The ecology and evolution of complex microbial communities

Katharine Z. Coyte

Jesus College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2016

Statement of Originality

I declare that this thesis was composed by myself, and that the work contained herein is my own except where explicitly stated. This work has not been submitted for any degree or professional qualification.

Katharine Z. Coyte, Hilary 2016

To my family

Acknowledgements

It is with deepest gratitude that I acknowledge the many people who have helped and supported me throughout my DPhil.

I would like to thank Sam Brown, Eamonn Gaffney, Seth Rakoff-Nahoum, Joao Xavier and the many, many others who have provided valuable discussion and feedback. Also Karina Xavier and everyone else at the inaugural IGC Host-Microbe summer school for a truly inspiring scientific experience.

I cannot thank enough everybody from the TEEB super-group, past and present – Sara, Danna, Nuno, and especially the residents of E140: Mel, Rene, Sandra, JB, Nick, and Chucky – for providing a great academic environment, and even greater comic relief. And, of course, Jonas – we make an awesome team, and I can't wait to reconvene in New York and continue our research adventures!

The EPSRC, ERC, and Jesus College provided vital financial support, for which I am very grateful. I'd also like to thank the DTC for providing me with funding, but even more so for providing me with a brilliant group of friends. I've been privileged to meet so many awesome people at Oxford, whom are far too numerous to name. In particular though, I'd like to thank Anne for being a great, and hilarious, teammate – from our first five-a-side in the park to cup finals (via the odd concussion). Lizzy, for never failing to persuade me that 'science rocks', even when Science is being a bit less friendly. And of course, Ellie, mainly for supporting my KitKat habit (but also for being a generally spectacular friend).

Most of all, I would like to thank my two most excellent supervisors. Mack, through our many hours covering the microscope in sellotape – and the occasional scientific discussion – you've taught me a huge amount (and I think we both learnt we suck at multitasking); Sheffield will be lucky to have you. And Kevin – from science chats in the Missing Bean to adventuring around Japan and Korea, you have been a great friend as well as supervisor – being your student has been a huge amount of fun, and I hope (/ demand) that we continue working together for many years to come!

Finally, I'd like to thank my family, for their unwavering support (and total madness), and to whom this thesis is dedicated.

Publications and statement of authorship

The majority of each chapter within this thesis has been my own work. I have received help from various collaborators, as outlined below. Chapter 2 is currently under review as:

- **Coyte K. Z.**, Tabuteau, H., Gaffney, E. A., Foster, K. R., and Durham, W. M. Flow mediates microbial evolution in porous environments.

William Durham assisted with the design and implementation of microfluidic experiments, and Hervé Tabuteau built the microfluidic devices. All authors commented upon the manuscript. Chapter 3 is published as:

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

Jonas Schluter aided in the mathematical analysis, and along with Kevin Foster also commented on the manuscript. Chapter 4 is solely my own work, but has benefited from guidance by Kevin Foster. Further, during my PhD I have also contributed to the data analysis in a project concerning the degradation and protection of heritage sites in southern Turkey, currently in press as:

- Wilhelm, K., Viles, H., Winter, E., Burke, O., Engelstaedter, S., and **Coyte, K. Z**. Catastrophic Limestone Decay At The Central Sanctuary Of Iupiter Dolichenus At Baba Duluk Tepesi In South Turkey: Causes And Implications For Future Conservation.

I have received permission for the submission of my thesis as a collection of papers. As such, each chapter has its own dedicated introduction and conclusion section.

Abstract

Ecology and evolution of complex microbial communities
Katharine Z. Coyte, Jesus College
Hilary 2016

Microbial communities colonise virtually every habitable environment on earth. They live on us, inside us, and all around us, and are key components of the earth's ecosystem. However, due to the complexity of microbial communities we still understand relatively little about the mechanisms that drive their ecology and evolution. The aim of this thesis is to develop novel experimental and mathematical approaches, in order to investigate complex microbial communities. Specifically, this thesis is split into two parts, focussing in turn on how environmental complexity, and interspecies interactions shape microbial ecology and evolution. In the first part I have studied microbial evolution in porous environments such as one might find in soil or aquifers. Here I have combined microfluidic experiments, mechanistic models, and game theory to study how hydrodynamics mediate competition between bacterial genotypes. With these I have demonstrated how even subtle environmental complexity can fundamentally affect microbial competition, enabling slow growing genotypes to outcompete their faster counterparts. In the second part of this thesis I have investigated how interspecies interactions influence the stability and initial assembly of microbial communities. Here I have developed new tools – from theoretical ecology – in order to study the microbial communities within the mammalian gastro-intestinal tract. With these I have demonstrated how strong, cooperative interactions between species can undermine both the stability, and the assembly of communities. Moreover, I have illustrated how these tools can be applied to real microbiome data, in order to better understand these important communities. In sum, this thesis explores two key drivers of microbial ecology and evolution, and presents a new set of tools for the future investigation of complex microbial communities.

Contents

1 Introduction

1.1 The diversity and complexity of microbial life

Microbial communities are ubiquitous on earth, colonising seemingly uninhabitable regions at the bottom of the ocean, through soils and sediments, to the nutrient rich habitats within our own bodies (Azam and Malfatti, 2007; Gill et al., 2006; Lozupone et al., 2012; Martiny et al., 2006; Whitman et al., 1998). These microbial communities are not only pervasive, but they are also vitally important. Which species colonise a particular environment and how these communities change over time impacts processes ranging from geochemical cycling to human health (Clemente et al., 2012; Hall-Stoodley et al., 2004; Tortell, 1999). Disentangling the drivers of microbial community dynamics thus represents a fundamental goal for many disparate disciplines. However, although sequencing technology now enables us to resolve the composition of microbial communities, we still lack a mechanistic understanding of how, and why, communities behave and respond the way that they do.

This major gap in understanding is due, in large part, to the sheer complexity of most microbial communities. Research in microbiology has tended to focus on single strains growing in shaking culture. However, the reality is that microbes typically live in genetically diverse communities, and in habitats with complex physical and chemical landscapes (Azam and Malfatti, 2007; Bever et al., 2009; Horner-Devine et al., 2004; Schluter and Foster, 2012; Young and Crawford, 2004). Habitat complexity is thought to influence microbial ecology in settings ranging from seemingly simple laboratory environments to the human lung (Azam and Malfatti, 2007; Bever et al., 2009; Horner-Devine et al., 2004; Markussen et al., 2014; Rainey and Travisano, 1998; Schluter and Foster, 2012). Moreover, the interaction between microbes and their environment is not a one-way process: microbial growth in turn shapes the environment, altering the physical conditions experienced by the community as a whole (Koza et al., 2011; Rainey and Travisano, 1998). If we wish to fully understand the mechanisms driving microbial community dynamics, then, it is vital that we take into account the interplay between microbes and their environment.

The challenge of understanding microbial community dynamics is heightened further by the sheer diversity, and interconnectivity, of most microbial communities. Far from being solitary organisms, microbes typically live in dense communities, in which cells interact with members both of their own species and of others (Cornforth and Foster, 2013; Mitri and Foster, 2013; West et al., 2006). These interactions range from the production of toxins that kill other cells, through simple competition for resources, to cross-feeding where one species feeds on the by-products of another (Belenguer et al., 2006; Ze et al., 2012, 2013). Such interactions have been intensively studied both mechanistically and theoretically. However, the typical focus has been on simple communities composed of only a few species and interactions. How ecological interactions scale up to drive the properties of complex, multispecies microbial communities remains poorly understood. If we wish to understand microbial community dynamics,

then, we must begin to disentangle the role of interspecies interactions within diverse, interconnected communities.

The goal of this thesis is thus to investigate the role of environmental complexity and interspecies interactions upon microbial ecology and evolution. Accordingly, the thesis is divided into two main parts. The first part asks how environmental complexity in porous environments shapes microbial ecology and evolution. The second part asks how interspecies interactions shape the stability and assembly of microbial communities, with a particular focus on the microbes that live inside us.

1.2 Mathematical modelling as a tool for understanding microbial communities

In particular, much of this thesis focuses on building novel mathematical models of microbial communities. Mathematical modelling can offer a unique insight into systems that are often too complex to manipulate experimentally. Such models provide us with a broad conceptual understanding of the dynamics underlying complex systems, but crucially, can also generate testable predictions. Yet despite this, much of modern microbiology has progressed without the integration of formalised mathematical theory. A key goal of this thesis is to address this gap – integrating mathematics, ecology, and evolution with traditional microbiology, in order to gain fundamental insights into the factors driving microbial community dynamics.

The role of environmental complexity and microbes in porous environments

I focus first on the role of environmental complexity, examining microbial communities within porous environments, such as one might see in soil or aquifers (e.g. see Figure 1.1). Bacteria in such porous environments constitute approximately one-half of the carbon within living organisms globally (Whitman et al., 1998) and play pivotal roles

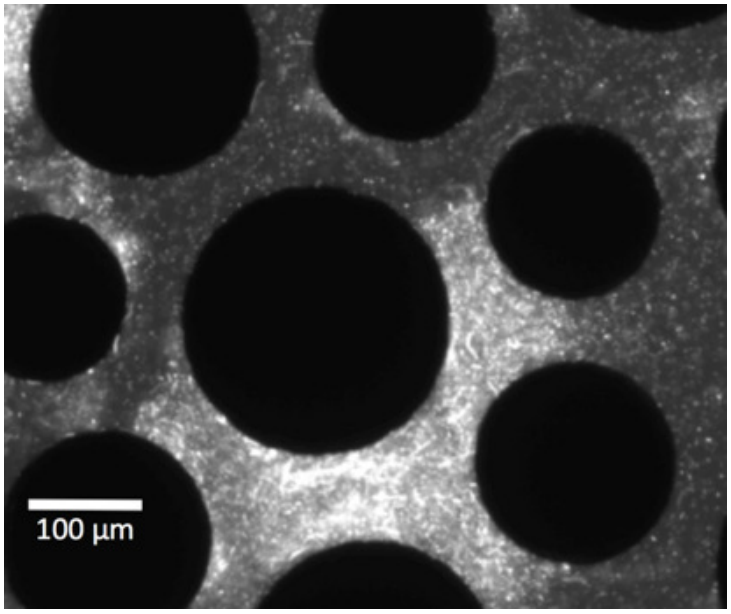

Figure 1.1: **A porous environment colonised by *Escherichia coli* cells.** Fluorescently labelled bacteria (white) colonise the spaces between silicon posts (black circles) in a microfluidic experiment mimicking bacterial colonisation of a porous environment. The large mass of white indicates an area where the bacteria have completely blocked their pore spaces, such that the cells at the centre will no longer be receiving nutrients, nor be able to disperse downstream.

in processes that range from nitrogen fixation through to microbially enhanced oil recovery (Danhorn and Fuqua, 2007; Nicolella et al., 2000; Pintelon et al., 2009; Wang and Hendry, 2012). In these settings, microbes colonise the spaces between grains of sand or sediment, typically existing as biofilms: close-knit associations of cells, surrounded by self-secreted matrices of polysaccharides (Costerton et al., 1995; Ghiorse and Wilson, 1988). Here, diverse bacterial genotypes live under intense competition for limited resources (Rothman and Forney, 2007; Torsvik, 2002) and the flow of fluid through the environment is of pivotal importance. Flow delivers nutrients, thus enabling cells to disperse to new habitats (Hall-Stoodley et al., 2004). However, as they grow, biofilms fundamentally alter the hydrodynamic properties of their environment – reducing its porosity and permeability, and potentially choking off their own access to nutri-

ents (Figure 1) (Cunningham et al., 1991; Shafahi and Vafai, 2009). A key challenge then is to understand how the interplay between the biotic and hydrodynamic features of an environment affects microbial ecology and evolution.

Engineers interested in bioremediation and wastewater treatment have modelled extensively how microbial populations colonise porous media. However, existing models have either been highly simplified (large-scale analytic models of broad environmental features, (Cunningham et al., 1991; Seifert and Engesgaard, 2012)), or very detailed (fine-scale individual based models of small, precisely defined porous geometries (Pintelon et al., 2009, 2012; von der Schulenburg et al., 2009)). Neither approach is appropriate for studying the ecology and evolution of microbial communities within complex environments. For this we require models detailed enough to capture the community dynamics occurring within the environment, but generalizable enough that we are not limited to a single, specific parameter regime. Here I have developed a mechanistic model that couples one-dimensional biofilm growth with traditional fluid mechanics. This model captures the interplay between microbes, their environment, and their neighbours, yet remains simple enough to be embedded within a theoretical framework based on the logic of evolutionary game theory. Each aspect of this model taken singly is relatively simple; however, in combination I am able to address an important biological question: "how do microbial communities compete and evolve within porous environments?"

The role of interspecies interactions and the microbial communities within us

Having focused on the environment, I next turn to communities themselves to examine how interspecies interactions affect community dynamics. I have chosen to concentrate primarily on the dense microbial communities within the mammalian gastrointestinal tract, which are currently a major focus of international research (Turnbaugh

et al., 2007). The human body contains approximately 100 trillion microbial cells that occur predominantly within the gastro-intestinal tract (Faith et al., 2013; Gill et al., 2006; Lozupone et al., 2012). Termed the gut microbiome, this microbial community is thought to be critical for our own health and wellbeing. A healthy gut community is implicated in processes ranging from the breakdown of complex molecules for food, through to healthy immune development (Bäckhed et al., 2005; Dethlefsen and Relman, 2011; Lee and Mazmanian, 2010; Mazmanian et al., 2008). However, while we have a vast and growing body of experimental data on these communities, we are still relatively ignorant about what governs their dynamics. In particular, we know little about how these communities initially assemble, nor do we understand what renders them stable once they have established.

It is clear, however, that to understand the microbiome we need to take careful account of how its members interact both with one another, and with the host. Microbiome communities are not only dense, but also highly interconnected; species interact with one another through a number of different mechanisms (Faust and Raes, 2012). In nature these interspecies interactions are thought to be predominantly competitive, yet the gut is often thought of as relatively cooperative, with microbes working together to break down complex food molecules [1] (Belenguer et al., 2006; Foster and Bell, 2012; Ze et al., 2012, 2013). These cooperative interactions are considered important for microbiome functioning. However, neither the nature, nor the effects of interspecies interactions within the microbiome have yet been substantially researched. Similarly, microbes interact in complex ways with their host species. These interactions include the host immune system targeting specific bacteria; epithelial secretions that have the potential to spatially segregate cells; and even the host actively feeding members of the microbiome (Derrien et al., 2010; Hooper, 2009; Johansson et al., 2008; Macpherson, 2000; Macpherson et al., 2005, 2008; Macpherson and Uhr, 2004; Petnicki-Ocwieja et al., 2009; Salzman et al., 2010; Wehkamp et al., 2004, 2005). The gut microbiome

thus represents a critical setting in which to understand the role of interspecies interactions in governing microbial community dynamics.

Despite this complexity, however, the field of microbiome research is lacking substantially in appropriate mathematical theory; rather, the majority of work to date has focused on gathering large amounts of empirical data. So far, most of the accompanying mathematical effort has therefore focused on developing tools to interpret this data – for example, mapping co-occurrence networks that determine whether certain bacterial groups are more or less likely to be observed together within experimental samples (Faust and Raes, 2012; Levy and Borenstein, 2013; Zhang et al., 2014). While useful, this approach does not tell us about the nature of the interactions between species, nor the underlying community dynamics. For example, while species might seem to co-occur because they are helping one another, it may actually be because both are attempting to occupy the same niche. More recent work has instead attempted to disentangle the precise nature of the interactions between species (Buffie et al., 2015; Marino et al., 2014; Stein et al., 2013) but these studies remain a small minority. Similarly, some work has set out to capture the interspecies dynamics within the microbiome using detailed individual based models (Schluter and Foster, 2012; Schluter et al., 2015), focusing on smaller sub-communities; however, such approaches are computationally intensive, and cannot easily be scaled up to complex microbiome communities. There is a clear need then for the development of predictive, systems level models that can capture the complexity of microbiome communities (Borenstein, 2012).

[1]Throughout chapters 3 and 4 we will refer to this cross-feeding process, where the net sign of the interaction is positive for both players (+/+) as "cooperation". This is so as to be consistent with its usage in the current literature in both theoretical ecology and microbiome research (e.g. (Allesina and Tang, 2012; Faust and Raes, 2012; Mougi and Kondoh, 2012)). However, it is important to note that the field of social evolution has invested a great deal in correctly defining terms such as cooperation. In particular, strictly speaking, cross feeding should not necessarily be considered cooperation in the sense that we are not strictly requiring that it incurs any fitness cost for either actor. Correctly defining this process will be especially important when moving from thinking about the ecology to the evolution of microbiome communities.

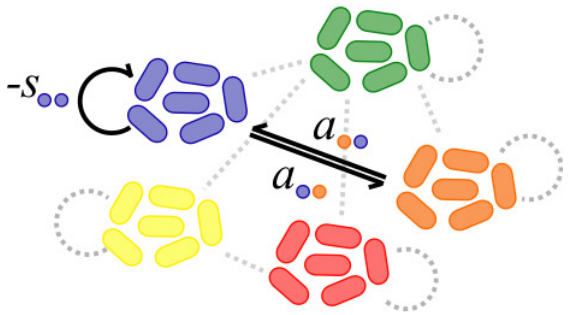

Figure 1.2: **Network models of microbial communities.** Theoretical ecology often models complex communities as networks, in which nodes represent individual species, while edges capture the interactions between them (whose strengths are captured by parameters a and s). Applying this technique to microbial communities offers a novel perspective with which to approach microbiome research.

In this thesis I draw upon modelling approaches developed in theoretical ecology where researchers for many years have been developing techniques to study diverse, interconnected communities. In particular, I model microbiome communities as ecological networks, in which each species is represented as a node in the network, while each edge captures the interactions between them (Figure 1.2). Rather than needing to simulate vast communities, I can ask instead how changing the biological aspects of a community influences this network, and the subsequent effects this will have on its ecological properties. This approach thus offers an excellent tool for interrogating communities that would otherwise be too complex to study using techniques traditionally associated with modelling microbial communities.

The greatest gains in our understanding of complex microbial communities will, however, come through the combination of mathematics with careful experimental work. Such a combination is vital for motivating initial modelling, forming predictions, and ultimately testing theories experimentally – iteratively repeating this process in order to continuously build our understanding of a given system. To this end, in Chapter 2 I

have conducted such experimental work myself, while in Chapters 3 and 4 I have performed some first tests of my predictions using a pre-existing dataset. However, there is much still to be done, and my concluding chapter discusses the approaches that will be critical to building a combined experimental and theoretical framework of microbial communities.

1.3 Thesis outline

Chapter 2: Flow mediates microbial evolution in porous environments

This chapter focuses on how bacteria interact and evolve within porous environments. Here I combine novel microfluidic experiments, mechanistic models, and game theory to study how hydrodynamics mediate competition between bacterial genotypes. With these I show that hydrodynamic interactions between competing genotypes give rise to an evolutionarily stable growth rate that stands in stark contrast with that observed in typical laboratory experiments: a strain can outcompete other genotypes by growing slowly. This work reveals that fluid dynamics can play a pivotal role in how bacteria compete and evolve within complex environments.

Chapter 3: The ecology of the microbiome: Networks, competition, and stability

Here I focus on developing predictive models to understand the factors driving ecological stability within the gut microbiome. I combine analytical and numerical mathematics with individual-based models in order to understand how interspecies interactions and host biology influence microbiome stability. This work suggests that cooperation – often thought to be important for microbiome function – can be destabilizing, and therefore that hosts may benefit from permitting inefficient competitive species.

Chapter 4: The assembly of microbiome communities

This chapter develops a theoretical framework to capture the principles underlying the assembly of microbiome communities. Specifically, I extend concepts from theoretical ecology to understand what promotes the robust assembly of diverse microbial communities. This shows again that cooperation between species can cause problems for microbial communities. Here, cooperation can render communities difficult to assemble and generate strict order effects, leading to hysteresis in the assembly process. Moreover, I illustrate how this modeling framework can be applied to real data in order to identify key processes or community members involved in the assembly of known microbiome communities.

2

Flow mediates microbial evolution in porous environments

2.1 Abstract

Microbes often live in dense communities called biofilms where competition between strains and species is fundamental to both evolution and community function. While biofilms are commonly found in soil-like porous environments, the study of microbial interactions has neglected a key currency for competition in these settings: fluid flow. Here we use novel microfluidic experiments, mechanistic models, and game theory to study how porous media hydrodynamics can mediate competition between bacterial genotypes. Our experiments reveal a fundamental challenge faced by microbial strains that live in porous environments: dividing too rapidly can block fluid flow and redirect resources to a competitor. To understand how these dynamics influence the evolution of bacterial growth rates we couple a model of flow-biofilm interactions with a game theory analysis. This reveals that hydrodynamic interactions between competing genotypes

give rise to an evolutionarily stable growth rate that stands in stark contrast with that observed in typical laboratory experiments: a strain can outcompete other genotypes by growing slowly. Our work reveals that hydrodynamics can profoundly affect how bacteria compete and evolve in porous environments, the habitat where most bacteria live.

Statement of significance

The overwhelming majority of bacteria live in porous environments, like soil, aquifers, and ocean sediments, where they facilitate many important processes. Despite this, we understand little about how these complex environments shape the composition of the microbial communities that live within them. Here we combine two major bodies of theory for the first time – fluid dynamics and game theory – to understand how bacteria evolve. We show that bacteria in porous environments face a fundamental dilemma: they rely on flow for nutrients and dispersal, however, as cells grow they tend to reduce their access to flow, diverting it instead to competitors. This means that slower-growing bacteria can have a competitive advantage and suggests that current paradigms of microbial competition need revision.

2.2 Introduction

Modern microbiology relies on growing cells in liquid cultures and agar plates. While these generate homogeneous conditions that enable high throughput and repeatability, they lack the complex physical and chemical landscapes that microbes experience in their natural environments. This environmental heterogeneity is increasingly recognized to exert a powerful influence on microbial ecology across a wide diversity of habitats, ranging from the ocean to the human gut (Azam and Malfatti, 2007; Bever et al., 2009; Horner-Devine et al., 2004; Schluter and Foster, 2012). While advances in sequencing technology now allow us to resolve how the genetic composition of microbial communities changes in response to environmental conditions (Allen and Banfield, 2005; Franzosa et al., 2015), we often lack a mechanistic understanding of the underlying processes. Novel empirical approaches, which simulate the conditions found in realistic microbial habitats, are needed to understand the strategies that cells use to gain an advantage over their competitors (Rusconi et al., 2014).

The overwhelming majority of bacteria live in porous environments, between the particles that compose soil, aquifers, and sediments, where they constitute approximately one-half the carbon within living organisms globally (Whitman et al., 1998). Cells in porous environments typically reside in surface attached structures known as biofilms (Ghiorse and Wilson, 1988), in which diverse bacterial genotypes live under intense competition for limited resources (Rothman and Forney, 2007; Torsvik, 2002). Recent efforts have identified specialized mechanisms that cells use to gain advantage over competing genotypes in biofilms, ranging from the secretion of toxins, to polymer production, and metabolic regulation (Hibbing et al., 2010; Kim et al., 2014; Klausen et al., 2003; Moons et al., 2009; Nadell et al., 2008; Rao et al., 2005; Xavier and Foster, 2007). While genotypic competition is most frequently studied in biofilms growing on simple flat surfaces, biofilms growing in the interstitial spaces within porous structures

face additional constraints. In porous environments space is much more limited, and biofilm growth tends to attenuate the fluid flow that supplies cells with nutrients and facilitates dispersal.

Biofilms typically reduce the flow through porous environments by orders of magnitude at the macroscopic Darcy scale, where the flow is averaged over many pore spaces (Thullner, 2010). Harnessing this effect, biofilms can be used to limit the transport of pollutants that have leaked into groundwater aquifers and to facilitate the extraction of petroleum from recalcitrant regions of reservoirs (Cunningham et al., 2003; Wang and Hendry, 2012). However, biofilm-induced clogging also generates unwanted effects: for example it severely limits the efficiency of porous filtration systems (Nicolella et al., 2000) and curtails the rate at which water infiltrates into aquifers, exacerbating droughts (Kim et al., 2010). Due to its importance, the attenuation of flow by biofilms has long been studied at the Darcy scale (Cunningham et al., 1991; Shafahi and Vafai, 2009) and more recent works have sought to resolve how this is in turn mediated by biofilm-hydrodynamic interactions at the microscopic pore scale (Drescher et al., 2013; Dupin et al., 2001; von der Schulenburg et al., 2009). However, it is unknown how these interactions influence the ecology and evolution of the bacteria themselves. Here, we combine experiments and models to show that porous media hydrodynamics can dramatically affect the principles of bacterial competition and evolution.

2.3 Results

A conceptual model to study hydrodynamic interactions between competing biofilms

Bacteria within biofilms tend to form patches of genetically-identical cells, even when the cells from which they are founded are initially mixed. This occurs because *in situ* growth, combined with the low mobility of cells within biofilms, tends to aggregate mul-

tiple generations of cells in the same place (Millet et al., 2014; Nadell et al., 2010). This clonal patchiness occurs because biofilms are often initiated from relatively sparse cells, and because nutrient limitation promotes population bottlenecks as a biofilm expands (Hallatschek et al., 2007; Mitri and Foster, 2013; van Gestel et al., 2014). Based on these observations, we focus here on the competition between localized biofilm patches that are each comprised of a single genotype.

To investigate how biofilm growth influences the flow through a porous environment we calculated the flow through a representative network of pore spaces that is driven by a difference in pressure at the boundaries (Figure 2.1 *A*, *Materials and Methods*). The addition of a small impermeable biofilm 'patch' sharply reduces the flow through the pore in which the biofilm resides, while concurrently increasing the flow through neighboring pores (Figure 2.1 *B, C*). While the magnitude of this flow diversion depends on the specific geometry of the pore space, this simulation shows that as a biofilm patch proliferates, it tends to decrease its access to flow, whilst increasing the flow to patches of biofilm that reside along other flow paths. This differs from bacterial competition in classical biofilm assays, where different genotypes growing together on flat surfaces typically have to be in close proximity to interact, for example through capturing one another's nutrients, or via secretions. In porous environments, biofilms can influence one another over much larger distances by affecting one another's access to flow.

Though porous substrates harbor many biofilm patches that can simultaneously perturb one another's flow environment, we idealize this network of interactions as a collection of its constituent pairwise interactions. We then resolve the dynamics of competition between a single pair of biofilm patches, each of which is composed of a different genotype. In this pairwise approximation, the proportion of the total volumetric flow rate, q_T, that passes each biofilm is a function of the hydrodynamic resistance of both its pore space and that of its competitor, which each are in turn a function of the thicknesses of the biofilms, k_1 and k_2 (Figure 2.1 *D, E*). In this model the growth of a biofilm

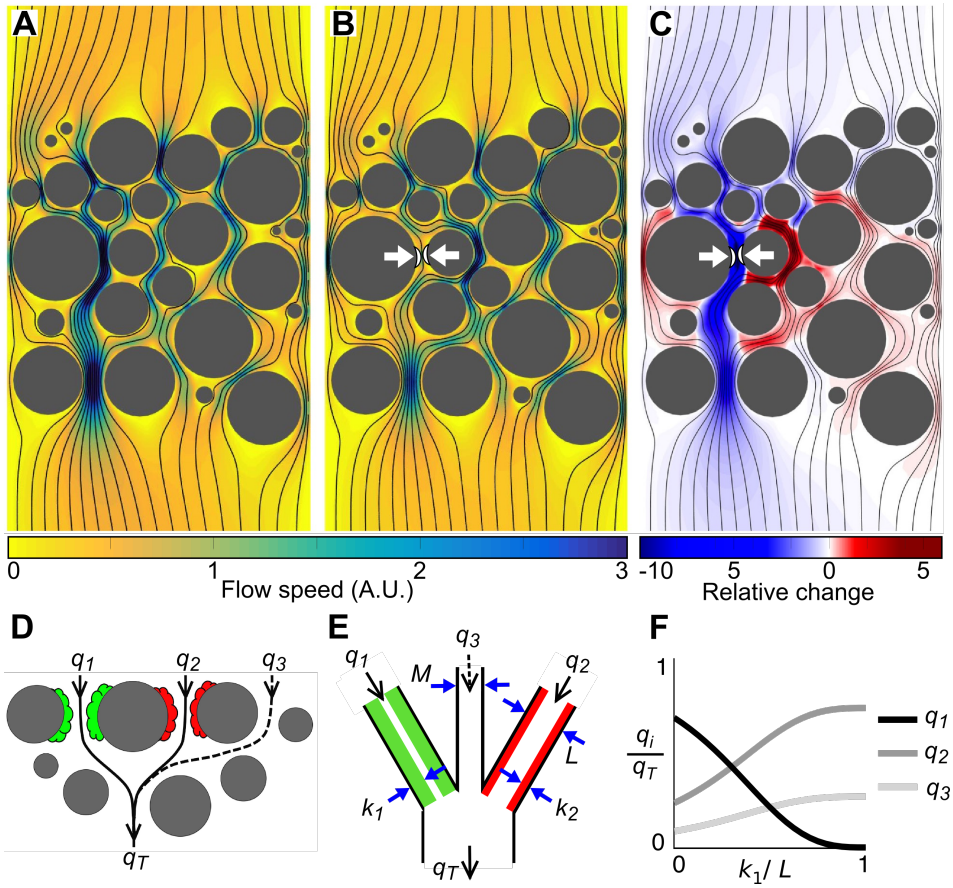

Figure 2.1: **A growing biofilm tends to decrease its access to flow whilst increasing the flow to its competitors.** (A) Viscosity dominates inertia in most porous environments, owing to the relatively small pore spaces (10 μm - 1 mm) and slow fluid velocities (1 - 1000 μm s^{-1}) (Biggar and Nielsen, 1976; Sinton et al., 2010), which allows flow to be modeled using the Stokes equations. Here, we numerically solved the Stokes equations within a representative two-dimensional porous geometry. Flow is driven by a fixed pressure difference between the top and bottom boundaries, whilst the left and right boundaries are impermeable (see *Materials and methods* for further details). Black lines show streamlines and the color map shows the flow speed in arbitrary units (A. U.). (B) The flow field after the addition of a small impermeable patch of biofilm (white arrows). All other parameters of the simulation remained constant. (C) The relative change in flow speed measured as $(s_a - s_b)/s_a$, where s_a is the initial flow speed and s_b is the flow speed after the addition of the biofilm patch, shows that the biofilm sharply decreases the flow through the pore in which it resides and increases the flow through neighbouring pore spaces. (D) A cartoon of two biofilm patches (green, red) that interact hydrodynamically. The proportion of the total flow, q_T, that moves past each biofilm changes as the biofilms grow and increase the hydrodynamic resistance of their respective pore spaces. A third flow path (dotted line) models the ability for flow to divert around the two competing biofilms. (E) Our conceptual model where two biofilms, with thicknesses k_1 and k_2, live in neighbouring pore spaces of width L, that are connected to a flow path of width M that does not harbor any biofilm. The proportion of the total volumetric rate flow, q_T, that passes along each of the three flow paths is calculated using Kirchoff's laws assuming planar Poiseuille flow in each pore space (see *Materials and Methods*). (F) Analogous to our Stokes flow simulations, if k_1 increases in thickness the proportion of the total flow rate through its pore space, q_1/q_T, decreases, whilst increasing the amount of flow, q_2/q_T, received by the neighbouring biofilm. Here $k_2/L = 0.3$ and $M/L = 1$.

tends to decrease its access to flow and increase the flow past its competitor (Figure 2.1 *F*). Importantly, our pairwise model captures the dynamics observed in our Stokes flow simulation, but is much more tractable and easily parameterized. Flow through a network of porous spaces can be modeled by fixing either the pressure gradient or the flow rate at the boundaries (Pintelon et al., 2009), with the former better characterizing flow through natural systems. However, localized biofilm growth in either of these scenarios will produce a flow diversion at the pore scale, as observed in our conceptual model.

Microfluidic experiments show bacteria that grow more slowly have a competitive advantage when flow is relatively weak

We next developed a microfluidic version of our pairwise flow model to experimentally test how pore scale hydrodynamics affects the competition between genotypes that grow at different rates (Figure 2.2). For this we chose a well-studied *Escherichia coli* experimental system. Specifically, we competed wild-type *E. coli* cells with $\Delta rpoS$ cells. The latter cells lack the ability to produce the sigma factor RpoS and, as a result, form biofilms at a much slower rate than the parental genotype (Figures 2.2 *E*, S2.5, (Adams and McLean, 1999; Ito et al., 2008)). The two genotypes were inoculated separately in each arm of the device, each of which represents a pore (Figure 2.2 *A, B*). One of the pores was irrigated with media mixed with dye, whilst the other was irrigated with clear media, which allowed us to measure the relative proportion of flow passing through each pore by tracking the dye interface downstream of their juncture. Control experiments showed that neither the dye, nor the fluorescent proteins used to differentially label the strains, had an appreciable effect on biofilm formation (Figure S2.5).

In porous environments biofilm growth is opposed by flow-induced detachment, which reduces the thickness of biofilms by shearing away cells from its surface (Stewart, 1993; Stoodley et al., 2002; Trulear and Characklis, 1982). To simulate different ambient flow conditions in our experiment, and thus the relative amount of detachment,

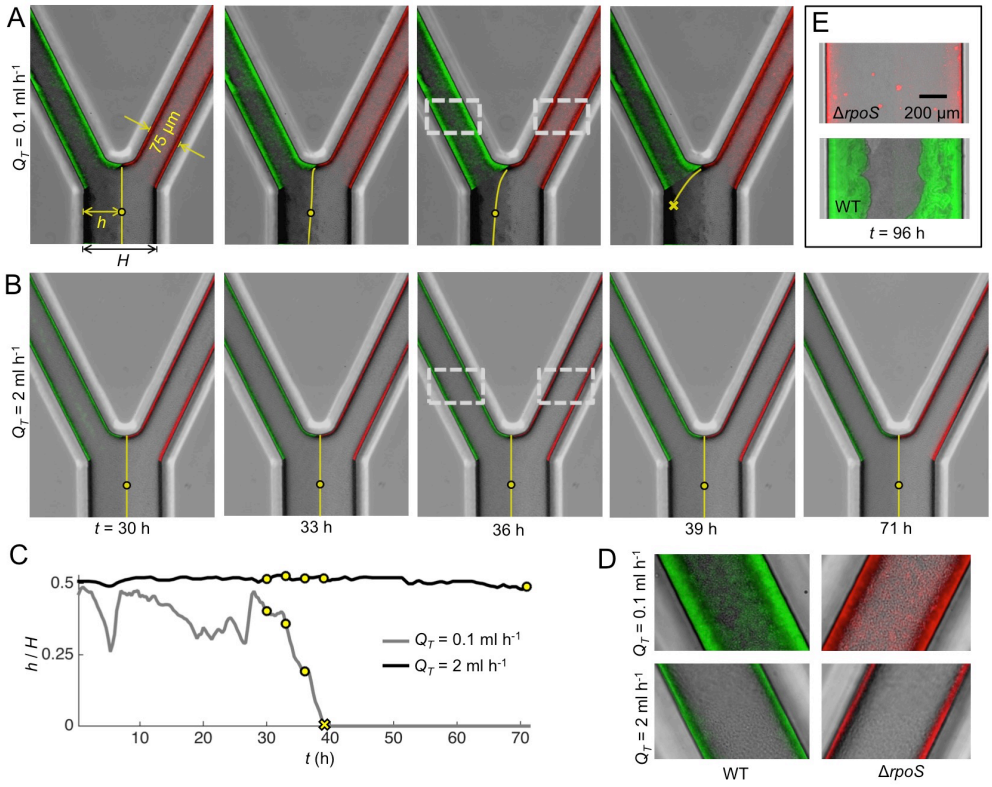

Figure 2.2: **Microfluidic competition experiments show that slower growing genotypes are favored when flow is relatively weak, whereas faster growing genotypes are favored when flow is relatively strong.** (A, B) The left pore of each device was seeded with wild-type cells (green), whereas the right pore was inoculated with $\Delta rpoS$ cells (red). Dyed media flows through the right pore, whereas clear media flows through the left pore. The dye interface downstream of the two pores (yellow line) allows us to dynamically track relative proportion of the total flow, q_T, moving through each pore space (Methods). (C) Following the movement of the dye interface (yellow circles in A, B) shows in the weak flow treatment (A) the wild-type biofilm fully blocked its pore space 39 hours after seeding. However, in the strong flow treatment (B) both biofilms are able to maintain access to flow for more than 70 hours after they are initiated. (D) In the weak flow treatment (dashed white squares in, A), the wild-type biofilm was thicker than the $\Delta rpoS$ null biofilm, which was responsible for the diversion of flow (see text). In strong flow, both biofilms were much thinner, such that the difference in biofilm thickness between the two strains was small. (E) The observation that wild type biofilms grow at a faster rate than $\Delta rpoS$ biofilms was confirmed in separate microfluidic experiments that exposed attached cells to much smaller shear stresses than in the competition experiment (see *Materials and Methods*), which minimized the effect of flow induced detachment.

we applied a total flow rate of either $q_T = 0.1$ ml h^{-1} or $q_T = 2$ ml h^{-1}. In the low flow treatment, the rapidly growing wild-type biofilm increased its pore's hydrodynamic resistance and diverted flow to the neighboring pore space containing the $\Delta rpoS$ biofilm. The reduction in flow experienced by the wild-type biofilm further reduced its detachment, driving a positive feedback loop that ultimately ended with the wild-type biofilm completely blocking its pore space (Figure 2.2 *A, C, D*). In contrast, under the high flow treatment, the growth of the wild-type biofilm was balanced by flow-induced detachment such that both genotypes were able to maintain access to flow for the entire duration of the experiment (Figure 2.2 *B, C, D*). Access to flow is essential for biofilms to acquire nutrients and disperse progeny downstream: these results suggest that a slow growing biofilm has a competitive advantage over fast growing biofilm when the magnitude of flow is below a critical value. However, fortunes reverse when flow is stronger, because faster growing genotypes are able to disperse more cells downstream than their slower growing competitors.

A mathematical model of two hydrodynamically coupled biofilms reveals a diversity of competitive regimes

Our microfluidic competition experiments represent a proof-of-principle that rapid proliferation within a pore space can lead to blocking and ultimately a redirection of resources to slower-growing genotypes. In order to understand this process better, we next moved to a model that resolves how two competing biofilms interact via hydrodynamics, enabling us to explore a much wider range of competitive scenarios. While the two pores in our experiment are strongly coupled, such that flow diverted from one pore is fully absorbed by the other pore, in a network of pores the strength of the hydrodynamic coupling between two competing biofilms will vary depending on the geometry of the pore space and their relative proximity to one another (Figure 2.1 *A, B*). To account for this, we consider two identical fluid pathways of width L colonized by

biofilms of thickness k_1 and k_2, which are connected in parallel to a channel of width M that does not contain biofilm. The dimensionless parameter $M^* = M/L$ then measures the ability for the two biofilms to influence one another via flow: $M^* = 0$ corresponds to the strong coupling observed in our experiments, whereas for increasing M^* flow is more likely to be diverted around the focal biofilms as they proliferate. Importantly, for $M^* > 0$, both biofilms are capable of clogging. This model then provides us a tractable way to resolve how changing the strength of the hydrodynamic interaction between two biofilms affects their dynamics, without requiring an explicit representation of the pore structure.

A wide range of physical and biological processes can affect biofilm development (Donlan, 2002), however, the thickness of biofilms in flowing environments is chiefly governed by the balance between cell division and flow induced detachment (Duddu et al., 2009; Van Loosdrecht et al., 1995). Cell division in biofilms is often confined to a layer at the exterior of the biofilm, where substrates are exposed to nutrients from the flow (Werner et al., 2004; Williamson and McCarty, 1976). The characteristic thickness, δ, of this metabolically active layer is set by the balance of the diffusion of the substrate into the biofilm with its consumption, which yields the expression $\delta = \sqrt{2c_0DY/\alpha}$ (Pirt, 1967), where c_0 is the substrate's concentration at the outer surface of the biofilm, D is the diffusion coefficient of the substrate in the biofilm, α is the bacterial growth rate, and Y the yield with which cells convert substrate to biomass. The rate at which the biofilm increases in thickness due to cell division is then given by the product of the growth rate, α, and $\min(\delta, k)$, such that $dk/dt = \alpha \min(\delta, k)$, which takes into account that the entire thickness of a biofilm is actively growing when $k < \delta$. Increases in biofilm thickness are countered by the detachment of cells due to mechanical forces exerted at the surface of the biofilm by fluid motion (Trulear and Characklis, 1982). While the literature contains a diversity of parameterizations to model flow-induced biofilm detachment (see (Horn and Lackner, 2014) for a comprehensive review), a formulation

based on the empirical study of (Rittmann, 1982) is one of the most widely used (Abbas, 2011; Abbas et al., 2012; Brovelli et al., 2009; Duddu et al., 2009; Ebigbo et al., 2010; Wanner and Gujer, 1986) and has been independently confirmed for biofilms growing in porous media (Brovelli et al., 2009; Ebigbo et al., 2010). Here the detachment rate is approximated by $dk/dt = -\chi k \tau^{1/2}$, where χ is an empirical parameter with units of $\sqrt{\text{length}/\text{mass}}$ that measures the ability of the biofilm to resist detachment, and τ is the shear stress exerted by the flow on the surface of the biofilm. Assuming planar Poiseuille flow, a biofilm in our model experiences a shear stress of $\tau = 6\mu q_i/h(L-k_i)^2$ where q_i is the flow rate through the pore in which the biofilm resides, h is the span wise depth of the pore, and μ is the dynamic viscosity of the fluid. As such, we model changes in the thickness of biofilm via the superposition of cell division growth and flow-induced detachment,

$$\frac{dk_i}{dt} = \alpha_i \min(\delta, k_i) - \chi k_i \tau^{1/2}. \tag{2.1}$$

We used our model to simulate the development of two biofilms, which grow at rates α_1 and α_2, respectively and are coupled using our simplified flow model (Figure 2.1 *E*, *Materials and Methods*). To reduce the number of tuneable parameters, we non-dimensionalized the coupled differential equations governing the biofilm thicknesses, k_i, to yield four dimensionless parameters: $\alpha^* = \alpha_2/\alpha_1$, the ratio of growth rate of the two biofilms, $\beta^* = \sqrt{6\mu q_T/hL^2\alpha_1^2}$, the magnitude of growth relative to flow induced detachment for the slower growing biofilm, $\delta^* = \sqrt{2c_0DY/\alpha_1 L^2}$ the non-dimensional growing edge thickness of the slower growing biofilm, and $M^* = M/L$ the strength of the hydrodynamic coupling between the two biofilms. Analogous to our experiments, we initialized each pore with a thin biofilm layer ($k_i = 0.01$) and calculated the thicknesses of the two biofilms until they reached steady state. To isolate how the relative strength of the flow affected the biofilms, we fixed $\delta^* = 0.3$ and $M^* = 1$ to focus our attention on the (α^*, β^*) phase plane.

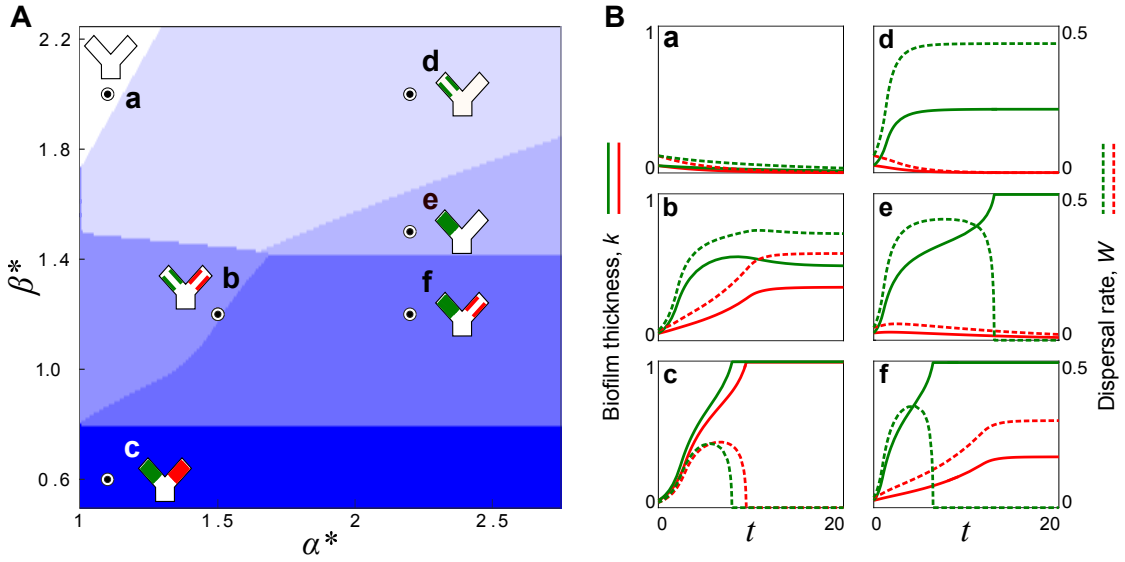

Figure 2.3: **Diverse ecological regimes emerge from a model of biofilm competition where two strains are coupled by flow.** (A) The phase space formed by α^*, the growth rate of a fast growing biofilm (green) divided by that of a slower growing biofilm (red), and β^*, the non-dimensional parameter that measures the importance of biofilm growth relative to flow induced detachment, reveals six different regimes at steady-state. (B, a-f) Here we plot the biofilm thicknesses, k_1, k_2 (solid lines), and the dispersal rates, W_1, W_2 (dashed lines), for a representative simulation in each of the regimes (circles in A). When a biofilm is fully scoured from the surface ($k_i = 0$) or completely blocks its pore space ($k_i = 1$) its dispersal goes to zero ($W = 0$). In contrast, if a biofilm thickness reaches a non-trivial fixed point $0 < k_i < 1$, it disperses cells downstream at steady-state. Here $M^* = 1, \delta^* = 0.3$. For clarity, we have omitted the third flow path from the cartoons in A.

Our model predicted a diversity of different ecological outcomes (Figures 2.3, S2.6). When flow was relatively weak ($\beta^* < 0.8$), biofilm growth dominated detachment, such that the positive feedback between increased flow diversion and reduced detachment ultimately led to both biofilms fully blocking their pore spaces (region c). In the opposite limit when flow was relatively strong ($\beta^* \geq 1.5$), detachment dominated the growth of the slower growing biofilm, which was completely scoured away from the surface. In this case, the faster growing genotype either fully detached (region a), reached a steady equilibrium thickness (region d), or blocked its own pore entirely (region e)

depending on the asymmetry in growth rates, α^*. When flow was at an intermediate level ($\beta^* \approx 0.8$ to 1.5), two outcomes were possible depending upon the value of α^*: if genotypes grew at a similar rate ($\alpha^* \approx 1$) the fast growing biofilm initially diverted flow away from its own pore space. However, as the thickness of slower growing biofilm increased over time, it diverted flow back towards the faster strain, and this stabilizing effect allowed both strains to access flow and disperse cells downstream at steady state, with the faster growing biofilm dispersing at a larger rate (Figures 2.3*b*, S2.6). If the asymmetry in the growth rates of the two strains was larger ($\alpha^* > 1$) in this intermediate flow regime, the slower growing strain was not able to stabilize the runaway growth of its neighbour and the faster growing genotype blocked its pore space (Figure 2.3 *f*). Using the biofilm's dispersal rate at steady state, W, (equivalent to the rate at which new biofilm is formed at steady state) as an objective measure of fitness, these results indicate that slower growing biofilms are favored when flow is relatively weak, whilst faster growing biofilms are favored when flow is relatively strong. This finding is in agreement with the two distinct flow regimes observed in our microfluidic experiments.

The impact of flow on the evolution of bacterial growth rate

Our model shows that a biofilm's fate depends not only on its growth rate, but also on the growth rate of other biofilms elsewhere within the porous network. But how do these hydrodynamic interactions impact the evolution of bacteria? To answer this, we implemented our mechanistic model in a game-theoretical invasion analysis that simulates how bacterial growth rate evolves over many successive rounds of competition. More specifically, we used adaptive dynamics (Brännström et al., 2013), which tests whether a "mutant" genotype that grows at rate α_M is able to invade a population of "resident" biofilms that grow at rate α_R, to determine if the flow conditions within a given porous environment will select for bacteria with a specific growth rate. Again we assume that the fitness of a genotype is directly proportional to its rate of cell dispersal at steady

state, which balances the rate of biofilm growth at steady state (Eqn. 2.1). The fitness of a biofilm, W, is a function of its own growth rate, its competitor's growth rate, and the properties of the pore space (Figure 2.3). Two conditions must hold for a mutant to supplant the resident genotype. First, the fitness of the mutant when competing against the resident, denoted $W(\alpha_M, \alpha_R)$, must be larger than or equal to that of the resident competing against itself, $W(\alpha_R, \alpha_R)$. This determines whether an initially rare mutant is able to gradually increase in frequency within the population. Second, the fitness of the resident when competing against the mutant, $W(\alpha_R, \alpha_M)$, must be smaller than that of the mutant competing against itself $W(\alpha_M, \alpha_M)$. This determines if the original resident growth strategy will be able to re-invade the system once the mutant has increased in frequency. More formally, the following two criteria must hold for a mutant to invade a resident population:

$$W(\alpha_M, \alpha_R) \geq W(\alpha_R, \alpha_R), \tag{2.2}$$

$$W(\alpha_R, \alpha_M) < W(\alpha_M, \alpha_M). \tag{2.3}$$

These criteria were then used to construct a pairwise invasibility plot (Figure 2.4, (Brännström et al., 2013)), with which we could then infer the evolutionary trajectory of the population's growth rate.

Our analysis reveals that competition between strains in porous environments will ultimately drive a population towards an evolutionarily stable biofilm growth rate, α_{ESS} (Figure 2.4 *A*). Intuitively, when a biofilm that grows at α_{ESS} competes against a faster growing strain, the latter will block its pore space. Conversely, when a biofilm that grows at α_{ESS} competes against a slower growing strain the latter will disperse cells downstream at a slower rate. We resolve how α_{ESS} varies as a function of the environmental conditions (Figure 2.4 *B-D*). Increasing β^* – for example by increasing the total flow rate – leads to a larger α_{ESS}, whilst increasing the non-dimensional growing

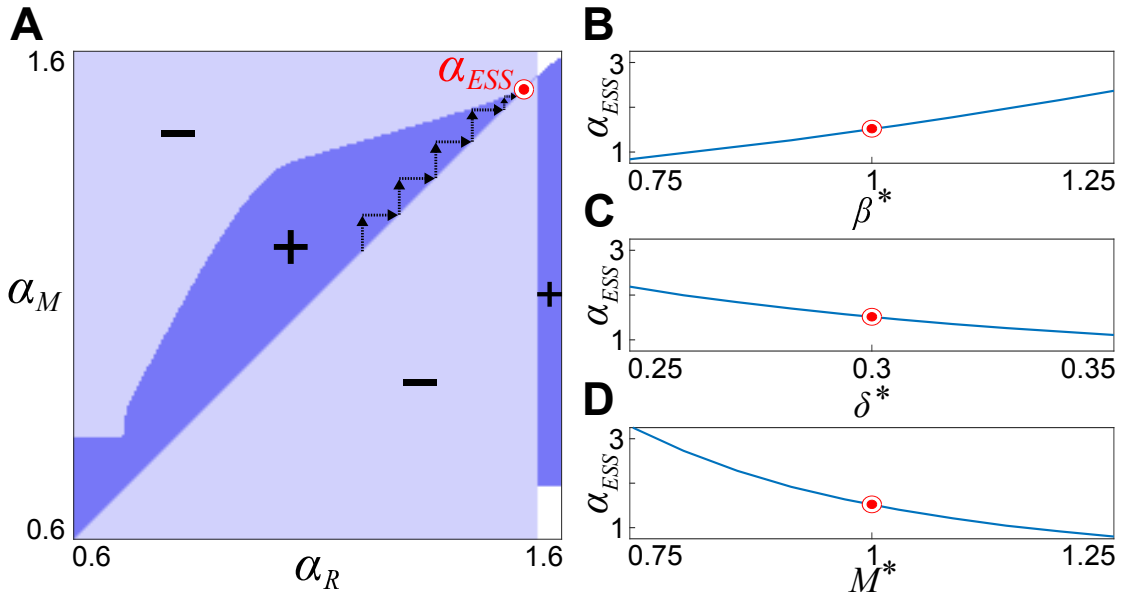

Figure 2.4: **A game theoretical analysis of the coupled biofilm model predicts an evolutionary stable growth rate.** (A) We used adaptive dynamics to construct a pairwise invasion plot, which maps the region of parameter space where a mutant, which grows at rate α_M, can invade a resident population of biofilms, which grow at rate α_R (see text). The mutant can invade in the dark blue regions ($+$) and cannot invade in the light blue regions ($-$). In the white regions the mutant and the resident biofilms both have a fitness of zero ($W = 0$) because they have either been fully detached by flow, or have blocked their pore space. Arrows show an example evolutionary trajectory where mutant genotypes successively replace the resident population, driving the growth rate towards the evolutionary stable growth rate, α_{ESS} (red circle). Here we set $\delta^* = 0.3\alpha_R^{-1/2}$ and $\beta^* = 1.1\alpha_R^{-1}$ and $M^* = 1$. (B, C, D) To determine the effect of β^*, δ^*, and M^* on α_{ESS}, we held two of these parameters constant and varied the third (red circles show fixed values).

edge thickness δ^* – as can occur when nutrients are more plentiful – leads to a smaller α_{ESS}. Moreover, the connectivity of the porous structure also influences this process: increasing M^*, which increases the ability of flow to bypass the focal biofilms, leads to a reduction in α_{ESS}. All of these trends are consistent with the idea that increasing the potential for blocking, whether through reduced detachment, increased growing layer thickness, or an increased ability for flow to divert around competing biofilms, would promote the evolution of slower-growing genotypes. These results suggest that pore

blocking places a fundamental physical limitation on the evolution of bacterial growth rates in porous environments, and stand in stark contrast with that observed in many laboratory assays, where evolution often selects for the fastest growing genotype (Hibbing et al., 2010).

2.4 Discussion

Biofilms growing in porous environments facilitate a wide range of important processes in the natural environment and industry (Danhorn and Fuqua, 2007; De Muynck et al., 2010; Kim et al., 2010; Leite et al., 2008; Michalakos et al., 1997; Mitchell et al., 2010; Nicolella et al., 2000; Wang and Hendry, 2012; Whitman et al., 1998). Our proof-of-principal experiments, mathematical modelling and game-theoretic analyses show that the feedback between biofilm proliferation and porous media hydrodynamics dramatically affects how different genotypes compete. We find that relatively fast and slow flow conditions favor fast and slow growing biofilms respectively, while intermediate flow rates allow biofilms with different growth rates to maintain access to flow (Figure 2.3).

In industrial settings, these new principles could be exploited to engineer microbial systems to favour a bacterial species with a particular growth rate, or keep multiple species with different growth rates active over longer time scales. For example, in porous wastewater reactors relatively fast growing species of bacteria convert ammonia to nitrite, but it is desirable to inhibit often slower growing species that further oxidize these products into nitrate, a potent environmental contaminant (Cho et al., 2011). Our work predicts then that using a larger flow rate may be a way to favour the former species of bacteria over the latter. In contrast, the remediation of mercury contaminated wastewater in porous reactors can be enhanced by maintaining multiple species of bacteria that grow at different rates (Von Canstein et al., 2002). Moreover, our findings

suggest that inoculating porous substrates with a community of cells from the effluent of a porous system would favor the growth of biofilms that do not block their pore space, whilst inoculating cells from communities that have evolved in homogeneous laboratory conditions would promote blocking. Such information has implications for the design of effective water treatment systems, where blocking reduces efficiency, or in the design of bio-barriers to stifle the movement of groundwater contaminants, where blocking is the main objective.

Our results may also shed light on how cells compete in natural environments. We expect that heterogeneity in pore size and temporal fluctuations in flow will promote diversity. These are expected to be common due to geological processes that generate localized patches of different sized particles (Rehfeldt et al., 1992) and episodic patterns of rainfall. However, some groundwater aquifers and packed bed bioreactors have a more uniform distribution of pore spaces and nearly constant rates of flow, which may promote competitive exclusion. In systems where blocking does occur, natural selection may favour cells periodically detaching en masse to regain access to flow. Broadly consistent with this, quorum sensing and nutrient deprivation induced biofilm weakening have both have been observed empirically (Kong et al., 2006; Sauer et al., 2004).

Bacteria are the subject of intense empirical and theoretical study. However, the vast majority of work on bacteria focuses on their behavior in liquid cultures or in simple biofilm assays. Here we have combined diverse bodies of theory, including fluid dynamics and game theory, to understand how bacteria compete and evolve within the complex porous environments where most bacteria live. Our assumptions greatly simplify the complexity of these systems, so there is considerable potential for useful extensions to our work. Nevertheless, our approaches are sufficient to show that porous habitats can have a profound impact on bacterial evolution. Consideration of fluid flow, and the competition that this brings, shows how slow growing-genotypes can evolve. While rapid division gives a microbe a distinct advantage in typical laboratory environ-

ments, our results suggest that this paradigm does not extend to the majority of bacterial habitats.

2.5 Materials and methods

Modelling Stokes flow through a representative network of pore spaces

The geometry of the pore space (Figure 2.1 *A*-*C*) was obtained using PFC2D (Particle Flow Code in Two Dimensions, ITSCA), which models the mechanical processes that form many porous substrates. The particle locations were then imported into COMSOL Multiphysics to model incompressible Stokes flow within the pore spaces between the particles using the finite element method. Zero flux, no slip boundary conditions were used at the left and right boundaries of the computational domain as well as on the surfaces of all the particles. At the top and bottom boundaries of the computational domain the pressure was fixed at two different values, such that the resulting pressure gradient was responsible for driving flow. The Stokes equations were solved with and without the presence of a biofilm 'patch' in one of the pore spaces. Results were then exported into Matlab 2015a (MathWorks) for further analysis and plotting.

Bacterial strains and culturing

Our experiments used *Escherichia coli* strain K12-W3110 and a mutant with a rpoS819 allele insertion (Keymer et al., 2008; Lambert et al., 2011). Each strain was labeled with either green fluorescent protein (GFP) or red fluorescent protein (RFP). Cell cultures were grown overnight in Tryptone broth (1 x TB, 10 g Tryptone per 1 L water) at 37°C, diluted to an optical density of 0.1 (at 600 nm), and then grown for a further hour at 37°C so that cells were in exponential phase when they were first introduced into the microfluidic devices.

Competition Experiments

Microfluidic device masters were fabricated from SU-8 on silicon wafers using standard soft lithography techniques (Whitesides and Stroock, 2001) and were cast with polydimethylsiloxane (PDMS, Sylgard 184, Dow Corning). The depth of these devices was 75 μm. Cured PDMS was bonded to glass coverslips (50 mm by 75 mm, No. 1 thickness, Agar Scientific) with a corona discharge system (BD-20AC, Electro-Technic Products) using previously described techniques (Haubert et al., 2006). Tygon tubing (Microbore, 0.51 mm inner radius) was used to plumb the inlets and outlets of the device.

For the high-flow rate treatments ($q_T = 2$ ml h^{-1}) the outlet was connected to a 140 mL syringe (Harvard Apparatus), which was mounted on a Harvard Apparatus 2000 syringe pump. In the low-flow treatments ($q_T = 0.1$ ml h^{-1}), the outlet was connected to a 20 mL syringe (Becton Dickinson) mounted on a Harvard Apparatus PhD Ultra syringe pump.

After the microfluidic device and tubing was primed with 1 x TB to remove air from the system, cells were introduced into the device by pulling a culture of wild type or $\Delta rpoS$ mutant cells through the device at $q_T = 0.1$ ml h^{-1}. Unlike many biofilm experiments, where cells are allowed to attach to surfaces in the absence of flow (e.g. (Drenkard and Ausubel, 2002; Klausen et al., 2003; Moller et al., 1998)), we inoculated cells under flow to keep the two strains confined to their respective pore spaces. Cultures of the wild type and $\Delta rpoS$ mutant were simultaneously flowed through the device for 20 hours to allow cells to attach and then we switched over to withdrawing tryptone broth (0.5 x TB, 5 g Tryptone per liter of water) through the device for a further 48 hours so that a thin biofilm was established in each of the pore spaces. We initiated the biofilms in the high and low flow treatments in the same way in the first three days of the experiment to ensure that the cell attachment was identical between the two treatments.

After this initial inoculation phase, we connected one of the inlets of the device to a reservoir containing tryptone broth (0.5 x TB) mixed with dye (Chicago Blue, Sigma Aldrich) that enabled us to dynamically track the relative proportion of flow passing through each side of the device.

After the initial inoculation phase, we applied a flow rate of $q_T = 2$ ml h^{-1} in the high flow rate treatment, whereas in the low flow rate treatment, we used a flow rate of $q_T = 0.1$ ml h^{-1}. We imaged the devices every thirty minutes for the next 70 hours, which allowed both treatments to reach a steady state. Each treatment was repeated 3 times and each of these yielded the same result at steady state. The time points shown in Figure 2.2 are measured from the end of the inoculation phase.

Control Experiments

We performed separate experiments to ascertain if the dye used in the competition experiments or the different fluorescent markers had an effect on biofilm development. We injected cells into straight microfluidic channels (1 mm width, 75 μm depth) and then clamped the tubing on either side of the channel to allow cells to attach in the absence of flow for 60 mins. We then started an injection of tryptone broth (0.5 x TB, 5 g Tryptone per liter of water, with or without dye) at a flow rate of 0.1 ml h^{-1} and imaged the biofilm within each of the channels every 30 minutes for 78 hours. These microfluidic devices were fabricated and plumbed using the same techniques described above. In these experiments we used 60 ml syringes mounted on a Harvard Apparatus PhD Ultra syringe pump. To minimize flow induced detachment, and thus more accurately measure biofilm growth, these experiments used relatively small shear stresses. Here the shear stress at the beginning of the experiment was $\tau = 0.030$ kg m^{-1} s^{-2}, approximately 8 times smaller than that of the low flow rate treatment in the competition experiment. See section below for calculation of shear stresses.

Microscopy and image processing

We imaged microfluidic experiments using a Zeiss Axio Observer inverted microscope with an AxioCam MRm camera, and a Definite Focus system. A Zeiss 20X Plan Apochromat objective was used for competition experiments (Figure 2.2), whilst a Zeiss EC Plan Neofluar 10X objective was used for control experiments (Figure S2.5). We used the software package Zen Blue (Zeiss) to automatically record brightfield, GFP, and RFP images at each time point. As our devices were larger than a single field of view we recorded multiple adjacent images and stitched them together in post processing.

To quantify the proportion of flow moving through each arm of the competition device over time, we manually tracked the location of the dye interface using the image analysis software Fiji (Schindelin et al., 2012) and applied a moving average filter to reduce high frequency fluctuations. To clarify the presentation of Figure 2.2*A, B, D*, we plotted bacterial fluorescence only in the arms of the device.

Biofilm model

Our mathematical model of biofilm competition considered the two dominant processes that affect biofilm development in flowing environments: biofilm growth and the detachment of biomass induced by shear stress at the surface of the biofilm (Abbas et al., 2012; Duddu et al., 2009). There are other processes that can potentially affect biofilm development, such as the formation of filamentous biofilm streamers (Drescher et al., 2013; Valiei et al., 2012). However, we note that streamer formation has not been observed in experiments that use the slow growing bacterial species and weak nutrient conditions that are typical of many porous environments (e.g. (Tang et al., 2013)).

The two differential equations that govern the thicknesses of the biofilms, k_i (Eqn 2.1, main text) are hydrodynamically coupled using Kirchoff's laws, which hold for the low Reynolds number flows typical of porous environments (Stone, 2007). The

amount of flow that each pore receives, q_i, is a function of both its own instantaneous hydrodynamic resistance and that of the other two pore spaces:

$$q_i = q_T \frac{1/R_i + 1/R_j + 1/R_M}{1/R_i}, \tag{2.4}$$

where R_i is the hydrodynamic resistance of pore i per unit length. Here $R_i = (6\mu)/h(L-k_i)^3$ where h is the span wise width of the pore space, and we have assumed planar Poiseuille flow and $L \ll h$. Substituting this expression into Eqn. 2.1 yields, after some algebra,

$$\frac{dk_1}{dt} = \alpha_1 \min(\delta, k_1) - k_1 \sqrt{\frac{6\mu Q_T \chi^2}{h}} \sqrt{\frac{(L-k_1)}{(L-k_1)^3+(L-k_2)^3+M^3}}, \tag{2.5}$$

$$\frac{dk_2}{dt} = \alpha_2 \min(\delta, k_2) - k_2 \sqrt{\frac{6\mu Q_T \chi^2}{h}} \sqrt{\frac{(L-k_2)}{(L-k_1)^3+(L-k_2)^3+M^3}}. \tag{2.6}$$

We non-dimensionalized the model above using the pore width L as our characteristic length scale and the growth rate of strain 1 as our characteristic time scale α_1, such that $k = LK^*$, and $t = T^*/\alpha_1$. This yields the dimensionless governing equations

$$\frac{dK_1^*}{dT^*} = \min(K_1^*, \delta^*) - \beta^* K_1^* \sqrt{\frac{(1-K_1^*)}{Z}}, \tag{2.7}$$

$$\frac{dK_2^*}{dT^*} = \alpha^* \min(K_1^*, \frac{\delta^*}{\alpha^*}) - \beta^* K_2^* \sqrt{\frac{(1-K_2^*)}{Z}}, \tag{2.8}$$

$$\text{where } Z = (1-K_1^*)^3 + (1-K_2^*)^3 + (M^*)^3, \tag{2.9}$$

which were solved numerically using Matlab.

Estimation of the shear stresses in experiments

The mean shear stresses within our microfluidic devices can be derived from the equations for Poiseuille flow through a rectangular channel (Stone, 2007). For a channel

width of L and depth h, where $L \leq h$, the mean shear stress on the side of length h is given by

$$\tau = \frac{6q\mu}{hL^2} f(\phi), \tag{2.10}$$

where μ is the dynamic viscosity of the media, q is the volume flux through the channel, and $f(\phi)$ is solely a function of the aspect ratio $\phi = L/h$. The function $f(\phi)$ can be determined by using separation of variables to solve the governing non-dimensional equations (Stone, 2007). The resulting expression is complicated so here we provide more compact quadratic interpolation for $f(\phi)$

$$f(\phi) \approx 1 + 0.06827\phi + 0.1173\phi^2, \tag{2.11}$$

where relative errors are no more than 10^{-3} for $L \leq h$.

For the competition channels (Figure 2.2), this expression yields a mean shear stress $\tau = 0.23$ kg m^{-1} s^{-2} for the low flow rate treatment and $\tau = 4.7$ kg m^{-1} s^{-2} for the high flow rate treatment condition. In the control experiment (Figure 2.5) $\phi = 0.075$ giving the mean shear stress as $\tau = 0.030$ kg m^{-1} s^{-2}. Here, $q, h,$ and L are defined as above and μ = 1.0 mPa s, the standard value for the viscosity of water at 20°C. These calculations estimate the shear stresses at the beginning of the experiments when the thickness of the biofilm is negligible and the flow through the two arms of the device are equally balanced.

2.6 Supplementary figures

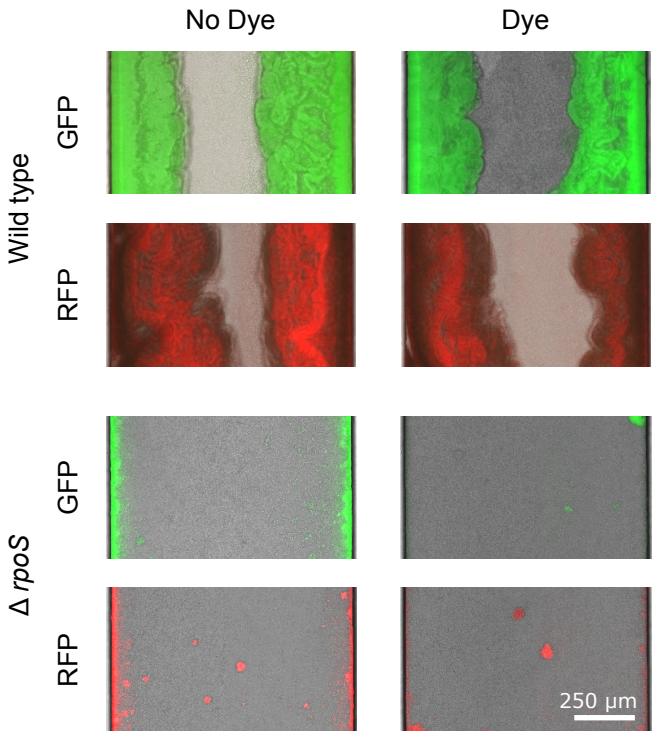

Figure 2.5: **Wild type biofilms grow at a faster rate than RpoS null biofilms.** We inoculated straight microfluidic channels (1 mm by 75μ m cross section) with either WT or $\Delta rpoS$ cells (*Materials and Methods*) and tracked biofilm formation over 96 hours. The wild type cells always formed much thicker biofilms than the mutant. Controls show that neither the fluorescent protein (RFP, GFP) nor the Chicago Blue dye was responsible for this difference.

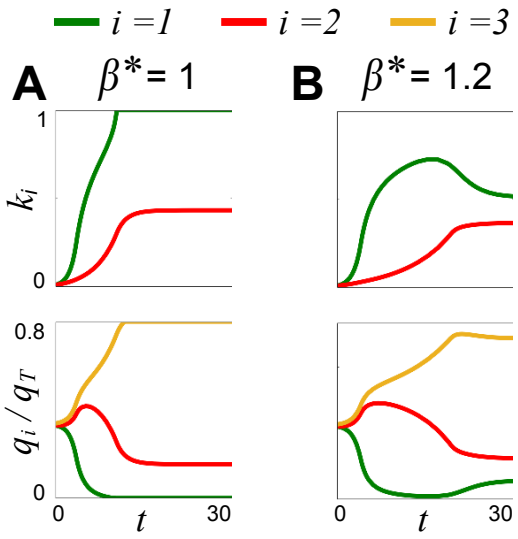

Figure 2.6: **The impact of flow on biofilm competition**. As a biofilm grows it increases the hydrodynamic resistance of its pore space, which tends to decrease both its access to flow (Figure 2.1) and flow-induced detachment. This creates a positive feedback loop because decreased detachment further increases its hydrodynamic resistance. When flow is relatively weak (A) this process can lead to the faster growing genotype (green, i = 1) completely blocking its pore space, such that it can no longer disperse cells downstream (Figure 2.3 *f*). However, when flow is stronger (B), increased detachment prevents the faster growing biofilm from blocking, but as the slower growing biofilm (red, i = 2) increases in thickness it diverts flow back to the faster growing biofilm, which then reduces in thickness. Here $M^* = 1, \delta^* = 0.3, \alpha^* = 1.5$, and the yellow line indicates flow through a third flow path (i = 3, Figure 2.1).

3

The ecology of the microbiome: Networks, competition, and stability

3.1 Abstract

The human gut harbors a large and complex community of beneficial microbes that remain stable over long periods. This stability is considered critical for good health but is poorly understood. Here we develop a new body of ecological theory to understand microbiome stability. While cooperating networks of microbes can be efficient, we find that they are often unstable. Counter-intuitively this finding suggests that hosts can benefit from microbial competition when this dampens cooperative networks and increases stability. More generally, stability is promoted by limiting positive feedbacks and weakening ecological interactions. We analyze host mechanisms for maintaining stability – including immune suppression, spatial structuring and feeding of community members – and support our key predictions with recent data.

3.2 Introduction

The human microbiome contains hundreds of species and trillions of cells that occur predominantly in the gastrointestinal tract (Gill et al., 2006; Lozupone et al., 2012). These microbes provide many health benefits, including the breakdown of complex molecules in food, protection from pathogens, and healthy immune development (Bäckhed et al., 2005; Dethlefsen and Relman, 2011; Lee and Mazmanian, 2010; Mazmanian et al., 2008). The gut microbiome is often noted for its ecological stability. Different people may carry different microbial species but any one individual tends to carry the same key set of species for long periods (Dethlefsen and Relman, 2011; Faith et al., 2013; Vanhoutte et al., 2004). This stability is considered critical for host health and wellbeing by ensuring that beneficial symbionts and their associated functions are maintained over time (De Cruz et al., 2012; Giongo et al., 2011; Lozupone et al., 2012; Relman, 2012). Consistent with this, major shifts in microbial community composition are often associated with ill health (Hsiao et al., 2013; Mazmanian et al., 2008).

Research into gut communities has been characterized by a large volume of empirical work. Despite this, we are far from a clear understanding of microbiome communities and, in particular, what promotes or disrupts their stability. There is a pressing need for complementary theory to identify overarching principles and patterns for the microbiome. Some progress has been made through the use of individual-based models (Schluter and Foster, 2012) and other analyses of two-species communities (Bucci et al., 2012). However, the microbiome contains many diverse species interacting with one another (Stein et al., 2013), which makes the full system complex and challenging to understand. The field of theoretical ecology has a long history of using network models that are specifically intended to deal with large and complex communities (Allesina and Tang, 2012; May, 1972). Here we develop ecological network theory to identify the general principles underlying microbiome stability. We then use these principles to

identify and analyze candidate mechanisms that a host can use to promote stability in their microbiome. Finally, we show that our key predictions are supported by recent data from the mammalian microbiome.

Seminal work by May suggests that species diversity can be problematic for community stability (May, 1972; McCann, 2000) . However, May's work focused on networks where the types of interactions between species are randomly distributed, meaning that $+/+$ (cooperation) and $-/-$ (competition) interactions occur with half the probability of $+/-$ (exploitation) interactions. And, while ecological competition is thought to be prevalent in natural microbial communities (Foster and Bell, 2012), it is commonly assumed in the literature that the functioning of microbiome communities rests upon species that engage in cooperative metabolism ($+/+$) and provide health benefits for the host (Bäckhed et al., 2005; Di Cagno et al., 2011; Eberl, 2010; Sleator, 2010; Van den Abbeele et al., 2011). There is a clear rationale for this assumption. Competition between microbes – captured by the number and magnitude of mutually negative interactions in our models – is associated both with antibiotic warfare (Clemente et al., 2015; Fiegna and Velicer, 2005) and a reduction in the cooperative secretions that promote community productivity (Oliveira et al., 2014). Both of these have the potential to severely reduce the efficiency of any cooperative metabolism that benefits the host (Bäckhed et al., 2005; Di Cagno et al., 2011; Eberl, 2010; Samuel and Gordon, 2006; Sleator, 2010; Van den Abbeele et al., 2011). However, while intuitive, this argument neglects the potential effects of cooperation or competition on the ecological stability of microbiome communities.

To understand ecological stability in the microbiome, therefore, we must consider the specific effects of cooperation, and how cooperation interacts with other community characteristics, such as microbiome diversity, to influence community dynamics. To do this, we have developed a dedicated body of theory based upon Wigner's semicircle law and subsequent analyses (Allesina and Tang, 2012; Sommers et al., 1988;

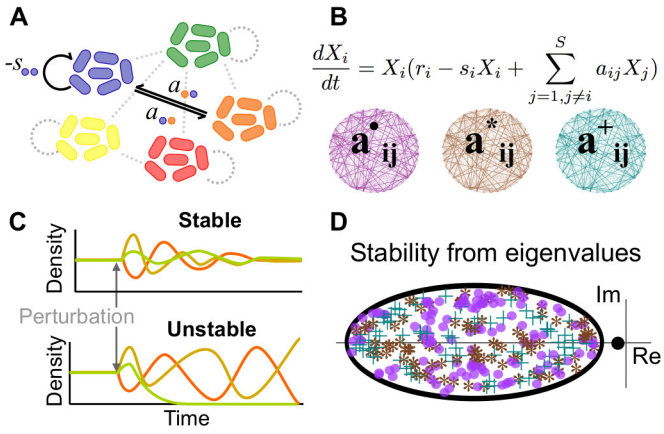

Figure 3.1: **Ecological theory and microbiota stability**. (**A**) Ecological network theory captures networks of microbial species that interact with themselves ($-s$) and other genotypes (a_{ij}). (**B**) Coupled ordinary differential equations capture all possible combinations of connectivity, interaction types (e.g., cooperation, competition, exploitation), and species numbers (S) covering any biologically feasible equilibrium microbiota. Three example networks are shown. (**C**) Communities that return to their previous densities following perturbation are classified as stable; those communities that return to their equilibrium faster are classified as more stable, and those that continue to diverge from the equilibrium as unstable (Allesina and Tang, 2012; May, 1972). (**D**) Linear stability analysis uses the eigenvalues' real (Re) and imaginary (Im) parts, shown plotted here. The largest real part of the eigenvalues underlying a community determines whether, and how fast, the community will return following perturbation. If negative, the community is stable, and more negative values mean the community returns to stability more quickly. The imaginary parts of the eigenvalues predict the extent of oscillations in species densities during a return to equilibrium – larger imaginary components predict more frequent oscillations. Eigenvalues are shown for the three example communities (purple, brown, green), and our analytic bound for their localization (black ellipse). Our analysis also derives one special eigenvalue location that for some parameters will lie outside of the ellipse (black dot).

Wigner, 1958). First, we derive a general analytic stability criterion that captures all potential community types, covering the full range of possible interactions and species diversities (Figure 3.1; Method 1 *Materials and methods*). We develop our theory for unstructured ecological networks because, unlike plant-pollinator communities or food webs (Mougi and Kondoh, 2012; Rohr et al., 2014), there is no evidence of strong structuring within microbial communities (Stein et al., 2013). However, while no single

structure type dominates in these communities, our mathematics concurrently analyzes all network arrangements and therefore naturally captures biologically critical network motifs, including chains of cross-feeding metabolic exchanges between species (Louis et al., 2014). Furthermore, a strength of studying unstructured networks is that they are amenable to comprehensive mathematical analysis. This means we can simultaneously analyze the ecological stability of all possible network permutations as a function of our focal parameters. Specifically, we can account for any variation in the proportion of co-operation $(+/+)$, competition $(-/-)$, exploitation $(+/-)$, commensalism $(+/0)$, and amensalism $(-/0)$, in networks with any combination of connectivity, C, and species number, S (Method 1 *Materials and methods*). Stability is assessed from the network's eigenvalues, which give three measures of stability: the probability that the community will return to its previous state after a small perturbation, the population dynamics during this return, and how long the return will take, which is a form of resilience ((Holling, 1973), Figure 3.1).

3.3 Results and discussion

We first show that May's (May, 1972) destabilizing effect of species diversity still holds in communities with any mixture of interaction types. From purely cooperative networks, through mixed interaction networks, to purely competitive networks (Figure S3.5), our model predicts that high species diversity leads to unstable microbiome communities. How though does altering the level of cooperation between species affect microbiome stability? We find that gradually increasing only the proportion of cooperative interactions within communities nearly always decreases the overall return rate and the likelihood of stability (Figure 3.2, Method 1b *Materials and methods*). We confirm our analytic results with numerical analyses (Figures 3.2 B, S3.7 – S3.10, Method 1f *Materials and methods*). This result contrasts with a recent analysis of macroscopic

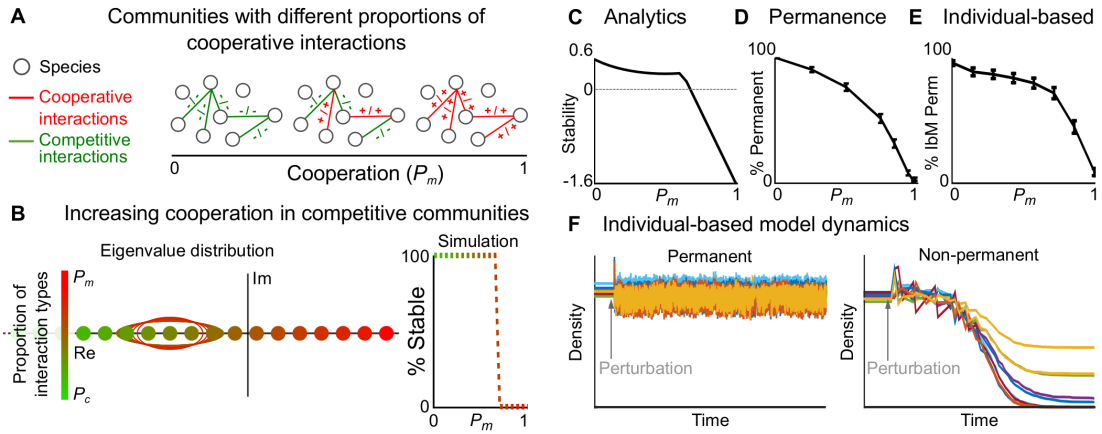

Figure 3.2: **Cooperation reduces community stability.** (**A**) Illustration of changing the proportion of cooperative links in networks (**B**) Linear stability analysis (also see Figure S3.6). Left hand plot shows solutions for eigenvalue locations as a function of increasing cooperation (shown as increasing red). The largest value of the real components (x-axis) determines whether, and how fast, a return to equilibrium occurs (stability), while the imaginary components (y-axis) determine the frequency of oscillations in population densities following perturbations (see Figure 3.1). The solutions give the position of all eigenvalues in the form of an ellipse, with the exception of a single eigenvalue that corresponds to the average row sum of the interaction matrix (represented by a dot that may lie outside of the ellipse). Increasing cooperation increases the largest eigenvalues, and therefore stability decreases. Solutions hold for any permutation of a community network with a given parameter set (here: $S = 100$, $C = 0.7$, $-s = -1$, $\sigma = 0.05$, see Method 1b *Materials and methods* for parameter sweeps that show that cooperation is nearly always destabilizing). The right hand side plots simulation results showing the proportion of communities that are stable and confirms our analytic results. (**C, D, E**) The effect of cooperation on community stability shown with linear stability analysis (C, also shown in B), permanence analysis (D) and individual-based modeling (E). The latter two methods are computationally expensive, which requires we study smaller networks than for linear stability analysis (see *Materials and methods*). Shown is: $S = 10$, $C = 0.7$, $-s = -0.2$, $\sigma = 0.05$ for 100 samples, errorbars: standard error of the mean (SEM). (**F**) Dynamics of individual-based model. Permanent communities maintain all their original species following a perturbation, even if species densities do not return back to their initial values. In non-permanent communities perturbations to species densities will lead to the extinction of some or all species.

communities, which predicts that ecological stability can be maximized for an intermediate frequency of cooperative interactions (Mougi and Kondoh, 2012). In the supplement we show that our analytical model recapitulates these numerical simulations, and

show that their predictions rest on assumptions that suit macroscopic communities but not the microbiome (Method 1g *Materials and methods*, Figures S3.11 – S3.13, and (Kondoh and Mougi, 2015; Mougi and Kondoh, 2012)).

This method, known as local stability analysis, has the benefits of both being extremely general and able to analyze communities with large numbers of species. However, this approach has the shortcoming of only analyzing whether viable communities are stable when they are close to their equilibrium (Chen and Cohen, 2001); it provides no information on how communities behave away from this equilibrium. We therefore also develop a second new analysis, based on a different method known as permanence analysis ((Jansen, 1987), Method 2 *Materials and methods*). Permanence analysis is considered one of the most rigorous methods of community analysis but the numerical analysis of permanence was, until now, limited solely to purely competitive or exploitative communities. Therefore, we expand traditional permanence methods here to study the effects of cooperation. While the method is very different to local stability analysis, we find the same prediction that cooperation is destabilizing. However, as discussed in the supplement, positive feedbacks arising from cooperative interactions can still constrain this analysis such that it may underestimate the number of permanent communities. We therefore also develop a third method – an individual-based model – to evaluate the relationship between cooperation and ecological stability in a microbial community (Method 3 *Materials and methods*). While less general than the other methods, our individual-based model is more explicit. It allows us to follow the dynamics of each species over time and examine additional factors, including explicit spatial structure that puts an upper limit on the community size. This third method again confirms our finding: cooperation is destabilizing for the community (Figure 3.2 E).

The reason for the destabilizing effect of cooperation in our models is that cooperation causes coupling between species and positive feedbacks. This means that if, for example, one species decreases in abundance, it will tend to pull others down with it

and destabilize the system. It has been hypothesized that cooperative and productive microbial communities are the reason for the remarkable stability of the microbiome (Van den Abbeele et al., 2011). However, even though cooperation means one species is helping another to survive and replicate, and might facilitate colonization, this does not equate to stability. This is because cooperation can create dependency and the potential for mutual downfall. Under these circumstances, therefore, our analyses suggest that the host faces a trade-off: while increased cooperation within communities is expected to promote overall metabolic efficiency (Bäckhed et al., 2005; Di Cagno et al., 2011; Fiegna and Velicer, 2005; Oliveira et al., 2014), it comes at the cost of decreasing ecological stability.

Our model predicts that ecological competition improves microbiome stability. However, a striking feature of the mammalian microbiome is species diversity, and we have seen that high species numbers tend to be destabilizing (Figure S3.5). Therefore, we next investigate how the stabilizing effects of ecological competition interact with the destabilizing effects of increasing diversity. Specifically, we ask: is a diverse and competitive community a route to a stable microbiome? We again turn to local stability analysis because this can deal with communities containing many species (Figure 3.3 A, Method 1 *Materials and methods*). While increasing species number is a destabilizing process (Figure S3.5), the increase in competition introduces negative feedback loops that are stabilizing. We find a wide range of diversities where this stabilizing effect dominates the destabilizing effect of increased species numbers (Figures 3.3 B, C, and S3.14). Even though competition may drive inefficiencies, therefore, it dampens the destabilizing effects of cooperation that can lead to the loss of community members (Figure 3.2 F). The key is that the new interactions from the introduction of additional species are competitive (or exploitative, Figure S3.15). The additional species need not necessarily be other bacteria; there is a growing recognition of the role of phage in microbial communities (Waller et al., 2014) that have the potential for similar effects to

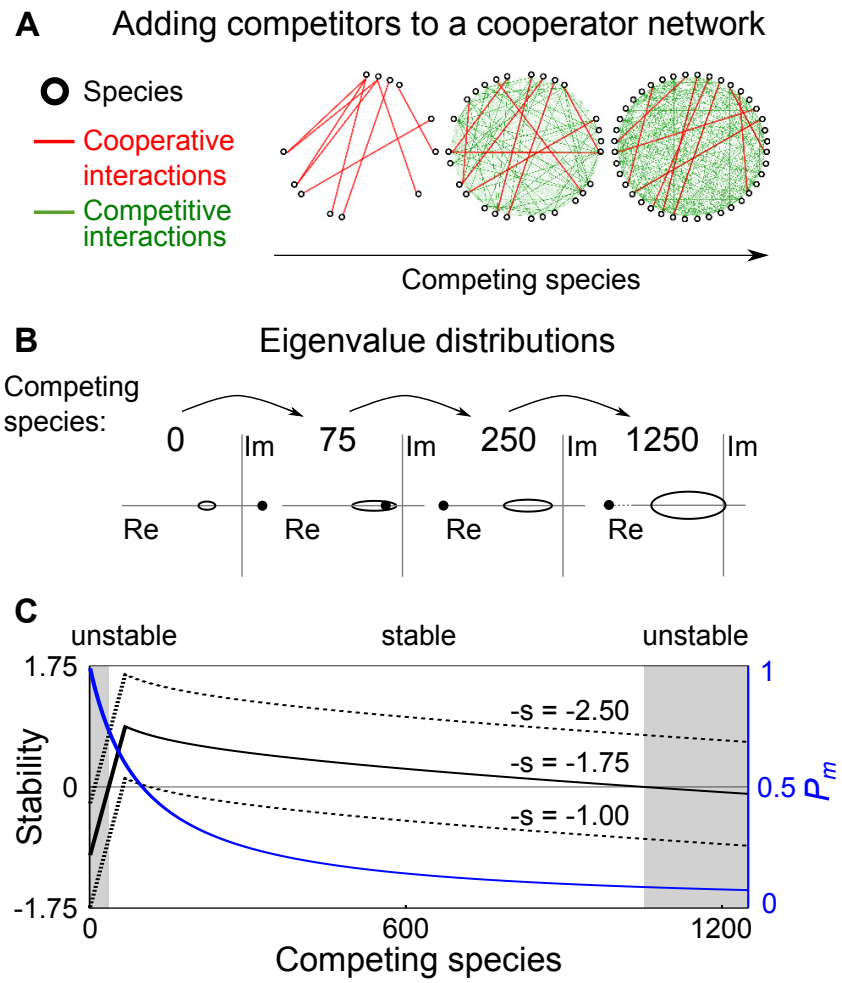

Figure 3.3: **Introducing species can stabilize a network of cooperators**. (**A**) Illustration of increasing species number and competition (green) in a network that originally only contained cooperating species (red). (**B**) Eigenvalue distribution changes following the addition of competitor species to 100 cooperating species (Connectivity, $C = 0.7$, $-s = -1.75$, $\sigma = 0.05$). As in Figures 3.1 and 3.2, the eigenvalues are contained within ellipses, except for one eigenvalue that may lie outside the ellipse. Crucially, this latter eigenvalue is reduced when we add competitive species (dot moves left), which is stabilizing. However, the ellipse containing all other eigenvalues becomes wider with additional species, leading to a reduction in stability with additional species once the ellipse contains the rightmost eigenvalue (once dot moves inside ellipse). (**C**) Ecological stability as a function of the number of competing species that are added (solid black line). Competitors are initially stabilizing because they reduce the proportion of cooperative interactions (P_m, blue line). Eventually, however, adding competitors becomes destabilizing, which corresponds to the widening of the ellipse in B. This behavior holds for a wide range of self-regulation strengths ($-s$, dashed black lines). Note that we do not apply permanence analysis or the individual-based model here as these analyses are limited to low species diversities (supplement *Materials and methods*).

exploitative species in our model (Figure S3.15, (Hofbauer and Sigmund, 1989)).

Our analyses allow us to identify key principles (Fredrickson, 2015) that are important for a stable microbiome community. In particular, the stabilizing effect of competition reflects the more general principle that dampening of positive feedback loops will promote stability. Can we use such principles to better understand host biology and, more specifically, how a host should interact with its symbionts? To answer this, we develop a further set of analyses dedicated to understanding how key features of host biology influence ecological stability. A clear candidate for promoting stability in the gut is the immune system. During dysbiosis and infection, adaptive immunity is thought to help re-establish a healthy microbiome by suppressing species whose abundance is causing harm (Bäckhed et al., 2005; Lee and Mazmanian, 2010; Mazmanian et al., 2008; Van den Abbeele et al., 2011). We can add such density-dependent regulation to our model and find that it is indeed stabilizing. Moreover, the reason this occurs is because immune regulation, like competition, will prevent run-away positive feedback loops (Method 1d *Materials and methods*).

Dampening positive feedbacks is an important route to stability. A second key principle affecting stability in our models is the strength of interactions between species. Stability will generally be promoted by processes that weaken interactions between species, because this reduces the coupling that drives instability. Redundancy can promote stability when a few strong cooperative interactions are replaced by several weaker ones (Method 1e

emphMaterials and methods, Figure S3.17). But a key candidate mechanism by which a host can actively weaken ecological interactions is through the introduction of spatial structure (Kim et al., 2008; Oliveira et al., 2014). In particular, introducing spatial structure (e.g., patchy growth with limited mixing, Figure S3.18) between species is known to reduce the strength of between-species interactions without reducing interactions within a species (Kim et al., 2008). Implementing this effect in our models greatly

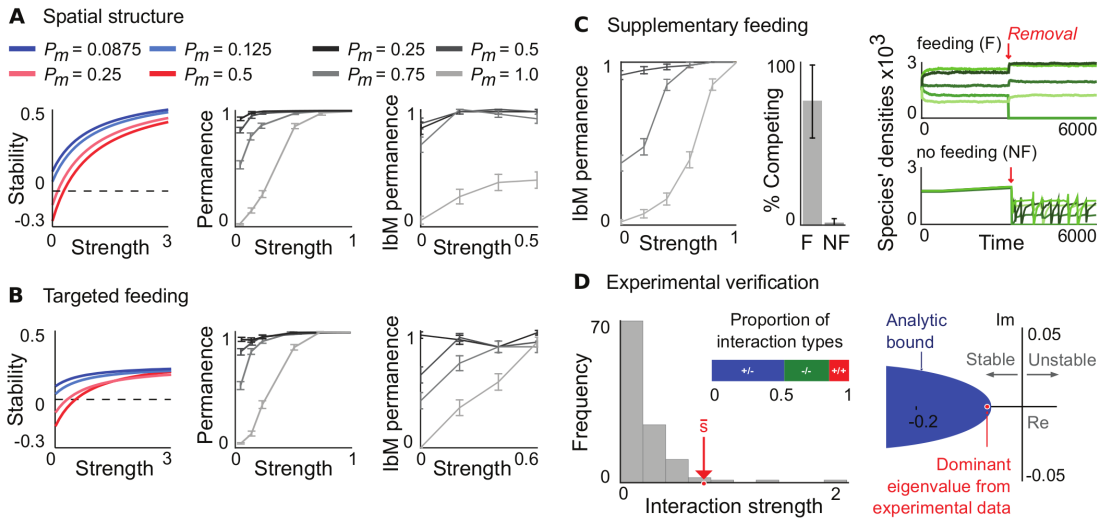

Figure 3.4: **Host strategies to promote ecological stability.** (**A**) Spatial structure promotes ecological stability in, from left to right, linear stability analysis (LAS), permanence analysis (PA), and an individual-based model (IbM). P_m is the proportion of cooperation. (**B**) Targeted feeding. Host-supplied nutrients that weaken cooperative interactions promote ecological stability. We observed an exception where targeted feeding may destabilize more competitive communities in the IbM ($P_m = 0.25$) but the effect is weak. (**C**) Non-targeted feeding may stabilize communities when it increases the growth rates of all species such that they become limited by the host's capacity (left). However, limited space means most previously cooperative species become competitors. Competition is demonstrated by removing one species from a community, and observing increased densities for most remaining species (bar chart, $P_m = 0.9$). IbM time-series plots show that feeding can stabilize communities upon species removal. (**D**) Left: empirical interaction parameters (mouse microbiome communities) (Stein et al., 2013) confirm our predictions: most interactions are weak (less than self-interaction, $\bar{s}$), and predominantly competitive or exploitative, with a small set of cooperative interactions (Experimental validation *Materials and methods*). Right: Our general analytical model predicts stability in a real community using community parameters from the mouse dataset ($S, C, P_m, P_c, \sigma, \bar{s}$). In A-C, LAS: $S = 300, C = 0.7, -s = -1, \sigma = 0.05$. PA, IbM: $S = 10, -s = -0.2$, errorbars: SEM). PA and IbM are computationally expensive and require smaller communities.

improves the potential for community stability (Figure 3.4). We predict, therefore, that a host can benefit from compartmentalizing species within gut communities to control interactions and limit the risk of extinctions.

Another candidate mechanism to influence microbial interactions is host epithe-

lial feeding. Nutrients are provided for symbionts from the epithelial surface, especially during starvation periods, when lumen nutrient concentrations are less abundant (Hooper et al., 1999). Recent work has shown that knockout mice lacking the ability to feed the microbiota with fucose show a significantly decreased community diversity (Kashyap et al., 2013). There is also the potential for some specificity in host feeding that occurs through fucose residues whose digestion relies on specific enzymes (Hooper et al., 1999; Sonnenburg et al., 2005; Tremaroli and Bäckhed, 2012). Studying feeding in our models indicates that feeding will indeed promote the stability of communities, provided such feeding can preferentially weaken interactions between cooperating species by, for example, providing an alternative carbon source to that involved in microbial cross-feeding (Figures 3.4 B, S3.16).

So far we have studied communities that are intrinsically limited by their own interactions. Feeding has the potential to alter this if it causes populations to grow until they are extrinsically limited by the capacity of the host. While high symbiont population sizes might harm a host for reasons independent of stability (Login et al., 2011), we can nevertheless examine the effect of this scenario on ecological stability using our individual-based model. Specifically, we ask: what happens if feeding, or indeed any process that drives high intrinsic growth rates, causes communities to expand until they become extrinsically limited by the capacity of the host environment. We find that such communities are stable to further perturbations (Figure 3.4 C), even those that start with many cooperative interactions. The possession of strongly growing communities that are extrinsically regulated by the host then is another potential route to ecological stability. However, after expansion, these communities are much less cooperative, because the increase in population densities means that species are now competing for the limited capacity of the host (Figure 3.4 C). Once again, we find that ecological stability is associated with competition.

Although the stability and composition of gut microbial communities is considered central to the benefits they provide to a host (De Cruz et al., 2012; Giongo et al., 2011; Lozupone et al., 2012; Relman, 2012), we lack a clear understanding of what underlies the striking stability of microbiome communities. Here, we have developed a body of theory to identify key principles underlying microbiome stability. We have also shown how these principles allow us to reinterpret and predict key features of host biology, including immune suppression, the potential for spatial structuring, and host feeding of microbes. Can we use existing data to test whether the principles we have identified are indeed at work in real communities? There are currently few data available on ecological interactions within the microbiome. Nevertheless, we can validate our approach and test our key predictions with recently published data on interactions in the mouse gut microbiome (Stein et al., 2013). The authors used time-resolved metagenomics and machine learning to infer the interactions within communities. In Figure 3.4 D we use these data to parameterize our general model and show that it correctly predicts stability within a real community. In addition we plot the distribution of ecological interactions within the mouse microbiome. Exactly as we predict, ecological interactions tend to be both non-cooperative and weak (Figure 3.4 D, Method 4 *Materials and methods*).

In conclusion, microbiome communities consist of many interacting species, making it difficult for a host to exercise control over each species individually. However, our work emphasizes that hosts can act as ecosystem engineers that manipulate general, system-wide properties of their microbial communities to their benefit. Identifying when and how hosts alter their symbionts' ecology will be important if we are to understand the dynamics of microbiome communities. More generally, while there is a vast and rapidly growing body of data on the mammalian microbiome, little attention has been paid to the strength or sign of ecological interactions within communities. To understand and manipulate the microbiome, we will need to both dissect and engineer the interactions within these critical communities.

3.4 Materials and methods

In the following we provide a detailed account of our analyses. The supplement is structured by methods; first linear stability analysis, second permanence analysis, third individual-based simulations, and fourth experimental validation.

Method 1

1a. Linear stability analysis

Our model considers networks of interacting species in which each species' growth is determined by its own intrinsic growth rate, interactions with other members of the same species, and interactions with members of other species. This is in accordance with previous models (Allesina and Tang, 2012; May, 1972; Mougi and Kondoh, 2012) and a generalization of classic Lotka-Volterra equations that allows us to incorporate multiple species and interaction types. Species are represented as densities averaging over many individuals – an approach that is well suited to the large population sizes of microbial species.

We analyze equilibrium populations and ask: will a population return to this equilibrium following a perturbation? This allows us to investigate the effect of different between-species interaction types on both the asymptotic stability of populations (May, 1972), and another measure of stability, the speed at which such a stable population returns to its equilibrium point (Haydon, 1994). The former definition of stability has been applied widely to analyze fundamental properties of dynamical systems, and generates similar predictions to other measures of ecological stability, which ask whether any species will be lost from a system following perturbation (Chen and Cohen, 2001). In later sections (Methods 2 and 3, below) we also consider alternative methods and definitions of stability, which show consistent results. Note that all our analyses are based on the logic of asking how community properties, like degree of connectedness or coop-

eration, affect ecological stability. This focus means we do not deal with the interesting corollary of how communities of a particular set of properties arise. However, we refer the reader to the literatures on both community assembly (Nemergut et al., 2013) and the evolution of species interactions (Mitri and Foster, 2013) that focus on this question. In addition, we acknowledge that there is the potential for important interactions between these processes and those that we study here, which are interesting targets for future work.

Our analytical approach is based upon calculating the eigenvalues of the Jacobian matrix of the dynamical system considered – a matrix that tells us how a change in the density of any of the species at equilibrium will affect the whole community. A population is asymptotically stable provided the real part of each of its associated eigenvalues lies below zero, and the community will return to its equilibrium following a small perturbation faster, the more negative its largest real part is. Both measures of stability, therefore, can be determined from the region in the complex plane to which all eigenvalues will be restricted, and we extend the recent analytic results of Allesina and Tang to derive these regions for an ecosystem with any potential mixture of interaction types (Allesina and Tang, 2012). Specifically, we show that given certain assumptions about the distribution from which interaction strengths are drawn, eigenvalues will be localized within an ellipse in the complex plane with horizontal radius r_e, centered about $-s$. One single eigenvalue may be localized outside of this ellipse; we denote this by r_s which is given by the average row sum of the Jacobian matrix minus the average self interaction s. The equilibrium community will therefore be asymptotically stable provided both the ellipse and r_s lie within the negative real part of the complex plane, equivalent to,

$$\text{Rightmost point} = \max(r_e, r_s) - s < 0.$$

We can further deduce that the more negative this rightmost point is, the quicker the community will return to equilibrium, and thus the stability of the system can be determined from the negative of this rightmost value. We are therefore able to study the effect of changing various community characteristics upon stability by examining the subsequent effects upon this rightmost value. In the following we discuss in detail how the eigenvalue bounds can be derived.

Problem outline

Throughout this work, we consider a microbiota community composed of S interacting species, whose population dynamics are determined by a Holling type 1 functional response, such that the change in the density of species i is given by,

$$\frac{dX_i}{dt} = X_i(r_i - s_i X_i + \sum_{j=1, j\neq i}^{S} a_{ij} X_j). \tag{3.1}$$

Here r_i represents the intrinsic species growth rate, s_i captures the effect of a species upon itself, and a_{ij} the effect of species j upon species i (for $i \neq j$). We assume that each species competes to the same extent with members of their own species (i.e. resulting in the same degree of self-regulation), such that $s_i = s, i = 1, \ldots, S$. The connectivity $C \in [0, 1]$ of the community determines the fraction of all S species that a single species i interacts with (i.e., the average number of links between species such that $a_{i\neq j} \neq 0$)

We assume that non-zero interactions between species i and j can take one of five possible forms based on the signs of a_{ij}/a_{ji}; $+/-$ (exploitative), $-/-$ (competitive), $+/+$ (cooperative), $+/0$, and $-/0$. For a given community, the proportion of interactions taking each of these forms are defined as,

$$\begin{aligned} +/- &= P_e, \\ -/- &= P_c, \\ +/+ &= P_m, \\ +/0 &= P_+, \\ -/0 &= P_-, \\ \text{where } P_e + P_c + P_m + P_+ + P_- &= 1. \end{aligned}$$

Following the same approach as Chen and Cohen and others (Allesina and Tang, 2012; Chen and Cohen, 2001; Mougi and Kondoh, 2012), we assume the system has a positive definite equilibrium given by the $S \times 1$ vector $\mathbf{X}^*$. This is equivalent to examining the system

$$\frac{dY_i}{dt} = Y_i(r_i + q_{ii}Y_i + \sum_{j=1, j\neq i}^{S} q_{ij}Y_j),$$

normalized by replacing each y_i by $x_i = \frac{y_i}{q_i}$. We ignore systems that do not have a positive definite equilibrium, as they will automatically be deemed unstable by our definition, and contradict experimental evidence suggesting equilibrium communities in the gut exist for extended periods of time. We can therefore determine the stability of a thus normalized equilibrium point by examining the distribution of the eigenvalues of the Jacobian of the system, evaluated at $\mathbf{X}^*$, whose entries are given by,

$$\begin{aligned} A_{ii} &= -s_i, \quad i = 1, \ldots, S \\ A_{ij} &= a_{ij}, \quad i, j = 1, \ldots, S. \end{aligned}$$

Derivation of stability criterion

The basis of our stability criterion can be traced back to the seminal work by Gerschgorin, who showed for a complex $S \times S$ matrix M with elements m_{ij} that each eigenvalue λ_i lies within at least one of S Gerschgorin disks, which are centered at the diagonal entries (m_{ii}) with radii

$$R_i = \sum_{j=1, j \neq i}^{S} |a_{ij}|,$$

(Gerschgorin, 1931). This was used by Sommers *et al* in 1988 (Sommers et al., 1988), who examined the eigenvalues of an $S \times S$ matrix M^* with elements drawn from a Normal distribution with mean,

$$\begin{aligned} \mathbb{E}(M^*_{ij}) &= 0, \\ Var(M^*_{ij}) &= \frac{1}{S}, \\ M^*_{ii} &= 0. \end{aligned}$$

The authors proved that as $S \to \infty$ then the eigenvalues of M^* will be distributed on an ellipse described by $(\frac{x}{a})^2 + (\frac{y}{b})^2 \leq 1$ centered at the origin, with $a = 1 + \tau$ and $b = 1 - \tau$ where $\tau = S\mathbb{E}(M^*_{ij}M^*_{ji})$. Allesina and Tang (Allesina and Tang, 2012), motivated by Tao *et al.* (Tao et al., 2010), showed that this conclusion is also applicable to non-Normal distributions, save for one eigenvalue that is approximately equal to the average row sum of M^*, and further noted that setting each of the diagonal entries a_{ii} to $-s$ will move the center of this ellipse to $(-s, 0)$, thus the system will be stable if the half-horizontal radius of the eigenvalue ellipse is less than s.

Here we follow a similar approach to that used in (Allesina and Tang, 2012), and use various rescaling arguments to derive a bound upon the eigenvalues of the Jacobian, A, corresponding to our dynamical system of the microbiota evaluated at the equilibrium

$\mathbf{X}^*$ for any mixture of different interaction types. We assume that between-species interaction strengths a_{ij} are drawn from a half Normal distribution $|N(0, \sigma^2)|$ such that the mean strength of realized interactions is given by $\mathbb{E}(|X|) = \sqrt{\frac{2\sigma^2}{\pi}}$, and the variance $Var(|X|) = \sigma^2\Big(1 - \frac{2}{\pi}\Big)$.

When S is large, the row sum of A is constant and equal to

$$\lim_{S\to\infty} m = S^{-1} \sum_{i,j=1;i\neq j}^{S} a_{ij},$$

thus $\mathbf{1}$ is an eigenvector of A. The mean of the off-diagonal entries of A, $E(A_{ij}) = m$, and thus A does not meet the requirements of Sommers's theorem, so we next define a new matrix $N = A + (-s + m)I - m.\mathbf{1}.\mathbf{1}^T$. Rothblum and Tan showed that for any $m \in \mathbb{R}$, the half horizontal radius of the eigenvalue ellipse for the matrix $N = A + (-s + m)I - m.\mathbf{1}.\mathbf{1}^T$ will have the same value as the half horizontal radius for A (Allesina and Tang, 2012; Rothblum and Tan, 1985). We can therefore calculate a bound on the distribution of the eigenvalues of A by first calculating the distribution of the eigenvalues of N.

We begin by taking into account the different interaction types present in our community, and decompose the mean of the off-diagonal entries according to these different types. The average off-diagonal row sum for mixed interaction type matrices is then given by,

$$\begin{aligned}\mathbb{E}(A_{ij}) &= C(P_m\mathbb{E}(|X|) - P_c\mathbb{E}(|X|) + 0.5P_+\mathbb{E}(|X|) - 0.5P_-\mathbb{E}(|X|)),\\ &= C\mathbb{E}(|X|)(P_m - P_c + 0.5(P_+ - P_-)),\end{aligned}$$

and we note that $m = \mathbb{E}(A_{ij})$, and the matrix N is therefore,

$$\mathbb{E}(N_{ij}) = \mathbb{E}(A_{ij} - m) = \mathbb{E}(A_{ij}) - m = 0,$$

such that N and any scalar multiple of this matrix will obey the first requirement of Sommers's theory, i.e. $\mathbb{E}(cN) = 0, c \in \mathbb{R}$. We next consider the variance of N, $Var(N_{ij}) = Var(A_{ij} - m)$. As m is simply a constant, we have,

$$\begin{aligned} Var(N_{ij}) &= Var(A_{ij}), \\ &= \mathbb{E}(A_{ij}^2) - \mathbb{E}(A_{ij})^2. \end{aligned}$$

Further, using the fact that $P_e + P_c + P_m + P_+ + P_- = 1$, and the definition of the second moment of the half-normal distribution from which the A_{ij}s are drawn, we have,

$$\begin{aligned} \mathbb{E}(A_{ij}^2) &= C(P_m\mathbb{E}(A_{ij}^2) + P_c\mathbb{E}(A_{ij}^2) + P_e\mathbb{E}(A_{ij}^2) + 0.5P_+\mathbb{E}(A_{ij}^2) + 0.5P_-\mathbb{E}(A_{ij}^2)), \\ &= C\mathbb{E}(A_{ij}^2)(1 - 0.5P_+ - 0.5P_-), \\ &= C\sigma^2(1 - \frac{P_+ + P_-}{2}), \text{and therefore,} \\ Var(N_{ij}) &= C\sigma^2(1 - \frac{P_+ + P_-}{2}) - m^2. \end{aligned} \tag{3.2}$$

We next rescale N, setting $N^* = \frac{N}{\beta}$ with $\beta = \sqrt{SVar(N_{ij})}$, such that the matrix N^* fulfills both of the requirements of Sommers's theory. That is:

$$\begin{aligned} \mathbb{E}(N^*) &= 0, \\ Var(N^*) &= \frac{1}{S}. \end{aligned}$$

Thus all eigenvalues of N^* will be contained in the ellipse with half-horizontal radius $a^* = 1 + \tau$ and half-vertical radius $b^* = 1 - \tau$ with $\tau = \mathbb{E}(N_{ij}^* N_{ji}^*) = \frac{\mathbb{E}(N_{ij}N_{ji})}{\beta^2}$. Further, we note that eigenvalues are closed under scalar multiplication, thus if λ is an eigenvalue of the matrix N^*, then $\beta\lambda$ will be an eigenvalue of N. This means that the eigenvalues of N will be contained in the ellipse with half-horizontal radius $a = \beta(1+\tau)$

and half-vertical radius $b = \beta(1-\tau)$. This gives us a bound on the half-horizontal radius of the ellipse containing the eigenvalues of N, and therefore of A of:

$$r_e = \sqrt{SVar(N_{ij})}\Big(1 + \frac{S\mathbb{E}(N_{ij}N_{ji})}{SVar(N_{ij})}\Big). \tag{3.3}$$

We must then calculate $\mathbb{E}(N_{ij}N_{ji})$, which we do by noting that:

$$\begin{aligned}\mathbb{E}(N_{ij}N_{ji}) &= \mathbb{E}(A_{ij}A_{ji}) - 2m\mathbb{E}(A_{ij}) + m^2, \\ &= \mathbb{E}(A_{ij}A_{ji}) - m^2\end{aligned} \tag{3.4}$$

where,

$$\begin{aligned}\mathbb{E}(A_{ij}A_{ji}) &= C(P_m\mathbb{E}(A_{ij}A_{ji}|+/+) + P_c\mathbb{E}(A_{ij}A_{ji}|-/-) + \dots \\ &\quad P_e\mathbb{E}(A_{ij}A_{ji}|+/-) + P_+\mathbb{E}(A_{ij}A_{ji}|+/0) + P_-\mathbb{E}(A_{ij}A_{ji}|-/0)), \\ &= C(P_m\mathbb{E}(A_{ij}A_{ji}|+/+) + P_c\mathbb{E}(A_{ij}A_{ji}|-/-) + P_e\mathbb{E}(A_{ij}A_{ji}|+/-)), \\ &= C(P_m\mathbb{E}(|X|)^2 + P_c\mathbb{E}(|X|)^2 - P_e\mathbb{E}(|X|)^2), \\ &= C\mathbb{E}(|X|)^2(2(P_m + P_c) + P_+ + P_- - 1).\end{aligned}$$

We can substitute (3.2) and (3.4) into (3.3) to give the expression for the half-horizontal radius of the eigenvalue ellipse of A as:

$$\begin{aligned}r_e &= \sqrt{SC(\sigma^2(1 - \frac{P_+ + P_-}{2}) - C\mathbb{E}(|X|)^2(P_m - P_c + \frac{P_+ - P_-}{2})^2)}\dots \\ &\quad \Big(1 + \frac{\mathbb{E}(|X|)^2((2P_m + 2P_c + P_+ + P_- - 1) - C(P_m - P_c + 0.5P_+ - 0.5P_-)^2)}{\sigma^2(1 - 0.5P_+ - 0.5P_-) - C\mathbb{E}(|X|)^2(P_m - P_c + 0.5P_+ - 0.5P_-)^2}\Big)\end{aligned}$$

Finally, we observe that the eigenvalue corresponding to the average row sum will be given by,

$$r_s = (S-1)C(P_m - P_c + \frac{P_+ - P_-}{2})\mathbb{E}(|X|).$$

From these we generate a stability criterion for a network with any mixture of interaction types,

$$\max(r_e, r_s) - s < 0.$$

1b. Analysis of stability criterion

The analyses presented in the main paper suggest that cooperation has a destabilizing influence on microbial communities (Figure 3.2). Here we provide supplementary analyses of our model to show that this prediction holds for the vast majority of conditions and parameters in our analytical model. We first show that cooperation has a universally destabilizing effect on exploitative or random communities for any realistic combinations of species number and connectivity. We then analyze the effect of cooperation on purely competitive networks. Here, increasing cooperation nearly always destabilizes communities except for a minor parameter range where increased cooperation can have a weakly stabilizing effect. Moreover, as we show below, the weakly stabilizing effect is further minimized when we consider realistic communities with a mixture of competitive and exploitative interactions.

To determine how cooperation affects community stability we take the derivative of our measure of stability, U, with respect to P_m. This tells us how stability changes with respect to the level of cooperation -– with a negative derivative meaning cooperation decreases stability. We can therefore examine the effect of cooperation by looking at the sign of this derivative, $\frac{dU}{dP_m}$, across parameter space.

In the analysis below we treat each type of community (exploitative / random / competitive) separately, and follow the same series of steps. For simplicity we consider communities with only $+/+$, $-/-$, and $+/-$ interactions, however, the same analysis can equivalently be extended to cover communities that also contain $+/0$ and $-/0$ interactions. We begin by recalling that stability of a community at equilibrium is defined

as the negative of the rightmost bound upon its underlying eigenvalues (that is, communities whose eigenvalues lie further to the right in the complex plane are less stable), which are contained within an ellipse centered at $-s$ with half-horizontal radius,

$$\begin{aligned} r_e &= \sqrt{SVar(N_{ij})}\Big(1+\frac{\mathbb{E}(N_{ij}N_{ji})}{Var(N_{ij})}\Big), \\ &= \sqrt{SC(\sigma^2 - C\mathbb{E}(|X|)^2(P_m-P_c)^2)}\ldots \\ &\quad \Big(1+\frac{\mathbb{E}(|X|)^2((2P_m+2P_c-1)-C(P_m-P_c)^2)}{\sigma^2 - C\mathbb{E}(|X|)^2(P_m-P_c)^2}\Big), \end{aligned} \tag{3.5}$$

save for a single eigenvalue that may lie outside the ellipse, and is approximated by,

$$r_s = -s + (S-1)C(P_m-P_c)\mathbb{E}(|X|). \tag{3.6}$$

Stability is therefore defined as $U = -(\max(r_e - s, r_s))$. As such, the derivative of stability, $\frac{dU}{dP_m}$ will be defined differently, depending upon whether the ellipse or the dot represents the rightmost bound. In our analysis, we therefore calculate separately the derivative of the ellipse, $\frac{dU^{ellipse}}{dP_m}$, and of the dot, $\frac{dU^{dot}}{dP_m}$, then for each combination of P_m, C, and S check which component is governing stability for that parameter set, to determine the appropriate derivative. We then plot a heatmap across parameter space, to indicate in which regions cooperation is stabilizing ($\frac{dU}{dP_m} > 0$) and in which it is destabilizing ($\frac{dU}{dP_m} < 0$).

We begin by noting that in all community types, the self-regulation, s, does not affect the impact of cooperation upon community stability and so our conclusions on how cooperation influences stability are general for all values of self-regulation. Specifically, in our measure of stability the self-regulation is a constant, independent of P_m, and as such, it does not appear in $\frac{dU}{dP_m}$ (equivalently, note that $\frac{d^2U}{dP_m ds} = 0$ for both the ellipse and the dot). Moreover, whilst increasing species number, S, will increase the magnitude of the role cooperation plays upon community stability, it will not affect the directionality

— thus if we find that increasing cooperation is destabilizing, increasing species number will not affect this conclusion. This can be seen from the derivative of $\frac{dU}{dP_m}$ with respect to S for the ellipse and dot. For the ellipse,

$$\frac{d^2U}{dP_m dS} = \frac{-1}{2\sqrt{S}}\frac{d}{dP_m}\Big(\sqrt{Var(N_{ij})}\Big(1+\frac{\mathbb{E}(N_{ij}N_{ji})}{Var(N_{ij})}\Big)\Big),$$

and so only the magnitude of the effect of increasing cooperation upon stability is affected. Specifically, the magnitude will decrease with the inverse of the square root of S when the ellipse governs stability. Similarly for the dot,

$$\frac{d^2U}{dP_m dS} = -\frac{d}{dP_m}C(P_m - P_c),$$

which is constant over S, so again only the magnitude of the effect of cooperation upon stability is affected. Crucially, these observations mean that any conclusions drawn about the effect of increasing cooperation upon stability for low species number will also hold for all larger communities (and any value of self-regulation). As such, in our analysis we can set species number and self-regulation to constants, $S = 100$ and $-s = -1$ respectively, and focus exclusively upon whether changing connectivity, C, will affect the role of cooperation upon stability.

Exploitative communities

We start by considering the case of increasing cooperation in communities in which all interactions are otherwise exploitative $(+/-)$. In this case there are never any competitive interactions, so we substitute $P_c = 0$ into our equations for stability, (3.5) and (3.6). We then take the derivative of each with respect to P_m to get the equation for the ellipse bound,

$$\frac{dU^{ellipse}_{exploit}}{dP_m} = -\frac{CS^{\frac{1}{2}}\mathbb{E}(|X|)^2(2C^2P_m^3\mathbb{E}(|X|)^2 - CP_m\mathbb{E}(|X|)^2 - 3CP_m\sigma^2 + 2\sigma^2)}{(-CP_m^2\mathbb{E}(|X|)^2 + \sigma^2)\sqrt{C(\sigma^2 - CP_m^2\mathbb{E}(|X|)^2)}},$$

and for the dot,

$$\frac{dU^{dot}_{exploit}}{dP_m} = -(S-1)C\mathbb{E}(|X|).$$

It is evident from the form of $\frac{dU^{dot}}{dP_m}$ that in cases where the dot governs community behavior stability will decrease linearly with cooperation. However, the form of $\frac{dU^{ellipse}}{dP_m}$ is less obvious, and we therefore plot $\frac{dU}{dP_m}$ across C/P_m parameter space for a community with 100 species (Figure S3.8). This illustrates how cooperation will always decrease stability ($\frac{dU}{dP_m}$ is always negative) and moreover, as explained above, this behavior will hold true for any equivalent communities with differing levels of self-regulation or greater species numbers. Figure S3.8 also illustrates how this destabilizing effect of increasing cooperation upon the community will be stronger in communities with higher network connectivity.

Random communities

We next examine the effect of increasing cooperation in communities that are otherwise "random". In these communities networks start with cooperative, exploitative, and competitive interactions in a $1:2:1$ ratio. As cooperation increases, exploitative and competitive interactions are replaced with equal probability, such that exploitation : competition remains at a ratio of $2:1$. To capture this, we substitute $P_c = \frac{1}{3}(1-P_m)$ in equations (3.5) and (3.6). We then take the derivative of each with respect to P_m to calculate the derivative for the ellipse,

$$\frac{dU^{ellipse}_{rand}}{dP_m} = \frac{4CS^{\frac{1}{2}}\mathbb{E}(|X|)^2\Big(2C^2\mathbb{E}(|X|)^2(64P_m^3-48P_m^2+12P_m-2)-27\sigma^2(4CP_m-C-1)\Big)}{3(9C\sigma^2-C^2\mathbb{E}(|X|)^2(4P_m-1)^2)^{1/2}(\mathbb{E}(|X|)^2(16CP_m^2-8CP_m+C)-9\sigma^2)},$$

and for the dot,

$$\frac{dU^{dot}_{rand}}{dP_m} = -\frac{4}{3}(S-1)C\mathbb{E}(|X|).$$

Again it is clear by inspection that when community behavior is determined by the dot, stability will decrease linearly with cooperation, and at a greater rate than for exploitative communities (the value of $\frac{dU^{dot}_{rand}}{dP_m}$ is constant and less than that of $\frac{dU^{dot}_{exploit}}{dP_m}$). As in the exploitative case, we plot $\frac{dU}{dP_m}$ across C/P_m for $S = 100$ (Figure S3.9). Again we see that in this case cooperation is always destabilizing, regardless of whether the dot or the ellipse determines community behavior (the derivative is always negative), and as outlined above, this will hold for any level of self-regulation and any larger communities. Moreover, again it is evident that this destabilizing effect is greater for communities with higher network connectivity.

Competitive communities

Finally we consider increasing cooperation in systems that are otherwise purely competitive, so substitute $P_c = 1 - P_m$ in equations (3.5) and (3.6). As before, we take the derivative of each with respect to P_m to find the change with cooperation to the ellipse,

$$\frac{dU^{ellipse}_{compete}}{dP_m} = \frac{2C^2S^{\frac{1}{2}}\mathbb{E}(|X|)^2(2P_m-1)(\mathbb{E}(|X|)^2(2C+1+8CP_m(P_m-1))-3\sigma^2)}{((C\sigma^2-C^2\mathbb{E}(|X|)^2(2P_m-1)^2))^{1/2}(4CP_m\mathbb{E}(|X|)^2(P_m-1)+C\mathbb{E}(|X|)^2-\sigma^2)},$$

and to the dot,

$$\frac{dU^{dot}_{compete}}{dP_m} = -2C\mathbb{E}(|X|)(S-1).$$

As before, we observe that when the behavior of the community is determined by the dot, stability will decrease linearly with cooperation, and at a rate twice as fast as equivalent exploitative communities. However, in this case, we find regions of C/P_m parameter space where increasing cooperation can in fact have a stabilizing effect upon the community (that is, $\frac{dU}{dP_m} > 0$), as illustrated by Figure S3.10. However, the magnitude of this effect is small such that it will rarely make an unstable system stable (Figure S3.10 C). Moreover, this effect is further reduced when one considers more

realistic communities that contain even just small amounts of exploitative interactions such that the total proportion of cooperative and competitive links are no longer kept constant (Figure S3.10 D – F). From these analyses, we conclude that cooperation between species is a destabilizing influence on ecological stability for the vast range of network compositions, species numbers, and connectivities.

1c. Independent scaling of interaction types

In order to study the effects of host manipulation (Figure 3.4), we want to independently scale the interaction strengths within communities. In this section, we outline how this is done. For simplicity, below we derive our new stability criterion with only $+/-, -/-$, and $+/+$ interactions. However, the same methods could be applied to also incorporate $+/0$ and $-/0$ interactions. We begin by weighting interactions such that the average magnitude of each interaction type is given by,

$$\begin{aligned} f\mathbb{E}(|X|) \quad &\text{for} \quad +/+, \\ g\mathbb{E}(|X|) \quad &\text{for} \quad -/-, \\ h\mathbb{E}(|X|) \quad &\text{for} \quad +/-, \end{aligned}$$

whilst self-regulation will now take the form $-s * k$. The mean of the off-diagonal entries of the Jacobian will now be given by,

$$\mathbb{E}(A_{ij}) = C\mathbb{E}(|X|)(fP_m - gP_c),$$

and similarly, the variance and covariance by,

$$\begin{aligned} Var(A_{ij}) &= C\sigma^2(f^2P_m + g^2P_c + h^2(1 - P_m - P_c)) - C^2\mathbb{E}(|X|)^2(fP_m - gP_c)^2, \\ \mathbb{E}(A_{ij}A_{ji}) &= C(f^2P_m\mathbb{E}(|X|)^2 + g^2P_c\mathbb{E}(|X|)^2 - h^2(1 - P_m - P_c)\mathbb{E}(|X|)^2). \end{aligned}$$

We can then use the same logic as in supplementary section 1 to get an expression for the half-horizontal radius of the ellipse,

$$r_e^w = \sqrt{SC(\sigma^2(f^2P_m + g^2P_c + h^2(1 - P_m - P_c)) - C\mathbb{E}(|X|)^2(fP_m - gP_c)^2)} \ldots \left(1 + \frac{\mathbb{E}(|X|)^2(P_m(f^2 + h^2) + P_c(g^2 + h^2) - h^2 - C(fP_m - gP_c)^2)}{\sigma^2(f^2P_m + g^2P_c + h^2(1 - P_m - P_c)) - C\mathbb{E}(|X|)^2(fP_m - gP_c)^2}\right),$$

and the average row sum,

$$r_s^w = C(S - 1)\mathbb{E}(|X|)(fP_m - gP_c).$$

From these our stability criterion for a network with any mixture of interaction types, each of which can now be weighted independently, is given by

$$\max(r_e^w, r_s^w) - sk < 0.$$

With this we can now investigate how different manipulations to the microbiota will affect the stability of the community. Specifically, we consider three distinct cases: in the first, we consider spatial structure that reduces between-species interactions while self-regulation stays the same, corresponding to a decrease in f, g, and h, whilst k remains the same (Figure 3.4). In the second, we consider the case where a host provides a generic food source with which all microbes interact instead of interacting with each other, corresponding to a decrease in f, g, h and k (down-weighting all interaction types, Figure S3.16). In the final case, only cooperative links are weakened by host nutrient secretion, for example when a nutrient source is diverting interaction efforts away from cross-feeding interactions towards consumption of host provided nutrients. This corresponds to a decrease in f only (Figure 3.4). See methods 2 and 3 below for further analyses of these three cases.

1d. Effect of host-mediated killing on community stability

The immune system is thought to act, via inflammation and other mechanisms, to help suppress pathogens or other species that reach densities where they are harmful to a host (Bäckhed et al., 2005; Lee and Mazmanian, 2010; Mazmanian et al., 2008; Van den Abbeele et al., 2011). We can capture this effect in our model as species-specific density dependent killing at a rate k_i where,

$$\begin{aligned} \frac{dX_i}{dt} &= X_i(r_i - s_i X_i + \sum a_{ij} X_j - k_i X_i), \\ &= X_i(r_i - s_i^* X_i + \sum a_{ij} X_j), \end{aligned} \tag{3.7}$$

with $s_i^* = s_i + k_i$. As such, density dependent killing acts in the same way as self-regulation, which has a stabilizing effect on the community.

We can also ask what happens should the host kill members of each species at a constant rate independent of the population sizes. Here, we observe no change in the stability of communities. Killing does put extra demands on community members, it will favor faster growing community members that can survive in the face of the killing. But for viable communities that can persist in the face of the killing, there is no net effect on ecological stability. This can be seen incorporating a global increased death rate in our model. The term r_i is the intrinsic growth rate of species i, which is a combination of the rates of birth, b_i, and death, d_i, of cells of species i, that is, $r_i = b_i - d_i$. In the absence of general host killing, stable communities with a given set of interactions, a_{ij}, fulfil,

$$r_i = s_i^* X_i - \sum a_{ij} X_j.$$

If the host kills members of each species at a constant rate, k_i, then this will be incorporated into our original system describing community dynamics as,

$$\begin{aligned} \frac{dX_i}{dt} &= X_i(r_i^* - s_i X_i + \sum a_{ij} X_j) - k_i X_i, \\ &= X_i((r_i^* - k_i) - s_i X_i + \sum a_{ij} X_j), \end{aligned} \tag{3.8}$$

Species members of a stable community must now have an intrinsic growth rate, r^*, such that,

$$r^* = s_i Xi - \sum aijX_j + k_i.$$

A community will therefore have to possess a stronger intrinsic growth rate when subject to a constant level of killing to be viable. However, we note that,

$$r^* - k_i = s_i X_i - \sum aijX_j = r_i,$$

such that the equations governing the dynamics of both populations are in fact identical, regardless of the level of killing, and thus will have an identical probability of being stable. While the evolution of non-responsive killing may cause shifts in which species are able to make a viable community, therefore, we find no net effect on the stability of the communities that can persist in the face of the killing.

1e. Effect of redundancy on community stability

Our models indicate that the instability of highly cooperative communities stems from having strong positive dependencies between species. What happens though if the interactions between cooperating partners is less specific and, instead of interacting strongly with few partners, cooperators interact weakly with several partners? Such redundancy is common in biological networks and we want to understand its effects here. One simple effect of such redundancy will be to weaken the strength of cooperative links, which has a stabilizing effect (Method 1b). However, redundancy is also expected to increase

the number of interactions. For example, rather than interacting with one species with an interaction strength of $a_{ij} = 1$, a species may instead interact with 3 others, where each interaction now has strength $\frac{1}{3}$.

We capture this in our analytic model with the parameter q, which represents the average number of new links that each original cooperative link is replaced with. Each new link has, on average, a strength $a_{ij}q$, where a_{ij} represents the strength of the original interaction. As redundancy increases so too will the connectivity of the community, C^{new}, as well as the overall proportion of cooperation, P_m^{new}. However, concurrently the average strength of cooperative interactions, and interactions in the community as a whole, will decrease. These changes are captured by,

$$\begin{aligned} C^{\text{new}} &= C^I(P_m^I q + P_c^I) \\ Pm^{\text{new}} &= \frac{qP_m^I}{qP_m^I + P_c^I} \end{aligned} \tag{3.9}$$

Where P_m^I, P_c^I, and C^I represent the initial levels of cooperation, competition, and connectivity respectively.

We can then examine the effect of increasing redundancy in communities with different initial levels of cooperation (Figure S3.17). We find that redundancy does not affect communities that start with high levels of cooperation, as any stabilizing effects of reduced link strength are canceled out by equivalent increases in the levels of cooperation and connectivity. However, in communities that begin with intermediate levels of cooperation, the stabilizing effect of reduced interaction strengths outweigh the destabilizing effects of increased cooperation and connectivity. To conclude, when redundancy means that cooperators interact with several other species weakly rather than few species strongly, ecological stability can increase.

1f. Numerical simulations of representative communities

In the sections below we present two fundamentally different methods for analysing ecological stability: permanence analysis and individual-based modelling. However, we first directly confirm our analytical method by simulating representative networks and explicitly calculating the eigenvalues of the corresponding Jacobians. As above, we consider the generalized Lotka-Volterra model with S species outlined in (3.1), and now follow the same approach as (Mougi and Kondoh, 2012) and (Allesina and Tang, 2012), whereby each species' intrinsic growth rate r_i is solved for so as to ensure a feasible equilibrium at $\mathbf{X} = \mathbf{1}$. The corresponding Jacobians will therefore have off-diagonal entries $J_{ij} = a_{ij}$, and diagonal entries $J_{ii} = -s_i$. We can then construct representative Jacobians as follows.

We first create an $S \times S$ matrix M with our desired connectivity, C. For each interaction pair (M_{ij}, M_{ji}) we draw a random variable p_1 from a Uniform $U([0,1])$ distribution, if $p_1 \leq C$ then we set M_{ij} and $M_{ji} = 1$, else, both are set to 0. We next impose the relevant distribution of interaction types on this network, where P_m represents the proportion of cooperative, $+/+$, interactions, P_c the proportion of competitive, $-/-$, interactions, and $1 - P_m - P_c$ the proportion of exploitative, $+/-$, interactions (for simplicity we confine ourselves to these types, however, this method can easily be extended to also incorporate $+/0$ and $-/0$ interactions). For each non-zero interaction pair (those where M_{ij} and $M_{ji} = 1$) we draw a new random variable p_2 from a $U([0,1])$ distribution. If $p_2 \geq 1 - P_m$ then (M_{ij}, M_{ji}) represents a cooperative interaction, and we draw values for the entries of our Jacobian J_{ij} and J_{ji} from a half-normal distribution $|N(0, \sigma^2)|$. If $p_2 \leq P_c$ then (M_{ij}, M_{ji}) represents a competitive interaction, and we draw the entries J_{ij} and J_{ji} from a negative half-normal distribution $-|N(0, \sigma^2)|$. If $P_c < p_2 < 1 - P_m$ then (M_{ij}, M_{ji}) represents an exploitative interaction, in this case we draw a further random variable p_3 from a $U([0,1])$ distribution. If $p_3 \leq 0.5$ then

species i benefits from the interaction whilst species j suffers and we draw J_{ij} from an $|N(0, \sigma^2)|$ distribution and J_{ji} from an $-|N(0, \sigma^2)|$ distribution. If $p_3 > 0.5$ then the converse is true, and we draw J_{ij} from an $-|N(0, \sigma^2)|$ distribution and J_{ji} from an $|N(0, \sigma^2)|$ distribution. Finally, we set each diagonal entry J_{ii} to $-s$. Thus we are able to generate Jacobians corresponding to networks with any mix of different interaction types, then solve to find the corresponding eigenvalues using Matlab (Mathworks, Natick, MA, USA).

The effects of feeding or spatial segregation on network stability are incorporated through altering the strengths of the interactions between species. Here, representative Jacobians are generated as outlined above, then, in the case of general feeding, all entries of the Jacobian are reduced by the weighting factor, in the case of targeted feeding, only the Jacobian entries corresponding to cooperative interactions are reduced, and in the case of spatial segregation, all Jacobian entries except for those on the diagonal, J_{ii}, are reduced.

1g. Comparison with previous numerical work

In this section we analyze a recent study of macroscopic communities by Mougi and Kondoh (Mougi and Kondoh, 2012) to show that our analytical method can recapitulate the behaviour of previous numerical analyses. In addition, we explain the relationship of our predictions to (Mougi and Kondoh, 2012), as we reach different conclusions owing to our focus on microbial communities. Both our work and that of (Mougi and Kondoh, 2012) investigate the stability of a community of S species whose dynamics are described by the system of Lotka-Volterra equations,

$$\frac{dX_i}{dt} = X_i(r_i - s_i X_i + \sum_{j=1, j\neq i}^{N} a_{ij} X_j).$$

Here X_i is the density of species i, r_i its intrinsic growth rate, s_i the strength of self-regulation, and a_{ij} the effect of species j on i. A proportion C of possible interactions between species are realized, and these interactions are split such that P_m are cooperative $(+/+)$ and $P_c = 1 - P_m$ are exploitative $(+/-)$. Our work also allows for $-/-, +/0$, and $-/0$ interactions, however, for comparison, we here limit ourselves to just $+/+$ and $+/-$ interactions.

Mougi and Kondoh numerically simulated communities with varying proportions of interaction types, and showed an increase in ecological stability for intermediate levels of cooperation (intermediate P_m). However, we find a monotonic decrease in ecological stability as cooperation is increased. Here we show that the finding of peak stability at intermediate cooperation is a consequence of specific assumptions in (Mougi and Kondoh, 2012) on how organisms in macroscopic communities function. As we discuss below, these assumptions are unlikely to generally apply to microbial communities. Nevertheless, we show here that our analytic approach can recapitulate these earlier numerical results.

The key assumption can be understood in terms of a focal species having a separate "capacity" (time budget) for each of its interaction types. Consider, for example, an ant species that is in in a mutualism with an acacia plant. Mougi and Kondoh assign the ant species a separate time budget for a) the exploitative interactions with other species and b) the cooperative interactions with other species. As a result, if the ant species takes on more mutualistic partners, a second acacia species, it is assumed that the strength of its interaction with the original plant is weakened but, critically, the strength of its interaction with prey species is not weakened. Therefore, if one adds more cooperative (mutualistic) interactions to a network, it will down-weight only the cooperative interactions, and if one adds more exploitative (predator) interactions to a network it will down-weight only the exploitative interactions. However, this particular form of

weighting, by interaction type, does not have a clear application to microbial communities. Interactions between microbial species can readily switch from cooperative to exploitative (or competitive) in a manner that is inconsistent with each interaction type being a distinct enterprise with its own capacity. For example, a bacterial mutualism can change to another interaction type if a focal species simply stops providing a metabolite for another, or if the focal species up-regulates the production of an antibiotic (Mitri and Foster, 2013). Such switches — effectively from a mutualist to a predator — highlight that different interaction types function within a single capacity, and give no justification for the weighting scheme illustrated with the ants and acacias above. And, as we show next, this assumption is key to the conclusion that community stability peaks at an intermediate proportion of cooperative interactions.

Mathematically, the weighting assumption is described by,

$$a_{ij} = \frac{e_i f_M A_{ij}}{\sum_{k \in \text{resources of mutualist } i} A_{ik}},$$

for cooperative (mutualistic) interactions, and,

$$\begin{aligned} a_{ij} &= \frac{g_i f_A A_{ij}}{\sum_{k \in \text{resources of predator } i} A_{ik}} \text{ and} \\ a_{ji} &= -\frac{f_A A_{ji}}{\sum_{k \in \text{resources of predator } i} A_{ik}}, \end{aligned}$$

for the predator and prey respectively in an exploitative interaction. Here A_{ij} represents the potential preference of species i to interact with species j, whilst e_{ij} and g_{ij} describe the efficiency of cooperative (mutualistic) and exploitative (predatory) interactions respectively. Each of these variables are drawn from Uniform $U(0,1)$ distributions with means $E(A_{ij}) = A$, $E(e_{ij}) = e$ and $E(g_{ij}) = g$. f_M and f_A describe the relative importance of mutualism versus predation, although these are simply set to $f_M = f_A = 1$ (Mougi and Kondoh, 2012). In the limit of large S, C, and P_m this can be approximated by,

$$\begin{aligned}
\text{Mutualistic, } (+/+)\ a_{ij} &= \frac{f_M e_{ij} A_{ij}}{P_m(S-1)CA}, \\
\text{Predation, } (+/-)\ a_{ij} &= \frac{f_A g_{ij} A_{ij}}{0.5 P_e(S-1)CA}, \\
\text{Prey, } (-/+)\ a_{ij} &= \frac{-f_A A_{ij}}{0.5 P_e(S-1)CA}.
\end{aligned}$$

In their supplementary material, Mougi and Kondoh showed numerically that the increase in stability at intermediate levels of cooperative interactions was not present when omitting this weighting in networks with a cascade structure. In Figures S3.11 and S3.12 we first confirm numerically that the increase in stability at intermediate levels of cooperation also disappeared without the frequency dependent scaling of interaction strength in unstructured networks, such as our microbial networks. We next derive a stability criterion that captures frequency-dependent scaling of interaction strengths analytically. We determine stability by analyzing the Jacobian, M corresponding to this system, whose off-diagonal entries are given by $M_{ij} = a_{ij}X_i$, whilst the on-diagonal entries are simply $M_{ii} = -s_i X_i$. In the following derivation, we assume a constant level of self-regulation and equilibrium species densities, such that $X_i = X \forall i$ and $s_i = s \forall i$. The general results still hold when these factors are also drawn from probability distributions, but the ellipses become less accurate.

We use the same analysis employed when deriving our previous stability criteria, which hinges upon the observation made by Sommers (Sommers et al., 1988), that a matrix N with mean 0, variance $\frac{1}{S}$, and correlation $E(N_{ij}N_{ji})$ (the mean of the product of all entries) centered at 0 will have eigenvalues uniformly distributed within an ellipse $(x/a)^2 + (y/b)^2 \leq 1$ where $a = 1 + \tau, b = 1 - \tau$ and τ is defined as $\tau = SE(N_{ij}N_{ji})$. Further, Rothblum and Tan showed that the only affect of adding the matrix $m\mathbf{1}.\mathbf{1}^{\mathbf{T}}$ to N will be to change the eigenvalue, λ^*, corresponding to the average row sum and the eigenvector $\mathbf{1}$ (which will become $\lambda + Sm$). Finally, adding a constant factor d to

each diagonal entry of the matrix will simply shift all eigenvalues by d along the real axis, such that each $\lambda_i = \lambda_i + d$. Thus, whilst the Jacobian, M, corresponding to our equilibrium community does not meet the requirements of Sommers's work, the network $N^* = \frac{N}{\sqrt{(SVarN))}}$, with $N = M - E(M)$, does, and therefore, the eigenvalues of our Jacobian will be contained within the ellipse centered at $-sX$, with half-horizontal radius, a, and half-vertical radius, b, given by,

$$a = \sqrt{SVar(N)}\Big(1 + \frac{E(N_{ij}N_{ji})}{Var(N)}\Big), \tag{3.10}$$

$$b = \sqrt{SVar(N)}\Big(1 - \frac{E(N_{ij}N_{ji})}{Var(N)}\Big). \tag{3.11}$$

We now derive expressions for $Var(N)$ and $E(N_{ij}N_{ji})$ in terms of the various community parameters (S, C, and the proportions of interaction types), such that we can determine how the eigenvalue distribution, and therefore stability, of our community varies as these parameters change. We first note that the mean of our Jacobian is given by,

$$\begin{aligned} E(M_{ij}) &= C\Big(P_m E(M_{ij}|+/+) + 0.5P_e E(M_{ij}|+/-) + 0.5P_e E(M_{ij}|-/+)\Big), \\ &= C\Big(P_m \frac{f_M e X}{P_m(S-1)C} + 0.5P_e \frac{f_A g X}{0.5P_e(S-1)C} - 0.5P_e \frac{f_A X}{0.5P_e(S-1)C}\Big), \\ &= \frac{X}{(S-1)}\Big(f_M e + f_A(g-1)\Big), \end{aligned}$$

and the second moment $E(M_{ij}^2)$ by,

$$\begin{aligned} E(M_{ij}^2) &= C\Big(P_m E(M_{ij}^2|+/+) + 0.5P_e E(M_{ij}^2|+/-) + 0.5P_e E(M_{ij}^2|-/+)\Big), \\ &= C\Big(P_m \frac{f_M^2 e_2 A_2 X^2}{(P_m(S-1)CA)^2} + 0.5P_e \frac{f_A^2 g_2 A_2 X^2}{(0.5P_e(S-1)CA)^2} \cdots \\ &\quad + \frac{f_A^2 A_2 X^2}{(0.5P_e(S-1)CA)^2}\Big), \\ &= \frac{X^2 A_2}{C(S-1)^2 A^2}\Big(\frac{f_M^2 e_2}{P_m} + \frac{2f_A^2(g_2+1)}{P_e}\Big). \end{aligned}$$

Where A_2, e_2 and g_2 are the second moments of A_{ij}, e_{ij} and g_{ij}, $A_2 = E(A_{ij}^2)$, $e_2 = E(e_{ij}^2)$ and $g_2 = E(g_{ij}^2)$ respectively. We can then calculate the variance of M, which is also the variance of $N = M - E(M_{ij})$, such that,

$$Var(M) = Var(N) = \frac{X^2}{C(S-1)^2}\Big(\frac{f_M^2 e_2 A_2}{P_m A^2} + \frac{2f_A^2(g_2+1)A_2}{P_e A^2} - C(f_M e + f_A(g-1))^2\Big) \quad (3.12)$$

From Sommers's theorem, we next require $E(N_{ij}N_{ji}) = E(M_{ij}M_{ji}) - E(M)^2$, for our stability criterion, and note that,

$$\begin{aligned} E(M_{ij}M_{ji}) &= C\Big(P_m E(a_{ij}a_{ji}X_iX_j|+/+) + P_e E(a_{ij}a_{ji}X_iX_j|+/-)\Big), \\ &= C\Big(P_m \frac{f_M^2 e^2 X^2}{(P_m(S-1)C)^2} - P_e \frac{4f_A^2 g X^2}{(P_e(S-1)C)^2}\Big), \\ &= \frac{X^2}{C(S-1)^2}\Big(\frac{f_M^2 e^2}{P_m} - \frac{4f_A^2 g}{P_e}\Big). \end{aligned}$$

We can, therefore, calculate the correlation of N,

$$E(N_{ij}N_{ji}) = \frac{X^2}{C(S-1)^2}\Big(\frac{f_M^2 e^2}{P_m} - \frac{4f_A^2 g}{P_e} - C(f_M e + f_A(g-1))^2\Big). \quad (3.13)$$

Combining this with equations (3.12) and (3.13) in (3.10) and (3.11) allows us to determine the coordinates of our eigenvalue ellipse. As before, there will also be an

eigenvalue that corresponds to the eigenvector $\mathbf{1}$ which does not necessarily lie within the ellipse, and is approximately equal to the average row sum, in this case equal to,

$$\begin{aligned}\text{Row sum} &= C(S-1)\Big(P_m\frac{f_M X e}{C(S-1)P_m} + 0.5P_e\frac{f_A X g}{0.5P_eC(S-1)} - 0.5P_e\frac{f_M X}{0.5P_eC(S-1)}\Big),\\ &= X(f_M e + f_A(g-1)).\end{aligned}$$

We can then see analytically how the distribution of eigenvalues changes as we increase the proportion of cooperative interactions, P_m, for the frequency-dependent weighted and unweighted cases. For example, in Figure S3.13, we plot the numerical stability (line) and average maximum eigenvalue from 10 networks, showing how our distributions match, and we reproduce the phenomenon of peak stability for intermediate P_m under the assumption of frequency-dependent, interaction specific weighting of interaction strengths.

Method 2

2a. Permanence analysis

We next consider a different method to analyze the ecological stability of communities. Permanence analysis asks whether a community will retain all its members, independent of the scale of any perturbation. Mathematically, this means that the boundary of the state space – the points where one or more species have gone extinct – behaves as a repeller, such that if the densities of any of the species within the community approach zero, these species will not collapse but instead grow, meaning no trajectory of the system leads to the extinction of any species. In contrast to the analytics that can deal with large community sizes, permanence analysis is computationally expensive and so we will analyze small communities with $S = 10$ and where interactions are either $+/+$ or $-/-$ in varying proportions. As we discuss below, permanence analysis also

has further restrictions in applicability such that it may under-estimate the proportion of permanent communities. Nevertheless, permanence analysis remains a much more complete study of the stability properties of a network, which makes it a valuable complement to the analytics. Moreover, despite its restrictions, we will show that it leads to the same conclusions as the analytical model and, subsequently, our individual-based model.

A sufficient condition for permanence was derived by Jansen in 1987 (Jansen, 1987). This derivation requires that not only must no species go extinct, but that it is not possible for any species to grow infinitely. In mutualistic Lotka-Volterra systems this is not always the case – some communities have the propensity for positive-feedbacks to drive species growth in such a way that some species will never stop growing. These situations are clearly unrealistic as microbiota populations will at least be capped by limited space within the gut. As this kind of cap is difficult to capture with first order differential equations, we investigate the effects of space limitation using an individual-based model (see supplementary section 3), and develop constraints for our permanence analysis so as to guarantee that we only analyze communities where infinite growth is not possible, while classifying those communities that permit infinite growth as non-permanent.

Specifically, the permanence analysis of a given community can be broken into the following parts. First, as in previous sections, we use Lotka Volterra equations to find a feasible equilibrium for the whole community by solving for the growth rates r using $X = \mathbf{1}$ (supplementary section 1a). We next check whether the community is unbounded – if the community is not bounded, and therefore has the capacity for infinite growth we classify it as non-permanent. If the community is bounded, we next check whether the boundary of the state space is also a repeller – if so then no species will ever go extinct, nor will any ever grow to infinity, and thus we classify the community as permanent. We outline these steps in more detail below.

2b. Boundedness of cooperative communities

In communities of mixed competitive and cooperative interactions, any growth towards infinity will be driven by the cooperative components of the community. Therefore, if the purely cooperative subsystem of a mixed community is bounded above then so too will be the mixed community.

In a purely cooperative community, interactions are described by an interaction matrix that is negative on the diagonal and non-negative elsewhere. Such a matrix is termed a Metzler matrix, and it has been shown that if a dynamical system described by a Metzler matrix is locally stable, then the it will also be globally stable (Goh, 1979). This means that, if a feasible equilibrium point of a purely cooperative community is globally stable, this community will be bounded, such that no species will grow to infinity. We can therefore determine whether a mixed community is bounded above by checking whether the cooperative subsystem of the community is locally stable.

To check the boundedness of a given community described by the matrix J, we therefore first determine the cooperative subsystem of the community by removing all competitive between species interactions – setting all the negative off-diagonal elements of J to zero, to create a new community matrix J^*. We then solve for a new equilibrium X^*, using the new community matrix J^*. J^* is now a Metzler matrix, as all diagonal entries are negative ($= -s$, $s > 0$) and all off-diagonal entries are non-negative. Therefore, if the new equilibrium is feasible, i.e. $X_i^* > 0$ for all i, and locally stable, i.e. the largest real part of the eigenvalues of J^* at equilibrium is negative, then we categorize the whole community as bounded.

Note, while global stability is a sufficient condition for boundedness, it is not necessary. This means we may underestimate the number of permanent communities as we discard those partially cooperative communities that do not permit a globally stable equilibrium yet do not grow to infinity. However, this confines our permanence anal-

ysis to bounded communities, and we employ the individual-based model to directly study properties of communities that we here classify as unbounded and non-permanent (Method 3, below).

2c. Permanence of bounded communities

The necessary conditions for permanence in Jansen's work yields the following linear programming problem (Jansen, 1987):

Minimize z subject to:

$$\sum_{i=1}^{S} h_i(r_i - s_i X_i^{B(k)} + \sum_{j=1, j\neq i}^{S} a_{ij} X_j^{B(k)}) + z \geq 0.$$

and $h_i > 0$ for all $i = 1, .., S$.

The h_i and z are variables of the linear programming problem taking into account the constraints arising from all feasible boundary equlibria X^B indexed by $k = 1, .., m$, where $m \leq 2^S$. We here perform permanence analysis in communities where $S = 10$. All other parameters and values, and the assembly of community matrices are the same as in the previous sections of this work with the exception that for the analysis of permanence, we choose a smaller self-regulation parameter, s. The small networks that we are studying here using permanence analysis are intrinsically more stable than the more realistically-large networks that we studied with the analytics above. All networks would simply be stable if we used the same self-regulation parameter as above. In order to study the effects of cooperation and other factors on stability, therefore, we lower s to 0.2 for the permanence analysis. Host manipulations such as feeding or spatial segregation are implemented as scaling factors for the entries of the Jacobian matrix, i.e. $a_{ij} \times f - 1$ where f is the strength of host manipulation. We simulate 200 independent communities for each set of parameters and plot the frequency of permanent systems.

Method 3

3a. Individual-based model

In addition to our permanence analysis, to verify our analytical work we also developed an individual-based model (IbM) of the microbiome. This cellular automaton explicitly models all cells in the community, and tracks their growth, death, and spatial location over time. The computationally intensive nature of our IbM limits the number of species that we are able to investigate, and in accordance with our permanence analysis, we again set species number $S = 10$.

This approach offers several important pieces of information. First, whilst our other models focus on the binary question of whether or not an equilibrium community will be stable to perturbation, this approach gives us added information as to how we expect a community to behave following perturbation, and allows us to track population sizes of all species over time. Additionally, we can explicitly take into account spatial structuring of a community. This allows us to more directly examine the effect of spatially segregated communities upon community stability. Similarly, we can more explicitly examine the effect of host-supplied nutrients on the manner in which cells interact with one another, and the subsequent effects on stability. Additionally, this approach introduces the realistic stochasticity of birth and death events, and cell-cell interactions.

As in our analytical work, we consider a community of S species, where species interact with one another with probability C (consistent with our previous work, here $C = 0.7$ throughout). These interactions may be competitive, exploitative, or cooperative (the model can also easily be extended to incorporate commensalism and ammensalism), and are defined by an interaction matrix, A. The interaction strengths are initially drawn from a half normal $|N(0, \sigma^2)|$ distribution. We then scale A such that overall a discrete approximation of the Jacobian associated with the community at equilibrium is of a similar magnitude and distribution to that of our analytic work. Although

each cell has a spatial location, we initially assume that diffusion in the environment is high, such that each cell can interact with every other cell within the environment. However, when we study the effects of spatial structure, cell interactions are constrained to a region around each cell (below).

We model the environment as a square lattice with periodic boundary conditions and sides of size N, such that the environment can contain a maximum of N^2 cells (throughout, $N = 100$). However if the population exceeds a certain cap (in our case, 90% of the maximum population size) then a random selection of cells are removed to reduce the overall population size to the level of the cap – this can be thought of as a density dependent sloughing that imposes an upper limit on population size whilst still enabling cells to grow.

Unless stated otherwise, each simulation begins by seeding the environment such that it is 80% full (initial population size $0.8 \times N^2$), with each species composed of a roughly equivalent number of cells, distributed at random. Once the environment has been seeded, we count the starting number of each species, and set this as our initial equilibrium point, X_{eq}. We then solve to find the intrinsic growth rate if each species, r_i, necessary to maintain this equilibrium,

$$r = -(A \times X_{eq}).$$

Simulations then consist of the following series of steps, described in further detail below.

1. Count how many cells of each species are present in the environment at the start of the time point, contained in the vector X_t.

2. Calculate the growth rate for each species as dependent on X_t. In our initial simulations, all cells of the same species will grow at the same rate.

3. Randomly list all of the cells present in the environment, consider each cell in turn according to this order.

4. Instigate growth / death of each cell depending upon its species' growth rate and whether there are free spaces such that growth is possible.

At the start of each time point we count the total number of each species, X_t, then calculate the growth / death rate of each species, defined by,

$$\text{growthRates} = r + AX_t - \text{diag}(A).$$

The $-\text{diag}(A)$ term removes the interaction of each specific cell with itself while maintaining interactions with other cells of the same species. We next rescale all the rates to be between $[0, 1]$ by dividing each rate by the magnitude of the maximum growth / death rate at that time point. This gives us an adaptive time-step that enables us to distinguish between species with very small differences in growth rates.

We next list all cells in random order, then assign an independent $U([0, 1])$ random number, q, to each cell. We consider each cell in the list in turn. For each cell, if its growth rate is positive and greater than its assigned random number, q, we assume growth will occur, and a new cell of the same type will be placed in an adjacent spot (up, down, left, or right, chosen at random). If there are no free spots immediately adjacent to the focal cell, then no growth occurs. Similarly, if the cell's growth rate is negative, and its magnitude greater than its assigned random number, q, then we assume death occurs, and that cell is removed from the environment.

Once each cell has been considered, we check to see whether the population has grown over the pre-defined population cap, and if so we remove a randomly chosen set of cells, such that the population is reduced down to the population cap. This marks the end of the time-point, and the process repeats again.

We initially allow the simulation to run for five time-steps, then perturb the community – reducing the population size of each species by 10%. The simulation process then returns to run as normal for 1000 time-steps, and we evaluate the stability of the community following this perturbation. Specifically, we define stability as the proportion of communities that retain all their initial species throughout the simulation.

3b. Effect of cooperation on community stability

We begin by investigating how changing the proportion of cooperation within a community affects its stability. As in our analytic work, we achieve this by altering the interaction matrix, A, to change the proportion of cooperative interactions. We investigate this in three different settings – increasing cooperation in initially purely competitive communities, in purely exploitative ones, and in communities where interactions are otherwise randomly assigned.

3c. Effect of spatial segregation on community stability

We are interested in what potential mechanisms the host might possess to promote a stable microbiome. As discussed in the main text, one possible mechanism is spatial segregation, whereby the host reduces the probability of species interacting with one another. Specifically, we consider a case where populations start from a varying initial density and then grow clonally to create patches of a single genotype. The lower the density of the initial inoculation, the stronger the spatial structure (figure S3.18). Moreover, whilst in our original simulations we assumed that each cell could interact with every other cell, we now assume that each cell only interacts with a subset of cells within a certain radius of it.

As in our initial simulations, we define the capacity for interactions between species by the $S \times S$ matrix A, the parameters of which are drawn from a half normal $|N(0, \sigma^2)|$

distribution, with signs chosen so as to achieve the desired levels of cooperation, competition, and exploitation. We next take the distribution of cells described above, and set this as our equilibrium community, and solve for the intrinsic growth rate of each cell necessary to achieve this. Due to the addition of spatial structure and limited interaction neighborhood each cell will interact with a different subset of the population, so we must now calculate the growth rate of each cell individually. We define the vector X_i^{nb} as the number of of cells of each species within the neighborhood of cell i. We then calculate the growth rate for cell i as,

$$r^i = -(A(l, \text{all}). * X_{nb}^i),$$

where l is the species type of cell i. For each species we then average the growth rates, r_i, of all the cells of that type, and set this as the intrinsic growth rate for that species. The simulations then progress in the same manner as in our first, non-structured model; the cells are randomly ordered, their growth rates calculated, then any action (growth / death) implemented according to this random order. Unlike our original simulations where all cells of the same species had the same instantaneous growth rate, each cell will now have a different overall growth rate depending upon the subset of the population that lies within its neighborhood. As in our initial simulations, we run the model for five time steps before perturbing all of the species densities, then observe the ability of the community to retain all of its initial species over time.

3d. Effect of feeding on community stability

Hosts further manipulate their microbiome via epithelial feeding of microbes within the gut community. Here we extend our simulations to capture this behavior, and examine the consequences for community stability. We focus on two potential effects of host provision of nutrients upon the behavior of the gut community – first, the scenario whereby host provided nutrients act as an alternative nutrient sources for the gut microbes, and

second, the case in which nutrients act as an extra source of energy by which to increase cell growth rates. We discuss each of these approaches in detail below.

Case 1: food as an alternative to interactions

One potential effect of host-provided nutrients is that they act as an alternative to bacteria-provided substrates – that is, the use of a host supplied nutrient weakens the interactions between the microbes in the community. For example, the host supplied nutrient may be an alternative carbon source to a nutrient supplied by another community member. This case is directly analogous to the case explored in our analytic work (Figures 3.4 B, S3.16). Here, we explore two possible options; in the first all interactions between cells can potentially be weakened by interacting with the host supplied nutrients (general feeding, Figure S3.16) whilst, in the second, host supplied nutrients only have the effect of weakening interactions between cooperative species (targeted feeding, Figure 3.4 B). In both cases we assume that each species' initial equilibrium density is determined from a combination of both its interactions with other microbes and with the host supplied nutrient.

In our simulation we capture this with the presence of an additional nutrient source, kept at a constant level, which can be metabolized by members of the gut community. Like cells themselves, this nutrient has a spatial component, and will only weaken interactions when a cell overlaps with the location of the nutrient. We define the effect of using the nutrient, ϕ, (here $\phi = 10^{-5}$ so as to be on the same order of magnitude as a single cell-cell interaction in our models) and the simulation is then composed of the following steps, listed below.

As in our initial simulations, the process begins by seeding the environment with roughly equal numbers of each species, placed at random (without any structure), and setting this distribution as the community equilibrium. In the same manner, set quantities of nutrients are also placed at random in the environment (to be redistributed each

timestep). We set γ to be the proportion of cells that will on average overlap in location with nutrient molecules, and of this, δ represents the proportion of these overlaps in which, on average, nutrients are metabolized in favor of an interaction with neighboring cells. This equates to whether cells have a preference for interacting with other microbes (when $\delta < 0.5$), or with the host provided nutrients (when $\delta > 0.5$). We then use this to calculate the intrinsic growth rates, r, of each species as,

$$r = -(A(X_{eq} - \delta\,\gamma\,X_{eq}) + 1\delta\,\phi\,\gamma\,N^2),$$

so as to set the initial population as the community equilibrium. The simulation then follows the same series of steps to our previous implementations of the model.

Case 2: Food as a supplement to cell-cell interactions

We next investigate the scenario where host-provided nutrients acts as an extra resource that acts in addition to pre-existing microbial interactions. In this case we consider a population that is already at equilibrium, then study the effect of adding nutrients to the environment. As in Case 1, we begin by seeding the environment with a random distribution of each species, and a set amount of nutrients. We then calculate the intrinsic growth rates of each species, r, independent of any nutrients present in the environment, such that,

$$r = -(A(X_{eq})).$$

Each time-step then begins by counting the population size of each species, contained in X, then calculating each species' growth rate, now given by,

$$\text{growthRates} = r + AX - diag(A) + \phi\,\gamma\,N^2,$$

where as before ϕ represents the strength of an interaction with the host supplied nutrients, and γN^2 the number of nutrient molecules that each cell interacts with. As

such, the last term on the right represents the contribution to the overall growth rate from interactions with host-supplied nutrient molecules. As before, growth rates are rescaled by the magnitude of the maximum rate of cell growth or death in the community, cells ordered at random, then each cell considered in turn, with any actions (growth / death / nothing) decided based on the value of their species' growth (or death) rate.

When the additional nutrients provided by the host remain at a constant level throughout the simulation, this increases intrinsic growth rates (growth not dependent upon microbial interactions) and therefore reduces the likelihood of species going extinct. By the measure of stability used to analyze our IbM results – whether a community is able to maintain all of its initial species over time, feeding therefore acts to increase the stability of the microbial community (Figure 3.4 C). This is a different form of stability to that above. As is typical in the literature, our earlier analyses consider communities maintained at a stable equilibrium by the interactions between the community members. However, the effect of increased growth rates here is to push the species' populations to their maximum size, such that they are now regulated extrinsically by the capacity of the host. This is critical because it can turn what were formerly cooperating species into competitors.

To see this, it is important to realize that the occurrence of competition versus cooperation can be more nuanced than simply examining the interaction parameters (a_{ij}) between species in the model. In the presence of an external population cap, these interaction terms only define the net effects of one species on another during early stages of growth, before the community has reached the cap. However, once the community has reached its steady state at the cap, the effect of one species on another is now a combination of both the direct interaction terms (a_{ij}) and competition for limited space. This latter effect can change previously cooperative interactions into competitive ones.

The change to competition can be demonstrated in our model by removing a previously cooperative species and observing how this removal affects the other species.

If the focal species is cooperating, removing it should harm the other species (Mitri and Foster, 2013). However, removing the focal species, we see that the other species typically both increase in their densities and their rate of cell production per unit time (Figure 3.4 C). The removed species was then, at this point, primarily a competitor with negative effects on most other species.

4. Experimental validation

While there is a rapidly growing body of data on species abundances in the mammalian microbiome, there have been few studies focused on deciphering the interactions between these species. However, the work by Stein *et al.* (Stein et al., 2013) is one such study. The authors used time-resolved metagenomics and machine learning to infer the interactions between community members – species or genera – within a mouse model. These data then allowed us to test our predictions about the profile of ecological interactions within the mammalian microbiome.

We first looked at the distribution of interaction types between the community members. Our analysis predicts that a stable community within the microbiome will contain only a small proportion of destabilizing cooperative interactions, amongst a larger number of competitive or exploitative links. We found that this is indeed the case in the Stein *et al.* dataset (Figure 3.4 D, bar chart), in which of all possible interactions, only 14.8 % are cooperative.

A second key prediction from our analysis is that, in a stable microbial community, the interactions between species should be predominantly weak relative to the self-regulation that each species experiences due to within-species competition, captured by s (see equation S1). Again we find our prediction is supported by the Stein et al. data – the histogram in Figure 3.4 D shows how the vast majority of ecological interactions between community members are weaker than that of the average interaction of a species with itself ($\bar{s}$), with only a few strong ecological interactions between members.

The models we have presented here are general, which makes them widely applicable and able to identify general principles. However, a criticism of such models is that generality might come at a cost of accuracy if important details are not accounted for. Do our models then make correct predictions when we parameterize them using real data? We evaluated this question using the Stein *et al.* data. Specifically, we obtained the general community parameters – $S, C, P_m, P_c, \sigma, \bar{s}$ – of a stable mouse microbiome community from Stein*et al.* Using these in our analytical model (supplementary section 1) we can predict the bound on the eigenvalues, and therefore the stability, of this community. Not only did our analysis correctly predict the community to be stable, but we also correctly estimated the location of the stability determining dominant eigenvalue in the community when this value was explicitly calculated using each individual member's densities and interaction strengths (Figure 3.4 D, right).

3.5 Supplementary Figures

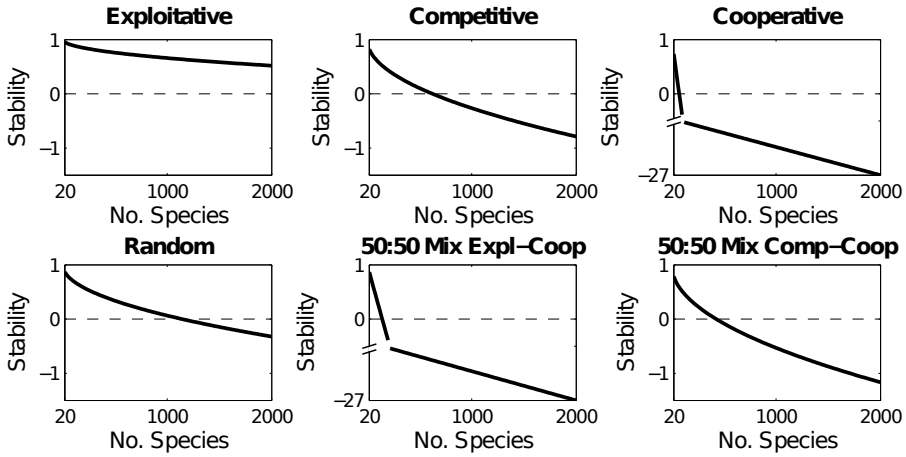

Figure 3.5: **Diversity destabilizes communities**. Here we plot our analytic measure of stability over increasing species number for random networks with different interaction types ($C = 0.7, -s = -1, \sigma = 0.05$). Increasing the species number has destabilizing effects across communities of all interaction types. Mathematically, adding new species to a network in this way always increases the largest real part of the community's eigenvalues and, therefore, the stability of the community decreases. The higher the proportion of positive interactions within the community, the larger this effect -– and thus diversity is most destabilizing for cooperative communities. Note that we use a break in the y axis of some plots in order to allow the early behaviour of all plots to be directly compared.

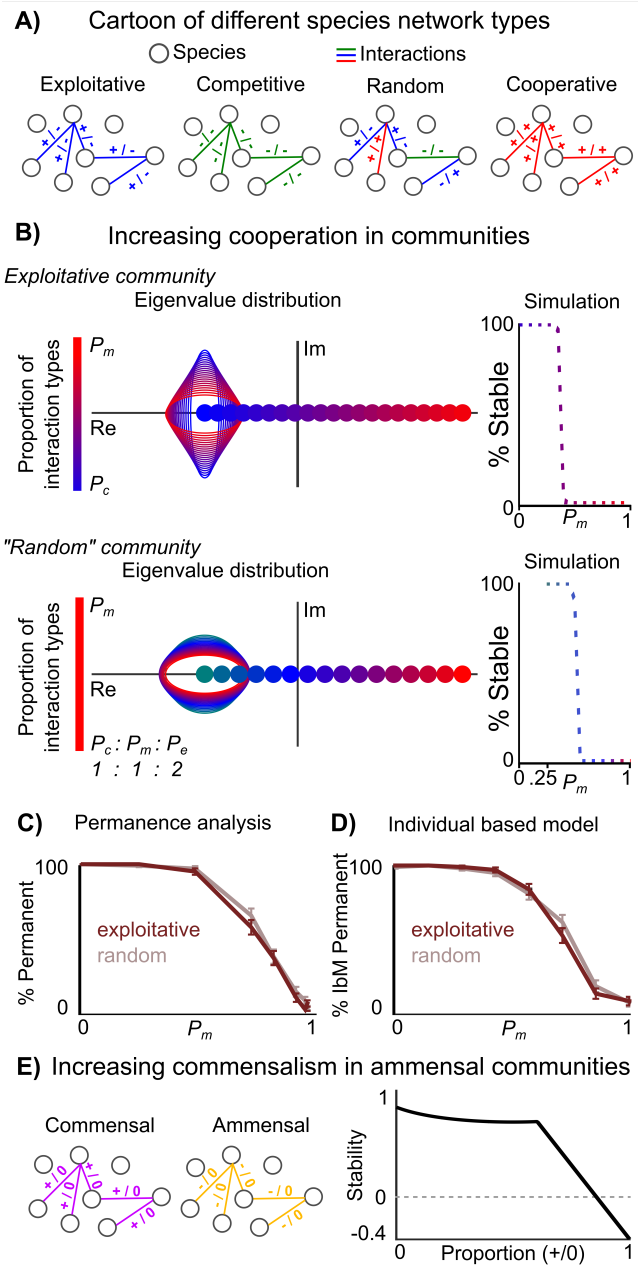

Figure 3.6: **Increased proportions of cooperative interactions decrease stability in exploitative and random communities.** A) Illustration of different network types; competitive (green), exploitative (blue), random, and cooperative (red). B) Linear stability analysis. Left hand plot shows analytical solutions for eigenvalue locations as a function of increasing cooperation (shown as increasing red). The largest value of the real components (x-axis) determines whether, and how fast, a return to equilibrium occurs (stability), whilst the imaginary components (y-axis) determine the frequency of oscillations in population densities following perturbations (Figure 1). The solutions give the position of all eigenvalues in the form of an ellipse, with the exception of a single eigenvalue that corresponds to the average row sum of the interaction matrix (represented by a dot that may lie outside of the ellipse). Increasing cooperation increases the largest eigenvalues, and stability decreases. Solutions hold for any permutation of a community network with a given parameter set (here: $S = 100$, $C = 0.7$, $-s = -1$, $\sigma = 0.05$, see supplementary section 1b for parameter sweeps that show that cooperation is nearly always destabilizing). Right hand plots are simulation results showing the proportion of communities that are stable to confirm our analytic results. C, D) Increasing cooperation also decreases community stability in our permanence analysis and individual-based model (here: $S = 10$, $C = 0.7$, $-s = -0.2$, $\sigma = 0.05$, errorbars: SEM for 100 samples). E) Increasing commensalism $(+/0)$ within an ammensal $(-/0)$ community has a similarly destabilizing effect.

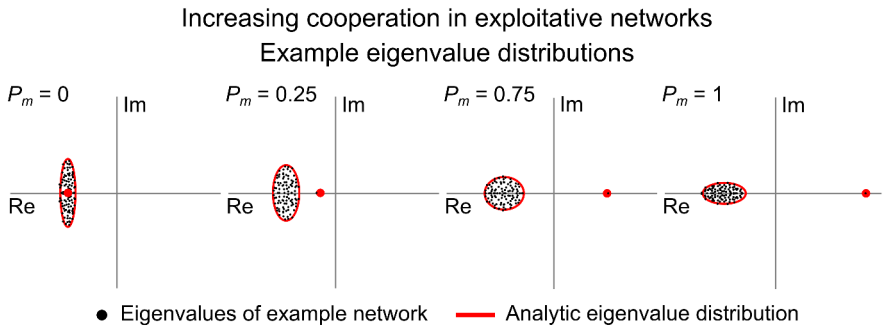

Figure 3.7: **Numerical confirmation of the analytical model for increasing cooperation in an exploitative community**. The stability of a community is dependent upon the localization of its eigenvalues within the complex plane, which we can predict using our analytic measure. We confirm the accuracy of this analytical approach by simulating representative communities ($S = 100, C = 0.7, -s = -1, \sigma = 0.05$) and explicitly calculating their eigenvalues (black dots), which we can see are contained within our analytically calculated bound (red, Figure 3.2). For example, we confirm that stability decreases (the largest real eigenvalue part becomes less negative) when we increase the proportion of cooperation in an otherwise exploitative network.

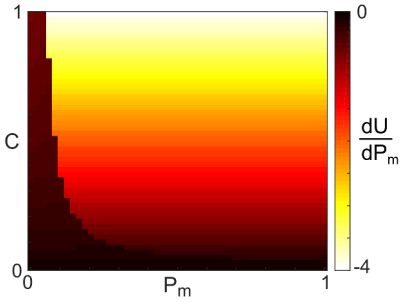

Figure 3.8: **Effect of increasing cooperation on stability in exploitative communities.** The derivative of stability with respect to the level of cooperation, P_m, across C/P_m parameter space reveals how increasing cooperation always destabilizes exploitative communities, regardless of the connectivity of the network – that is, $\frac{dU}{dP_m} < 0$ for all values of connectivity, C, and cooperation, P_m. Here we illustrate this in a community with species number $S = 100$, and self-regulation, $-s = -1$, however, the same behavior holds true for equivalent communities with different levels of self-regulation, and higher species numbers. The two visually distinct regions of the heatmap indicate the regions where the ellipse determines stability (left) and where the dot does (right), illustrating how the dot decreases stability at a linear rate, whilst this rate is much slower in regions where community behavior is driven by the ellipse. The magnitude of $\frac{dU}{dP_m}$ is larger for higher C, indicating how increasing cooperation is more destabilizing in communities with higher connectivity.

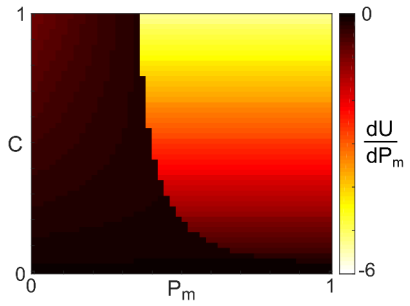

Figure 3.9: **Effect of increasing cooperation on stability in random communities.** The derivative of stability with respect to the level of cooperation, P_m, across C/P_m parameter space reveals how increasing cooperation always destabilizes random communities, regardless of the connectivity of the network – that is, $\frac{dU}{dP_m} < 0$ for all values of connectivity, C, and cooperation, P_m. Here we illustrate this in a community with species number $S = 100$, and self-regulation, $-s = -1$, however, the same behavior holds true for equivalent communities with different levels of self-regulation, and higher species numbers. The two visually distinct regions of the heatmap indicate the regions where the ellipse determines stability (left) and where the dot does (right), illustrating how the dot decreases stability at a linear rate, whilst this rate is much slower in regions where community behavior is driven by the ellipse. The magnitude of $\frac{dU}{dP_m}$ is larger for higher C, indicating how increasing cooperation is more destabilizing in communities with higher connectivity.

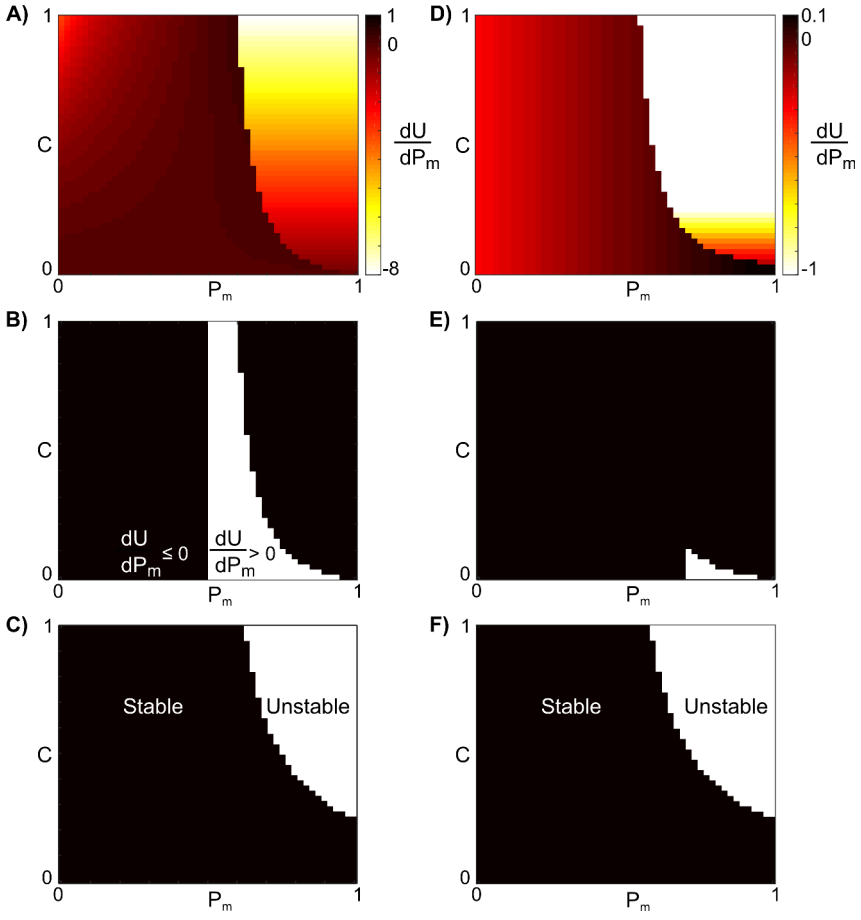

Figure 3.10: **Effect of cooperation on stability in competitive communities.** A) The derivative of stability with respect to the level of cooperation, P_m, across C/P_m parameter space reveals how the effect of cooperation upon stability in competitive communities changes across C/P_m parameter space. When $\frac{dU}{dP_m} < 0$ cooperation is destabilizing, whilst when $\frac{dU}{dP_m} > 0$ it is stabilizing. Note that the magnitude of the rate of stabilization is far smaller than the average rate of destabilization. B) Black regions highlight parameter space where increasing cooperation is destabilizing, whilst white regions mark where it is stabilizing. The size of the region where increasing cooperation is stabilizing decreases as connectivity, C, or species number, S, increase (here $S = 100$). C) The stabilizing influence of cooperation increase does not typically affect the transition from stable to unstable communities. This is illustrated by plots of ecological stability for a community with $S = 100, \sigma = 0.05$. D, E) In realistic communities that also contain exploitative interactions (here at 20%), the range of parameter space where increasing cooperation can be stabilizing is further reduced. F) Plot of ecological stability for communities that also contain exploitative interactions; again the stabilizing effect of increasing cooperation does not affect the stability threshold.

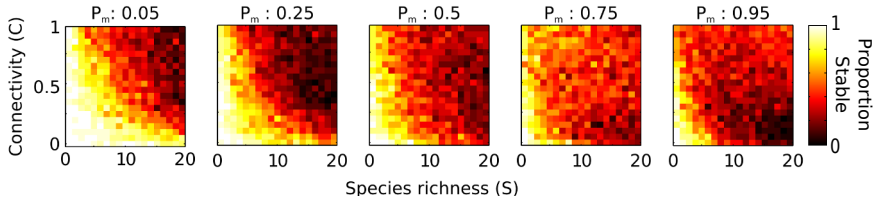

Figure 3.11: **Numerical confirmation that intermediate levels of cooperation (mutualism) can promote stability in a model of a macroscopic community.** We reproduce the numerical model of (Mougi and Kondoh, 2012) in an unstructured network, and demonstrate how certain interaction dependencies can lead to increased stability for intermediate levels of cooperative interactions. Specifically, here different interaction types (cooperative versus exploitative) are weighted independently in a frequency dependent manner – such that a given species always dedicates a constant level of effort towards cooperative interactions, and a separate constant level towards exploitative interactions. As in (Mougi and Kondoh, 2012) we numerically simulate such communities for varying values of species number, S, and connectivity, C, and plot the proportion of communities that are stable for each parameter combination in 25 independent repeats.

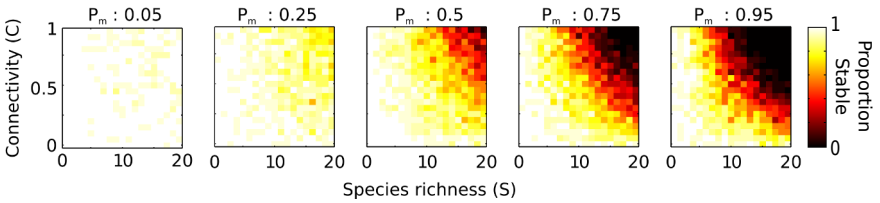

Figure 3.12: **Numerical confirmation that increased stability at intermediate levels of cooperation (mutualism) is not observed when each interaction type is not weighted individually.** We numerically simulate communities in S, C parameter space for varying levels of cooperative interactions (mutualism) without the frequency dependent weighting of interaction types. We simulate 25 independent communities for each parameter combination, with interaction strengths drawn from a Uniform $U([0, 0.1])$ distribution so as to, on average, be of a similar magnitude to those in Figure 3.11, and plot the proportion of communities that are stable.

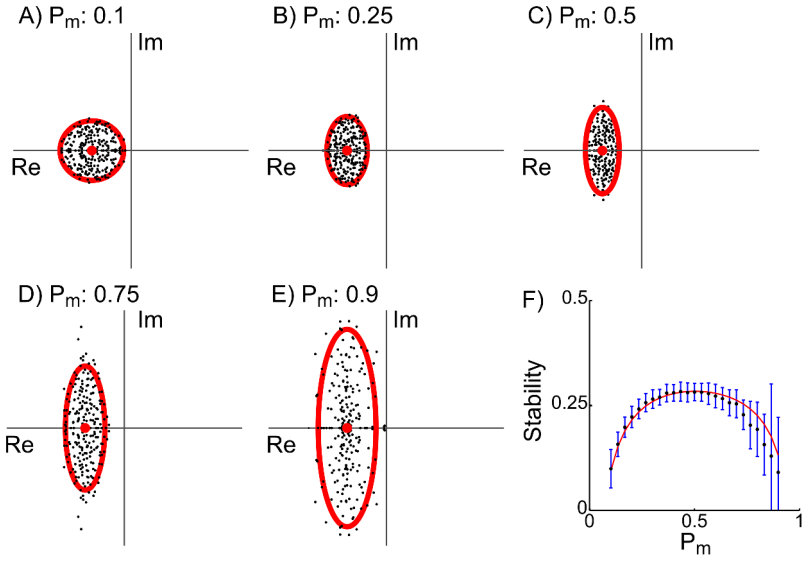

Figure 3.13: **Our analytical model captures the stability behavior seen numerically in (Mougi and Kondoh, 2012).** Ecological stability is dependent upon the localization of the community's underlying eigenvalues in the complex plane. Here we demonstrate that our analytic approach can predict the change in eigenvalue localization for communities in which interactions between species are weighted in a frequency dependent manner, as in (Mougi and Kondoh, 2012). A-E) Eigenvalues (black dots) of 4 simulated communities ($S = 100, C = 0.7$) with the frequency dependent weighting on interaction types as implemented in Figure 3.11 and (Mougi and Kondoh, 2012), alongside our analytical prediction of the eigenvalue distribution (red line and dot). Some eigenvalues lie outside the ellipse because we sample interaction strengths from a random distribution, but this does not affect our ability to capture the general behavior. Panel F compares our analytical prediction of stability (red line) against the average observed stability for 100 representative networks (black dots), with blue lines representing the standard deviation. This demonstrates how our analysis can capture the behavior seen in (Mougi and Kondoh, 2012), and our numerical simulations, when each interactions type is weighted individually.

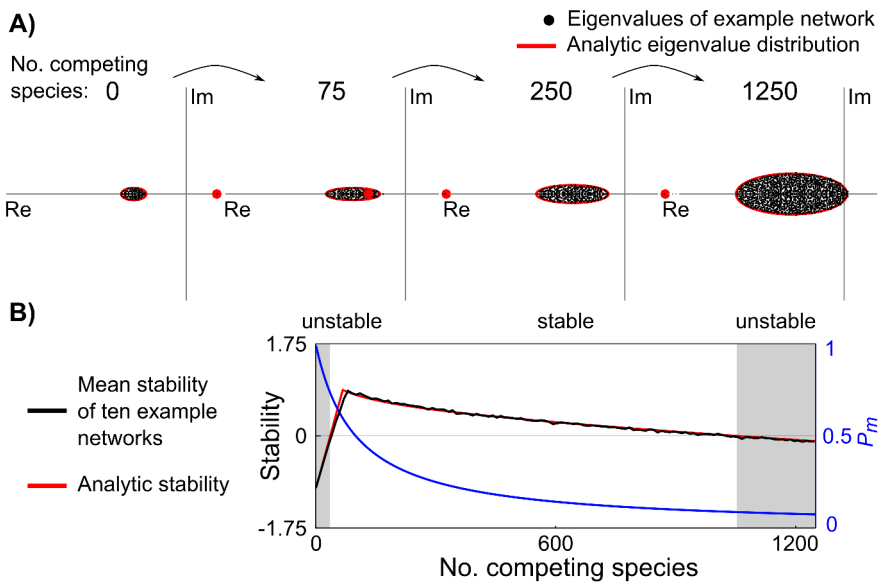

Figure 3.14: **Numerical confirmation of the analytical model for introducing competitive species to a network of cooperators** The stability of a community is dependent upon the localization of its eigenvalues within the complex plane, which we can predict using our analytic measure. A) We confirm the accuracy of our analytical results by numerically simulating communities in which competitive species are added to an initially purely cooperative community of 100 species ($C = 0.7, -s = -1.75, \sigma = 0.05$). We plot the eigenvalues of these simulated communities (black dots) and show how their localization is accurately predicted by our analytic bound (red line) B) We then simulate independent sample communities with different numbers of competing species, calculate their average stability (10 each, black line), and compare this with our analytical results (red line). The addition of competitors changes the proportion of cooperative links (P_m, blue line) which has a stabilizing effect that initially is stronger than the destabilizing effect of larger species numbers.

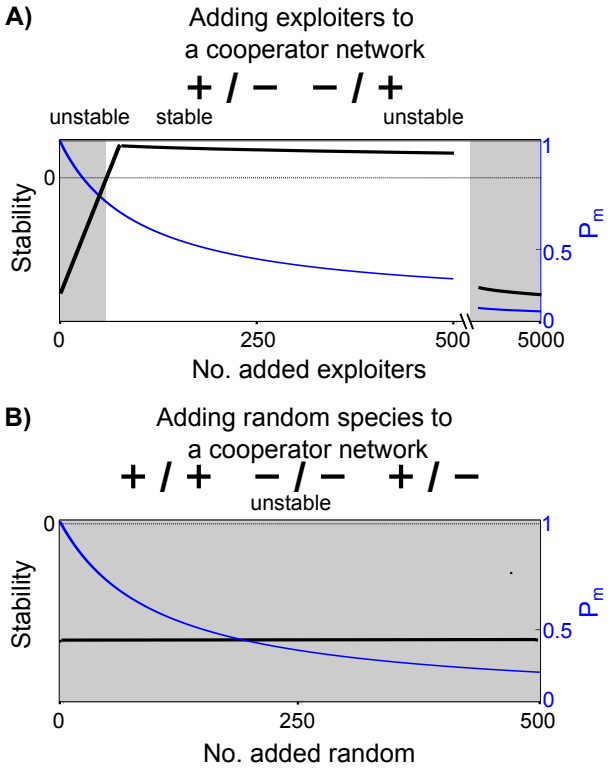

Figure 3.15: **Introducing exploitative species to a network of cooperators is also stabilizing, but adding random species is not**. We consider an unstable cooperative network in which all interactions are concentrated between a small number of species ($S_{\text{init}} = 100, C = 0.7, -s = -1, \sigma = 0.05$), then gradually add new species such that the original cooperative interactions are now distributed between a larger number of species. The new species interact with community members in a manner that is either exploitative (A) or "random" (B). We plot our analytic measure of stability as a function of the number of new species added (black line), and (A) demonstrate how adding exploitative species can initially stabilize the focal community, although, at high numbers of additional species, the destabilizing effect of high species numbers (Figure 3.5) will dominate, and the community will again destabilize. To capture this we break the plot into two sections (note the x-axis break) as the destabilizing behavior occurs at a much lower rate than the initial stabilization that occurs at low species numbers. (B) Adding species that interact in a "random" manner (25 % cooperatively, 25 % competitively, and 50 % exploitatively) does not have a stabilizing effect, as it does not sufficiently decrease the overall proportion of destabilizing cooperative interactions (blue line).

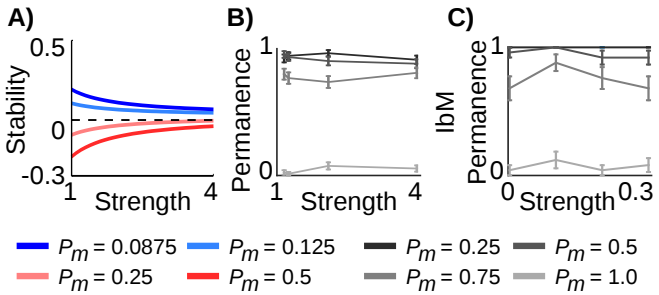

Figure 3.16: **Non-targeted feeding does not make communities stable.** Host-supplied nutrients that weaken all interaction types equally do not increase community stability in linear stability analysis (A), permanence analysis (B), or an individual-based model (C). In order to capture any effects of host manipulation, we study parameters for which some communities are stable and some are unstable. For linear stability analysis: $S = 300$, $C = 0.7$, $-s = -1$, $\sigma = 0.05$. For permanence analysis and IbM: $S = 10$, $-s = -0.2$, errorbars: SEM). We use different parameters for the second two methods because they are computationally expensive and can only be applied to smaller communities.

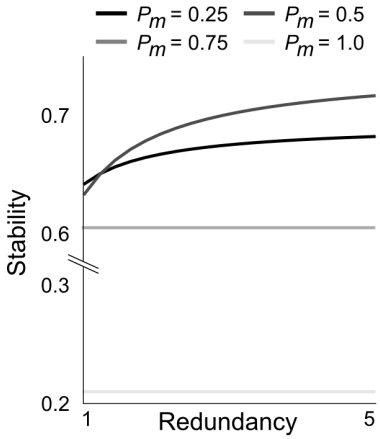

Figure 3.17: **Redundancy in cooperation can increase ecological stability.** We use linear stability analysis to examine the effect of replacing strong cooperative links with multiple weaker cooperative interactions. For example, one interaction of strength 1 may be replaced with three interactions, each of strength $1/3$. Whilst this manipulation does not affect highly unstable, highly cooperative communities, it can increase the stability of communities with lower levels of cooperation. The cross in the two curves for $P_m = 0.25$ and $P_m = 0.5$ occurs because the former has fewer cooperative links to make redundant and so the effects making cooperative links highly redundancy is overall weaker. We examine a network with a low initial level of connectivity, $C = 0.2$, to allow a wide range of redundancy to be examined (here: $S = 100$, $-s = -1$, $\sigma = 0.05$).

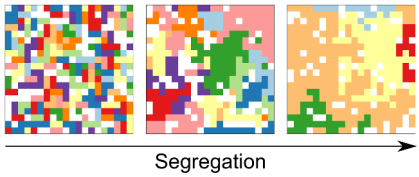

Figure 3.18: **Implementation of increasing spatial structure in the individual-based model.** To increase structure, we decrease the number of randomly arranged cells that start communities, which leads to the emergence of larger clonal patches. In these models, we also assume that each cell can only interact with those cells within a certain local neighborhood.

4

The assembly of microbiome communities

4.1 Abstract

Every generation, humans and many other host species must establish a diverse community of symbiotic microbes in order to function properly. Yet despite their importance, we currently lack a general understanding of what promotes, or inhibits, the assembly of these vital communities. Here we develop ecological theory to identify key factors underlying microbial community assembly. Specifically, we develop a new method for mapping out ecological landscapes that has strong analogies to the study of fitness landscapes in evolutionary biology. We show that diverse communities will most reliably assemble when their constituent species interact both weakly and competitively. By contrast, strong or cooperative interactions can create barriers to community assembly whereby many species must arrive simultaneously for a community to establish. Moreover, in these communities, the loss of one or a few species can lead to the system

collapsing down to a state of very low diversity. Communities dominated by cooperative and strong interactions, therefore, also tend to be less diverse on average. Our methods allow us to identify the most probable communities from any given set of microbiome species and open up the way for the rational design of communities from the bottom up.

4.2 Introduction

The human gastro-intestinal tract harbours a dense community of microbes that are vital for our own health, providing benefits ranging from protection from pathogens, through to priming our immune systems (Bäckhed et al., 2005; Dethlefsen and Relman, 2011; Lee and Mazmanian, 2010; Mazmanian et al., 2008). The initial acquisition of a healthy microbiome community is, therefore, a critical period in infant development (Bergström et al., 2014; Cox et al., 2014; Jost et al., 2012). Typically this assembly process is assumed to follow a relatively predictable pattern, with the infant gut progressing from initially uncolonised to possessing a stable, diverse microbiome that resembles that of an adult within the first years of life (Koenig et al., 2011; La Rosa et al., 2014; Lim et al., 2015). This development process is considered crucial for good health; a stable, diverse microbiome is thought to correlate to host wellbeing, and disruptions to this process contribute to a number of diseases, including necrotising enterocolitis – the second most common cause of death in premature infants (Decker et al., 2010; Neu and Walker, 2011; Zhou et al., 2015). However, despite its clear importance, we still understand little about the causative factors that enable a healthy microbiome community to initially assemble and persist.

A key barrier to establishing a general understanding of microbiome communities is their complexity. There is a large and rapidly growing body of empirical data on microbiome communities. However, on its own this data is poorly placed to disentan-

gle the drivers of microbiome assembly, because so many species and processes are involved that can potentially be conflated. There is a clear need then for complementary theory to identify general principles and testable hypotheses underlying the rules of microbiome community assembly. The field of theoretical ecology has a long history of using mathematical approaches in order to understand the ecological dynamics of complex macroscopic communities (Allesina and Tang, 2012; Campbell et al., 2011; Capitán and Cuesta, 2010; Law and Morton, 1996; May, 1972; Morton and Law, 1997; Warren et al., 2003). However, while these approaches have been widely applied to the study of plant and animal communities, they have received little attention in the context of the microbiome. Based on these methods, we recently developed ecological theory to understand the factors underlying the stability of established microbiome communities (Chapter 3, (Coyte et al., 2015)). This identified a number of general ecological properties that are predicted to influence the stability of microbiome communities, which we verified using published empirical data. However, our work was silent on how these communities assemble and establish in the first place.

Here we develop a new theoretical framework in order to investigate the key drivers underlying the robust assembly of diverse microbiome communities. Our framework captures the different stages of microbiome assembly as a series of ecological networks that can contain different numbers and combinations of species. We evaluate what happens when a species is added to, or removed from each one of these networks, in order to examine the potential pathways by which any state of the microbiome can move to a position of increased, or decreased diversity. With this, we ask how species properties affect the ability of a microbiome community to reliably assemble up to a diverse state as new species successively arrive. Our work predicts that diverse communities can most robustly assemble when the constituent species have strong within-species interactions, weak between-species interactions and when the interactions between species

are predominantly non-cooperative. These simple and testable predictions provide a basis to both understand and manipulate the assembly of microbiome communities.

4.3 Results and discussion

To explore the dynamics of microbiome assembly we develop a method to build complete *assembly landscapes* for communities (Figure 4.1), inspired by studies in theoretical ecology of plant and animal communities (Campbell et al., 2011; Capitán and Cuesta, 2010; Law and Morton, 1996; Morton and Law, 1997; Warren et al., 2003). We choose viable communities of a certain size then identify all the possible pathways by which they could have assembled. This is analogous then to observing a particular microbiome community in nature and trying to determine the characteristics of its original assembly. Specifically, we identify all the sub-communities of a given total species pool that are able to persist within the environment, then determine all the possible transitions between them that can occur through the addition or removal of a species – thus building the assembly landscape for the focal community (Figure 4.1, *Materials and methods* Section 1). We focus throughout on systems where species arrive or are lost one by one, but the same principles may be used to identify the possible transitions that would occur were two or more species to arrive at one time.

We capture the dynamics of each of the sub-communities using unstructured generalised Lotka-Volterra models – systems of ordinary differential equations that describe each species' growth within a complex community (*Materials and methods*, Section 2). This approach naturally captures biologically critical network motifs such as chains of cross-feeding metabolic exchanges between species (Louis et al., 2014), and, crucially, has been shown to well capture microbiome dynamics (Buffie et al., 2015; Marino et al., 2014; Stein et al., 2013). Current microbiome data do not indicate strong deviations from such unstructured communities, making them a good starting point for our anal-

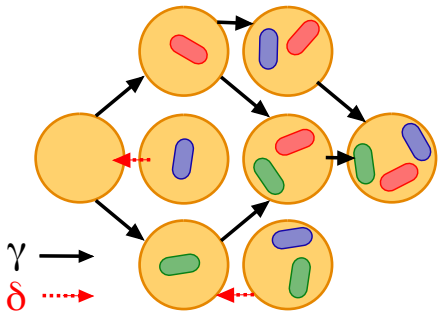

Figure 4.1: **Schematic of an assembly landscape.** The assembly landscape maps out all viable sub-communities of a given species pool, as well as all the possible transitions (via gain or loss of species) between them. Here orange circles represent individual sub-communities, while solid black lines indicate the ability of a community to transition from one state to another via the successful invasion of a new species. Dashed red lines indicate transitions due to the loss of species. By assigning rates to each of these transition events, given by γ and δ, (for example, based on the likelihood of a new species arriving) we can predict features such as the likely long-term state of the community.

ysis of community assembly. However, there remains the possibility to incorporate the effects of structured ecological networks should the data warrant it.

We embed our models of community dynamics in an assembly framework that analyses whether the addition, or removal, of a species to a given sub-community leads to a viable new community (*Materials and methods*, Sections 3 and 4). We define viable communities as those that are robust to species fluctuations, such that the community will not collapse if any one species is lost, and every constituent species is individually capable of invading when rare (Figure 4.2). We can then identify all of the viable sub-communities for a focal community. This method is built upon the logic of *permanence* in ecological theory (Jansen, 1987). Permanence analysis is a method that identifies extremely robust communities, which once assembled will retain all of their constituent species even in the face of incredibly strong perturbations. In our method, these extremely robust communities are identified but so too are communities that are still highly robust, but do not meet the stringent mathematical requirements of a tradi-

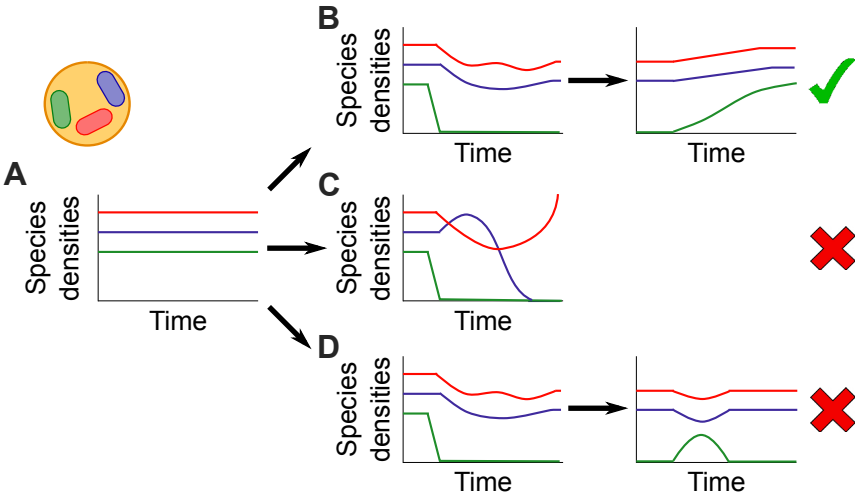

Figure 4.2: **Classification of viable communities.** For a given sub-community (**A**) to be viable it must meet the criteria that (**B, left**) if any of its members are removed, the remaining species can achieve new, non-negative equilibrium densities and (**B, right**) the missing species should be able to reinvade when rare. If the sub-community cannot achieve a new equilibrium when any one member is removed (**C**) or the removed species is unable to reinvade (**D**) then the sub-community will be classed as unviable.

tional permanence analysis. With this, we can ask how key ecological factors influence the number and nature of paths to the assembly of diverse communities.

Strongly interacting and highly cooperative communities are difficult to assemble, and prone to collapse

Systematically varying the ecological properties of species reveals clear effects on community assembly. We find that strong ecological interactions between species, cooperation, and the connectivity of the ecological network all reduce the probability that a focal community can reliably assemble (Figure 4.3). Even though all the communities we are studying are viable when all constituent species are present, we find that strong or positive ecological interactions reduce the probability that there is a continuous path through the assembly landscape. Put another way, we see an increase in the number of

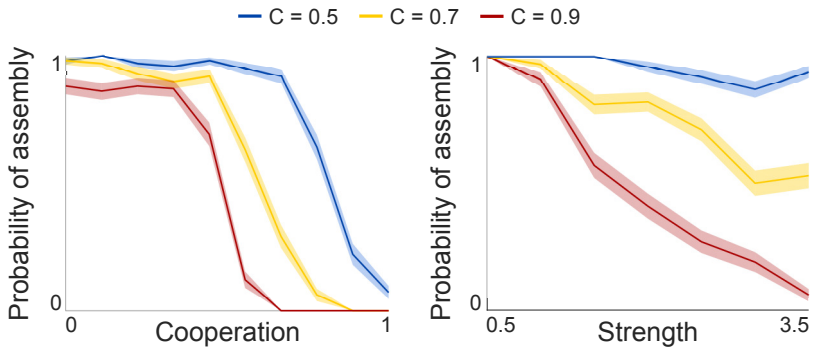

Figure 4.3: **Increasing cooperation or interaction strength reduces the probability that a community will be able to assemble via stochastic species arrival and loss alone.** Here we generate pools containing a total of S species and construct each of their assembly landscapes. *Probability of assembly* refers to the likelihood that a path exists from the initial uncolonised environment to the fully assembled community containing all S species, and we see that this decreases with increasing levels of cooperation or interaction strength. This trend is consistent for varying levels of connectivity. When varying cooperation (left) the remaining interactions within the species pool are distributed randomly, giving approximately $2/3$rds exploitative and $1/3$ competitive interactions. When varying interaction strength (right) then interaction types are distributed as $1/3$ cooperative, $2/3$ exploitative, and $1/3$ competitive. Here, $S = 7, -s = -0.1, \sigma = 0.05$, repeats $= 100$, errorbars $= SEM$.

communities with fractured assembly landscapes (Figure 4.4 A, B). In these cases, the diverse final community may be relatively resilient to small to medium perturbations – such that if one or two species were lost the community would be able to reassemble – but it cannot be reached from the initial uncolonised state.

What is the intuitive reason for the formation of fractured assembly landscapes? Previous work has shown that interaction strength, cooperation, and connectivity can all destabilise communities (Allesina and Tang, 2012; Coyte et al., 2015; May, 1972). The reason for this is that cooperation and strong interactions increase the coupling between species and the potential for positive feedbacks, whereby the loss of one species can drive the loss of others. In the context of community assembly, this effect takes on new significance, as co-dependence can mean that some species cannot invade a community

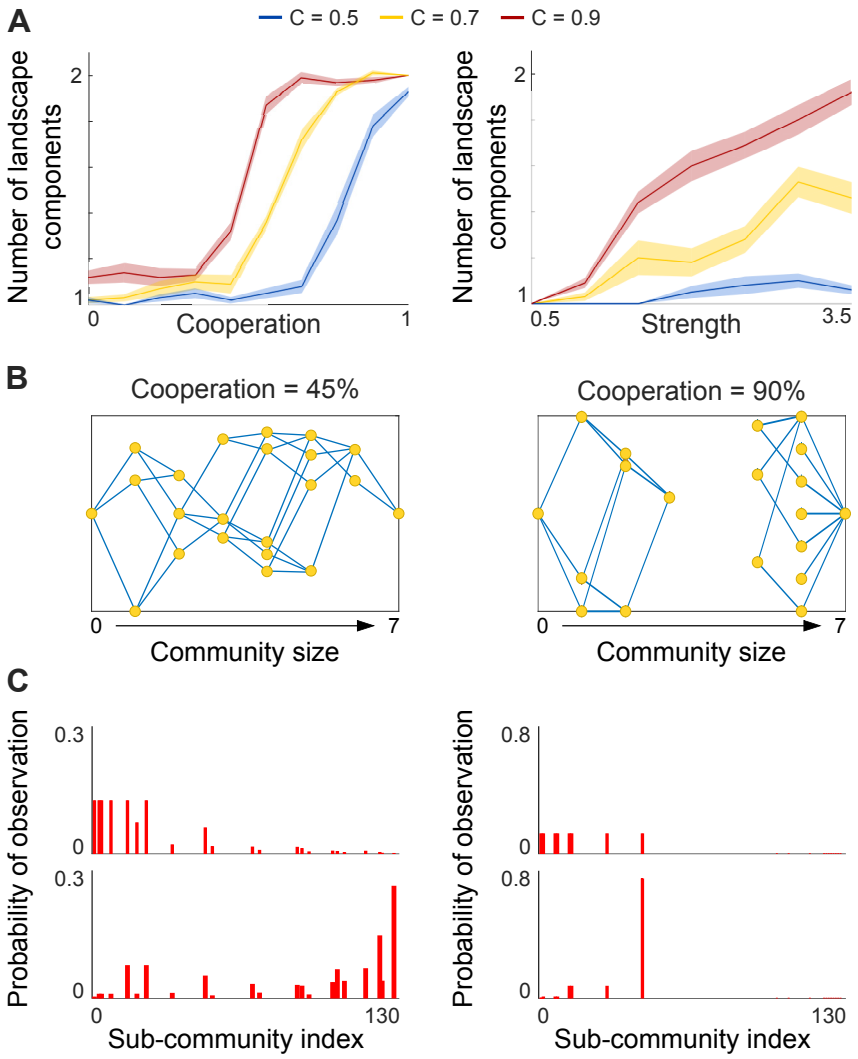

Figure 4.4: **Increasing cooperation or interaction strengths fractures the assembly landscape. (A)** When the total species pool contains little cooperation or weak interactions then the assembly landscape generally forms a single connected component, as illustrated in (**B, left**) (cooperation $= 45\%$, strength $= 1$). Here orange circles represent viable sub-communities and blue lines capture transitions between them due to the addition of new species (for clarity, transitions due to species loss are not shown). At higher levels of cooperation or interaction strengths the landscape breaks into multiple components, as illustrated in (**B, right**) (cooperation $= 90\%$ strength $= 1$). (**C**) Treating the assembly landscapes in (**B**) as Markov processes we can examine the long-term probability of observing different sub-communities. (**C, top**) when the species arrival rate equals that of species loss we are more likely to observe less diverse communities. (**C, bottom**) when the species arrival rate is much higher ($\times 10$) than that of loss then more diverse communities are more probable. However, even though highly diverse communities are viable for high levels of cooperation, these are unlikely to be observed in the long-term due to fractures in the assembly landscape (**B, C, right**). Here $S = 7, -s = -0.1, \sigma = 0.05$, repeats $= 100$, errorbars = SEM.

without the presence of certain other species that support their growth. Accordingly, even though a final community may be viable – with any one species able to bounce back from low numbers – this does not mean that the subsets of this community are similarly robust. A key implication of such fractured landscapes then is that even once assembled communities can be prone to collapse, whereby the loss of one or two species can drive a runaway effect where many further species are lost, and diversity drops precipitously. This phenomenon is analogous with the "Humpty-Dumpty" effect observed by Law and Morton when examining macroscopic food-web systems (Law and Morton, 1996). In this work species were strictly classified as producers or consumers and thus the community contained only exploitative ($+/-$) interactions (with no direct competition or cooperation) but similarly the authors observed certain communities that were stable once present, yet could not be reassembled from scratch.

Cooperative communities have a stereotyped assembly process

A well-known property of community assembly in many contexts, including the microbiome, is that species tend to establish in a relatively predictable order (Begon et al., 1996; Koenig et al., 2011; La Rosa et al., 2014; Palmer et al., 2007). We therefore next examined how community properties influenced the predictability of assembly within our landscape framework. The goal of this analysis has overlaps with the study of fitness landscapes in evolutionary biology, where viable paths of increasing fitness are sought through mutational space. There, a general property observed is that there are only a few pathways through a mutational landscape that increase fitness with each new mutation (Poelwijk et al., 2007; Weinreich et al., 2006) . Analogously, we observe that assembly landscapes are also often relatively sparse. That is, even though there may be several different paths through a landscape, these paths are a small subset of the total number of possible paths were all sub-communities viable.

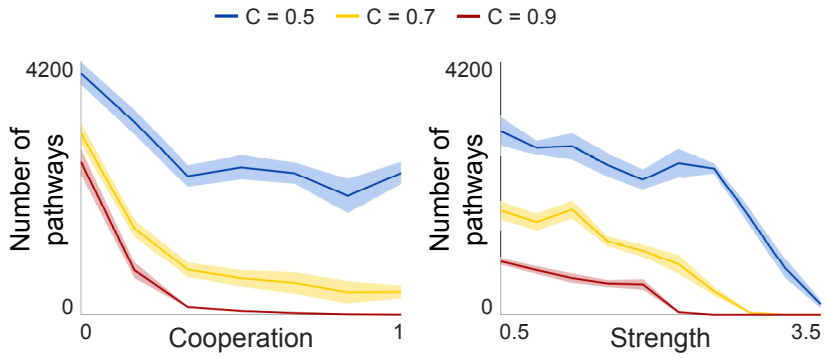

Figure 4.5: **Communities with strong or cooperative interactions have fewer viable pathways to full assembly.** Here we plot the number of feasible assembly pathways in landscapes for which there exists at least one pathway to the fully diverse state. A $S = 7$ species community will have a total of $7! = 5048$ possible assembly pathways, however, the unviability of some sub-communities will reduce the proportion of these pathways that are actually achievable. Increasing cooperation (left) or interaction strength (right) reduces the number of viable sub-communities, and thus the number of feasible assembly pathways. Here $S = 7, -s = -0.1, \sigma = 0.05$, repeats $= 100$, errorbars = SEM.

There is often a degree of predictability, therefore, in how communities assemble, because many of the potential steps in assembly do not lead to viable communities. We can then ask how the ecological properties of species affects this measure of predictability. Consistent with our earlier results, we observe that communities with strong or cooperative interactions tend to have fewer viable paths through them, such that assembly when it is possible is more predictable and stereotyped (Figure 4.5). A related signature of order effects within community assembly landscapes is the existence of species that are unable to invade initially, and require the presence of another species in order to colonise. Intuitively, such species are those that rely on strong positive ecological interactions from other species in order to be viable. As expected then, we observe that the number of secondary colonising species during assembly increases with both positive and strong interactions (Figure 4.6). However, this trend reverses at high levels of cooperation (Figure 4.6 A). Here the members of the community have become so

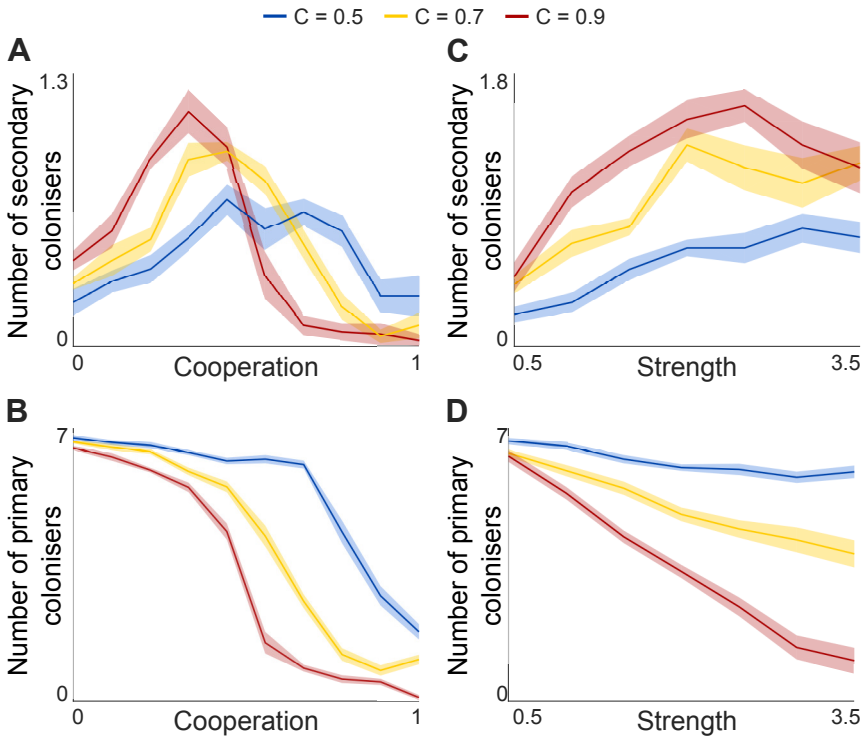

Figure 4.6: **Changing the level of cooperation or interaction strength exerts order effects upon the assembly landscape.** (**A**) Increasing cooperation initially increases observations of species that cannot invade until others have established themselves (termed "secondary colonisers"), due to increasing dependency between species. However, this trend reverses at high levels of cooperation. At this point members of the community have become so dependent upon one another than few will be capable of initial invasion at all ("primary colonisers"), reducing the probability of observing one species facilitating the other, and rendering the community unable to form via stochastic species arrival alone (**B**). (**C**) Increasing the strength of interspecies interactions also increases the probability of observing secondary colonisers, with this trend only reversing at high levels of connectivity. While increasing interaction strength also increases the potential level of dependency between species, only in the case of high connectivity does it do so enough to substantially reduce the number of primary colonisers, and prevent the entire community from assembling (**D**). Here $S = 7, -s = -0.1, \sigma = 0.05$, repeats $= 100$, errorbars = SEM.

dependent upon one another that few will be capable of initial invasion (Figure 4.6 B), reducing the ability of a diverse community to assemble at all, and thus reducing the probability of observing one species facilitating another. In sum, while cooperation and strong ecological interactions both generally inhibit the potential for community assembly, an important corollary of this is that when such communities can assemble, they will do so following relatively predictable sequences.

Community landscapes as dynamic and hysteretic systems

So far, we have only considered the structure of assembly landscapes. However, in reality, they represent dynamic systems where species can continually arrive or go extinct. In order to study these dynamics and understand their implications for community assembly, we next explicitly study the arrival and extinction of species (as a Markov process, *Materials and methods* Section 6). This approach makes it clear that communities can spend significant time at a wide range of possible diversities. Moreover, the highest diversity state need not be the most probable state in a community landscape, as species can be lost as well as gained. More formally, the probability of spending time at the highest diversity will depend critically on the rate of supply of new species. Increasing the rate at which new species arrive increases the chance of observing the most diverse communities (Figure 4.7). By contrast, increasing cooperation or interaction strengths decreases the probability of observing larger communities (Figure 4.7). This is expected, as these communities often have fractured landscapes that lack a continuous path from the initial state to a fully assembled community (Figure 4.4). Such communities will spend more time at a low diversity, therefore, than comparable communities with weaker or more competitive interactions.

Our Markovian analyses also reveals more nuanced effects, such as the potential for non-fractured landscapes to get stuck in sub-communities of low diversity (Figure 4.8 A). As just discussed, the rate at which species arrive and go extinct is central to the

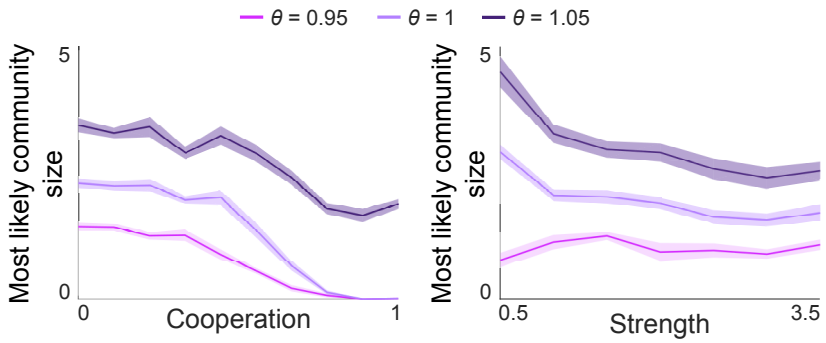

Figure 4.7: **Changing species migration rates or interspecies cooperation influences the likely state of the community.** Increasing the relative rate at which new species arrive ($theta = \gamma/\delta$) increases the probability of observing larger communities when the system is allowed to assemble over time, even when the bias towards arrival is small. Increasing the level of cooperation decreases the probability of observing large communities, as does increasing the strength of interactions (though here the effect is weaker). Here $S = 7$, $-s = -0.1$, $\sigma = 0.05$, repeats $= 100$ $C = 0.7$, errorbars $= SEM$.

average diversity of a given community landscape (Figure 4.7). However, even when the rate of arrivals far outstrips extinctions we observe landscapes in which there are multiple low-diversity community states, which once assembled cannot be changed, no matter how frequently new species arrive (Figure 4.8 B). In these situations, even though there is a clear pathway to peak diversity, the particular route that is taken through the assembly landscape will determine whether or not the community can attain this fully diverse state. Such hysteretic effects again have interesting analogies in the study of evolutionary landscapes. There, a long-standing question has been whether such local peaks exist and, if so, how often systems are stuck at them. Similarly, in ecology, the potential for hysteresis has been noted in the context of aquatic microcosms, as well as in simpler two-species microbial communities (Bucci et al., 2012; Drake, 1991). And, though we observe this phenomenon rarely, we most often see such locally absorbing, low diversity states for intermediate levels of cooperation and relatively strong interspecies interactions (Figure 4.8 C, D).

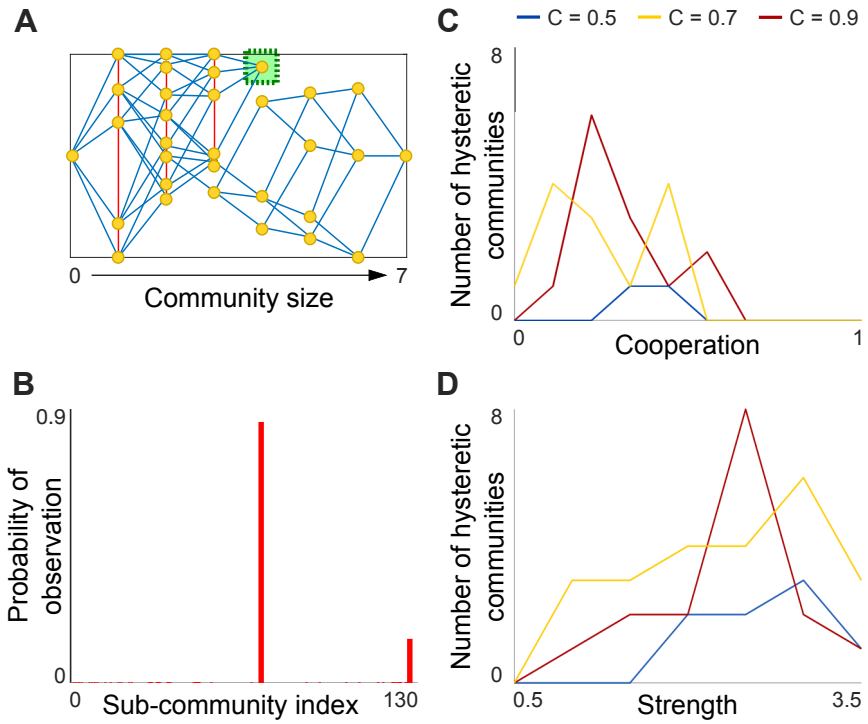

Figure 4.8: **Hysteresis in community assembly.** (**A**) An assembly landscape with multiple potentially absorbing communities. Orange circles represent sub-communities, blue lines indicate transitions between them due to the addition of new species, while red lines capture cases where a new species replaces an existing one. Though there are feasible pathways to the assembly of the full community, this system can get stuck at a state of lower diversity (dashed green box). (**B**) A Markovian analysis of this landscape indicates that this lower diversity sub-community is more probable, even when the species arrival rate is much greater than the rate of loss (here cooperation $= 55\%$, strength $= 1$, $\theta = 100$). (**C, D**) Such examples of multiple absorbing states (hysteretic communities) are rare, but occur most often in communities with intermediate levels of cooperation, or relatively high interspecies interaction strengths. Here $S = 7, -s = -0.1, \sigma = 0.05$, repeats $= 100$

Do these attracting low-diversity states occur in natural microbiome communities? Currently, there are few available data on ecological interactions within the microbiome. However, one key study to date has attempted to map the ecology of microbiome communities to generalised Lotka-Volterra models (Stein et al., 2013). While this dataset contains 11 potential phylogenetic groups, our analysis predicts that the largest com-

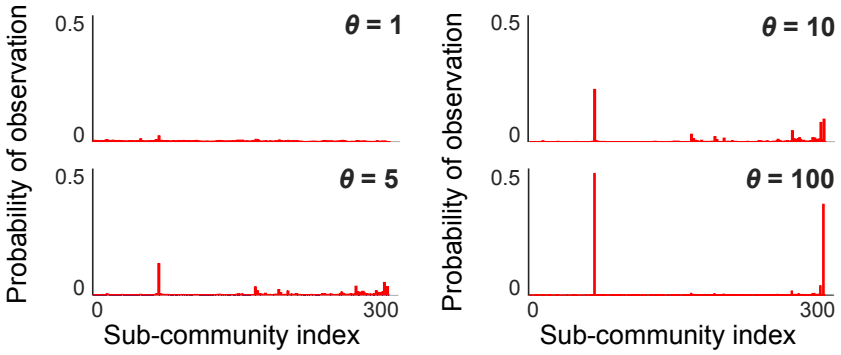

Figure 4.9: **Identifying key sub-communities in real microbiome data.** Applying our analysis to real microbiome data (Stein et al., 2013) identifies a robust sub-community of three groups that is more likely to be observed than any others, even when new species arrive frequently and extinctions are rare (indicated by the largest red bar). However, increasing the rate of species arrival to loss (increasing $\theta = \gamma/\delta$) can none-the-less increase the probability of observing more diverse communities.

munity that can be assembled contains only 7 of these groups. In addition, applying a Markov process identifies a particularly robust combination of three groups that is an attractor in this system, even when new species arrive frequently and extinctions are rare (Figure 4.9). This includes two groups that cooperate with one another (*Barnesiella* and unclassified *Lachnospiraceae*), and a third mixed group from the analysis that exploits the first two ("*Other*"). Applying parameters from microbiome communities, therefore, again appears to warn of the potential for assembly landscapes to have attracting peaks of low diversity. Such low diversity states have the potential to be problematic should the host rely on a diverse community for correct microbiome function. However, crucially in these data this low-diversity community is not fully absorbing, meaning it can be altered provided new species arrive frequently enough. As such, our model also provides an intuitive solution to avoiding these low diversity states. Increasing the rate at which new species arrive can gradually increase the probability of the community achieving a state of maximal viable diversity (Figures 4.9, 4.4 C).

4.4 Conclusions

Here we have presented new ecological theory in order to investigate the factors governing the assembly of microbiome communities. Our work emphasizes how the ecological properties of species, and in particular their interactions, are central to their ability to robustly form a diverse community. Strong interactions between species, and positive interactions between species, can both form barriers to community assembly. Such interactions lead to reduced and fractured assembly landscapes that can get stuck at states of low diversity. However, when cooperative and strongly interacting communities can assemble, they do so much more predictably, because of the relatively few viable routes by which assembly can proceed. Applying our methods to data from the mouse microbiome highlights how our theory can be used to predict which species, or groups of species, can robustly coexist. Such considerations will become increasingly important as we seek to rationally engineer microbiome communities by introducing species that perform a missing function for a host (Aroniadis and Brandt, 2013; Preidis and Versalovic, 2009). More generally, our work emphasizes the need to increase our understanding of the ecological interactions within microbiome communities. Only by taking into account the ecology of the microbiome in this way, both theoretically and experimentally, will we be able to understand and ultimately manipulate the microbiome.

4.5 Materials and methods

1. An introduction to assembly landscapes

As outlined in the main text, this work focuses on determining the *assembly landscapes* for given communities. These structures first calculate all the possible sub-communities of a focal species pool that are capable of persisting within the environment (termed *viable sub-communities*), and then determine all the possible transitions between them.

As such, they capture all possible developmental pathways of a habitat from initially uncolonised, through various sub-communities, to a final state in which no further species can be added (Figure 4.1). Specifically, we assume that there is a limited pool, T, of possible species [1], and that new species arrive rarely enough that we only need to consider the arrival of species one at a time. With these assumptions in place the problem collapses down into establishing which of the possible sub-communities of the species pool are viable, and determining the transitions between these communities based on the gain or loss of individual species. This is an approach that has been used a number of times in theoretical ecology in order to investigate specific community types, thus here we adapted existing methodologies in order to better suit the microbiome (Campbell et al., 2011; Capitán and Cuesta, 2010; Law and Morton, 1996; Morton and Law, 1997; Warren et al., 2003).

Below we outline the details of each of the core components of our assembly framework in turn. First we describe the underlying model used to capture community dynamics, then the manner in which we define viable sub-communities. Next we discuss the algorithm for building the assembly landscape itself, and finally detail the way in which these assembly landscapes can be analysed.

2. Underlying community models

In order to properly capture the dynamics of microbiome assembly we require an underlying community dynamics model that is not constrained by strict assumptions about community structure. To achieve this we capture microbiome community dynamics using the generalised Lotka-Volterra equations introduced in Chapter 3. Under this model the rate of change of species i is given by,

[1] An alternative here is to make no initial assumption about the total species pool, but rather repeatedly draw new species from some pre-determined distribution. While this has the advantage of not limiting the analysis ahead of time, this approach only allows us to analyse a single developmental pathway, and it may therefore miss rare but important aspects of community assembly.

$$\frac{dX_i}{dt} = X_i(r_i - s_i X_i + \sum_{j=1}^{S}(a_{ij}X_j)), \tag{4.1}$$

where r_i represents the intrinsic species growth rate, s_i captures the effect of a species upon itself (termed the self-regulation), and a_{ij} the effect of species j upon species i. These inter-species interactions (a) can take one of five possible forms based on the signs of a_{ij}/a_{ji}; $+/-$ (exploitative), $-/-$ (competitive), $+/+$ (cooperative), $+/0$, and $-/0$, and crucially, unlike previous work on community assembly, we can manipulate the proportions of each of these forms in order to fully examine how specific interaction types affect community assembly. The magnitude of these interspecies-interactions are drawn from a half-normal distribution, with standard deviation σ.

Importantly, we also modify the canonical generalised Lotka-Volterra model with the addition of a cap on growth once individual species' populations reach a certain size. This is to prevent unrealistic infinite growth, which can otherwise arise in cooperative Lotka-Volterra models. Specifically, we assume that each species, i, will grow according to Lotka-Volterra growth dynamics unless it exceeds a threshold population size, K^i, after which its population cannot increase any further. Moreover, we also assume that the community as a whole has a maximum population size, K^T, for example, analogous with there being a limited amount of space within the environment. Importantly, these caps not only prevent infinite growth, but also introduce the potential for second order competition, similar to that observed in the individual based models used in Chapter 3 (see Figure 3.4 C). Specifically, in the event that the total community cap is lower than the sum of the individual caps, that is, $K^T < \sum(K^i)$ then at high population sizes species will begin to compete for the limited space within the environment, even when their intrinsic interactions are classified as cooperative (i.e. a_{ij} and a_{ji} are both positive).

The ratio of these two caps – and the subsequent introduction of second-order competition – is likely to have an effect on community assembly dynamics. In particular,

if communities are fully saturated this may reduce the ability of new species to invade, stifling the assembly process. For this initial investigation we therefore set K^I and K^T so as to prevent this second order competition, however, fully investigating the role of such behaviour will be an important component to understanding community assembly, thus we highlight this as a key next direction for this work.

3. Determining community feasibility

We then determine whether or not communities are viable based on their robustness to species loss (Figure 4.2). Specifically, for a community to be considered viable and thus included within the assembly landscape it must fulfil the criteria that if any of its resident species are removed then,

1. The remaining community must still be able to persist – i.e. no further species should immediately go extinct.

2. The removed species would be capable of reinvading if reintroduced (i.e. the removed species has a positive growth rate when reintroduced at a low density).

The aim here is to include communities within the assembly landscape that are relatively robust to small scale perturbations (affecting just one species) without enforcing stringent requirements on the entirety of the state space. This measure can thus intuitively be thought of as a weaker form of the permanence concept described in Chapter 3 (see Chapter 3 *Materials and methods* Section 2), which formally checks whether the boundary of state space is always a repeller (and thus once assembled no species will ever go extinct). In practice, we find that communities classified as viable under our framework are typically also classified as permanent. However, our less stringent requirements enable us to examine larger communities with higher levels of cooperation, which we find in simulations are capable of persisting over long periods of time but would be discounted when using traditional permanence as a metric.

4. Constructing the assembly landscape

A species pool of size S will have 2^S possible sub-communities, including the uncolonised state. Therefore, the assembly landscape can be described by a single ($2^S \times 2^S$) matrix, Q, in which each entry Q_{ij} represents the ability of sub-community i to transition to sub-community j. In order to construct the assembly landscape, Q, for a given species pool we go through the following process, using the total species pool $T = [ABCD]$ as an example. First, to calculate the transitions within the assembly landscape due to the arrival of new species:

1. Calculate which of the sub-communities of T are viable (see above, "3. Determining community feasibility").

2. For each viable sub-community, identify each of the sub-communities that could be reached via the arrival of a single new species – termed the augmented communities. For example, for sub-community $i = [AB]$ these would be $j = [ABC]$ and $k = [ABD]$. For those augmented communities that are also viable, add a connection in the assembly landscape, i.e., if j is also viable, set $Q_{ij} = 1$.

3. For those augmented communities that are not viable, calculate the growth rate of the new species when introduced to the initial community at a low density, i.e., calculate whether this species would be capable of invading. For example, in our case, if community $k = [ABD]$ was not permanent, check the growth rate of $[D]$ when introduced to $[AB]$. If this growth rate is negative then set $Q_{ii} = Q_{ii+1}$, if not:

4. Check which of the sub-communities of the augmented community are viable, this will include the initial community but may also contain others. For example, for $k = [ABD]$ both $i = [AB]$ and $l = [AD]$ may be viable. As such, the invasion of species D could either be transient such that the community eventually returns

to state $[AB]$ or it could trigger the replacement of species B with species D. As such, set $Q_{il} = 1$ and $Q_{ii} = Q_{ii+1}$.

5. Repeat this process for each of the viable sub-communities of T.

Next, we wish to determine the transitions within the assembly landscape that are a consequence of stochastic extinction events (i.e. the random loss of a single species) within a community.

1. For each viable state, determine the sub-communities that would result from the loss of a single species, termed the reduced communities. For example, for $j = [ABC]$ the reduced communities would be $i = [AB]$ $m = [AC]$ and $n = [BC]$. For those reduced communities that are viable, add a link in the assembly landscape. For example, if i and m are viable, set $Q_{ji} = 1$ and $Q_{jm} = 1$.

2. If any of these reduced communities are themselves not viable, check to see which of their own reduced communities are viable, iterating downwards until you find a viable community. For example, if $n = [BC]$ were not viable, we would check the viability of $p = [B]$ and $q = [C]$. If these two reduced communities were viable, we would set $Q_{jp} = 1$ and $Q_{jq} = 1$, if not, we would check the next level downward (in this case, the empty set). This corresponds to a community randomly losing one species, and this event triggering further extinctions.

5. Which communities to examine?

In our main text analysis we focus on the assembly of communities that are viable once assembled, and whose dynamics are governed by their own interspecies interactions. However, what happens if we instead focus on communities that are chosen entirely at random, with no requirements on the interactions between species or viability of the final community? Specifically, we can generate species pools in which the interaction

parameters and intrinsic growth rates are each drawn from a normal distribution, scaled to be of a similar magnitude to those in the main analysis. Applying our community assembly framework to these species pools we see domination by species that are rarely competitive, and are no longer at all dependent on one another. Members of these communities will continue to grow until they hit the external cap on their population size, and the full assembly landscape will typically be almost fully connected – meaning there is very little required order to the invasion of new species. Moreover, this independence means that the loss of one species from the species pool would have little if any effect on the assembly of the remaining community members, and we do not see the 'Secondary coloniser' species types described in the main text. This system thus represents the opposite scenario to that examined in our main text – a world in which community assembly is relatively independent of the interactions between members, and is therefore relatively unpredictable.

6. The successional landscape as a transition matrix

To predict the long-term assembly behaviour of given communities, we treat the assembly landscape as a Markov chain process in which transitions between states are determined by the gain or loss of individual species. Specifically, we assume that new species arrive at a rate γ per arbitrary time unit, and that species are lost from the community due to random extinctions at a rate δ. Moreover, we assume that each of these gain or loss events occur rarely enough that only one such event will ever occur during any given time-step. We can then use our assembly landscape, Q, to generate a transition matrix that describes the rate at which each sub-community can transition to each other sub-community. Each entry in the transition matrix can take one of four possible values:

1. Entries that correspond to the simple addition of a single species to an existing community are set at γ/S – the rate at which the new species in question arrives.

2. Entries corresponding to cases where the addition of a new species can have b possible outcomes (for example where a new species either invades transiently and then disappears, or replaces another), are set to $\gamma/(b \times S)$ – as above, but now assuming that each of the b possible events occurs with equal probability.

3. Entries that correspond to the simple loss of a single species are set to δ/S – the rate at which the species in question goes extinct (note, here the total extinction rate δ corresponds to the rate at which any of the species in the total pool go extinct, not the rate of loss of species within the current community).

4. Entries corresponding to cases where the loss of a single species can have b possible outcomes (i.e. where the loss of a species triggers the potential loss of multiple others) are set to $\delta/(b \times S)$ - as above, but now assuming that each of the b possible events occurs with equal probability.

We can then use this matrix to find the stationary distribution of the assembly process – and therefore predict the most likely states in which to observe the community. Setting $\delta \approx 0$ allows us to determine any absorbing states for the community.

5 Conclusions

My thesis explores microbial communities in two relatively disparate settings: in porous media like soil, and within the human gut. Each chapter has its own detailed discussion, thus here I take the opportunity to briefly summarise the major conclusions before discussing the general observations and future directions for this work.

In Chapter 2, I combined novel microfluidics experiments with mechanistic models and game theory to study how the interplay between microbes and their environment (in this case, via the influence of fluid flow) affects microbial competition and, ultimately, evolution. This work is intended as a proof-of-principle to show how subtle changes in an environment can fundamentally affect community ecology and evolution. Specifically, I demonstrated how a slight change in the geometry of an environment could enable slow growing genotypes, which would typically be outcompeted in simple settings, to supplant their faster growing counterparts. Microbiology to date has focused predominantly on communities within simple environments; however, this work demonstrates how crucial it can be to take into account environmental complexity when

thinking about microbial community dynamics.

My analysis is inevitably a simplification of the realities of porous media, in which microbes interact in complex three-dimensional geometries, likely via a variety of different mechanisms. A natural extension is to compare experimentally the outcomes of genotypic competition within yet more complex porous environments. To date, our ability to interrogate microbial communities within such settings has been limited primarily by technical restrictions; it is harder to achieve the high-throughput, repeatable experimentation possible in simple laboratory systems once considering environments that are both dynamic and spatially heterogeneous. It will be interesting then to see how the continuing improvements in microfluidics will change this, and thus provide novel insights into the role of environmental complexity in microbial ecology and evolution (Ho et al., 2015; Sackmann et al., 2014).

In Chapters 3 and 4, I changed setting, moving to consider the dense microbial communities within the mammalian gastro-intestinal tract. In both chapters I focused on the role of interspecies interactions – and in particular cooperation between species – in governing microbiome ecology. In Chapter 3, I developed a new body of ecological theory in order to investigate what makes microbiome communities so surprisingly stable. Contrary to suggestions in the literature, this work revealed that cooperation between species is likely to have a destabilising effect upon communities as a whole. Moreover, this work reinforced the intuition that hosts should benefit from any mechanisms that can keep their microbial communities in check. Though simple, this work has demonstrated the important ways in which interspecies interactions (both microbe-microbe and host-microbe) can influence microbiome dynamics. My hope is that this will help to move empiricists away from simply cataloguing the members of microbiome communities to more explicitly determining the ecology of these systems.

In Chapter 4, I extended this modelling framework yet further to investigate the dynamics of initial community assembly within the gut microbiome. Here I demonstrated

the fundamental ways in which ecological interactions can influence community assembly, and in particular, impact the predictability of assembly processes. Specifically, I illustrated that cooperation between species can render communities difficult to assemble, and enforce strict order effects on the community assembly. This work also lays out a number of testable predictions for empirical work, and a key next step is to integrate this modelling with relevant empirical data to allow empirical tests. More generally, I hope that Chapters 3 and 4 illustrate the advances that can be made through the application of ecological theory to our microbiome communities.

This thesis has presented one of the first theoretical models of the microbiome; however, the application of mathematical theory to the microbiome is still very much in its infancy, and there are a number of avenues yet to be explored. These fall broadly into two categories: improving the underlying mathematical framework, particularly to incorporate further aspects of biology, and integrating these models with relevant biological data.

The focus of my work has been on constructing models that are detailed enough to capture the relevant aspects of biology, but still simple enough to be mathematically tractable. In the case of the microbiome this has been aided greatly through the use of Lotka-Volterra models, which are amenable to mathematical analysis, but crucially can still capture microbiome behaviour. There is value then in using simple models as a first step to understand complex systems. Nevertheless, there remains a broad scope for considering more complex models in order to incorporate particular details of microbiome biology, or indeed the biology of other microbial communities. This applies both to the study of community stability and community assembly processes. In particular, incorporating host biology – for example, aspects of the host immune system – into our assembly framework is a key priority for the immediate future. Other important avenues for future work include allowing plasticity in how species interact (for example, explicitly enabling species to switch between interaction types depending on

their environmental context) a phenomenon that is predicted to play an important role in community ecology and evolution (Agrawal, 2001). And, similarly, more explicitly taking into account how microbial communities respond to antibiotic perturbations or dietary changes – two key components of host lifestyles that are already known to significantly impact microbiome communities (Bucci et al., 2012; Turnbaugh et al., 2009). Finally, in the longer-term a key goal will be to embed this ecological modelling framework within one that can be used to study the evolutionary dynamics of microbiome communities.

It is also crucial that these theoretical investigations are further integrated with experimental data. While there is a vast body of data concerning the composition of the microbiome, there is relatively little information available about underlying community dynamics – particularly concerning the nature of the interactions within these communities. There has been a recent rise in studies attempting to disentangle these interactions (Buffie et al., 2015; Marino et al., 2014; Stein et al., 2013), and the results stemming from this work will provide valuable information for theoreticians, both for informing modeling decisions and to ultimately test hypotheses. More fundamentally however, the most progress will be made when experiments and theory are tightly linked from the start. Accordingly, my plan is to now work closely with empiricists interested in neonatal microbiome community assembly. My aim is to parameterize the assembly models presented in Chapter 4 using combined data from both neonatal and adult microbiome communities, in order to try and predict the neonatal assembly process and identify key members, or interspecies interactions, involved in microbiome assembly. We hope to then test these predictions using community reconstruction experiments *in vitro* and ultimately *in vivo*, with the results then feeding back to further improve the modelling framework. This is clearly an ambitious goal, but one that I believe is essential if we wish to understand, and ultimately manipulate microbial communities.

Bibliography

Abbas, F. (2011), 'Mathematical Contributions to One-dimensional Biofilm Modeling, PhD Thesis', *The University of Guelph* .

Abbas, F., Sudarsan, R. and Eberl, H. J. (2012), 'Longtime behavior of one-dimensional biofilm models with shear dependent detachment rates', *Mathematical Biosciences and Engineering* **9**(2), 215–239.

Adams, J. L. and McLean, R. J. (1999), 'Impact of rpoS deletion on Escherichia coli biofilms.', *Applied and environmental microbiology* **65**(9), 4285–7.

Agrawal, A. A. (2001), 'Phenotypic plasticity in the interactions and evolution of species.', *Science* **294**(5541), 321–6.

Allen, E. E. and Banfield, J. F. (2005), 'Community genomics in microbial ecology and evolution.', *Nature reviews. Microbiology* **3**(6), 489–98.

Allesina, S. and Tang, S. (2012), 'Stability criteria for complex ecosystems.', *Nature* **483**(7388), 205–8.

Aroniadis, O. C. and Brandt, L. J. (2013), 'Fecal microbiota transplantation: past, present and future', *Current opinion in gastroenterology* **29**(1), 79–84.

Azam, F. and Malfatti, F. (2007), 'Microbial structuring of marine ecosystems.', *Nature reviews. Microbiology* **5**(10), 782–91.

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. and Gordon, J. I. (2005), 'Host-bacterial mutualism in the human intestine.', *Science* **307**(5717), 1915–20.

Begon, M., Townsend, C. R. and Harper, J. D. (1996), *Ecology: individuals, populations and communities*, Oxford: Blackwell Science, ISBN 0-632-03801-2.

Belenguer, A., Duncan, S. H., Calder, A. G., Holtrop, G., Louis, P., Lobley, G. E. and Flint, H. J. (2006), 'Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut.', *Applied and environmental microbiology* **72**(5), 3593–9.

Bergström, A., Skov, T. H., Bahl, M. I., Roager, H. M., Christensen, L. B., Ejlerskov, K. T., Mølgaard, C., Michaelsen, K. F. and Licht, T. R. (2014), 'Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of Danish infants.', *Applied and environmental microbiology* **80**(9), 2889–900.

Bever, J. D., Richardson, S. C., Lawrence, B. M., Holmes, J. and Watson, M. (2009), 'Preferential allocation to beneficial symbiont with spatial structure maintains mycorrhizal mutualism.', *Ecology letters* **12**(1), 13–21.

Biggar, J. W. and Nielsen, D. R. (1976), 'Spatial variability of the leaching characteristics of a field soil', *Water Resources Research* **12**(1), 78.

Borenstein, E. (2012), 'Computational systems biology and in silico modeling of the human microbiome.', *Briefings in bioinformatics* **13**(6), 769–80.

Brännström, Å., Johansson, J. and von Festenberg, N. (2013), 'The Hitchhiker's Guide to Adaptive Dynamics', *Games* **4**(3), 304–328.

Brovelli, A., Malaguerra, F. and Barry, D. A. (2009), 'Bioclogging in porous media: Model development and sensitivity to initial conditions', *Environmental Modelling & Software* **24**(5), 611–626.

Bucci, V., Bradde, S., Biroli, G. and Xavier, J. B. (2012), 'Social Interaction, Noise and Antibiotic-Mediated Switches in the Intestinal Microbiota', *PLoS Comput Biol* **8**(4), e1002497.

Buffie, C. G., Bucci, V., Stein, R. R., McKenney, P. T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., Littmann, E., van den Brink, M. R. M., Jenq, R. R., Taur, Y., Sander, C., Cross, J. R., Toussaint, N. C., Xavier, J. B. and Pamer, E. G. (2015), 'Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile', *Nature* **517**(7533), 205–208.

Campbell, C., Yang, S., Albert, R. and Shea, K. (2011), 'A network model for plant-pollinator community assembly.', *Proceedings of the National Academy of Sciences of the United States of America* **108**(1), 197–202.

Capitán, J. A. and Cuesta, J. A. (2010), 'Catastrophic regime shifts in model ecological communities are true phase transitions', *Journal of Statistical Mechanics: Theory and Experiment* **2010**(10), P10003.

Chen, X. and Cohen, J. E. (2001), 'Global stability, local stability and permanence in model food webs.', *Journal of Theoretical Biology* **212**(2), 223–235.

Cho, S., Fujii, N., Lee, T. and Okabe, S. (2011), 'Development of a simultaneous partial nitrification and anaerobic ammonia oxidation process in a single reactor.', *Bioresource technology* **102**(2), 652–9.

Clemente, J. C., Pehrsson, E. C., Blaser, M. J., Sandhu, K., Gao, Z., Wang, B., Magris, M., Hidalgo, G., Contreras, M., Noya-Alarcon, O., Lander, O., McDonald, J., Cox, M., Walter, J., Oh, P. L., Ruiz, J. F., Rodriguez, S., Shen, N., Song, S. J., Metcalf, J., Knight, R., Dantas, G. and Dominguez-Bello, M. G. (2015), 'The microbiome of uncontacted Amerindians', *Science Advances* **1**(3), e1500183–e1500183.

Clemente, J. C., Ursell, L. K., Parfrey, L. W. and Knight, R. (2012), 'The impact of the gut microbiota on human health: an integrative view.', *Cell* **148**(6), 1258–70.

Cornforth, D. M. and Foster, K. R. (2013), 'Competition sensing: the social side of bacterial stress responses.', *Nat Rev Microbiol* **11**(4), 285–93.

Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R. and Lappin-Scott, H. M. (1995), 'Microbial biofilms.', *Annual Review of Microbiology* **49**, 711–45.

Cox, L. M., Yamanishi, S., Sohn, J., Alekseyenko, A. V., Leung, J. M., Cho, I., Kim, S. G., Li, H., Gao, Z., Mahana, D., Zárate Rodriguez, J. G., Rogers, A. B., Robine, N., Loke, P. and Blaser, M. J. (2014), 'Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences.', *Cell* **158**(4), 705–21.

Coyte, K. Z., Schluter, J. and Foster, K. R. (2015), 'The ecology of the microbiome: Networks, competition, and stability', *Science* **350**(6261), 663–666.

Cunningham, A. B., Sharp, R. R., Hiebert, R. and James, G. (2003), 'Subsurface Biofilm Barriers for the Containment and Remediation of Contaminated Groundwater', *Bioremediation Journal* **7**(3-4), 151–164.

Cunningham, A., Characklis, W., Abedeen, F. and Crawford, D. (1991), 'Influence of biofilm accumulation on porous media hydrodynamics', *Environment Science Technology* pp. 1305–1311.

Danhorn, T. and Fuqua, C. (2007), 'Biofilm formation by plant-associated bacteria.', *Annual review of microbiology* **61**, 401–422.

De Cruz, P., Prideaux, L., Wagner, J., Ng, S. C., McSweeney, C., Kirkwood, C., Morrison, M. and Kamm, M. A. (2012), 'Characterization of the gastrointestinal microbiota in health and inflammatory bowel disease.', *Inflamm Bowel Dis* **18**(2), 372–90.

De Muynck, W., De Belie, N. and Verstraete, W. (2010), 'Microbial carbonate precipitation in construction materials: A review', *Ecological Engineering* **36**(2), 118–136.

Decker, E., Engelmann, G., Findeisen, A., Gerner, P., Laass, M., Ney, D., Posovszky, C., Hoy, L. and Hornef, M. W. (2010), 'Cesarean Delivery Is Associated With Celiac Disease but Not Inflammatory Bowel Disease in Children', *Pediatrics* **125**(6), e1433–e1440.

Derrien, M., van Passel, M. W., van de Bovenkamp, J. H., Schipper, R. G., de Vos, W. M. and Dekker, J. (2010), 'Mucin-bacterial interactions in the human oral cavity and digestive tract.', *Gut Microbes* **1**(4), 254–268.

Dethlefsen, L. and Relman, D. A. (2011), 'Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation.', *Proceedings of the National Academy of Sciences of the United States of America* **108 Suppl**, 4554–61.

Di Cagno, R., De Angelis, M., De Pasquale, I., Ndagijimana, M., Vernocchi, P., Ricciuti, P., Gagliardi, F., Laghi, L., Crecchio, C., Guerzoni, M. E., Gobbetti, M. and Francavilla, R. (2011), 'Duodenal and faecal microbiota of celiac children: molecular, phenotype and metabolome characterization.', *BMC microbiology* **11**(1), 219.

Donlan, R. M. (2002), 'Biofilms: microbial life on surfaces.', *Emerging infectious diseases* **8**(9), 881–90.

Drake, J. A. (1991), 'Community-assembly mechanics and the structure of an experimental species ensemble', *American Naturalist* **137**(1), 1–26.

Drenkard, E. and Ausubel, F. M. (2002), 'Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation.', *Nature* **416**(6882), 740–3.

Drescher, K., Shen, Y., Bassler, B. L. and Stone, H. A. (2013), 'Biofilm streamers cause catastrophic disruption of flow with consequences for environmental and medical systems.', *Proceedings of the National Academy of Sciences of the United States of America* **110**(11), 4345–50.

Duddu, R., Chopp, D. L. and Moran, B. (2009), 'A two-dimensional continuum model of biofilm growth incorporating fluid flow and shear stress based detachment.', *Biotechnology and bioengineering* **103**(1), 92–104.

Dupin, H. J., Kitanidis, P. K. and McCarty, P. L. (2001), 'Pore-scale modeling of biological clogging due to aggregate expansion: A material mechanics approach', *Water Resources Research* **37**(12), 2965.

Eberl, G. (2010), 'A new vision of immunity: homeostasis of the superorganism.', *Mucosal Immunol* **3**(5), 450–60.

Ebigbo, A., Helmig, R., Cunningham, A. B., Class, H. and Gerlach, R. (2010), 'Modelling biofilm growth in the presence of carbon dioxide and water flow in the subsurface', *Advances in Water Resources* **33**(7), 762–781.

Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., Clemente, J. C., Knight, R., Heath, A. C., Leibel, R. L., Rosenbaum, M. and Gordon, J. I. (2013), 'The long-term stability of the human gut microbiota.', *Science* **341**(6141), 1237439.

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

Fiegna, F. and Velicer, G. J. (2005), 'Exploitative and hierarchical antagonism in a cooperative bacterium.', *PLoS biology* **3**(11), e370.

Foster, K. R. and Bell, T. (2012), 'Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species', *Current Biology* **22**(19), 1845–1850.

Franzosa, E. A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X. C. and Huttenhower, C. (2015), 'Sequencing and beyond: integrating molecular 'omics' for microbial community profiling', *Nature Reviews Microbiology* **13**(6), 360–372.

Fredrickson, J. K. (2015), 'Ecological communities by design', *Science* **348**(6242), 1425–1427.

Gerschgorin, S. (1931), 'Über die Abgrenzung der Eigenwerte einer Matrix', *B Acad Sci USSR* **25**(1900), 750–754.

Ghiorse, W. C. and Wilson, J. T. (1988), *Advances in Applied Microbiology Volume 33*, Vol. 33 of *Advances in Applied Microbiology*, Elsevier.

Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M. and Nelson, K. E. (2006), 'Metagenomic analysis of the human distal gut microbiome.', *Science* **312**(5778), 1355–9.

Giongo, A., Gano, K. A., Crabb, D. B., Mukherjee, N., Novelo, L. L., Casella, G., Drew, J. C., Ilonen, J., Knip, M., Hyöty, H., Veijola, R., Simell, T., Simell, O., Neu, J., Wasserfall, C. H., Schatz, D., Atkinson, M. A. and Triplett, E. W. (2011), 'Toward defining the autoimmune microbiome for type 1 diabetes.', *The ISME journal* **5**(1), 82–91.

Goh, B. S. (1979), 'Stability in Models of Mutualism', *The American Naturalist* **113**(2), 261.

Hall-Stoodley, L., Costerton, J. W. and Stoodley, P. (2004), 'Bacterial biofilms: from the natural environment to infectious diseases.', *Nature reviews. Microbiology* **2**(2), 95–108.

Hallatschek, O., Hersen, P., Ramanathan, S. and Nelson, D. R. (2007), 'Genetic drift at expanding frontiers promotes gene segregation.', *Proceedings of the National Academy of Sciences of the United States of America* **104**(50), 19926–30.

Haubert, K., Drier, T. and Beebe, D. (2006), 'PDMS bonding by means of a portable, low-cost corona system.', *Lab on a chip* **6**(12), 1548–9.

Haydon, D. (1994), 'Pivotal assumptions determining the relationship between stability and complexity: an analytical synthesis of the stability-complexity debate', *American Naturalist* **1**(144), 14–29.

Hibbing, M. E., Fuqua, C., Parsek, M. R. and Peterson, S. B. (2010), 'Bacterial competition: surviving and thriving in the microbial jungle.', *Nature reviews. Microbiology* **8**(1), 15–25.

Ho, C. M. B., Ng, S. H., Li, K. H. H. and Yoon, Y.-J. (2015), '3D printed microfluidics for biological applications', *Lab Chip* **15**(18), 3627–3637.

Hofbauer, J. and Sigmund, K. (1989), 'On the stabilizing effect of predators and competitors on ecological communities', *J Math Biol* **27**(5), 537–548.

Holling, C. S. (1973), 'Resilience and Stability of Ecological Systems', *Annual Review of Ecology and Systematics* **4**(1), 1–23.

Hooper, L. V. (2009), 'Do symbiotic bacteria subvert host immunity?', *Nat Rev Microbiol* **7**(5), 367–74.

Hooper, L. V., Xu, J., Falk, P. G., Midtvedt, T. and Gordon, J. I. (1999), 'A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem.', *P Natl Acad Sci USA* **96**(17), 9833–8.

Horn, H. and Lackner, S. (2014), 'Modeling of biofilm systems: a review.', *Advances in biochemical engineering/biotechnology* **146**, 53–76.

Horner-Devine, M. C., Carney, K. M. and Bohannan, B. J. M. (2004), 'An ecological perspective on bacterial biodiversity.', *Proceedings of the Royal Society B: Biological sciences* **271**(1535), 113–22.

Hsiao, E. Y., McBride, S. W., Hsien, S., Sharon, G., Hyde, E. R., McCue, T., Codelli, J. A., Chow, J., Reisman, S. E., Petrosino, J. F., Patterson, P. H. and Mazmanian, S. K. (2013), 'Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders.', *Cell* **155**(7), 1451–63.

Ito, A., May, T., Kawata, K. and Okabe, S. (2008), 'Significance of rpoS during maturation of Escherichia coli biofilms.', *Biotechnology and bioengineering* **99**(6), 1462–71.

Jansen, W. (1987), 'Mathematical Biology and Lotka-Volterra systems', *J Math Biol* **25**, 411–422.

Johansson, M. E. V., Phillipson, M., Petersson, J., Velcich, A., Holm, L. and Hansson, G. C. (2008), 'The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria.', *Proceedings of the National Academy of Sciences of the United States of America* **105**(39), 15064–9.

Jost, T., Lacroix, C., Braegger, C. P. and Chassard, C. (2012), 'New insights in gut microbiota establishment in healthy breast fed neonates.', *PloS one* **7**(8), e44595.

Kashyap, P. C., Marcobal, A., Ursell, L. K., Smits, S. A., Sonnenburg, E. D., Costello, E. K., Higginbottom, S. K., Domino, S. E., Holmes, S. P., Relman, D. A., Knight, R., Gordon, J. I. and Sonnenburg, J. L. (2013), 'Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota.',

Proceedings of the National Academy of Sciences of the United States of America **110**(42), 17059–64.

Keymer, J. E., Galajda, P., Lambert, G., Liao, D. and Austin, R. H. (2008), 'Computation of mutual fitness by competing bacteria.', *Proceedings of the National Academy of Sciences of the United States of America* **105**(51), 20269–73.

Kim, H. J., Boedicker, J. Q., Choi, J. W. and Ismagilov, R. F. (2008), 'Defined spatial structure stabilizes a synthetic multispecies bacterial community.', *Proceedings of the National Academy of Sciences of the United States of America* **105**(47), 18188–93.

Kim, J.-W., Choi, H. and Pachepsky, Y. a. (2010), 'Biofilm morphology as related to the porous media clogging.', *Water Research* **44**(4), 1193–201.

Kim, W., Racimo, F., Schluter, J., Levy, S. B. and Foster, K. R. (2014), 'Importance of positioning for microbial evolution.', *Proceedings of the National Academy of Sciences of the United States of America* **111**(16), E1639–47.

Klausen, M., Heydorn, A., Ragas, P., Lambertsen, L., Aaes-Jørgensen, A., Molin, S. and Tolker-Nielsen, T. (2003), 'Biofilm formation by Pseudomonas aeruginosa wild type, flagella and type IV pili mutants', *Molecular Microbiology* **48**(6), 1511–1524.

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenent, L. T. and Ley, R. E. (2011), 'Succession of microbial consortia in the developing infant gut microbiome.', *Proceedings of the National Academy of Sciences of the United States of America* **108 Suppl**(1), 4578–85.

Kondoh, M. and Mougi, A. (2015), 'Interaction-type diversity hypothesis and interaction strength: the condition for the positive complexity-stability effect to arise', *Population Ecology* **57**(1), 21–27.

Kong, K.-F., Vuong, C. and Otto, M. (2006), 'Staphylococcus quorum sensing in biofilm formation and infection.', *International journal of medical microbiology : IJMM* **296**(2-3), 133–9.

Koza, A., Moshynets, O., Otten, W. and Spiers, A. J. (2011), 'Environmental modification and niche construction: developing O2 gradients drive the evolution of the Wrinkly Spreader.', *The ISME journal* **5**(4), 665–73.

La Rosa, P. S., Warner, B. B., Zhou, Y., Weinstock, G. M., Sodergren, E., Hall-Moore, C. M., Stevens, H. J., Bennett, W. E., Shaikh, N., Linneman, L. A., Hoffmann, J. A., Hamvas, A., Deych, E., Shands, B. A., Shannon, W. D. and Tarr, P. I. (2014), 'Patterned progression of bacterial populations in the premature infant gut.', *Proceedings of the National Academy of Sciences of the United States of America* **111**(34), 12522–7.

Lambert, G., Liao, D., Vyawahare, S. and Austin, R. H. (2011), 'Anomalous spatial redistribution of competing bacteria under starvation conditions.', *Journal of Bacteriology* **193**(8), 1878–83.

Law, R. and Morton, R. D. (1996), 'Permanence and the Assembly of Ecological Communities', *Ecology* **77**(3), 762.

Lee, Y. K. and Mazmanian, S. K. (2010), 'Has the microbiota played a critical role in the evolution of the adaptive immune system?', *Science* **330**(6012), 1768–73.

Leite, J., Fernandes, B. and Pozzi, E. (2008), 'Application of an anaerobic packed-bed bioreactor for the production of hydrogen and organic acids', *International Journal of Hydrogen Energy* **33**(2), 579–586.

Levy, R. and Borenstein, E. (2013), 'Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules.', *Proceedings of the National Academy of Sciences of the United States of America* **110**(31), 12804–9.

Lim, E. S., Zhou, Y., Zhao, G., Bauer, I. K., Droit, L., Ndao, I. M., Warner, B. B., Tarr, P. I., Wang, D. and Holtz, L. R. (2015), 'Early life dynamics of the human gut virome and bacterial microbiome in infants', *Nature Medicine* **21**(10), 1228–1234.

Login, F. H., Balmand, S., Vallier, A., Vincent-Monégat, C., Vigneron, A., Weiss-Gayet, M., Rochat, D. and Heddi, A. (2011), 'Antimicrobial peptides keep insect endosymbionts under control.', *Science* **334**(6054), 362–5.

Louis, P., Hold, G. L. and Flint, H. J. (2014), 'The gut microbiota, bacterial metabolites and colorectal cancer', *Nature Reviews Microbiology* **12**(10), 661–672.

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. and Knight, R. (2012), 'Diversity, stability and resilience of the human gut microbiota.', *Nature* **489**(7415), 220–30.

Macpherson, A. J. (2000), 'A Primitive T Cell-Independent Mechanism of Intestinal Mucosal IgA Responses to Commensal Bacteria', *Science* **288**(5474), 2222–2226.

Macpherson, A. J., Geuking, M. B. and McCoy, K. D. (2005), 'Immune responses that adapt the intestinal mucosa to commensal intestinal bacteria.', *Immunology* **115**(2), 153–62.

Macpherson, A. J., McCoy, K. D., Johansen, F.-E. and Brandtzaeg, P. (2008), 'The immune geography of IgA induction and function.', *Mucosal Immunol* **1**(1), 11–22.

Macpherson, A. J. and Uhr, T. (2004), 'Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria.', *Science* **303**(5664), 1662–5.

Marino, S., Baxter, N. T., Huffnagle, G. B., Petrosino, J. F. and Schloss, P. D. (2014), 'Mathematical modeling of primary succession of murine intestinal microbiota.', *Proceedings of the National Academy of Sciences of the United States of America* **111**(1), 439–44.

Markussen, T., Marvig, R. L., Gómez-Lozano, M., Aanæs, K., Burleigh, A. E., Høiby, N., Johansen, H. K., Molin, S. and Jelsbak, L. (2014), 'Environmental heterogeneity drives within-host diversification and evolution of Pseudomonas aeruginosa.', *mBio* **5**(5), e01592–14.

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Krumins, J. A., Kuske, C. R., Morin, P. J., Naeem, S., Ovreå s, L., Reysenbach, A.-L., Smith, V. H. and Staley, J. T. (2006), 'Microbial biogeography: putting microorganisms on the map.', *Nat Rev Microbiol* **4**(2), 102–12.

May, R. (1972), 'Will a large complex system be stable?', *Nature* **238**, 413–414.

Mazmanian, S. K., Round, J. L. and Kasper, D. L. (2008), 'A microbial symbiosis factor prevents intestinal inflammatory disease.', *Nature* **453**(7195), 620–5.

McCann, K. S. (2000), 'The diversity-stability debate.', *Nature* **405**(6783), 228–33.

Michalakos, G., Nieva, J., Vayenas, D. and Lyberatos, G. (1997), 'Removal of iron from potable water using a trickling filter', *Water research* **31**(5), 991–996.

Millet, Y. A., Alvarez, D., Ringgaard, S., von Andrian, U. H., Davis, B. M. and Waldor, M. K. (2014), 'Insights into Vibrio cholerae intestinal colonization from monitoring fluorescently labeled bacteria.', *PLoS pathogens* **10**(10), e1004405.

Mitchell, A. C., Dideriksen, K., Spangler, L. H., Cunningham, A. B. and Gerlach, R. (2010), 'Microbially enhanced carbon capture and storage by mineral-trapping and solubility-trapping.', *Environmental science & technology* **44**(13), 5270–6.

Mitri, S. and Foster, K. R. (2013), 'A Genotypic View of Social Interactions in Microbial Communities', *Annual Review of Genetics* **47**, 265–91.

Moller, S., Sternberg, C., Andersen, J. B., Christensen, B. B., Ramos, J. L., Givskov, M. and Molin, S. (1998), 'In Situ Gene Expression in Mixed-Culture Biofilms: Evidence of Metabolic Interactions between Community Members', *Appl. Envir. Microbiol.* **64**(2), 721–732.

Moons, P., Michiels, C. W. and Aertsen, A. (2009), 'Bacterial interactions in biofilms', *Critical Reviews in Microbiology* **35**(3), 157–168.

Morton, R. and Law, R. (1997), 'Regional Species Pools and the Assembly of Local Ecological Communities', *Journal of Theoretical Biology* **187**(3), 321–331.

Mougi, a. and Kondoh, M. (2012), 'Diversity of interaction types and ecological community stability.', *Science* **337**(6092), 349–51.

Nadell, C. D., Foster, K. R. and Xavier, J. B. (2010), 'Emergence of spatial structure in cell groups and the evolution of cooperation.', *PLoS computational biology* **6**(3), e1000716.

Nadell, C. D., Xavier, J. B., Levin, S. A. and Foster, K. R. (2008), 'The evolution of quorum sensing in bacterial biofilms', *PLoS Biology* **6**(1), 0171–0179.

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., Knelman, J. E., Darcy, J. L., Lynch, R. C., Wickey, P. and Ferrenberg, S. (2013), 'Patterns and processes of microbial community assembly.', *Microbiology and molecular biology reviews : MMBR* **77**(3), 342–56.

Neu, J. and Walker, W. A. (2011), 'Necrotizing enterocolitis.', *The New England journal of medicine* **364**(3), 255–64.

Nicolella, C., van Loosdrecht, M. C. and Heijnen, J. J. (2000), 'Wastewater treatment with particulate biofilm reactors.', *Journal of biotechnology* **80**(1), 1–33.

Oliveira, N. M., Niehus, R. and Foster, K. R. (2014), 'Evolutionary limits to cooperation in microbial communities', *Proceedings of the National Academy of Sciences* **111**(50), 201412673.

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. and Brown, P. O. (2007), 'Development of the human infant intestinal microbiota', *PLoS Biol* **5**(7), e177.

Petnicki-Ocwieja, T., Hrncir, T., Liu, Y.-J., Biswas, A., Hudcovic, T., Tlaskalova-Hogenova, H. and Kobayashi, K. S. (2009), 'Nod2 is required for the regulation of commensal microbiota in the intestine.', *Proceedings of the National Academy of Sciences of the United States of America* **106**(37), 15813–8.

Pintelon, T. R. R., Graf von der Schulenburg, D. a. and Johns, M. L. (2009), 'Towards optimum permeability reduction in porous media using biofilm growth simulations.', *Biotechnology and bioengineering* **103**(4), 767–79.

Pintelon, T. R. R., Picioreanu, C., Loosdrecht, M. C. M. V. and Johns, M. L. (2012), 'The effect of biofilm permeability on bio-clogging of porous media.', *Biotechnology and bioengineering* **109**(4), 1031–42.

Pirt, S. (1967), 'A kinetic study of the mode of growth of surface colonies of bacteria and fungi', *Journal of General Microbiology* **47**, 181–197.

Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M. and Tans, S. J. (2007), 'Empirical fitness landscapes reveal accessible evolutionary paths.', *Nature* **445**(7126), 383–386.

Preidis, G. A. and Versalovic, J. (2009), 'Targeting the human microbiome with antibiotics, probiotics, and prebiotics: gastroenterology enters the metagenomics era', *Gastroenterology* **136**(6), 2015–2031.

Rainey, P. B. and Travisano, M. (1998), 'Adaptive radiation in a heterogeneous environment.', *Nature* **394**(6688), 69–72.

Rao, D., Webb, J. S. and Kjelleberg, S. (2005), 'Competitive interactions in mixed-species biofilms containing the marine bacterium Pseudoalteromonas tunicata.', *Applied and environmental microbiology* **71**(4), 1729–36.

Rehfeldt, K. R., Boggs, J. M. and Gelhar, L. W. (1992), 'Field study of dispersion in a heterogeneous aquifer: 3. Geostatistical analysis of hydraulic conductivity', *Water Resources Research* **28**(12), 3309–3324.

Relman, D. A. (2012), 'The human microbiome: ecosystem resilience and health.', *Nutrition reviews* **70 Suppl 1**, S2–9.

Rittmann, B. (1982), 'The effect of shear stress on biofilm loss rate', *Biotechnology and Bioengineering* **24**(2), 501–6.

Rohr, R. P., Saavedra, S. and Bascompte, J. (2014), 'On the structural stability of mutualistic systems', *Science* **345**(6195), 1253497.

Rothblum, U. G. and Tan, C. P. (1985), 'Upper bounds on the maximum modulus of subdominant eigenvalues of nonnegative matrices', *Linear Algebra and its Applications* **66**, 45–86.

Rothman, D. H. and Forney, D. C. (2007), 'Physical model for the decay and preservation of marine organic carbon.', *Science* **316**(5829), 1325–8.

Rusconi, R., Garren, M. and Stocker, R. (2014), 'Microfluidics Expanding the Frontiers of Microbial Ecology', *Annual Review of Biophysics* **43**(1), 65–91.

Sackmann, E. K., Fulton, A. L. and Beebe, D. J. (2014), 'The present and future role of microfluidics in biomedical research.', *Nature* **507**(7491), 181–9.

Salzman, N. H., Hung, K., Haribhai, D., Chu, H., Karlsson-Sjöberg, J., Amir, E., Teggatz, P., Barman, M., Hayward, M., Eastwood, D., Stoel, M., Zhou, Y., Sodergren, E., Weinstock, G. M., Bevins, C. L., Williams, C. B. and Bos, N. A. (2010), 'Enteric defensins are essential regulators of intestinal microbial ecology.', *Nat Immunol* **11**(1), 76–83.

Samuel, B. S. and Gordon, J. I. (2006), 'A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism.', *Proceedings of the National Academy of Sciences of the United States of America* **103**(26), 10011–6.

Sauer, K., Cullen, M. C., Rickard, A. H., Zeef, L. A. H., Davies, D. G. and Gilbert, P. (2004), 'Characterization of nutrient-induced dispersion in Pseudomonas aeruginosa PAO1 biofilm.', *Journal of bacteriology* **186**(21), 7312–26.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. and Cardona, A. (2012), 'Fiji: an open-source platform for biological-image analysis.', *Nature methods* **9**(7), 676–82.

Schluter, J. and Foster, K. R. (2012), 'The evolution of mutualism in gut microbiota via host epithelial selection.', *PLoS biology* **10**(11), e1001424.

Schluter, J., Nadell, C. D., Bassler, B. L. and Foster, K. R. (2015), 'Adhesion as a weapon in microbial competition.', *The ISME journal* **9**(1), 139–49.

Seifert, D. and Engesgaard, P. (2012), 'Sand box experiments with bioclogging of porous media: hydraulic conductivity reductions.', *Journal of contaminant hydrology* **136-137**, 1–9.

Shafahi, M. and Vafai, K. (2009), 'Biofilm affected characteristics of porous structures', *International Journal of Heat and Mass Transfer* **52**(3-4), 574–581.

Sinton, L. W., Mackenzie, M. L., Karki, N., Braithwaite, R. R., Hall, C. H. and Flintoft, M. J. (2010), 'Transport of Escherichia coli and F-RNA bacteriophages in a 5m column of saturated pea gravel', *Journal of Contaminant Hydrology* **117**(1), 71–81.

Sleator, R. D. (2010), 'The human superorganism - of microbes and men.', *Med Hypotheses* **74**(2), 214–5.

Sommers, H., Crisanti, A., Sompolinsky, H. and Stein, Y. (1988), 'Spectrum of Large Random Asymmetric Matrices', *Physical Review Letters* **60**(19), 1895–1898.

Sonnenburg, J. L., Xu, J., Leip, D. D., Chen, C.-H., Westover, B. P., Weatherford, J., Buhler, J. D. and Gordon, J. I. (2005), 'Glycan foraging in vivo by an intestine-adapted bacterial symbiont.', *Science* **307**(5717), 1955–9.

Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Rätsch, G., Pamer, E. G., Sander, C. and Xavier, J. B. (2013), 'Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota.', *PLoS Comput Biol* **9**(12), e1003388.

Stewart, P. S. (1993), 'A model of biofilm detachment.', *Biotechnology and bioengineering* **41**(1), 111–7.

Stone, H. A. (2007), *Introduction to fluid dynamics for microfluidic flows*, Series on Integrated Circuits and Systems, Springer US, Boston, MA.

Stoodley, P., Cargo, R., Rupp, C. J., Wilson, S. and Klapper, I. (2002), 'Biofilm material properties as related to shear-induced deformation and detachment phenomena.', *Journal of industrial microbiology & biotechnology* **29**(6), 361–7.

Tang, Y., Valocchi, A. J., Werth, C. J. and Liu, H. (2013), 'An improved pore-scale biofilm model and comparison with a microfluidic flow cell experiment', *Water Resources Research* **49**(12), 8370–8382.

Tao, T., Vu, V. and Krishnapur, M. (2010), 'Random matrices: Universality of ESDs and the circular law', *The Annals of Probability* **38**(5), 2023–2065.

Thullner, M. (2010), 'Comparison of bioclogging effects in saturated porous media within one- and two-dimensional flow systems', *Ecological Engineering* **36**(2), 176–196.

Torsvik, V. (2002), 'Prokaryotic Diversity–Magnitude, Dynamics, and Controlling Factors', *Science* **296**(5570), 1064–1066.

Tortell, P. (1999), 'Marine bacteria and biogeochemical cycling of iron in the oceans', *FEMS Microbiology Ecology* **29**(1), 1–11.

Tremaroli, V. and Bäckhed, F. (2012), 'Functional interactions between the gut microbiota and host metabolism.', *Nature* **489**(7415), 242–9.

Trulear, M. and Characklis, W. (1982), 'Dynamics of biofilm processes', *Journal (Water Pollution Control Federation)* **54**(9), 1288–1301.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C., Knight, R. and Gordon, J. I. (2007), 'The human microbiome project: exploring the microbial part of ourselves in a changing world', *Nature* **449**(7164), 804.

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R. and Gordon, J. I. (2009), 'The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice.', *Science translational medicine* **1**(6), 6ra14.

Valiei, A., Kumar, A., Mukherjee, P. P., Liu, Y. and Thundat, T. (2012), 'A web of streamers: biofilm formation in a porous microfluidic device.', *Lab on a chip* **12**(24), 5133–7.

Van den Abbeele, P., Van de Wiele, T., Verstraete, W. and Possemiers, S. (2011), 'The host selects mucosal and luminal associations of coevolved gut microorganisms: a novel concept.', *FEMS Microbiol Rev* **35**(4), 681–704.

van Gestel, J., Weissing, F. J., Kuipers, O. P. and Kovács, A. T. (2014), 'Density of founder cells affects spatial pattern formation and cooperation in Bacillus subtilis biofilms.', *The ISME journal* **8**(10), 2069–79.

Van Loosdrecht, M. C. M., Eikelboom, D., Gjaltema, A., Mulder, A., Tijhuis, L. and Heijnen, J. J. (1995), 'Biofilm structures', *Water Science and Technology* **32**(8), 35–43.

Vanhoutte, T., Huys, G., Brandt, E. and Swings, J. (2004), 'Temporal stability analysis of the microbiota in human feces by denaturing gradient gel electrophoresis using universal and group-specific 16S rRNA gene primers.', *FEMS Microbiol Ecol* **48**(3), 437–46.

Von Canstein, H., Kelly, S., Li, Y. and Wagner-Döbler, I. (2002), 'Species diversity improves the efficiency of mercury-reducing biofilms under changing environmental conditions.', *Applied and environmental microbiology* **68**(6), 2829–37.

von der Schulenburg, D. A. G., Pintelon, T. R. R., Picioreanu, C., Van Loosdrecht, M. C. M. and Johns, M. L. (2009), 'Three-dimensional simulations of biofilm growth in porous media', *AIChE Journal* **55**(2), 494–504.

Waller, A. S., Yamada, T., Kristensen, D. M., Kultima, J. R., Sunagawa, S., Koonin, E. V. and Bork, P. (2014), 'Classification and quantification of bacteriophage taxa in human gut metagenomes.', *ISME J* **8**(7), 1391–402.

Wang, X. and Hendry, P. (2012), 'Effect of Nutrient Addition on an Oil Reservoir Microbial Population: Implications for Enhanced Oil Recovery', *Journal of Petroleum & Environmental Biotechnology* **3**(3).

Wanner, O. and Gujer, W. (1986), 'A Multispecies Biofilrn Model', *Biotechnology and Bioengineering* **28**, 314–328.

Warren, P. H., Law, R. and Weatherby, A. J. (2003), 'Mapping the assembly of protist communities in microcosms', *Ecology* **84**(4), 1001–1011.

Wehkamp, J., Harder, J., Weichenthal, M., Schwab, M., Schäffeler, E., Schlee, M., Herrlinger, K. R., Stallmach, A., Noack, F., Fritz, P., Schröder, J. M., Bevins, C. L., Fellermann, K. and Stange, E. F. (2004), 'NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression.', *Gut* **53**(11), 1658–64.

Wehkamp, J., Salzman, N. H., Porter, E., Nuding, S., Weichenthal, M., Petras, R. E., Shen, B., Schaeffeler, E., Schwab, M., Linzmeier, R., Feathers, R. W., Chu, H., Lima, H., Fellermann, K., Ganz, T., Stange, E. F. and Bevins, C. L. (2005), 'Reduced Paneth cell alpha-defensins in ileal Crohn's disease.', *Proceedings of the National Academy of Sciences of the United States of America* **102**(50), 18129–34.

Weinreich, D. M., Delaney, N. F., DePristo, M. A. and Hartl, D. L. (2006), 'Darwinian evolution can follow only very few mutational paths to fitter proteins', *Science* **312**(5770), 111–114.

Werner, E., Roe, F., Bugnicourt, A., Franklin, M. J., Heydorn, A., Molin, S., Pitts, B. and Stewart, P. S. (2004), 'Stratified growth in Pseudomonas aeruginosa biofilms.', *Applied and environmental microbiology* **70**(10), 6188–96.

West, S. A., Griffin, A. S., Gardner, A. and Diggle, S. P. (2006), 'Social evolution theory for microorganisms.', *Nature reviews. Microbiology* **4**(8), 597–607.

Whitesides, G. and Stroock, A. (2001), 'Flexible methods for microfluidics', *Phys Today* **54**(6), 42–48.

Whitman, W. B., Coleman, D. C. and Wiebe, W. J. (1998), 'Perspective Prokaryotes : The unseen majority', *Proceedings of the National Academy of Sciences of the United States of America* **95**(June), 6578–6583.

Wigner, E. P. (1958), 'On the Distribution of the Roots of Certain Symmetric Matrices', *The Annals of Mathematics* **67**(2), 325.

Williamson, K. and McCarty, P. L. (1976), 'A model of substrate utilization by bacterial films.', *Journal - Water Pollution Control Federation* **48**(1), 9–24.

Xavier, J. B. and Foster, K. R. (2007), 'Cooperation and conflict in microbial biofilms.', *Proceedings of the National Academy of Sciences of the United States of America* **104**(3), 876–81.

Young, I. M. and Crawford, J. W. (2004), 'Interactions and self-organization in the soil-microbe complex.', *Science* **304**(5677), 1634–7.

Ze, X., Duncan, S. H., Louis, P. and Flint, H. J. (2012), 'Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon.', *The ISME journal* **6**(8), 1535–43.

Ze, X., Mougen, F. L., Duncan, S. H., Louis, P. and Flint, H. J. (2013), 'Some Are More Equal Than Others', *Gut Microbes* **4**(3), 236–240.

Zhang, Z., Geng, J., Tang, X., Fan, H., Xu, J., Wen, X., Ma, Z. S. and Shi, P. (2014), 'Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota.', *The ISME journal* **8**(4), 881–93.

Zhou, Y., Shan, G., Sodergren, E., Weinstock, G., Walker, W. A. and Gregory, K. E. (2015), 'Longitudinal analysis of the premature infant intestinal microbiome prior to necrotizing enterocolitis: a case-control study.', *PloS one* **10**(3), e0118632.