):75–78, July 2006.

[93] Benjamin Kerr, Margaret A. Riley, Marcus W. Feldman, and Brendan J. M. Bohannan. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature*, 418(6894):171–174, July 2002.

[94] Hyun Jung Kim, Hu Li, James J Collins, and Donald E Ingber. Contributions of microbiome and mechanical deformation to intestinal bacterial overgrowth and inflammation in a human gut-on-a-chip. *Proceedings of the National Academy of Sciences*, page 201522193, 2015.

[95] Motoo Kimura and George H. Weiss. The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. *Genetics*, 49(4):561, April 1964.

[96] Thomas Kiørboe, Hans-Peter Grossart, Helle Ploug, and Kam Tang. Mechanisms and rates of bacterial colonization of sinking aggregates. *Applied and environmental microbiology*, 68(8):3996–4006, August 2002.

[97] Thomas Kiørboe, Kam Tang, Hans-Peter Grossart, and Helle Ploug. Dynamics of Microbial Communities on Marine Snow Aggregates: Colonization, Growth, Detachment, and Grazing Mortality of Attached Bacteria. *Applied and environmental microbiology*, 69(6):3036–3047, June 2003.

[98] David L Kirchman. *Microbial Ecology of the Oceans*. John Wiley & Sons, June 2010.

[99] Jeremy E Koenig, Ayme Spor, Nicholas Scalfone, Ashwana D Fricker, Jesse Stombaugh, Rob Knight, Largus T Angenent, and Ruth E Ley. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences*, 108 Suppl 1(Supplement_1):4578–4585, March 2011.

[100] A.M. Kolmogorov. Selected Works of A. N. Kolmogorov: Volume I: Mathematics and Mechanics (Mathematics and its Applications): Vladimir M. Tikhomirov: 9789027727961: Amazon.com: Books.

[101] K. S. Korolev, Mikkel Avlund, Oskar Hallatschek, and David R. Nelson. Genetic demixing and evolution in linear stepping stone models. *Reviews of Modern Physics*, 82(2):1691–1718, May 2010.

[102] Kirill S. Korolev. The Fate of Cooperation during Range Expansions. *PLoS Computational Biology*, 9(3):e1002994, March 2013.

[103] Adam Lampert and Tsvi Tlusty. Density-dependent cooperation as a mechanism for persistence and coexistence: density-dependent cooperation. *Evolution*, 65(10):2750–2759, October 2011.

[104] R. Levins. Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bulletin of the Entomological Society of America*, 15(3):237–240, September 1969.

[105] Ainslie E F Little, Courtney J Robinson, S Brook Peterson, Kenneth F Raffa, and Jo Handelsman. Rules of engagement: interspecies interactions that regulate microbial communities. *Annual review of microbiology*, 62:375–401, 2008.

[106] Jintao Liu, Arthur Prindle, Jacqueline Humphries, Marcal Gabalda-Sagarra, Munehiro Asally, Dong-Yeon D Lee, San Ly, Jordi Garcia-Ojalvo, and Gurol M Suel. Metabolic co-dependence gives rise to collective oscillations within biofilms. *Nature*, 523(7562):550–554, July 2015.

[107] George Livingston, Miguel Matias, Vincent Calcagno, Claire Barbera, Marine Combe, Mathew A. Leibold, and Nicolas Mouquet. Competition-colonization dynamics in experimental bacterial metacommunities. *Nature Communications*, 3:1234, December 2012.

[108] R A Long and F Azam. Antagonistic interactions among marine pelagic bacteria. *Applied and environmental microbiology*, 67(11):4975–4983, November 2001.

[109] R. Craig MacLean, Ayari Fuentes-Hernandez, Duncan Greig, Laurence D. Hurst, and Ivana Gudelj. A Mixture of "Cheats" and "Co-Operators" Can Enable Maximal Group Benefit. *PLoS Biology*, 8(9):e1000486, September 2010.

[110] Andrea Manica, William Amos, Francois Balloux, and Tsunehiko Hanihara. The effect of ancient population bottlenecks on human phenotypic variation. *Nature*, 448(7151):346–348, July 2007.

[111] Jessica L. Mark Welch, Blair J. Rossetti, Christopher W. Rieken, Floyd E. Dewhirst, and Gary G. Borisy. Biogeography of a human oral microbiome at the micron scale. *Proceedings of the National Academy of Sciences*, page 201522149, January 2016.

[112] M Matsushita, J Wakita, H Itoh, K Watanabe, T Arai, T Matsuyama, H Sakaguchi, and M Mimura. Formation of colony patterns by a bacterial cell population. *Physica A: Statistical Mechanics and its Applications*, 274(1-2):190–199, December 1999.

[113] Diane McDougald, Scott A Rice, Nicolas Barraud, Peter D Steinberg, and Staffan Kjelleberg. Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal. *Nature Reviews Microbiology*, 10(1):39–50, January 2012.

[114] Shawn E. McGlynn, Grayson L. Chadwick, Christopher P. Kempes, and Victoria J. Orphan. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature*, 526(7574):531–535, September 2015.

[115] C B Miller, J A Crain, and C A Morgan. Oil storage variability in Calanus finmarchicus. *ICES Journal of Marine Science: Journal du Conseil*, 57(6):1786–1799, December 2000.

[116] Charles B. Miller, Timothy J. Cowles, Peter H. Wiebe, Nancy J. Copley, and Helen Grigg. Phenology in Calanus finmarchicus; hypotheses about control mechanisms. May 1991.

[117] Eva Friis Mller, Lasse Riemann, and Morten Sndergaard. Bacteria associated with copepods: abundance, activity and community composition. *Aquatic Microbial Ecology*, 47(1):99–106, April 2007.

[118] Pia H Moisander, Andrew D Sexton, and Meaghan C Daley. Stable Associations Masked by Temporal Variability in the Marine Copepod Microbiome. *PloS one*, 10(9):e0138967, September 2015.

[119] Eva Friis Moller, Peter Thor, and Torkel Gissel Nielsen. Production of DOC by Calanus finmarchicus, C. glacialis and C. hyperboreus through sloppy feeding and leakage from fecal pellets. *Marine Ecology Progress Series*, 262:185–191, November 2003.

[120] C. Moreau, C. Bherer, H. Vezina, M. Jomphe, D. Labuda, and L. Excoffier. Deep Human Genealogies Reveal a Selective Advantage to Be on an Expanding Wave Front. *Science*, 334(6059):1148–1150, November 2011.

[121] J Jeffrey Morris, Richard E Lenski, and Erik R Zinser. The black queen hypothesis: evolution of dependencies through adaptive gene loss. *MBio*, 3(2):e00036–12, 2012.

[122] J. D. Murray. *Mathematical Biology II: Spatial Models and Biomedical Applications*. Springer New York, May 2013.

[123] Carey D. Nadell, Kevin R. Foster, and Joao B. Xavier. Emergence of Spatial Structure in Cell Groups and the Evolution of Cooperation. *PLoS Computational Biology*, 6(3):e1000716, March 2010.

[124] Diana R Nemergut, Suzanne P Anderson, Cory C Cleveland, Andrew P Martin, Amy E Miller, Anton Seimon, and Steven K Schmidt. Microbial community succession in an unvegetated, recently deglaciated soil. *Microbial Ecology*, 53(1):110–122, January 2007.

[125] John Novembre and Anna Di Rienzo. Spatial patterns of variation due to natural selection in humans. *Nature Reviews Genetics*, 10(11):745–755, November 2009.

[126] M. A. Nowak. Five Rules for the Evolution of Cooperation. *Science*, 314(5805):1560–1563, December 2006.

[127] Anthony G O'Donnell, Iain M Young, Steven P Rushton, Mark D Shirley, and John W Crawford. Visualization, modelling and prediction in soil microbiology. *Nature Reviews Microbiology*, 5(9):689–699, September 2007.

[128] Kabir G. Peay, Matteo Garbelotto, and Thomas D. Bruns. Evidence of dispersal limitation in soil microorganisms: Isolation reduces species richness on mycorrhizal tree islands. *Ecology*, 91(12):3631–3640, December 2010.

[129] Pascal Pons and Matthieu Latapy. Computing Communities in Large Networks Using Random Walks. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Dough Tygar, Moshe Y. Vardi, Gerhard Weikum, Pinar Yolum, Tunga Gungor, Fikret Gurgen, and Can Ozturan, editors, *Computer and Information Sciences - ISCIS 2005*, volume 3733, pages 284–293. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[130] Sarah P Preheim, Allison R Perrotta, Antonio M Martin-Platero, Anika Gupta, and Eric J Alm. Distribution-based clustering: using ecology to refine the operational taxonomic unit. *Applied and environmental microbiology*, 79(21):6593–6603, November 2013.

[131] Arthur Prindle, Jintao Liu, Munehiro Asally, San Ly, Jordi Garcia-Ojalvo, and Gurol M. Suel. Ion channels enable electrical communication in bacterial communities. *Nature*, 527(7576):59–63, October 2015.

[132] Jonathan N. Pruitt and Susan E. Riechert. Frequency-dependent success of cheaters during foraging bouts might limit their spread within colonies of a socially polymorphic spider. *Evolution*, 63(11):2966–2973, November 2009.

[133] Adrian E. Raftery. Bayesian Model Selection in Social Research. *Sociological Methodology*, 25:111, 1995.

[134] Adrian E. Raftery, David Madigan, and Jennifer A. Hoeting. Bayesian Model Averaging for Linear Regression Models. *Journal of the American Statistical Association*, 92(437):179–191, March 1997.

[135] M G Ramirez, LIR Avelizapa, and NGR Avelizapa. Colloidal chitin stained with Remazol Brilliant Blue, a useful substrate to select chitinolytic microorganisms and to evaluate chitinases. *Journal of Microbiological Methods*, 56(2):213–219, 2004.

[136] Xavier Raynaud and Naoise Nunan. Spatial ecology of bacteria at the microscale in soil. *PloS one*, 9(1):e87217–e87217, January 2014.

[137] Misty R Riddle, Bonnie K Baxter, and Brian J Avery. Molecular identification of microorganisms associated with the brine shrimp Artemia franciscana. *Aquatic Biosystems*, 9(1):7, 2013.

[138] Lasse Riemann and Farooq Azam. Widespread N-acetyl-D-glucosamine uptake among pelagic marine bacteria and its ecological implications. *Applied and environmental microbiology*, 68(11):5554–5562, November 2002.

[139] L. Roques, J. Garnier, F. Hamel, and E. K. Klein. Allee effect promotes diversity in traveling waves of colonization. *Proceedings of the National Academy of Sciences*, 109(23):8828–8833, June 2012.

[140] Adin Ross-Gillespie, Andy Gardner, Angus Buckling, Stuart A. West, and Ashleigh S. Griffin. Density dependence and cooperation: theory and a test with bacteria. *Evolution*, 63(9):2315–2325, September 2009.

[141] Adin Ross-Gillespie, Andy Gardner, Stuart A. West, and Ashleigh S. Griffin. Frequency Dependence and Cooperation: Theory and a Test with Bacteria. *The American Naturalist*, 170(3):331–342, September 2007.

[142] Alvaro Sanchez and Jeff Gore. Feedback between Population and Evolutionary Dynamics Determines the Fate of Social Microbial Populations. *PLoS Biology*, 11(4):e1001547, April 2013.

[143] Lorenzo A. Santorelli, Christopher R. L. Thompson, Elizabeth Villegas, Jessica Svetz, Christopher Dinh, Anup Parikh, Richard Sucgang, Adam Kuspa, Joan E. Strassmann, David C. Queller, and Gad Shaulsky. Facultative cheater mutants reveal the genetic complexity of cooperation in social amoebae. *Nature*, 451(7182):1107–1110, February 2008.

[144] Franz Josef Sartoris, David N. Thomas, Astrid Cornils, and Sigrid B. Schnack-Schiel. Buoyancy and diapause in Antarctic copepods: The role of ammonium accumulation. *Limnology and Oceanography*, 55(5):1860–1864, 2010.

[145] Rosemary E Scavotto, Claudia Dziallas, Mikkel Bentzon-Tilia, Lasse Riemann, and Pia H Moisander. Nitrogen-fixing bacteria associated with copepods in coastal waters of the North Atlantic Ocean. *Environmental Microbiology*, February 2015.

[146] Patrick D Schloss, Sarah L Westcott, Thomas Ryabin, Justine R Hall, Martin Hartmann, Emily B Hollister, Ryan A Lesniewski, Brian B Oakley, Donovan H Parks, Courtney J Robinson, Jason W Sahl, Blaz Stres, Gerhard G Thallinger, David J Van Horn, and Carolyn F Weber. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, 75(23):7537–7541, December 2009.

[147] Sabine Schrunder, Sigrid B. Schnack-Schiel, Holger Auel, and Franz Josef Sartoris. Control of Diapause by Acidic pH and Ammonium Accumulation in the Hemolymph of Antarctic Copepods. *PLoS ONE*, 8(10):e77498, October 2013.

[148] Orr H. Shapiro, Esti Kramarsky-Winter, Assaf R. Gavish, Roman Stocker, and Assaf Vardi. A coral-on-a-chip microfluidic platform enabling live-imaging microscopy of reef-building corals. *Nature Communications*, 7:10860, March 2016.

[149] Katyanne M Shoemaker and Pia H Moisander. Microbial diversity associated with copepods in the North Atlantic subtropical gyre. *FEMS Microbiology Ecology*, June 2015.

[150] M Simon, H P Grossart, and B Schweitzer. Microbial ecology of organic aggregates in aquatic ecosystems. *Aquatic Microbial Ecology*, 2002.

[151] D C Smith, M Simon, A L Alldredge, and F Azam. Intense hydrolytic enzyme activity on marine aggregates and implications for rapid particle dissolution. *Nature*, 359(6391):139–142, 1992.

[152] Steven Smriga, Vicente I. Fernandez, James G. Mitchell, and Roman Stocker. Chemotaxis toward phytoplankton drives organic matter partitioning among marine bacteria. *Proceedings of the National Academy of Sciences*, page 201512307, January 2016.

[153] M R Sochard, D F Wilson, B Austin, and R R Colwell. Bacteria associated with the surface and gut of marine copepods. *Applied and environmental microbiology*, 37(4):750–759, April 1979.

[154] Kwangmin Son, Douglas R Brumley, and Roman Stocker. Live from under the lens: exploring microbial motility with dynamic imaging and microfluidics. *Nature Reviews Microbiology*, 13(12):761–775, November 2015.

[155] Claire E. Stanley, Guido Grossmann, Xavier Casadevall i Solvas, and Andrew J. deMello. Soil-on-a-Chip: microfluidic platforms for environmental organismal studies. *Lab Chip*, 16(2):228–241, 2016.

[156] James C. Stegen, Xueju Lin, Jim K. Fredrickson, and Allan E. Konopka. Estimating and mapping ecological processes influencing microbial community assembly. *Frontiers in Microbiology*, 6, May 2015.

[157] R Stocker. The 100 Âţm length scale in the microbial ocean. *Aquatic Microbial Ecology*, 76(3):189–194, December 2015.

[158] Roman Stocker and Justin R Seymour. Ecology and physics of bacterial chemotaxis in the ocean. *Microbiology and molecular biology reviews : MMBR*, 76(4):792–812, December 2012.

[159] Roman Stocker, Justin R Seymour, Azadeh Samadani, Dana E Hunt, and Martin F Polz. Rapid chemotactic response enables marine bacteria to exploit ephemeral microscale nutrient patches. *Proceedings of the National Academy of Sciences*, 105(11):4209–4214, March 2008.

[160] J. E. Strassmann and D. C. Queller. Evolution of cooperation and control of cheating in a social microbe. *Proceedings of the National Academy of Sciences*, 108(Supplement_2):10855–10862, June 2011.

[161] K W Tang. Copepods as microbial hotspots in the ocean: effects of host feeding activities on attached bacteria. *Aquatic Microbial Ecology*, 38:31–40, 2005.

[162] K W Tang, S L Bickel, C Dziallas, and H P Grossart. Microbial activities accompanying decomposition of cladoceran and copepod carcasses under different environmental conditions. *Aquatic Microbial Ecology*, 57:89–100, September 2009.

[163] K W Tang, V Turk, and H P Grossart. Linkage between crustacean zooplankton and aquatic bacteria. *Aquatic Microbial Ecology*, 61(3):261–277, 2010.

[164] Kam W Tang, Ronnie N Glud, Anni Glud, Søren Rysgaard, and Torkel Gissel Nielsen. Copepod guts as biogeochemical hotspots in the sea: Evidence from microelectrode profiling of Calanus spp. *Limnology and Oceanography*, 56(2):666–672, March 2011.

[165] Anna Tarakanova and Markus J. Buehler. A Materiomics Approach to Spider Silk: Protein Molecules to Webs. *JOM*, 64(2):214–225, February 2012.

[166] A M Tarrant, M F Baumgartner, and T Verslycke. Differential gene expression in diapausing and active Calanus finmarchicus (Copepoda). *Marine Ecology Progress Series*, 355:193–207, 2008.

[167] Ann M Tarrant, Mark F Baumgartner, Bjørn Henrik Hansen, Dag Altin, Trond Nordtug, and Anders J Olsen. Transcriptional profiling of reproductive development, lipid storage and molting throughout the last juvenile stage of the marine copepod Calanus finmarchicus. *Frontiers in Zoology*, 11(1):1, December 2014.

[168] Hanno Teeling, Bernhard M Fuchs, Dorte Becher, Christine Klockow, Antje Gardebrecht, Christin M Bennke, Mariette Kassabgy, Sixing Huang, Alexander J Mann, Jost Waldmann, Marc Weber, Anna Klindworth, Andreas Otto, Jana Lange, Jorg Bernhardt, Christine Reinsch, Michael Hecker, Jorg Peplies, Frank D Bockelmann, Ulrich Callies, Gunnar Gerdts, Antje Wichels, Karen H Wiltshire, Frank Oliver Glockner, Thomas Schweder, and Rudolf Amann. Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. *Science*, 336(6):608–611, May 2012.

[169] Stefan Thiele, Bernhard M. Fuchs, Rudolf Amann, and Morten H. Iversen. Colonization in the Photic Zone and Subsequent Changes during Sinking Determine Bacterial Community Composition in Marine Snow. *Applied and Environmental Microbiology*, 81(4):1463–1471, February 2015.

[170] B J Tibbles and D E Rawlings. Characterization of nitrogen-fixing bacteria from a temperate saltmarsh lagoon, including isolates that produce ethane from acetylene. *Microbial Ecology*, 27(1):65–80, January 1994.

[171] Patrick C. Tobin, Ludek Berec, and Andrew M. Liebhold. Exploiting Allee effects for managing biological invasions: Exploiting Allee effects. *Ecology Letters*, 14(6):615–624, June 2011.

[172] Jefferson T. Turner. The Importance of Small Planktonic Copepods and Their Roles in Pelagic Marine Food Webs. *Zoological Studies*, 43(2):255–266, January 2004.

[173] Jefferson T. Turner. Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Progress in Oceanography*, 130:205–248, January 2015.

[174] Paul E Turner and Lin Chao. Prisoner's dilemma in an rna virus. *Nature*, 398(6726):441–443, 1999.

[175] I. Van Bocxlaer, S. P. Loader, K. Roelants, S. D. Biju, M. Menegon, and F. Bossuyt. Gradual Adaptation Toward a Range-Expansion Phenotype Initiated the Global Radiation of Toads. *Science*, 327(5966):679–682, February 2010.

[176] K. Gerald van den Boogaart and R. Tolosana-Delgado. "compositions": A unified R package to analyze compositional data. *Computers & Geosciences*, 34(4):320–338, April 2008.

[177] Amy S Van Wey, Adrian L Cookson, Nicole C Roy, Warren C McNabb, Tanya K Soboleva, and Paul R Shorten. Bacterial biofilms associated with food particles in the human large bowel. *Molecular Nutrition & Food Research*, 55(7):969–978, July 2011.

[178] Andrea Velenich and Jeff Gore. Synthetic approaches to understanding biological constraints. *Current Opinion in Chemical Biology*, 16(3-4):323–328, August 2012.

[179] P G Verity and V Smetacek. *Organism life cycles, predation, and the structure of marine pelagic ecosystems.* 1996.

[180] Alan W Walker, Sylvia H Duncan, Hermie J M Harmsen, Grietje Holtrop, Gjalt W Welling, and Harry J Flint. The species composition of the human intestinal microbiota differs between particle-associated and liquid phase communities. *Environmental Microbiology*, 10(12):3275–3283, December 2008.

[181] Gian-Reto Walther, Eric Post, Peter Convey, Annette Menzel, Camille Parmesan, Trevor J. C. Beebee, Jean-Marc Fromentin, Ove Hoegh-Guldberg, and Franz Bairlein. Ecological responses to recent climate change. *Nature*, 416(6879):389–395, March 2002.

[182] Qiong Wang, George M Garrity, James M Tiedje, and James R Cole. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and environmental microbiology*, 73(16):5261–5267, August 2007.

[183] Daniel Wangpraseurt, Mathieu Pernice, Paul Guagliardo, Matt R Kilburn, Peta L Clode, Lubos Polerecky, and Michael Kuhl. Light microenvironment and single-cell gradients of carbon fixation in tissues of symbiont-bearing corals. *The ISME Journal*, 10(3):788–792, March 2016.

[184] Stuart A. West, Ashleigh S. Griffin, and Andy Gardner. Evolutionary Explanations for Cooperation. *Current Biology*, 17(16):R661–R672, August 2007.

[185] Elizabeth G. Wilbanks, Ulrike Jaekel, Verena Salman, Parris T. Humphrey, Jonathan A. Eisen, Marc T. Facciotti, Daniel H. Buckley, Stephen H. Zinder, Gregory K. Druschel, David A. Fike, and Victoria J. Orphan. Microscale sulfur cycling in the phototrophic pink berry consortia of the Sippewissett Salt Marsh. *Environmental Microbiology*, 16(11):3398–3415, November 2014.

[186] Ned S Wingreen and Simon A Levin. Cooperation among Microorganisms. *PLoS Biology*, 4(9):e299, September 2006.

[187] Benjamin E Wolfe and Rachel J Dutton. Fermented foods as experimentally tractable microbial ecosystems. *Cell*, 161(1):49–55, March 2015.

[188] J. B. Xavier. Social interaction in synthetic and natural microbial communities. *Molecular Systems Biology*, 7(1):483–483, April 2014.

[189] Joao B. Xavier, Wook Kim, and Kevin R. Foster. A molecular mechanism that stabilizes cooperative secretions in Pseudomonas aeruginosa: Metabolic prudence in P. aeruginosa. *Molecular Microbiology*, 79(1):166–179, January 2011.

[190] Yutaka Yawata, Otto X Cordero, Filippo Menolascina, Jan-Hendrik Hehemann, Martin F Polz, and Roman Stocker. Competition-dispersal tradeoff ecologically differentiates recently speciated marine bacterioplankton populations. *Proceedings of the National Academy of Sciences of the United States of America*, 111(15):5622–5627, April 2014.

[191] Aleksej Zelezniak, Sergej Andrejev, Olga Ponomarova, Daniel R. Mende, Peer Bork, and Kiran Raosaheb Patil. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences*, 112(20):6449–6454, May 2015.

[192] Feng Zhang and Cang Hui. Eco-Evolutionary Feedback and the Invasion of Cooperation in Prisoner's Dilemma Games. *PLoS ONE*, 6(11):e27523, November 2011.

[193] Feng Zhang, Cang Hui, Xiaozhuo Han, and Zizhen Li. Evolution of cooperation in patchy habitat under patch decay and isolation. *Ecological Research*, 20(4):461–469, July 2005.

[194] J Zhou, M A Bruns, and J M Tiedje. DNA recovery from soils of diverse composition. *Applied and environmental microbiology*, 62(2):316–322, February 1996.