Modelling the ecology of host-associated microbiomes

Dissertation in fulfilment of the requirements for the degree

Doctor rerum naturalium

of the Faculty of Mathematics and Natural Sciences

at the Christian-Albrechts University of Kiel

Submitted by

Román Ulises Zapién Campos

Evolutionary Theory Department,

Max Planck Institute for Evolutionary Biology, Plön

Kiel, 2021

First examiner:	Prof. Dr. Arne Traulsen MPI for Evolutionary Biology, Plön, Germany
Second examiner:	Prof. Dr. Hinrich Schulenburg CAU, Kiel, Germany
Additional examiner:	Prof. Dr. Ute Hentschel Humeida GEOMAR, Kiel, Germany
Chairperson:	Prof. Dr. Frank Kempken CAU, Kiel, Germany

Date of oral examination:	December 10, 2021
Date of approval for publication:	December 10, 2021

"The good life is one inspired by love and guided by knowledge."
— Bertrand Russell

Le dedico esta tesis a mi familia.
Laura y Román, por crear el ambiente en el que todo es posible;
Christian, por tu capacidad de convertir retos en oportunidades.

I dedicate this thesis to my family.
Laura and Román, for creating the environment where everything is possible.
Christian, for your ability to turn challenges into opportunities.

Summary

Since the turn of the century, great developments in our understanding of the microbial world have occurred. The study of microbes associated with animals, plants, and fungi – here referred to as host-associated microbiome – is one that is modifying our understanding of ecology and evolution. While empirical evidence has accumulated rapidly, the theory to explain it has been lagging behind. During the time of my PhD, new theoretical advances have been introduced. In this thesis, I present my contribution to understanding the effect of host-specific ecological processes on the microbiome.

Each chapter is focused on a specific process with eco-evolutionary consequences. To gain certainty about our understanding, I simplify the complex biology of real-world hosts and microbes. In the models, hosts only provide a space for microbes to inhabit without being affected by them. On the other hand, the microbiome is affected by the processes imposed by their hosts – specifically the microbial composition. I rely on the theory of stochastic processes and focus on birth, death, and immigration as the drivers of the microbial dynamics. All in all, the results are discussed in a broad context, where potentially general microbiome patterns are identified.

The first project stems from observing that compared to classical habitats, hosts have a lifespan that might interfere with the microbial dynamics. For hosts which acquire all their microbes from the environment, the lifespan indeed interferes in the long run, even leading to the coexistence of subpopulations with distinct microbiome. In the second project I address the possible "inheritance" of microbes from parents to their newborn hosts. I observe that its effects depend on life-history traits including immigration and host lifespan. In the last project, I question one of the critical assumptions. Rather than equal growth and death rates for all microbes, I assume differences in a simplified model. I find a surprising result – regardless of the assumption on these rates, if immigration and biodiversity are sufficiently large, differences at the population level do not modify the emergent community pattern.

This thesis is a glance of the many questions surrounding the ecology and evolution of the microbiome. The results provided here build upon a theory of the drivers at the microbial scale and its interaction with the macro scale.

Zusammenfassung

Seit der Jahrhundertwende haben sich große Entwicklungen in unserem Verständnis der mikrobiellen Welt vollzogen. Die Untersuchung von Mikroben, welche Tiere, Pflanzen und Pilze besiedeln — hier als wirtsassoziiertes Mikrobiom bezeichnet – verändert unser Verständnis von Ökologie und Evolution. Während sich empirische Daten schnell angesammelt haben, hinkte die erklärende Theorie hinterher. In dieser Dissertation präsentiere ich meinen Beitrag zum Verständnis der Wirkung wirtsspezifischer ökologischer Prozesse auf das Mikrobiom.

Jedes Kapitel konzentriert sich auf einen bestimmten öko-evolutionären Prozess. Dabei vereinfache ich zum besseren Verständnis die komplexe Biologie realer Wirte und Mikroben in mathematischen Modellen. In den Modellen bieten Wirte einen Lebensraum für Mikroben ohne von ihnen selbst beeinflusst zu werden. Das Mikrobiom wird dagegen von Wirtsprozessen beeinflusst – insbesondere in seiner mikrobiellen Zusammensetzung. Ich wende die Theorie stochastischer Prozesse an und konzentriere mich auf Reproduktion, Tod und Einwanderung als Treiber der mikrobiellen Dynamik. Die Ergebnisse werden in einem breiten Kontext diskutiert und potenzielle, allgemeine Mikrobiomstrukturen identifiziert.

Das erste Projekt basiert auf der Beobachtung, dass Wirte im Vergleich zu klassischen Habitaten eine begrenzte Lebensdauer haben, was die mikrobielle Dynamik beeinträchtigen könnte. Bei Wirten, die ihre gesamten Mikroben aus der Umwelt aufnehmen, wechselwirkt die begrenzte Wirtslebensdauer tatsächlich auf lange Sicht mit der mikrobiellen Dynamik. Dies führt zur Koexistenz von Wirtssubpopulationen mit unterschiedlichen Mikrobiomzusammensetzungen. Im zweiten Projekt beschäftige ich mich mit der möglichen "Vererbung" von Mikroben von den Eltern an neugeborenen Wirte. Ich zeige, dass die Auswirkungen von Mikrobiom-Vererbung von Merkmalen des Wirts-Lebenzyklus abhängen, einschließlich der Einwanderung von Mikroben und der Lebensdauer des Wirts. Im letzten Projekt hinterfrage ich eine der kritischen Annahmen der vorherigen Modelle. Anstelle gleicher Wachstums- und Todesraten für alle Mikroben gehe ich von unterschiedlichen Raten in einem vereinfachten Modell aus. Ich finde ein überraschendes Ergebnis —

wenn die Einwanderung von Mikroben und die Biodiversität ausreichend groß sind, ändern Unterschiede in den mikrobiellen Wachstumsraten nicht das entstehende Muster der ökologischen Gemeinschaft auf Populationsebene.

Diese Arbeit ist ein Blick auf die vielen Fragen rund um die Ökologie und Evolution des Mikrobioms. Die hier vorgelegten Ergebnisse bauen auf einer Theorie der Prozesse auf der mikrobiellen Ebene und ihren Wechselwirkungen mit der Makroskala auf.

Contents

Introduction

"Our imagination is struck only by what is great; but the lover of natural philosophy should reflect equally on little things."

– Alexander von Humboldt

Using current technologies, we can find microbes almost everywhere. Microbial life prevails in the vastly diverse habitats of the biosphere (Thompson et al., 2017). This is more meaningful once we consider that ancient microbes were the first living organisms and that their large abundance has even triggered global-scale transformations (Battistuzzi et al., 2004). Nonetheless, the central role of microbes in the biosphere has not always been clear.

Throughout human history, our perception of microbes has changed – especially of those microbes found in the human body. Microbes have been in our minds since the pioneering observations of Van Leeuwenhoek in the 17th century (Leeuwenhoek, 1677). At the time, their nature was unclear; however, some generations later, microbes had become tiny enemies – the agents of disease. The author Paul de Kruif gives us a vivid account of this time in Microbe Hunters (De Kruif, 1926) – where characters like Pasteur and Koch were at the front line to identify the pathogens behind some of the most severe human diseases. Only new technologies allowed us to see that microbes are more than pathogens. As Sergei Winogradsky would find out, microbes are an integral part of ecosystems, not just pathogens – microbial ecology had just been born (Winogradsky et al., 1949; Dworkin and Gutnick, 2012).

For microbes living in or on hosts, the extent of such integration has crystalized, in recent years, in the concept of the *host-associated microbiome* (Adair and Douglas, 2017). This view, where hosts and microbes are closely associated, asks for the reinterpretation of ecological and evolutionary concepts (Bordenstein and Theis, 2015). Instead of thinking of animals, plants, or fungi as isolated individuals challenged by the environment, microbes might actively participate in their eco-evolutionary processes. Correspondingly, hosts might influence the dynamics of their microbiome. Some even consider the concept of the *holobiont* (from the Greek *hólos* - whole, and *biont* - unit

of life) as the best way to describe what hosts and microbes experience in nature (McFall-Ngai et al., 2013).

Motivation and structure of the thesis

In this thesis, I investigate processes experienced by microbes living in or on hosts. While it could be argued that theories developed in a macro-ecological context could explain the ecology and evolution of host-associated microbiomes (Prosser et al., 2007), empirical evidence suggests that new theoretical developments are needed (Miller et al., 2018). My motivation is to provide theoretical insights into the ecological processes giving rise to microbiome patterns in nature. Although each project addresses a distinct possibility, the following questions connect them all: (*i*) Is there a process exclusively (or mostly) experienced by host-associated microbes? And (*ii*) how does this process affect the microbiome composition? Inspired by its insights to microbiome research (Sieber et al., 2019) and simple principles of birth, death, and immigration, I build upon neutral ecological theory (Hubbell, 2001; Sloan et al., 2006).

Classic neutral models assume that habitats are everlasting (Hubbell, 2001). Although this is a sensible assumption if habitats change on geological timescales, when animal, fungi, or plant hosts are the habitats, their timescale of survival might interfere with the microbial dynamics. Here I investigate the effect of host lifespan on the microbiome composition – where the transient dynamics and the microbiome of newborns become essential. First, I consider hosts born free of microbes in Chapter 1, then, the parental transfer of microbes to newborns in Chapter 2. Both cases are of great interest for microbiome research but challenging to address experimentally so far (Hammer et al., 2019; Funkhouser and Bordenstein, 2013). Throughout, classic neutral theory is my baseline for comparison (Sloan et al., 2006), however in Chapter 3 I relax the assumption of neutrality. Each time, the observations are connected to empirical evidence. I finish stating the overarching conclusions of my work and identifying avenues of future research.

In the remainder of this section I provide necessary definitions and concepts, including host-associated microbiome and neutral ecological theory. Then, I present the theory of stochastic modeling relevant for the analyses.

What is a microbiome?

Ecology is the science that studies the interactions that determine the distribution and abundance of organisms in the environment (Begon and Townsend, 2020). Although its levels of study range from populations to the entire biosphere, conventionally, studies focus on a specific ecological level,

where particular processes might come at play (Begon and Townsend, 2020).

A concept that transverses ecological levels is *biome*. A biome is a group of individuals adapted to a particular environment (Berg et al., 2020). The concept is sometimes used as synonym to community (a group of populations) or ecosystem. A biome composed of microscopic individuals is called *microbiome* (Berg et al., 2020). Microbiomes can be classified in several ways. One of considerable interest is according to the habitat where the microbes are contained. A distinction is made between microbes located in an abiotic environment – *environmental microbiome* – and those hosted in a larger, living organism – e.g. plants, fungi, or animals – the *host-associated microbiome*. In this thesis I refer to microbiome in the sense of community, where various microbial populations coexist within the same environment.

Ecology of communities

Ecology and evolution have been intertwined since their conception as scientific fields (Watts et al., 2019). Nowadays, we know this is especially true in microbial communities, where the ecological and evolutionary time-scales can overlap (Koskella et al., 2017). Although each chapter of this thesis focuses on an ecological question, these have important evolutionary consequences. Moreover, ecological and evolutionary definitions are blurred in Chapter 2, where I consider a process of microbial "inheritance" to hosts.

Community ecology has a long history of dialogue between empirical and theoretical studies. While new theories have been developed from empirical insights, new experiments have emerged from theoretical predictions as well. In the next paragraphs I briefly summarize the empirical and theoretical knowledge of what Vellend (2010) calls the main *drivers of community ecology*. These are useful to build from a solid conceptual ground.

The first driver is *selection*. Within a population, individuals naturally show variation in their traits. In an ecological time-scale, such variation biases the success of individuals in favour of those better adapted (Vellend, 2010). Although the relevant traits for selection are contingent to the environment, a commonality observed is that in large populations the strength of selection increases (Vellend, 2010).

The second driver is ecological noise, often called *ecological drift*. This occurs when random environmental or demographic fluctuations lead to changes in the community composition (Zhou and Ning, 2017). Evidence shows this is particularly important for smaller communities, where chance plays a larger role (Vellend, 2010). Chapters 1 and 2 operate in a context where all individuals have the same traits, so without selection, ecological drift is the main driver acting upon the microbial types.

The third driver is *migration*. In nature, habitats are very often partitioned (Sieber et al., 2021). For example, the islands of an archipelago or the crypts of the intestinal epithelium. If these partitions are sufficiently isolated from each other, each could contain an isolated community. Migration is the movement of individuals between such partitions – which has proved to be important to sustain biodiversity (Macarthur and Wilson, 1967).

The last driver is *speciation*. In nature, numerous species coexist within the same community. Speciation is the process by which new diversity is generated from existing individuals (Gavrilets, 2003).

The drivers introduced above have been used extensively to develop theories explaining the community patterns observed in nature (Gravel et al., 2006). Although specific discoveries have paved the way, two theories unify many of the findings – the so called niche and neutral theories of ecology.

Niche theory is developed from the premise of differential adaptation of taxa to the habitat. Then, according to this perspective, the observed diversity in nature is the result of different taxa being able to occupy different ecological niches – e.g., food, space, time, etc (Chase et al., 2003).

Neutral theory is developed from an opposite premise, where selection is considered negligible compared to ecological drift and migration. According to this perspective, the observed diversity in nature results from continuous demographic and environmental fluctuations, as well as migration between patchy habitats (Hubbell, 2001).

Much knowledge has been gained about the ecology at the macro-scale (Vellend, 2010). However, less is known about the microbial scale, and particularly about the association of microbes to hosts (Zhou and Ning, 2017) – among many questions, whether hosts provide more stable or challenging habitats (Bansept et al., 2021), the effect of biological clocks (Thaiss et al., 2016a), the fluctuation of resources based on feeding rhythms (Thaiss et al., 2014), and the interaction with the immune system (Thaiss et al., 2016b). Each of these topics is subject of active empirical and theoretical research.

In this thesis, I combine the ecological drivers and theories introduced above with other mechanisms that could specifically drive the ecology of host-associated microbes. For this, I primarily rely on stochastic models.

Stochastic modeling of communities

Stochastic fluctuations are one of the main drivers of ecological change. These fluctuations can have an environmental or demographic origin (Zhou and Ning, 2017). In this thesis, I use the concept of stochasticity in its demographic sense – i.e. any fluctuation in the abundance of individuals due to the probabilistic nature of birth, death, or migration.

Stochastic change is specially important if, at any point, the number of individuals is close to zero (Vellend, 2010). In this case, fluctuations of abundance can lead to a divergence of community compositions that can propagate in time, even when the number of individuals is far from zero (Vellend, 2010). In an extreme case, fluctuations can even lead to the extinction of whole populations or communities (Vellend, 2010).

In the next paragraphs I detail the process to write down a stochastic model. In general, this starts by identifying the subjects of study and the events that change them at an individual – *microscopic* – level. Then, an equation accounting for the change of the community composition through time is derived. Finally, if needed, the microscopic equation can be translated to a *mesoscopic* or *macroscopic* description. The resulting model can then be used to address specific biological questions.

States, events, and rates

Given a collection of states, a *stochastic model* tracks either the state that occurs or the probability of observing a given one. In a community, a *state* is one of the possible community compositions (Otto and Day, 2007). States are described by variables that can be discrete or continuous. The choice between a discrete or continuous description primarily depends on the properties of the community and the observables of interest.

At any given time, the composition can be maintained or changed. The occurrence of a specific *event* depends on the processes driving the community dynamics. The *rates of events* determine which of the possible events occurs and the time between each of them (Otto and Day, 2007).

The knowledge of the states and rates of events is enough to give a probabilistic account of the community composition through time. This probabilistic description can be obtained in several ways. One possibility are stochastic simulations. A simulation produces a time-course realization of a single community, that upon collection with other realizations can be subjected to statistical analyses (Gillespie, 1976). Another, sometimes more powerful, possibility is to use an algebraic representation of the stochastic model (Gardiner, 2004). Nonetheless, commonly, obtaining explicit solutions to algebraic representations is not possible – thus, equations, their numerical solution, and simulations become complementary.

Stochastic simulations

The aim of a stochastic simulation is to use the rates of events to produce an individual realization of the model. Running a simulation is straightforward once an initial state is specified – based on the current state a future state

is obtained, then the new rates of events computed, and the steps repeated. The most direct simulations use all the rates of events.

In general, a large number of realizations are needed before robust statistical analyses can be performed. Depending on the complexity of the model, this could lead to a large demand of computational resources and computing time. Thus, various simulation algorithms have been developed to decrease this burden. The *Gillespie algorithm*, also called *stochastic simulation algorithm*, is the most common (Gillespie, 1976). This algorithm focuses only on the rates of events that cause a change – called *transition rates*. Each time-step, an event is randomly chosen based on the relative weight of its transition rate. Then, the time length of the time-step is drawn based on the sum of transition rates. Other algorithms build upon Gillespie to further speed up simulations – for example, by exploiting particular properties of the model or by sacrificing resolution at short time scales. *Tau-leaping* is one of such algorithms, where the transition rates are only updated after a chosen number of time-steps (Gillespie, 2001).

In some systems, the probability distribution reaches stationarity after a transient dynamics. In this case, the collection of realizations with sufficiently many simulated time-steps can lead to such a stationary distribution.

Algebraic representations

The possible realizations of a community can be "compressed" into a dynamical equation. Then, the methods for differential equations can be used to analyse it or even solve it. Similarly to the Gillespie algorithm, these equations only track changes in the composition (Gardiner, 2004).

Under proper conditions or simplifying assumptions, closed form equations of relevant quantities, such as probabilities and statistical moments (e.g. mean and variance) can be obtained. Alternatively, numerical methods can be used to compute solutions for chosen parameters.

Sometimes, a specific algebraic representation could simplify the computation of certain observables. That is the case in this thesis – thus, in the upcoming Chapters I analyse the models using various representations.

When a stochastic model is developed from its microscopic description – focusing on birth, death, and migration – transforming the set of transition rates into a dynamical equation is the first step (Otto and Day, 2007). The premise here is that for a given state – let us called it $\mathbf{n}$ – there is a set of transition rates that lead to this state $R(\mathbf{n}' \to \mathbf{n})$, but also a set of transition rates that lead out of it $R(\mathbf{n} \to \mathbf{n}')$. Then, the probability of composition $\mathbf{n}$

at time t, $P(\mathbf{n}, t)$, changes according to the following dynamical equation

$$\underbrace{\frac{\partial P(\mathbf{n},t)}{\partial t}}_{\text{probability change}} = \underbrace{\sum_{\mathbf{n}'\neq\mathbf{n}} R(\mathbf{n}'\to\mathbf{n})P(\mathbf{n}',t)}_{\text{probability influx}} - \underbrace{\sum_{\mathbf{n}'\neq\mathbf{n}} R(\mathbf{n}\to\mathbf{n}')P(\mathbf{n},t)}_{\text{probability outflux}}. \tag{1}$$

This equation is called the *master equation* of the system (Gardiner, 2004). Its usefulness is that it contains the time-dependent information about every possible state of the system. I use this equation as the final subject of study in Chapter 3 and as an intermediate step in Chapters 1 and 2.

The states in the master equation are discrete. However, in some systems the number of states is so large and their distance so short, that assuming the states as a continuum might simplify the analysis. This is the premise of the *differential Chapman-Kolmogorov equation* (dCK) (Gardiner, 2004). In particular, if the number of states is bounded to a domain, these can be normalized and approximated as a continuous variable. Let us define the frequency $\mathbf{x} = \mathbf{n}/N$, where N is the number of individuals in the community, which is constant. Then, the dCK equation is given by

$$\begin{aligned}\underbrace{\frac{\partial P(\mathbf{x},t)}{\partial t}}_{\text{probability change}} = &- \underbrace{\sum_i \frac{\partial}{\partial x_i} a_i(\mathbf{x})P(\mathbf{x},t)}_{\text{drift component}} + \underbrace{\frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j} b_{i,j}(\mathbf{x})P(\mathbf{x},t)}_{\text{diffusion component}} \\ &+ \underbrace{\int \left(R(\mathbf{x}'\to\mathbf{x})P(\mathbf{x}',t) - R(\mathbf{x}\to\mathbf{x}')P(\mathbf{x},t)\right) d\mathbf{x}'}_{\text{jump component}}.\end{aligned} \tag{2}$$

Here, the first and second terms describe local changes, while the last term describes non-local changes. Each of these terms scale differently. In particular, in contrast to the other terms (order $1/N$), the diffusion term (order $1/N^2$) implies that noise is more relevant for smaller communities, N. I use the dCK equation in Chapter 1, where I account for the effect of host death – described as jumps – on the microbiome composition.

A particular case of the dCK equation is when no jumps occur – so the last term is equal to zero. This is called the *Fokker-Planck equation* (FP) (Risken, 1996). This equation has been useful to investigate the dynamics of environmental microbiomes – where no sudden changes occur (Sloan et al., 2006). In Chapters 1 and 2 I use the FP equation as a baseline to quantify the effect of host dynamics on the microbiome composition.

The dCK equation describes the change of probability for an infinite number of realizations. Similarly to simulations, single realizations can be

obtained using the equivalent *stochastic differential equation* (SDE) (Gardiner, 2004). According to the theory of stochastic processes, the dynamics of a system is generated by two contributing factors. Like in an ordinary differential equation, in a SDE the change of frequency $\mathbf{x}$ is proportional to a deterministic rate. However, in addition, the change depends on a noise term (Allen, 2007),

$$\underbrace{d\mathbf{x}(t)}_{\text{frequency change}} = \underbrace{\mathbf{a}(\mathbf{x})dt}_{\text{deterministic component}} + \underbrace{c(\mathbf{x})d\mathbf{W}(t)}_{\text{noise component}} . \tag{3}$$

The stochastic fluctuations lead to deviations from the deterministic case, so resulting realizations are different from each other. I use this representation in Chapter 2, where a set of more complicated transition rates account for the transfer of microbes from parents to newborns. There, I apply the *Euler-Maruyama algorithm*, a first-order numerical method to solve SDEs (Allen, 2007).

Probability distribution at equilibrium

In this thesis I primarily focus on the long term state of the microbiome. More concretely, on its *stationary probability distribution*. If stationarity exists, this occurs in the limit of infinite time – mathematically, this can be found when the derivative with respect to time in the master equation or the dCK equals zero (Gardiner, 2004).

In particular, I model cases where regardless of the initial conditions, the same stationary probability distribution is reached. This condition is called *ergodicity* (Gardiner, 2004). Stochastic systems have this property if the collection of all states cannot be separated in subcollections. In other words, if all states are directly or indirectly connected in the space of states.

Exploiting ergodic properties, I use two approaches to compute the stationary probability distribution. Firstly, in Chapter 1, I analyse the following time-discrete equation of probabilities,

$$\underbrace{P(\mathbf{n}, t+1)}_{\text{prob. at } t+1} = \underbrace{T(\mathbf{n})}_{\text{transition r.}} \underbrace{P(\mathbf{n}, t)}_{\text{prob. at } t}, \tag{4}$$

This equation is stationary when the probability $P(\mathbf{n}, t+1) = P(\mathbf{n}, t)$, so effectively $P(\mathbf{n}, t) = P(\mathbf{n})$. To find this stationary vector, I use the *Perron-Frobenius theorem* (Caswell, 2001). This theorem indicates that we can write $T(\mathbf{n})P(\mathbf{n}) = \lambda P(\mathbf{n})$, where λ is the associated eigenvalue of matrix $T(\mathbf{n})$. Thus, $P(\mathbf{n})$ corresponds to the eigenvector of $T(\mathbf{n})$ with eigenvalue $\lambda = 1$.

A second method, used in Chapter 3, relies on the master equation instead. At stationarity, its derivative with respect to time equals zero. Then, for

each state the influx and outflux of probability are exactly equal. This is called a *detailed balance condition* (Gardiner, 2004). We are left with a set of equalities, that after some algebraic manipulations leads to a recurrence equation. This equation allows us to compute the stationary probability distribution starting from one of the states at the boundaries.

Modeling the microbiome

At the core of each stochastic model lie its transition rates. Regardless of the algebraic representation or simulation method chosen, the transition rates are responsible for the resulting dynamics. In fact, they contain information about the variables that we consider relevant, their interactions, and importantly, the assumptions that we make on them. In this thesis, I refer to *ecological process* as the mathematical form of the transition rates (Figure 0).

There are many ways in which the transition rates could be written. However, historically, some ways have been preferred since, mathematically, they make the analyses feasible. A *Markov process* is probably the most common way (Gardiner, 2004). In this process, we assume that the future state of the community only depends on its present state – thus, no time memory is considered. This is a strong simplification, that nonetheless, has proven to be very effective to model diverse physical systems – including ecological systems (Otto and Day, 2007).

A further simplification, common in ecological models, is to maintain the number of individuals within a community constant. One conventional way to achieve this is the *Moran process* (Moran, 1962). In a Moran process, deaths and births – the drivers of dynamic change – occur consecutively, such that within a single time-step either birth follows death or death follows birth. The models presented in this thesis are Markovian following a Moran process, where local microbial death is followed by either birth from a local microbe or immigration of an environmental microbe (Figure 0).

In their most general form, the ecological processes that I analyse here are described by transition rates like the following, where one microbial type decreases by one individual while another increases by one individual,

$$\underbrace{\phi_j n_j}_{\text{death}}\Big(\underbrace{f_i n_i}_{\text{birth}} + \underbrace{m p_i}_{\text{migration}}\Big), \tag{5}$$

where n_j and n_i are the abundances of two microbial types, and ϕ_j and f_i are death rates and growth rates, respectively. The immigration from the environment has magnitude m, each type with a fraction of immigrants p_i.

Within each chapter, the research focus of the model leads to certain assumptions on this transition rate (Figure 0). In Chapters 1 and 2, I assume

neutral death and growth rate for all types, thus $\phi_i = f_i = 1$ for all microbial types, focusing on the effect of host lifespan. In Chapter 3, I assume infinite host lifespan, rather focusing on the effect of type-specific growth and death rates – i.e. relaxing the assumption of neutral rates.

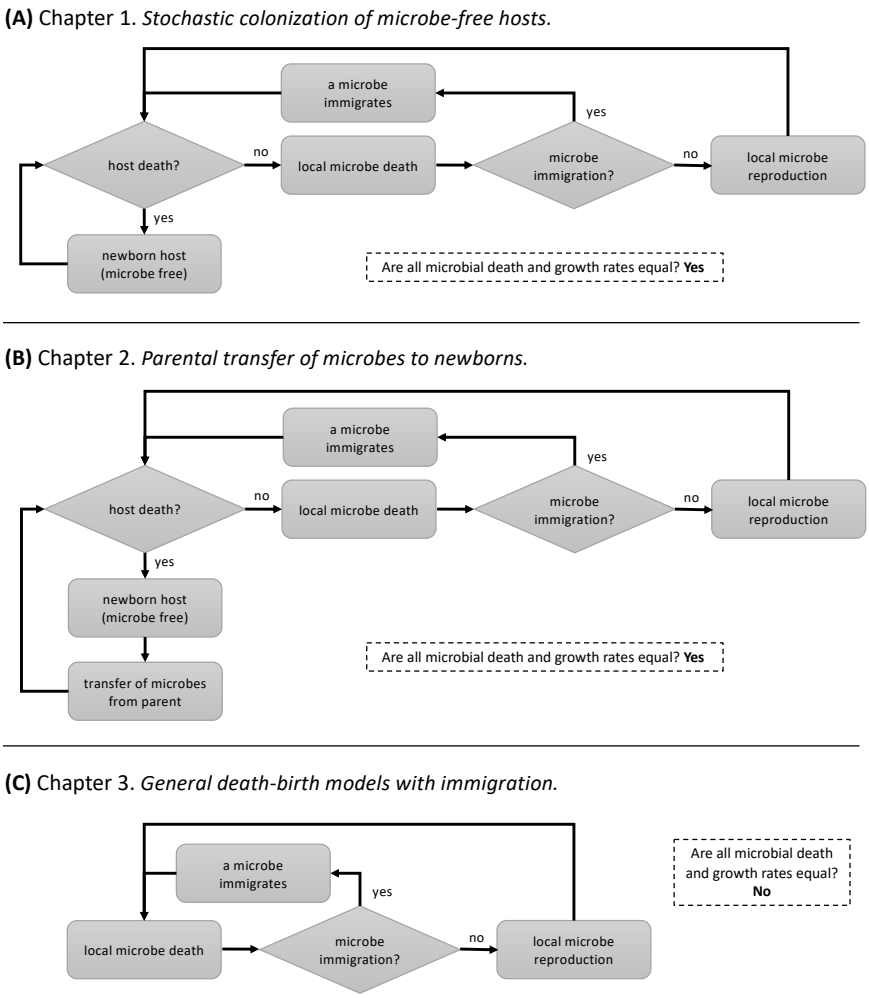

Figure 0: **Ecological processes modeled in this thesis.** A flowchart indicates the sequence of events occurring during a single time-step leading to the dynamics of the microbiome composition. In the diamonds only one of the possible events occurs. Which one occurs depends on the probability of each event. (**A**) All growth and death rates are neutral and the research focus is the effect of the host lifespan on the microbiome composition. (**B**) All growth and death rates are neutral and the research focus is the effect of the parental transfer of microbes to newborn hosts. (**C**). The growth and death rates are not strictly neutral, this relaxation becoming the research focus of the simplified model.

CHAPTER 1

Stochastic colonization of microbe-free hosts

The history of models in macro-ecology is rich. Models that track the abundance of types within a local community have been particularly successful. When these models are applied to host-associated microbiomes, an immediate realization is that compared to conventional habitats, hosts are also living organisms with a finite lifespan. In this Chapter, I investigate how the host lifespan can interfere with the microbiome dynamics and its composition.

This Chapter has been published under the title *Stochastic colonization of hosts with a finite lifespan can drive individual host microbes out of equilibrium* (Zapien-Campos et al., 2020), coauthored by Michael Sieber and Arne Traulsen. The authors' contributions are detailed at the end of the thesis.

1.1 Abstract

Macroorganisms are inhabited by microbial communities that often change through the lifespan of an individual. One of the factors contributing to this change is colonization from the environment. The colonization of initially microbe-free hosts is particularly interesting, as their microbiome depends entirely on microbes of external origin. We present a mathematical model of this process with a particular emphasis on the effect of ecological drift and a finite host lifespan. Our results indicate the host lifespan becomes especially relevant for short-living organisms (e.g. *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Danio rerio*). In this case, alternative microbiome states (often called enterotypes), the coexistence of microbe-free and colonized hosts, and a reduced probability of colonization can be observed in our model. These results unify multiple reported observations around colonization and suggest that no selective or deterministic drivers are necessary to explain them.

1.2 Introduction

Microbial communities inhabit every available habitat on this planet, including the tissues of macroorganisms. For such host-associated communities every host animal constitutes a distinct habitat. Migration between these individual habitats and ecological drift within them play important roles in structuring these communities (Zhou and Ning, 2017). This idea is formalized in the Unified Neutral Theory of Biodiversity where individuals within a community are regarded as ecologically equivalent (Hubbell, 2001). First developed in a macro-ecological context, its application has been extended to microbial populations (Sloan et al., 2006; Woodcock et al., 2007) and host-associated microbiomes (Burns et al., 2016; Adair et al., 2018; Sieber et al., 2019).

What sets host-associated microbiomes apart is that their habitat – the host animal – is itself subject to demographic processes such as reproduction and death. Previous applications of neutral models to microbiome data have generally ignored these host-level processes by assuming essentially static hosts with infinite lifespans, allowing convergence to a long-term equilibrium distribution of microbial abundances, see e.g. (Sieber et al., 2019). However, any real host species will have a finite lifespan that may not allow for the community to settle down on a potential long-term composition. Moreover, differences in the lifespan across host species could obscure comparisons of neutrality across different species. Several authors have fitted one of these 'static host' neutral models (Sloan et al., 2006) to microbiome datasets across multiple host species, finding an overall high resemblance (Burns et al., 2016; Adair et al., 2018; Sieber et al., 2019). However, one of these studies has found much less resemblance for the gut microbiome of *C. elegans* compared to sponges and hydra and speculated that this may be explained by the shorter lifetime of *C. elegans* (Sieber et al., 2019). Others have noted that the worm microbiome might be neutrally assembled, obscured by a transient state far from the neutral long-term equilibrium (Vega and Gore, 2017).

Few studies have explored the effect of host life cycles on the microbiome. Zeng et al. studied the change of microbiome composition under neutrality and discrete host generations, but did not consider microbial dynamics (Zeng et al., 2015). The effect of microbial symbionts, particularly the coevolution under multilevel selection (Van Vliet and Doebeli, 2019), and the effect of the horizontal and vertical acquisition of such microbes have been studied elsewhere (Roughgarden, 2020).

An additional assumption of current neutral models is that hosts contain the same abundance of microbes throughout their lives (Sloan et al., 2006). This is not the case in the gut of important model organisms like

C. elegans (Zhang et al., 2017), *D. melanogaster* (Chandler et al., 2011), and *D. rerio* (Stephens et al., 2016), which are initially microbe-free and only colonized from the environment later. According to the "sterile womb" hypothesis (Perez-Muñoz et al., 2017), also human newborns may be initially microbe-free.

By modelling the change from a microbe-free to a fully microbe-populated state, we study the transients of colonization, and their implications as the lifespan of hosts shortens. Existing models have suggested that neutral models can explain microbial abundances within hosts. Extending these neutral models of a host's microbiome to capture microbial community dynamics during the finite lifespan of a host seems thus natural. We analyse such a model, including the colonization from a microbe-free state and the finite lifespan of hosts. We discuss the dynamical consequences and the connection to experimental observations.

1.3 Model and methods

A nearly-neutral model

We consider multiple hosts (habitats) connected to a pool of microbes. This pool is the subset of environmental microbes capable of colonizing the hosts. Microbial abundances within each host change by three processes: (i) the death of a microbe, giving rise to empty space (ii) a birth-immigration process, when the new empty space is replaced by a microbe, and (iii) host renewal, when a host dies with its microbiome and a new host appears that does not contain any microbe. An illustration of this host-microbiome model is shown in Fig 1.1.

We consider a population of hosts that is sufficiently large to draw statistical conclusions. The microbial community in each host grows dynamically, but with a fixed maximum capacity N. To make this more precise, let n_i be the number of individuals of the *i-th* microbial taxon within a host ($i \geq 1$) and M be the number of taxa. At any time we have $\sum_{i=1}^{M} n_i \leq N$. We reserve the index $i = 0$ for the unoccupied space, namely $n_0 = N - \sum_{i=1}^{M} n_i$. We define $x_i = n_i/N$ as the frequency of the *i-th* taxon within a host and assume $N \gg 1$, such that x_i is continuous and $N-1 \approx N$. Note, that x_0 then denotes the fraction of available space within a host. We assume the death of hosts can be approximated as an event occurring each time step with probability τ, given by the probability of host death-birth events per microbial death-birth event. The limiting case $\tau = 0$ corresponds to infinitely living hosts (as in (Sloan et al., 2006; Sieber et al., 2019)), while $\tau = 1$ corresponds to hosts whose lifespan is as short as the average lifespan of a microbe, leading to almost entirely empty hosts.

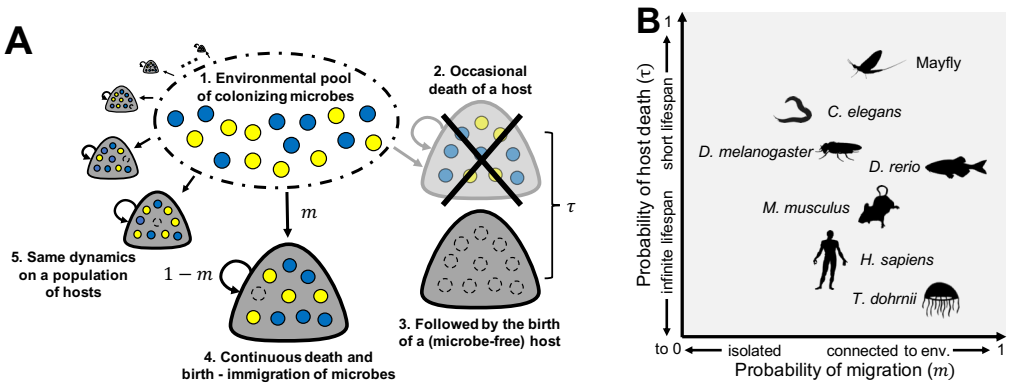

Figure 1.1: **The host-microbiome dynamics in our model and the coordinates of host species in a map of its parameters.** (**A**) Gray blobs indicate hosts, coloured- and empty circles indicate microbes and empty space, respectively. Within the host, microbes go through a death-birth process, with newborns migrating from a pool of colonizing microbes with probability m or being chosen from the same host with probability $1-m$. The pool of microbes includes all microbes capable of living within hosts. Hosts are identical habitats, each with a finite, geometrically distributed lifespan. Each time step, there is a probability τ of a host death followed by the birth of a new host. Newborn hosts contain no microbes. The probability of a host death-birth event is relative to a microbe death-birth event. (**B**) A space of m and τ provides a life-history map on top of which hosts can be located and variables interrogated (see Fig 1.3-1.6). We sketch hypothetical coordinates of multiple host species. In practice however, m and τ may depend on physiological, environmental and behavioural factors. Silhouettes from PhyloPic (`http://phylopic.org`) licensed under Public Domain Dedication 1.0 licenses.

Let us focus on the events within a single host. In each time step, a randomly selected site is changed. This site is either unoccupied space or a microbe. Death is followed by replacement via immigration or birth of a new type. With probability m, its content is replaced by a random microbe from the environment, selected proportionally to its frequency in the pool of colonizers, p_i (note that $p_0 = 0$). With probability $1-m$, it is replaced by a microbe from the same host, selected proportionally to the fitness $(1+\alpha_i)x_i$ of the reproducing microbe – or it is replaced by unoccupied space with probability proportional to $(1+\alpha_0)x_0$. The fitness parameter α_i describes deviations from strict neutrality, where proliferation of microbe i is promoted ($\alpha_i > 0$) or impeded ($\alpha_i < 0$). The parameter α_0 controls how rapidly unoccupied space within a host is filled with microbes. This determines the level of resistance a host poses to be occupied by microbes, or in other words, how favourable the host environment is for microbial reproduction and persistence. For $\alpha_0 > 0$, hosts pose an increased resistance to the internal microbes, while $\alpha_0 < 0$ decreases such resistance. The acceptable range of

α_i and α_0 ranges from -1 to infinity. The resulting stochastic process for a given host can be described by the probabilities of events after one time step,

$$P[x_i \to \delta_{i,0}] = \tau \tag{1.1a}$$

$$P[x_i \to x_i + \tfrac{1}{N}] = (1-\tau)\,(1-x_i)\left(mp_i + (1-m)\frac{(1+\alpha_i)\,x_i}{\sum_j (1+\alpha_j)\,x_j}\right) \tag{1.1b}$$

$$P[x_i \to x_i - \tfrac{1}{N}] = (1-\tau)\,x_i\left(m(1-p_i) + (1-m)\left(1 - \frac{(1+\alpha_i)\,x_i}{\sum_j (1+\alpha_j)\,x_j}\right)\right) \tag{1.1c}$$

$$P[x_i \to x_i] = 1 - P[x_i \to \delta_{i,0}] - P[x_i \to x_i + \tfrac{1}{N}] - P[x_i \to x_i - \tfrac{1}{N}], \tag{1.1d}$$

where Eq (1.1a) describes the probability of a host death event: All microbial frequencies are set to zero, i.e. $x_i \to 0$ for $i \geq 1$. At the same time, a new empty host arises, corresponding to $x_0 \to 1$. This is captured by $\delta_{i,0}$, the Kronecker delta (1 for $i = 0$ and 0 otherwise). The three other probabilities require that the host survives, which occurs with probability $1 - \tau$. For a microbial taxon i, Eq (1.1b) describes the probability of increase by immigration or reproduction within the host, and Eq (1.1c) describes the probability of decrease derived from other taxa immigration, reproduction within the host, or their inability to reproduce. For $i = 0$, Eq (1.1b) and Eq (1.1c) describe the probability of increasing and decreasing the unoccupied space, respectively. Finally, Eq (1.1d) indicates the probability of no change. Focusing on the effect of ecological drift we fix the microbial fitness $\alpha_i = 0$ (for $i \geq 1$) for the remainder of the manuscript.

Probabilities in Eq (1.1) change considerably through time. For example, because hosts are largely empty at birth, unoccupied space decreases rapidly as $P[x_0 \to x_0 - \frac{1}{N}] \gg P[x_0 \to x_0 + \frac{1}{N}]$, while the microbial frequencies increase because $P[x_i \to x_i - \frac{1}{N}] \ll P[x_i \to x_i + \frac{1}{N}]$.

For $\tau = 0$ the probabilities are as in Sloan et al.'s (Sloan et al., 2006), which becomes a good approximation when the time scale of reproduction on the microbial level is much faster than the time scale of reproduction on the host level. We focus on the dynamics of the probability density of x_i, $\Phi_i[x_i, t]$. Due to the differences in p_i and α_i, $\Phi_i[x_i, t]$ can be different for all microbial taxa i. This can be approximated in the large N limit by a Fokker-Planck equation (see Appendix B.1), with t being measured in the number of microbial death-birth events. Writing down the equations for

unoccupied space x_0 and microbes separately we have

$$\frac{\partial}{\partial t}\Phi_0[x_0,t] = \frac{\partial}{\partial x_0}\left[-a_0[x_0]\Phi_0[x_0,t] + \frac{1}{2}\frac{\partial}{\partial x_0}b_0^2[x_0]\Phi_0[x_0,t]\right] + \tau\left(\delta_{x_0,1} - \Phi_0[x_0,t]\right) \tag{1.2a}$$

$$\frac{\partial}{\partial t}\Phi_i[x_i,t] = \frac{\partial}{\partial x_i}\left[-a_i[x_i]\Phi_i[x_i,t] + \frac{1}{2}\frac{\partial}{\partial x_i}b_i^2[x_i]\Phi_i[x_i,t]\right] + \tau\left(\delta_{x_i,0} - \Phi_i[x_i,t]\right), \tag{1.2b}$$

where $a_i[x_i]$ describes the deterministic part of the change and $b_i^2[x_i]$ describes changes due to randomness (Gardiner, 2004). The term $a_i[x_i]$ is calculated as the first moment of Δx_i, the expectation $\langle \Delta x_i \rangle$,

$$a_0[x_0] = (1-\tau)\left(-mx_0 - (1-m)x_0\left(1 - \frac{1+\alpha_0}{(1+\alpha_0)x_0 + (1-x_0)}\right)\right)\frac{1}{N} \tag{1.3a}$$

$$a_i[x_i] = (1-\tau)\left(m(p_i - x_i) - (1-m)x_i\left(1 - \frac{1}{(1+\alpha_0)x_0 + (1-x_0)}\right)\right)\frac{1}{N}. \tag{1.3b}$$

The term $b_i^2[x_i]$ is calculated as the second moment of Δx_i, the expectation $\langle (\Delta x_i)^2 \rangle$,

$$b_0^2[x_0] = (1-\tau)\left(mx_0 + (1-m)x_0\left(1 + \frac{(1+\alpha_0)(1-2x_0)}{(1+\alpha_0)x_0 + (1-x_0)}\right)\right)\frac{1}{N^2} \tag{1.4a}$$

$$b_i^2[x_i] = (1-\tau)\left(m(p_i + x_i - 2p_i x_i) + (1-m)x_i\left(1 + \frac{(1-2x_i)}{(1+\alpha_0)x_0 + (1-x_0)}\right)\right)\frac{1}{N^2}. \tag{1.4b}$$

For $\tau \to 0$, the last terms in Eq (1.2) vanish, recovering the usual Fokker-Planck equation of the neutral model without host death (Sloan et al., 2006), while for $\tau > 0$ these additional terms describe the change due to host death, where a new, microbe-free host appears.

Although individual hosts constantly change their microbiome through the process of microbial death birth-immigration and host death, the collection of transient host states becomes stationary at the population level. This stationary distribution is found setting the time derivative of Eq (1.2) equal

to zero,

$$0 = \frac{d}{dx_0}\left[-a_0[x_0]\Phi_0[x_0] + \frac{1}{2}\frac{d}{dx_0}\left[b_0^2[x_0]\Phi_0[x_0]\right]\right] + \tau\left(\delta_{x_0,1} - \Phi_0[x_0]\right) \quad (1.5a)$$

$$0 = \frac{d}{dx_i}\left[-a_i[x_i]\Phi_i[x_i] + \frac{1}{2}\frac{d}{dx_i}\left[b_i^2[x_i]\Phi_i[x_i]\right]\right] + \tau\left(\delta_{x_i,0} - \Phi_i[x_i]\right) \quad (1.5b)$$

The Fokker-Planck approximation has several benefits: It provides an intuition of the stochastic process at the population level and the effect of host death (τ), a direct connection to models not considering finite host lifespans (Sloan et al., 2006), and the possibility to frame the process in the broader stochastic processes literature (Gardiner, 2004).

An alternative interpretation of the stochastic process is provided by (Evans and Majumdar, 2011)

$$\Phi_i[x_i] = \int_0^\infty \Phi_i[x_i, t_r]|_{\tau=0}\Psi[t_r]dt_r,$$

where $\Phi_i[x_i]$ results from considering all the possible distributions of the time-dependent death-birth process of microbes without host dynamics, $\Phi_i[x_i, t_r]|_{\tau=0}$, influenced by the distribution of death-birth time of hosts, $\Psi[t_r]$. The distribution of these resetting events is given by

$$\Psi[t_r] = \tau e^{-\tau t_r} \quad (1.6)$$

This equation will help us to compare our model and individual-based simulations.

Now we aim to solve Eq (1.5), where a major challenge arises from the additional terms capturing the host death-birth events, which correspond to a resetting of the local microbial community. Such resetting events are often referred to as "catastrophes" in the Mathematics literature and research has focused on finding closed form solutions of the corresponding discrete problem derived from the master equation using first order transition probabilities (Kyriakidis, 1994; Swift, 2001; Chao and Zheng, 2003). In physics, this is called diffusion-drift with resetting and its Fokker-Planck approximation and zero order transition probabilities have been used to find closed form solutions and compute quantities of interest (Meylahn et al., 2015; Evans and Majumdar, 2011). Our model considers density-dependent transition probabilities, i.e. second order effects. Although these provide a well defined system at the boundaries $x_i = \{0, 1\}$, they complicate finding a closed form

solution of $\Phi_i[x_i]$ tremendously. Approximating the solutions numerically using the finite differences and finite element methods (Kumar and Narayanan, 2006) is possible.

We solved this equation numerically to query the parameter space (Kumar and Narayanan, 2006). However, we found our implementation could lead to numerical errors that were large and inconsistent in some cases, especially as $\tau \to 0$. As it proved more robust numerically (Fig Sup B.1, Fig Sup B.2 and Fig Sup B.3), we used the master equation (see Appendix B.1) to produce our figures instead,

$$\Delta\vec{\Phi}_i[\vec{x_i}, t+1] = T_i\vec{\Phi}_i[\vec{x_i}, t], \tag{1.7}$$

where $\Delta\vec{\Phi}_i$ is the change of the distribution during one time step. In this case the distribution at a given time is represented by the vector $\vec{\Phi}_i[\vec{x_i}, t]$, whose entries correspond to the probability densities of $x_i \in \{0, 1/N, 2/N, ..., 1\}$. Upon multiplying by the matrix of transition probabilities, T_i, the time change of the distribution is obtained. Because only transitions are considered, the main diagonal of T_i equals zero, while the upper and lower diagonals equal Eq (1.1b) and Eq (1.1c), respectively. Host death is reflected in additional non-zero probabilities, τ, at the first column for microbial taxa ($i \geq 1$) or last column for unoccupied space ($i = 0$). The non-trivial stationary distribution $\vec{\Phi}_i[\vec{x_i}]$ occurs for $\Delta\vec{\Phi}_i[\vec{x_i}, t+1] = \vec{0}$, corresponding to the eigenvector of T_i with eigenvalue zero. We used this method to compute the stationary distribution in Python 3.6.

If numerical problems emerged solving Eq (1.7), we focused on solving $\vec{\Phi}_i[\vec{x_i}, t+1] = R_i\vec{\Phi}_i[\vec{x_i}, t]$ for $\vec{\Phi}_i[\vec{x_i}, t+1] = \vec{\Phi}_i[\vec{x_i}, t]$ instead. Here R_i, the probability matrix, is identical to T_i, except at the main diagonal where it equals Eq (1.1d). The stationary distribution corresponds to the eigenvector of R_i with eigenvalue one.

Stochastic simulations

To study the transient dynamics of colonization and test our stationary estimation, we performed individual-based simulations. These were performed for 500 hosts, $N = 10^4$, two equally abundant microbial taxa in the pool of colonizers, $p_1 = p_2 = 0.5$, and initially sterile hosts ($x_0 = 1$ and $x_1 = x_2 = 0$ as initial condition). We varied the values of migration (m) and rate of occupation of empty space (α_0).

Difference between models

To compare models considering finite ($\tau > 0$) and infinite host lifespan ($\tau = 0$), we calculated the total difference between their stationary distributions,

$\Phi_i[x_i]|_{\tau>0}$ and $\Phi_i[x_i]|_{\tau=0}$, for all x_i

$$\frac{1}{2}\sum_{x_i}\left|\Phi_i[x_i]|_{\tau>0} - \Phi_i[x_i]|_{\tau=0}\right| \tag{1.8}$$

This difference, ranging from 0 to 1, will equal zero only if for all x_i, the two distributions are identical, $\Phi_i[x_i]|_{\tau>0} = \Phi_i[x_i]|_{\tau=0}$.

Probability of microbe-free, colonized and fully-colonized hosts

To analyse when a particular microbial taxon will not be observed in a host, i.e. its probability of non-colonization, we calculated

$$P\left[x_i < \frac{1}{N}\right] = \Phi_i[0], \tag{1.9}$$

where $1/N$ is the minimum observation limit and $P[x_i \geq 1/N] = 1 - P[x_i < 1/N]$ is the probability of colonization by microbe i.

On the other hand, to analyse when a particular microbial taxon will fully occupy a host, we calculated

$$P\left[x_i > \frac{N-1}{N}\right] = \Phi_i[1], \tag{1.10}$$

where $\frac{N-1}{N}$ is the maximum observation limit, and $P[x_i \leq (N-1)/N] = 1 - P[x_i > (N-1)/N]$ is the combined probability of partial and non-colonization.

Finally, the quantities $P\left[x_0 < 1/N\right]$ and $P\left[x_0 > (N-1)/N\right]$, indicate the probability of hosts full of microbes and the probability of hosts free of microbes, respectively.

Alternative microbiome states

To assess the modality of the distribution $\Phi_i[x_i]$, i.e. alternative microbiome states, we identified the maxima of the distribution of its numerical solution for varying parameters. The distribution can be unimodal, with the maximum located at one of the boundaries or between them, $x_i = \{0, x_i^*, 1\}$, or bimodal, by a combination of the former. We classified these states and calculated the magnitude of their maxima.

Comparison between the model and simulated data

In order to evaluate our model, we compared it to stochastic simulations (Fig Sup B.1, Fig Sup B.2 and Fig Sup B.3). As mentioned above, we

simulated hosts individually. However, our model provides a population description for overlapping generations. Therefore, we sampled single time steps of the colonization trajectories according to Eq (1.6), which indicates the probability of a host death-birth event through time. The distribution of the simulated sampled set was then compared to our theoretical model predictions.

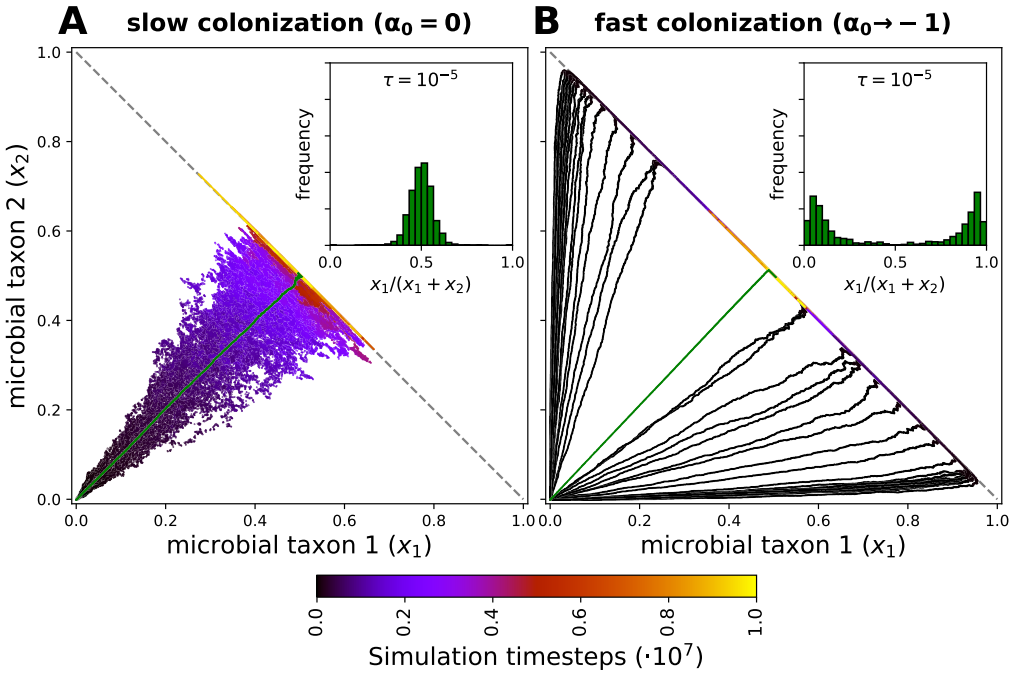

Figure 1.2: **Individual-based simulations of colonization for two neutral microbial taxa.** The colonization trajectory of 50 microbe-free hosts is shown, with colors indicating the time of sampling. Each trajectory is composed of 10^4 points, with the green line indicating the mean and the dashed line indicating full colonization. Insets show the distribution of $x_1/(x_1 + x_2)$ for time steps sampled according to a host death-birth probability $\tau = 10^{-5}$ in Eq (1.6). (**A**) When hosts are colonized slowly, the trajectories maintain a mean frequency given by the pool of colonizers (p_i), but show an increased standard deviation before full colonization, which decreases later to reach $x_i \approx p_i$ at the long-term equilibrium. (**B**) When hosts are colonized rapidly, the mean frequency across many hosts is conserved, but the distribution becomes bimodal as a result of the fast proliferation of the first colonizer, and a slower convergence to the long-term equilibrium. The inset shows the distribution of taxon 1 over the simulated time. For finite host lifespans and fast colonization, such dynamics can produce alternative microbiome states at the population level (inset in B). Sampling was performed every 10 time steps in a simulation of 10^7 time steps. Other parameters: $N = 10^4$, $m = 0.01$, $p_1 = p_2 = 0.5$, and $\alpha_1 = \alpha_2 = 0$.

Code availibility

The Python code for simulations, numerical solution of the model and figures is available at `https://github.com/romanzapien/microbiome-hostspan`.

1.4 Results

The dynamics of colonization affects the microbiome of finite-living hosts, but not of infinite-living habitats

The formation of a microbiome goes through several stages. Analytically, much of the focus has been on its long-term equilibrium, assuming hosts with infinite lifespan. Much less is known about the transient stage. Fig 1.2 shows two illustrative individual-based simulations, where hosts are colonized by two neutral microbial taxa, going from a microbe-free to a microbe-occupied state. The dynamics is qualitatively different depending on α_0: For $\alpha_0 = 0$, the host is colonized by the two microbes at the same time, leading to a unimodal distribution that is similar to the long term equilibrium even during the transient. For $\alpha_0 < 0$ empty space is occupied more rapidly compared to the dynamics between microbes. This leads to a situation where one microbial strain dominates the host until the host is fully colonized, leading to a bimodal distribution in the colonization of hosts. Only on a much longer timescale, this distribution is replaced by the unimodal distribution characteristic for the long term equilibrium.

Given a low rate of external colonization ($m \to 0$), the time required for full colonization will be shorter than that to reach the long-term equilibrium. Such difference will increase even further for rapid colonization, $\alpha_0 < 0$. When considering a finite host lifespan ($\tau > 0$), this difference in timescales will influence the expected microbiome composition. Interestingly, for shorter lifespans, the host population might be multimodal and only partially colonized (Fig 1.2B). Moreover, for sufficiently small external colonization and short host lifespan, coexistence of colonized and microbe-free individuals is expected (Fig Sup B.4).

From a microbial point of view, the results shown here occur in a completely neutral context. They can also be generalized to cases with many microbial taxa. A non-neutral dynamics of the microbes ($\alpha_i \neq 0$) will modify the stationary distribution, i.e. they will not only depend on the frequency in the pool of colonizers (p_i) and host lifespan (via τ). Instead, asymmetries of the multimodality and differential colonization are expected once $\alpha_i \neq 0$ is assumed.

A short host lifespan influences the microbiome

We quantified the change of the stationary distribution caused by a finite host lifespan systematically. Using the stationary distribution of the frequency, $\Phi_i[x_i]$, we compared the predictions assuming hosts with infinite lifespan ($\tau = 0$) against those with hosts with finite lifespan ($\tau > 0$). Such comparisons were done for multiple migration probabilities (m), frequencies in the pool of colonizers (p_i), and rates of empty space occupation (α_0). As explained in the Methods, Eq (1.8), we express the results as the difference between the stationary distributions.

Fig 1.3 and 1.4 show the results of the microbial load (total microbial frequency) and frequency of a particular microbe, respectively. Within the range of m and τ analysed, the difference is always greater than zero, indicating the importance of τ in our model and the predictions arising from it. Only for $\tau \to 0$, full agreement is expected.

Regarding the microbial load, infinitely living hosts ($\tau = 0$) provide enough time for them to be fully colonized and for the distribution of microbes to reach an equilibrium. In contrast, a finite lifespan ($\tau > 0$) might not allow full colonization before host death. For a slow occupation of empty space ($\alpha_0 = 0$) the difference increases with shorter lifespan (large τ) and reduced migration (small m), Fig 1.3A. In this case, the model with $\tau = 0$ predicts a distribution centered at frequency 1 decaying towards 0, while the model with $\tau > 0$ predicts a sharp maximum centered at frequency 0 decaying towards 1. In contrast, rapid occupation of empty space ($\alpha_0 < 0$) causes the difference to decrease and to become increasingly independent of m (Fig 1.3B). This occurs because the time for colonization, i.e. host lifespan, becomes more relevant than migration, as successful migrants are increasingly likely to proliferate within hosts.

For a specific microbial taxon, infinitely living hosts ($\tau = 0$) allow the frequency in the hosts to reach that in the pool of colonizing microbes (p_i). However, a restricted, finite lifespan ($\tau > 0$) might not allow to reach this value. In our model, the relevance of τ increases with its magnitude, but not independently of m. The maximum difference between the two distributions occurs for short lifespan (large τ) and large migration (larger m) as $p_i \to 0$ (Fig 1.4B-C). In this region, the model with $\tau = 0$ predicts a distribution centered at $x_i \approx p_i$, while the model with $\tau > 0$ predicts a distribution centered at $x_i = 0$ decaying towards 1. Finally, for a single colonizing taxon ($p_i = 1$, Fig 1.4A) the difference increases analogously to Fig 1.3A, i.e. the difference increases for smaller migration and shorter lifespan.

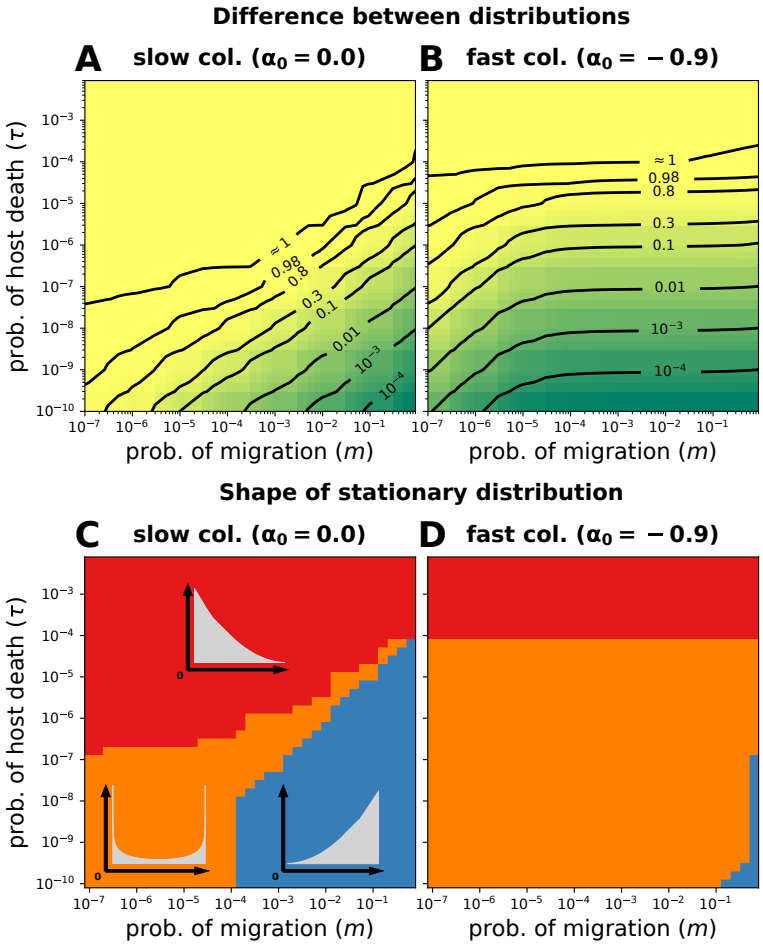

Figure 1.3: **Microbial load.** (**A-B**) The difference between models with finite ($\tau > 0$) and infinite ($\tau = 0$) host lifespan is shown, Eq (1.8). (**A**) For a slow occupation of empty space, the difference is maximal for small migration (m) and large τ as the model with $\tau = 0$ predicts a distribution centred at frequency 1 decaying towards 0, whereas the model with $\tau > 0$ predicts a distribution centred at frequency 0 decaying towards 1. For a fixed τ the difference is always greater for smaller m. Only for $\tau \gtrsim 10^{-4}$ the difference is maximal and independent of m. Finally, a smaller τ always approximates the models; nonetheless within the range analysed the difference is always greater than zero. (**B**) A faster occupation of empty space decreases the difference and makes it increasingly independent of m, as τ dominates the predictions of the model. (**C-D**) The distributions are classified according to their number of maxima (unimodal or bimodal) and location (0 and 1). (**C**) A slow occupation of empty space results in microbe-free hosts being the maximum for short host lifespans (large τ), fully colonized hosts for large migration (m) and small τ, or microbe-free and microbe-occupied hosts simultaneously for small m and τ. The bimodality results from a limited migration preventing all the hosts from being colonized but over a host lifespan sufficient for successful colonizers to occupy host fully. (**D**) A faster occupation of empty space increases the bimodality region at the expense of the unimodal cases. In this case, $\alpha_0 \to -1$ favours the microbe-occupied maximum. When classifying the distributions, any probability smaller than 10^{-9} was considered as zero. Other parameters: $N = 10^4$. We use Eq (1.5a) where no definition of p_i and a_i is required.

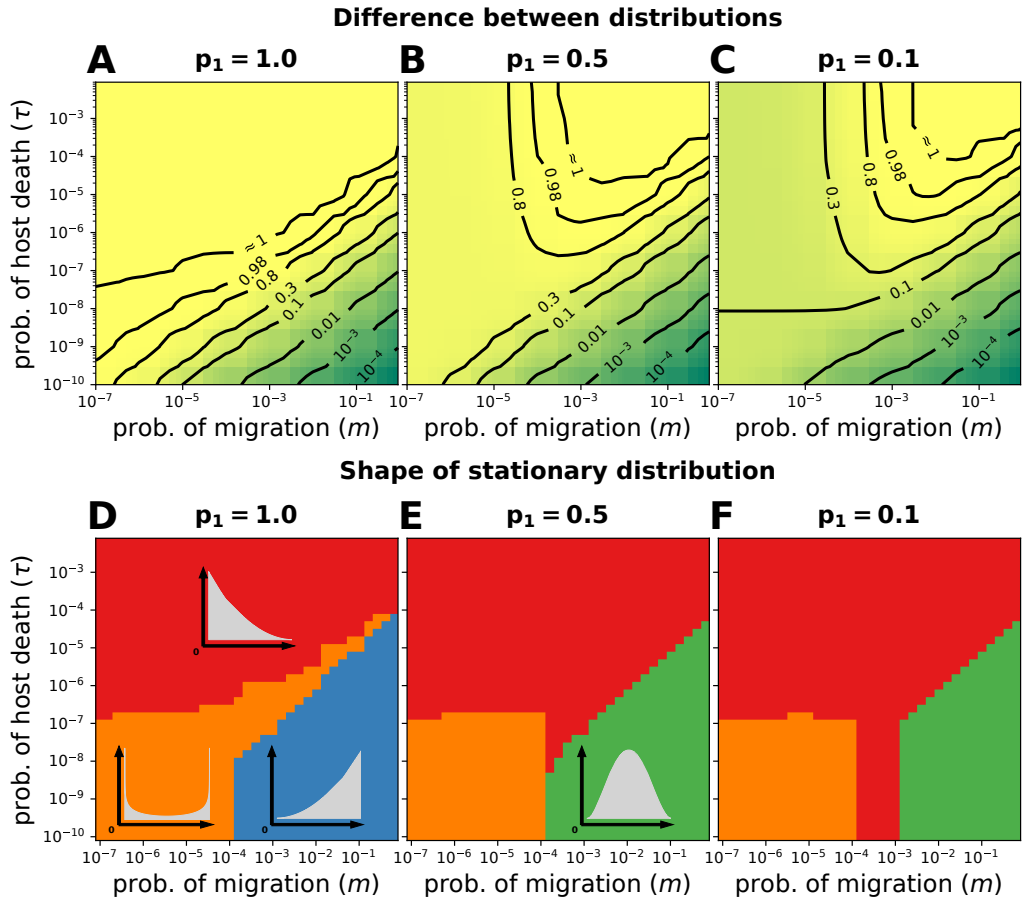

Figure 1.4: **Microbial taxon 1.** p_1 indicates the frequency of microbial taxon 1 in the pool of colonizers. (**A**-**C**) The difference between models with finite ($\tau > 0$) and infinite ($\tau = 0$) host lifespan is shown, Eq (1.8). (**A**) A single colonizing taxon follows the same pattern shown in Fig 1.3A. (**B**-**C**) The maximal difference of a less abundant colonizing microbe changes in the direction of larger m. (**D**-**F**) The distributions are classified according to their number of maxima (unimodal or bimodal) and location (0, 1, and an internal maximum). (**D**) A single colonizing taxon mirrors Fig 1.3C, and bimodality is prevalent. (**E**-**F**) Less abundant taxa have a decreased probability of colonization, and an internal maximum emerges for large m and long host lifespan (small τ), whose location is influenced by the frequency in the pool of colonizers (p_1). When classifying the distributions, any probability smaller than 10^{-9} was considered as zero (Other parameters $N = 10^4$ and $\alpha_0 = \alpha_1 = 0$). Fig Sup B.7 shows how the frequency x_1 changes as we increase τ for $m = 10^{-3}$.

Microbe-free, colonized hosts, and their coexistence are expected

A major consequence of a host finite lifespan is the coexistence of hosts with various degrees of colonization, including microbe-free hosts. We calculated the

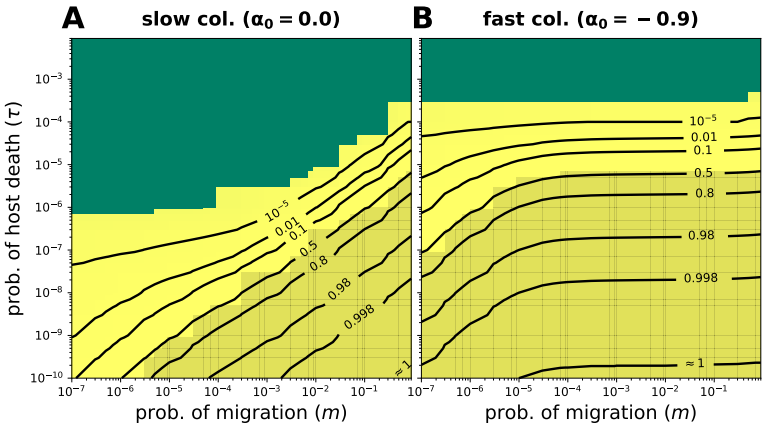

Figure 1.5: **Probability of full colonization in the stationary distribution.** The probability of full colonization $P[x_0 < 1/N]$ is shown, Eq (1.9), where $1/N$ is the minimum observation limit. For short host lifespans (large τ), partial colonization is more common than full colonization. (**A**) For long host lifespans (small τ) or large migration (m), m has an effect on the probability, but this is lost as τ is larger and m smaller. (**B**) A faster occupation of empty space increases the probability of full colonization, but migration (m) influence is now restricted to long host lifespans (small τ) and small m. I.e. the host lifespan (τ) becomes the most relevant parameter. Other parameters: $N = 10^4$. We use Eq (1.5a) where no definition of p_i and a_i is required.

probability of full colonization in the stationary distribution, i.e. $P[x_0 < 1/N]$ (Eq (1.9)), for different parameters given a certain capacity for microbes (N).

Fig 1.5 shows the effect of m, τ, and α_0 on the probability of full colonization. Different parameter combinations can result in the same probability of full colonization. Partial colonization is the most likely state for short host lifespans (large τ). Only for long living hosts (small τ), both death probability τ and migration m are important, with m having a larger impact on the distribution when it is larger (Fig 1.5A). Finally, a faster occupation of empty space ($\alpha_0 < 0$) makes the probability of full colonization less dependent on m and increases it for shorter living hosts (larger τ), i.e. the coexistence with partially colonized hosts becomes less likely (Fig 1.5B).

The results shown in Fig 1.5 depend heavily on the capacity for microbes of the host (N). Decreasing N causes the hosts to be fully colonized quicker; thus partially colonized hosts will be observed for shorter host lifespans (larger τ), slower occupation of empty space (larger α_0), and less migration (smaller m), Fig Sup B.5. The opposite is expected for larger N.

As shown by our calculations, Fig Sup B.6, we argue that even microbe-free hosts might not be an experimental artefact, but an inherent outcome of the host colonization process in some host-microbiome systems (Hammer et al., 2017; Obadia et al., 2017), even under neutral (i.e. non-selective) conditions

(Vega and Gore, 2017). This might be evident for short living hosts, but less so for longer lifespans. In such case, its experimental observation might be possible only for large samples of hosts.

Rapid proliferation of the first colonizer can result in alternative microbiome states

We have noted previously the existence of multimodal distributions in the transient colonization, and how these prevail in the stationary distribution due to the finite lifespan of hosts (Fig 1.2). A particular microbial taxon might either succeed or fail to colonize a host, leading to the coexistence of hosts with alternative microbiome states. Moreover, in specific cases all possible microbes could succeed or fail to colonize a host, allowing the coexistence of microbe-free and microbe-occupied hosts. These extremes can have similar or different magnitudes, as shown in Fig 1.5 and Fig Sup B.6.

Fig 1.3C-D shows the stationary distribution of microbial load for different rates of empty space occupation, α_0. Firstly, a large host death-birth probability (τ) causes hosts to be rarely colonized; hence most remain microbe-free, so $x_0 = 1$ is the only maximum. Secondly, a large migration (m) and small τ provides enough time for hosts to be fully colonized, so $x_0 = 0$ is the only maximum. Finally, the processes of limited migration and long host lifespan combine to define a region where bimodality is expected (Fig 1.3C). The magnitude of the maxima and region of bimodality are influenced by α_0 (Fig 1.3D), with $\alpha_0 \to -1$ favouring the microbe-occupied over the microbe-free state (Fig 1.5 and Fig Sup B.6).

Similarly, Fig 1.4D-F shows the stationary distribution for various frequencies of a microbial taxon in the pool of colonizers (p_i) and $\alpha_0 = 0$. A qualitative description of the complete distributions (see Fig Sup B.7) is shown. Again, bimodality is expected for small m and large τ. Many microbes do not colonize, but successful colonizers proliferate to occupy hosts entirely. The bimodality region is shaped by p_i. A single colonizer ($p_i = 1$, Fig 1.4D) mirrors Fig 1.3C. In contrast, $p_i < 1$ has the effect of vanishing the bimodality if m or τ are larger (Fig 1.4E-F). Outside this region, a large τ causes most hosts to be microbe-free, so $x_i = 0$ is the only maximum. However, a larger m and smaller τ make $x_i = 1$ the single maximum if $p_i = 1$, or an internal maximum if $p_i < 1$. Finally, the split into alternative states might be reinforced if empty space is occupied more rapidly, $\alpha_0 < 0$ (Fig 1.2 and Fig Sup B.4). This results from a limited migration and rapid proliferation of the first colonizer. Although the alternative states could be transient for long-living hosts, they might persist for short-living ones.

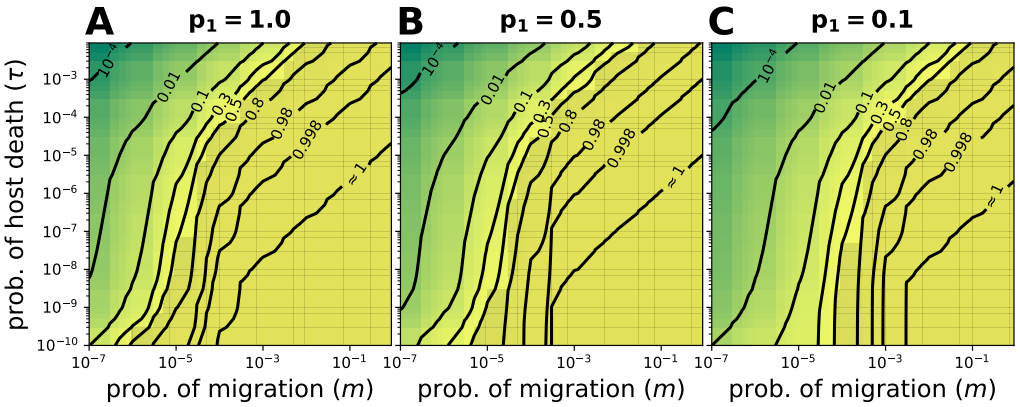

Figure 1.6: **Probability of colonization of microbial taxon 1 in the stationary distribution.** p_1 indicates the frequency of microbial taxon 1 in the pool of colonizers. The probability that a particular microbe is present, $P[x_1 \geq 1/N]$, is shown, where $1/N$ is the minimum observation limit. (**A**) A single microbial taxon colonizes for a large combination of migration (m) and probability of host death-birth (τ). The probability increases with m and with longer lifespan (small τ). (**B-C**) For less abundant colonizing microbes, the probability is reduced. Fig Sup B.8 shows the effect of p_1 on $P[x_1 \geq 1/N]$. Other parameters: $N = 10^4$ and $\alpha_0 = \alpha_1 = 0$.

By reducing the colonization probability, the finite host lifespan makes the core microbiome context-dependent

Previous research has focused on defining the set of microbial taxa consistently observed in a given host species. This is often called the core microbiome. In our model, stochastic colonization reduces the probability of observing a taxon in all hosts (Fig 1.6). Importantly, this is not caused by any kind of selection or competition, but by migration (m), time for colonization (via τ), capacity for microbes (N), and the frequency of a colonizing taxon (p_i) alone. Fig 1.6 shows the probability of observing a microbial taxon within a host, $P[x_i \geq 1/N]$, for different values of m, τ, and a fixed N. For the values of p_i shown, the contour lines increasingly depend on τ for larger τ. Successful colonization is more prevalent whenever m is larger and τ smaller, for microbes down to a frequency of $p_i = 0.1$. Nonetheless, even a single colonizing taxon could not consistently be observed for some m and τ (Fig 1.6A and Fig Sup B.8). Finally, a smaller microbial frequency in the pool of colonizers (p_i) reduces the overall colonization probability (Fig 1.6B-C, smaller values are shown in Fig Sup B.8).

These results suggest that under neutral dynamics, the observed frequency of microbes within hosts, i.e. the colonization probability, cannot be universally used to define a core microbiome, as the frequency of readily colonizing

taxa depends on host and microbial features.

1.5 Discussion

Although microbes are ubiquitous in nature (Bar-On et al., 2018), including the human body (Sender et al., 2016), it remains to be answered which microbes not only transit from the environment to the hosts but also persist in or on them. Our understanding of these processes relies on identifying the factors underlying host colonization.

We have introduced a stochastic model along the lines suggested by (Miller et al., 2018), where migration and death-birth processes of microbes within hosts with finite lifespans can produce a range of colonization dynamics and distinctly different microbiomes – even when there is no selection at all (Fig 1.1). A key assumption in our model is the absence of inheritance of microbes (van Opstal and Bordenstein, 2015), as hosts are colonized after birth from the environment only. In this context, the microbiome is driven by the frequency in the pool of colonizers. This frequency (which is constant in our model) does not need to be the frequency of an environmental microbe, but can more generally be a function of it. Several organisms including *D. rerio* (Stephens et al., 2015), *C. elegans* (Vega and Gore, 2017), and *D. melanogaster* (Obadia et al., 2017) might be colonized from the environment only. Others have weak inheritance (Björk et al., 2019), or might be microbe-free prior to birth, like humans (Perez-Muñoz et al., 2017). Many host species will also inherit their microbes from their parents.

Critical to colonization in our model is the magnitude of the microbial migration from the environment to the hosts (m) (Miller et al., 2018). As observed in the gut of *D. rerio* (Burns et al., 2017), microbial migration could overwhelm other host selective and non-selective processes. In addition, we have combined the host lifespan with a constant microbial cell doubling time (Gibson et al., 2018) to define τ as the parameter of timescale separation between hosts and microbes. This serves as an indicator of the relevance of a host population dynamics for the microbiome dynamics. In agreement with (Stephens et al., 2015), we observe that a limited migration imposes a bottleneck on the colonizers, which combined with a finite host lifespan might produce complicated colonization patterns (Fig 1.2 and Fig Sup B.4). The parameters m and τ have allowed us not only to classify the stationary colonization distributions (Fig 1.3-1.4), but also to quantify the relevance of the finite host lifespan in our model (Fig 1.3-1.4).

The parameters m and τ can be inferred from data. Alternatively, prior knowledge of the host lifestyle can give us intuition. For example, given the short lifespan of *C. elegans* a large τ is expected; while its feeding mechanism

might pose a bottleneck, suggesting a small m. In principle, m can range from 0 (no environmental microbes going in) to 1 (only external migration and no internal reproduction). This range is spanned by previous studies that estimated this parameter for multiple species (Burns et al., 2016; Adair et al., 2018; Sieber et al., 2019).

Sloan et al. (Sloan et al., 2006) developed a neutral model to estimate the equilibrium distribution of a microbiome in an infinite-living habitat. Several studies have fit this model to data of different host species (Burns et al., 2016; Adair et al., 2018; Sieber et al., 2019). However, based on our results for hosts with varying lifespans, we predict that Sloan et al.'s model will perform poorly for hosts with short lifespans, e.g. *D. rerio*, *D. melanogaster*, and *C. elegans*, impeding comparisons of neutrality between host species (Fig 1.3-1.4). On top of that, the average microbiome of all sampled hosts might be a transient state, not the long-term equilibrium that is assumed when fitting the model. These problems are expected to be even more pronounced for low frequency microbial taxa (Fig 1.6 and Fig Sup B.8), and small host populations samples.

As going from a microbe-free to a colonized state might affect the expected stationary distribution in hosts with finite lifespans, we included the occupation of empty space by microbes in our model explicitly. Then, we computed the probabilities of observing microbe-free (Fig Sup B.6), fully colonized hosts (Fig 1.5), and their coexistence (Fig 1.3). Interestingly, there is building evidence of individuals with microbe-free guts coexisting in *D. melanogaster* (Obadia et al., 2017), *C. elegans* (Vega and Gore, 2017), and caterpillars where a microbe-free state might be prevalent (Hammer et al., 2017) – supporting our results. We argue that in such host species, both a low microbial migration and short host lifespan might be causative (Hammer et al., 2019).

We have also observed alternative microbiome states. In other words, subsets of hosts whose microbiome is dominated by different microbial taxa (Fig 1.2). Our results suggest this might occur for low microbial migration and short host lifespan (Fig 1.4). Recently, (Vega and Gore, 2017) have observed alternative microbiome states occurring in *C. elegans* when this is colonized by two neutral *Escherichia coli* strains. The implications of our results go beyond colonization, as they predict priority effects (Sprockett et al., 2018), life history (Martínez et al., 2018), and timing to be important conditions for any host control mechanism. Furthermore, we provide a generative process for the emergence of different microbiome states in the gut (Arumugam et al., 2011), that does not rely on selection, interaction networks or environmental change (Gibson et al., 2016; Gonze et al., 2017). Our results support the current view that the enterotypes often discussed are indeed states contained in a continuum of colonization (Costea et al., 2018).

Finally, we have addressed the issue of identifying a core microbiome. In contrast to the present interest on identifying this subset of microbes (Björk et al., 2018), we argue that intrinsic features of the colonization process might impede finding a consistent subset. Specifically if the observed frequency within hosts is the criterion (Fig 1.6). More informative, however, would be distinguishing potential from factual colonizers, with members of the latter depending on the context where the colonization happens. We stress the relevance of regarding the colonization and coexistence ahead of the coevolution of hosts and microbes. Let alone, their organismic nature and implications (Bordenstein and Theis, 2015; Moran and Sloan, 2015).

As a consequence of the neutral assumption (fitness $\alpha_i = 0$ in Eq (1.1) for $i \geq 1$), our results extend to microbiomes with an arbitrary number of taxa. Although we first illustrate the process with two of them (Fig 1.2), analogously to (Vega and Gore, 2017), we move on to focus on the perspective of a single taxon (x_i in Eq (1.3b), Eq (1.4bb), and Eq (1.5b)). In this view, the collection of other taxa can be arbitrarily complicated. This is particularly important in conditions leading to alternative microbiome states, where the frequency in the pool of colonizers, p_i, becomes extremely relevant. While symmetric p_i across taxa will result in as many alternative states as taxa, asymmetries will make those with larger p_i appear more prominent, giving the impression of a reduced number of alternative states (Costea et al., 2018).

Future empirical work could focus on characterizing the prevalence of effects associated with the short lifespan - slow immigration regime (Fig 1.2). Although this depends on the timescale of the microbial dynamics also (resulting from the quality of the host as a habitat), host life-history might provide direction (Fig 1.1B). For example, a short lifespan together with a reduced amount of microbes reaching the gut, indicate the potential of observing such regime in nematodes (Vega and Gore, 2017) and some insects (Obadia et al., 2017; Hammer et al., 2017, 2019). Moreover, different tissues within a host might provide different conditions. Other hosts might be subtler. As our model indicates, different life-histories might lead to similar results (contours in Fig 1.3-1.6).

We have presented a minimal neutral model. More complex processes could build upon it. Among others, the influence of the prenatal microbiome on the dynamics and stationary distribution in a neutral context is largely unknown (Zeng et al., 2015; Van Vliet and Doebeli, 2019; Roughgarden, 2020). Additionally, after an initial stochastic assembly, hosts might actively influence their microbiome via immunity and development (Stephens et al., 2016). This might have general or taxa specific effects. Particularly relevant as well, is the role that first colonizers (Fig 1.2) might have in modifying the internal host, influencing the arrival of upcoming microbes (Koenig et al.,

2011). This could reinforce the difference between alternative microbiome states, at taxonomic and functional levels. Finally, as reported in some hosts (Nyholm and McFall-Ngai, 2004), non-smooth changes of the microbiome could occur. These changes, of intrinsic (e.g. microbial succession (Koenig et al., 2011), host and metabolic rhythms (Thaiss et al., 2014)) or extrinsic (e.g. diet change (Koenig et al., 2011), disease, and antibiotics (Bokulich et al., 2016)) origin might be more akin to a Lévy walk (Gardiner, 2004).

Although previous models have studied signatures of ecological neutrality and selection in microbiome data (Li and Ma, 2016; Sala et al., 2016), as well as its evolution (Zeng et al., 2015, 2017), they have not described the ecological effects that we have described here. We share Roughgarden et al.'s (Roughgarden et al., 2018) view that an eco-evolutionary approach is needed, but our results emphasize that colonization in a neutral context might already be sufficient to unify important and disconnected experimental observations, often implictly attributed to selection. Non-neutral processes might then build on top of such patterns.

1.6 Conclusion

We have introduced a stochastic model of the colonization of microbe-free hosts. After considering the environmental colonization and finite lifespan of hosts, our model recapitulates patterns reported experimentally. Namely, the coexistence of microbe-free and partially colonized hosts, as well as alternative microbiome states; both depending especially on the host lifespan. Crucially, our observations occur under non-selective conditions at the level of microbes or hosts. The model and results presented here aim to provide a null model for studying the host-microbiome formation by assuming the neutrality of microbial taxa – without ruling out that also selection will be important for these processes in nature. But even in the absence of any selective differences, our model explains a wide range of recent observations in microbiomes, from the observation of non-colonized hosts to alternative microbiome states.

CHAPTER 2

Parental transfer of microbes to newborns

In Chapter 1 I assumed that newborn hosts only get microbes from the environment. Although this might be true for some host species, more generally, newborns could get microbes from their parents. In this Chapter, I consider this possibility and its effect on the microbiome composition. I investigate changes in the total microbial load and on single microbial taxa.

This Chapter, available as a preprint (Zapién-Campos et al., 2021a), has been submitted for peer-review under the title *On the effect of inheritance of microbes in commensal microbiomes*, coauthored by Florence Bansept, Michael Sieber, and Arne Traulsen. The authors' contributions are detailed at the end of the thesis.

2.1 Abstract

Our current view of nature depicts a world where macroorganisms dwell in a landscape full of microbes. Some of these microbes not only transit but establish themselves in or on hosts. Although hosts might be occupied by microbes for most of their lives, a microbe-free stage during their prenatal development seems to be the rule for many hosts. The questions of who the first colonizers of a newborn host are and to what extent these are obtained from the parents follow naturally. We have developed a mathematical model to study the effect of the transfer of microbes from parents to offspring. Even without selection, we observe that microbial inheritance is particularly effective in modifying the microbiome of hosts with a short lifespan or limited colonization from the environment, for example by favouring the acquisition of rare microbes. Thus, we suggest that in an eco-evolutionary context, the impact of microbial inheritance is of particular importance for some specific life histories.

2.2 Introduction

Microbial life is ubiquitous in the biosphere (Thompson et al., 2017). The human body is no exception, as first described by van Leeuwenhoek in the 17th century. We are among the many macroorganisms where diverse microbiomes – microbial communities living in or on hosts – have been observed (Colston and Jackson, 2016; Hammer et al., 2019). As part of their life cycle, members of the microbiome may migrate between hosts and the environment. The migration process has been studied using experimental (Johnke et al., 2020) and theoretical approaches (Miller and Bohannan, 2019; Sieber et al., 2021). However, some microbes have been found exclusively in hosts (Johnke et al., 2020; Almeida et al., 2019). How do such microbes persist in the population?

One possibility is the vertical transfer of microbes from parents to offspring (McDonald and McCoy, 2019). Although there is ample literature about transmission of endosymbionts (e.g. *Buchnera* and *Wolbachia* in insects (Bright and Bulgheresi, 2010)), less is known about extracellular – possibly transient – microbes. Quantifying the low microbial loads in newborns (Eisenhofer et al., 2019) and deciphering the true origin of microbes (Perez-Muñoz et al., 2017) remains experimentally challenging (Funkhouser and Bordenstein, 2013; Russell, 2019). A few experimental studies have explored the vertical transfer of the microbiome in specific species across the tree of life – including sponges (Björk et al., 2019), mice (Moeller et al., 2018), cockroach eggs (Renelies-Hamilton et al., 2021), and wheat seedlings (Özkurt et al., 2020). For many others, including humans, there is an ongoing debate on when and how inherited microbes are obtained (Perez-Muñoz et al., 2017). Together, these studies suggest there is no universal reliance on microbial inheritance across host species, raising the possibility that even if such associations matter to the host, certain life-history traits may limit their inheritance (Russell, 2019; Bruijning et al., 2020). Relevant traits may include, among others, the extent of environmentally acquired microbes and host lifespan.

Previous theoretical work has studied microbial inheritance in the context of symbiosis – where microbes affect the host fitness. In these models, depending on whether the interaction is positive (mutualism) or negative (parasitism) the presence of symbionts is promoted or impeded, respectively. Using multilevel selection arguments, Van Vliet and Doebeli have shown that a symbiosis that is costly for microbes can be sustained only when the host generation time is short and the contribution of inheritance exceeds that of environmental immigration (Van Vliet and Doebeli, 2019). Following up, in addition to individual inheritance (single contributing parent), Roughgarden analyzed scenarios of collective inheritance (multiple contributing parents) (Roughgarden, 2020); while Leftwich et al. found a weak influence of the host

reproductive mode (sexual or asexual) and mate choice (based on symbiont presence) on the symbiont occurrence (Leftwich et al., 2020). If these host-symbiont interactions persist over evolutionary timescales, they are said to lead to phylosymbiosis – where microbiomes recapitulate the phylogeny of their hosts (Lim and Bordenstein, 2020).

Not all co-occurrences between hosts and microbes reflect a fitness impact, however. As suggested by Bruijning et al., the selection on the host-microbiome pair and the microbial inheritance might change with the environment (Bruijning et al., 2020). Moreover, despite taxonomic differences, functional equivalence of microbes in localized host populations could prevail (Renelies-Hamilton et al., 2021). Microbes might not always influence host fitness (Bruijning et al., 2020) nor benefit from influencing it (Leftwich et al., 2020). In this context where there is no active selection of the microbes by the host, the role of microbial inheritance remains largely unexplored (Zeng et al., 2015).

Using a stochastic model, we study the effect of microbial inheritance on the commensal microbiome – microbes living in hosts but not affecting their fitness. First, we introduce different models of inheritance representative of diverse host species. Then we discuss their effect on microbes present in both hosts and environment, or only present in hosts. We see that inheritance might influence the within-host occurrence and abundance in some cases. However, within the same microbiome, microbial types could be affected differently – while inheritance causes some microbes to increase in frequency, others decrease from it. Moreover, the effects may be transient, rendering life history parameters crucial. Altogether, we highlight the potential and limits of microbial inheritance to modify the composition of commensal microbiomes under different life-history scenarios.

2.3 Model and methods

Consider the host-microbiome system depicted in Fig. 2.1A. A population of hosts is colonized by a set of microbes, and each microbial taxon i has a constant frequency p_i in the environment. The total number of microbes a host can contain is finite and given by N. Each newborn empty host inherits a set of microbes from its parent, chosen at random within the host population. The inherited sample, taken off the parental microbiome, is drawn according to a probability distribution (Fig. 2.1B). After this initial seeding, only the death, immigration and replication of microbes can modify the host microbiome. Through these processes, the microbial populations within the host can decrease or increase by one individual each time step. After one microbe is selected to die, migration from the pool of colonizers occurs with

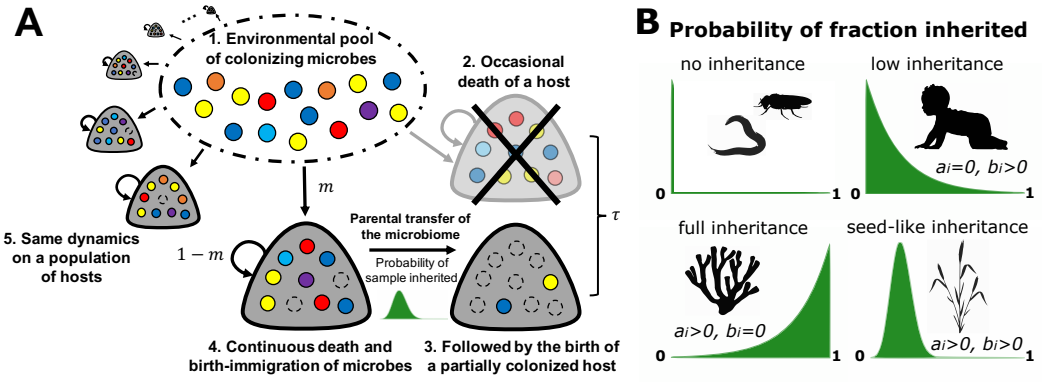

Figure 2.1: **Host-microbiome dynamics and microbial inheritance in our model.** (**A**) Dark blobs indicate hosts, coloured- and empty-circles indicate microbes and empty-space, respectively. Within the hosts, microbes go through a death and immigration-birth process, with new residents migrating from the pool of colonizing microbes with probability m or replicating within a host with probability $1 - m$. For microbes, each host is an identical habitat. The host population is at a dynamic equilibrium, every timestep there is a probability τ that a host death occurs, immediately followed by the birth of a new one. The newborn obtains a sample of its parent microbiome according to a probability distribution. (**B**) The probability distribution of the fraction of the parental microbiome inherited vary across host taxa – among others, influenced by development, reproduction and delivery mode. Certain hosts might not transfer microbes (eg. *C. elegans* (Zhang et al., 2017) or *D. melanogaster* (Blum et al., 2013)). Others might provide minimal (eg. humans (Perez-Muñoz et al., 2017)) or large fractions of their microbes (eg. fragmentation of some sponges, corals, fungi and plants (Cancino and Rodger, 1985; Frey and Kürschner, 2011)), while others might be centred around a fixed value (eg. seeds of plants (Özkurt et al., 2020)). In our model, we control this probability distribution through the parameters a_i and b_i in Eq. (2.4).

probability m, while duplication of a resident microbe, or non-replacement, occurs with probability $1 - m$. This process ends with the host death, which occurs with probability τ at each time step. We assume that the number of hosts does not change, so that a host death is followed by the birth of a new empty host, for which the process described above is repeated.

Transition probabilities

Our aim is to describe the dynamics of the microbiome load and composition, focusing in particular on how a certain microbial taxon experiences it. Within a specific host, the frequency of the *i-th* taxon is denoted by x_i (for $i \geq 1$), and of the remaining other microbes by $o_i = \sum_{j \neq i} x_j$. The frequency of available space is then given by $x_0 = 1 - x_i - o_i$. The transition probabilities from state $\{x_i, o_i\}$ that are due to the microbial dynamics are given by the product of the probability of host survival, $1 - \tau$, by the probability of death

of a certain microbial type followed by an immigration or birth event. These events produce changes in the frequencies of magnitude $\frac{1}{N}$. First, microbial taxa can replace each other when a microbe dies and is replaced by another one,

$$T_{x_i-}^{o_i+} = (1-\tau)\, x_i \left(m(1-p_i) + (1-m)\frac{o_i}{\alpha_0 x_0 + x_i + o_i} \right) \quad (2.1a)$$

$$T_{x_i+}^{o_i-} = (1-\tau)\, o_i \left(m p_i + (1-m)\frac{x_i}{\alpha_0 x_0 + x_i + o_i} \right). \quad (2.1b)$$

In Eq. (2.1a), a microbe of type i dies and is replaced by another microbe, either by immigration from the environmental pool or by replication within the same host. Similarly, in Eq. (2.1b), a microbe of another type dies and is replaced by a microbe of type i.

Alternatively, microbes may occupy previously available space, such that the microbial abundance increases,

$$T_{x_i}^{o_i+} = (1-\tau)\, x_0 \left(m(1-p_i) + (1-m)\frac{o_i}{\alpha_0 x_0 + x_i + o_i} \right) \quad (2.1c)$$

$$T_{x_i+}^{o_i} = (1-\tau)\, x_0 \left(m p_i + (1-m)\frac{x_i}{\alpha_0 x_0 + x_i + o_i} \right). \quad (2.1d)$$

Finally, microbes may decrease in abundance, when a microbe selected for death is not replaced,

$$T_{x_i-}^{o_i} = (1-\tau)\, x_i \left((1-m)\frac{\alpha_0 x_0}{\alpha_0 x_0 + x_i + o_i} \right) \quad (2.1e)$$

$$T_{x_i}^{o_i-} = (1-\tau)\, o_i \left((1-m)\frac{\alpha_0 x_0}{\alpha_0 x_0 + x_i + o_i} \right). \quad (2.1f)$$

In these equations, α_0 controls the establishment of microbes in hosts – the ability to occupy available space – going from fast for $\alpha_0 = 0$, to slow if α_0 is positive. For $\alpha_0 > 1$ and without migration, microbes cannot be maintained in hosts.

The transition probabilities due to the hosts dynamics are given by the product of the probability of host death and birth of an empty host (τ), by the probability to inherit certain microbes,

$$T_{\Delta x_i}^{\Delta o_i} = \tau \sum_p \frac{1}{H-1} \omega_i[\Delta x_i, x_i^{(p)}] \omega_i[\Delta o_i, o_i^{(p)}], \quad (2.2)$$

where $1/(H-1)$ is the probability of choosing a parent p in the population of $H-1$ potential parents, and $\omega_i[\Delta x_i, x_i^{(p)}]$ and $\omega_i[\Delta o_i, o_i^{(p)}]$ are the probabilities of transfer of Δx_i and Δo_i microbes from the parent to the offspring,

respectively. Because the frequencies within the parent are $x_i^{(p)}$ and $o_i^{(p)}$, the probability to transfer more microbes than the parent can provide is zero.

Finally, for completeness, the probability of staying in state $\{x_i, o_i\}$ without host death is

$$T_{x_i}^{o_i} = 1 - T_{x_i-}^{o_i+} - T_{x_i+}^{o_i-} - T_{x_i}^{o_i+} - T_{x_i+}^{o_i} - T_{x_i-}^{o_i} - T_{x_i}^{o_i-} - \tau, \tag{2.3}$$

where the last term includes all possible transitions due to parental transfer of microbes, $\int\int T_{\Delta\tilde{x}_i}^{\Delta\tilde{o}_i} d\Delta\tilde{x}_i d\Delta\tilde{o}_i = \tau$.

Distribution of inherited microbes

In our model, parents can seed the microbiome of their offspring. A sample of the parental microbiome is vertically transmitted according to a probability distribution function, Eq. (2.2). In addition to the case without inheritance, which we have previously analyzed elsewhere (Zapien-Campos et al., 2020), at least three qualitatively distinct cases may be observed (Fig. 2.1B), depending on host development, reproduction, and mode of delivery.

Firstly, inheritance could be low. For example in animals, newborns get microbes attached to epithelia or fluids during delivery (Perez-Muñoz et al., 2017; McDonald and McCoy, 2019). These represent a small fraction of the parental microbiome, leading to distributions centred at frequency zero decaying towards one. Secondly, certain hosts, including some sponges, corals, fungi and plants (Cancino and Rodger, 1985; Frey and Kürschner, 2011), are able to reproduce by fragmentation, where a breaking body part generates a new individual. Such fragments could carry a faithful microbiome composition, leading to distributions centered at frequency one decaying towards zero. Finally, hosts that produce embryos that can disperse, eg. seeds, might transfer a microbiome sample contained within these physical structures (Özkurt et al., 2020).

We modelled such diverse parental microbiome samplings (Δx_i) using the beta distribution for the probability $\omega_i[\Delta x_i, x_i^{(p)}]$ to inherit Δx_i microbes from parent p. This probability distribution can take arguments in the range from zero to the current frequency of a microbe in the parent p, $x_i^{(p)}$,

$$\omega_i[\Delta x_i, x_i^{(p)}] = \frac{1}{B[a_i+1, b_i+1]} \left(\frac{\Delta x_i}{x_i^{(p)}}\right)^{a_i} \left(1 - \frac{\Delta x_i}{x_i^{(p)}}\right)^{b_i}, \tag{2.4}$$

where B is the beta function (Gradshteyn and Ryzhik, 1994), $1/B$ a normalization constant, and a_i and b_i are shape parameters. The expected value of our beta distribution is $\frac{a_i+1}{a_i+b_i+2}$. The special case of $a_i, b_i = 0$ leads to a uniform distribution, where the parental microbes are distributed randomly

between parent and offspring. Other combinations of $a_i, b_i \geq 0$ produce different unimodal distributions (Fig. 2.1B). The case of $a_i > b_i$ skews the distribution towards full inheritance of the parental microbes, $\Delta x_i = x_i^{(p)}$ – all the *i-th* microbes from the parent could be transferred to the offspring. The case of $a_i < b_i$ skews the distribution towards non-inheritance of microbes of type i to offspring, $\Delta x_i = 0$. Finally for $a_i = b_i$, the distribution is symmetric and the parental microbes are likely to be equally distributed between parent and offspring. In most of our analyses a_i and b_i are the same for all microbial taxa. Only for non-neutral, asymmetric inheritance, we will set different a_i and b_i for the focal taxon (x_i) and the set of others (o_i). To illustrate the effect of a_i and b_i, on average, an offspring inherits $\approx 9\%$ of the parental microbes of taxon 1 for $a_1 = 0$ and $b_1 = 9$, while only $\approx 1\%$ is inherited for $a_1 = 0$ and $b_1 = 99$.

Throughout the results, we focus on distributions with a maximum at microbial frequency zero decaying towards $x_i^{(p)}$, which we call 'low inheritance' (Fig. 2.1B). In our model, the low inheritance and the 'full inheritance' scenarios (distributions with maximum at frequency $x_i^{(p)}$ decaying towards zero) are equivalent. This stems from the fact that the number of microbes is conserved, so that inheritance happens through the splitting of the parental microbiome between the parent and the offspring. Thus, since in our model, the probability to die of a host does not depend on its age, the splitting of microbes in the low inheritance scenario - where a small fraction is transmitted - and in the full inheritance scenario - where most of the microbiome is transmitted - are equivalent. Finally, we address under which circumstance a 'seed-like inheritance' leads to different results.

Stochastic simulations

In order to simulate the microbiome dynamics of individual hosts we formulated the model as a stochastic differential equation. We solved this equation numerically using the Euler-Maruyama method (Gardiner, 2004). Starting from state $\mathbf{x} = \{x_i, o_i\}$ at time t the new state after an interval Δt is given by

$$\mathbf{x}[t + \Delta t] = \mathbf{x}[t] + \mathbf{A}[\mathbf{x}[t]]\Delta t + B[\mathbf{x}[t]]\Delta \mathbf{W}[\Delta t], \tag{2.5}$$

where $\mathbf{A}[\mathbf{x}[t]]$ is the vector of expected changes of $\mathbf{x}$, the deterministic contribution; while $B[\mathbf{x}[t]]$ is a matrix that has the property $B[\mathbf{x}[t]]^T B[\mathbf{x}[t]] = V[\mathbf{x}[t]]$, where $V[\mathbf{x}[t]]$ is the covariance matrix of the change of $\mathbf{x}$. Further, $\Delta \mathbf{W}$ is a vector of uncorrelated random variables sampled from a normal distribution with mean 0 and variance Δt, the stochastic contribution. That $\Delta \mathbf{W}$ is normally distributed arises from the time independence and identical distribution

of the noise. A detailed description connecting Eq. (2.1) and Eq. (2.5) is provided in Appendix C.1.

For most of their life, hosts are independent of each other, only newborns are influenced by others when they acquire their initial microbiome. A given host lives for a duration sampled from an exponential distribution $\tau e^{-\tau t}$, with mean $1/\tau$. We solve Eq. (2.5) for that interval. Immediately after a host dies, the microbiome of a newborn is assembled according to Eq. (2.2). We repeat these steps for all hosts until the total simulation time is reached.

As a result of stochasticity, each host trajectory is different. We look into the statistical description of the microbiome composition across the host population.

Code availability

The simulation code in Python is available at
https://github.com/romanzapien/microbiome-inheritance

2.4 Results

Inheritance can increase the occurrence of microbes in hosts with low microbial loads

Without microbial inheritance, which will be our reference case throughout, any microbe occurring inside a host has to have migrated from the environment during the host lifespan. As a result, a low environmental migration or short host lifespan can be limiting (Zapien-Campos et al., 2020). The transfer of microbes from parents to offspring during birth could increase the probability of observing any microbes in hosts, $P[x_i + o_i > 0]_{\text{inh.}}$. We quantified the change in the probability of occurrence relative to its microbe-free birth condition $P[x_i + o_i > 0]_{\text{no inh.}}$,

$$\Delta P[x_i + o_i > 0] = P[x_i + o_i > 0]_{\text{inh.}} - P[x_i + o_i > 0]_{\text{no inh.}}. \tag{2.6}$$

Using this observable, we investigated the role of life history in modulating the effect that inheritance has on the microbiome. We quantified this for a single microbial taxon, x_i, as well.

Fig. 2.2 shows a condition where, in the absence of inheritance, hosts are not fully occupied by microbes. This results from a short host lifespan (τ) and low microbial immigration from the pool of colonizers (m). We tested the effect of the 'low inheritance' mode (Fig. 2.1B) for increasing rates of establishment of microbes ($\alpha_0 \to 0$) and other life-history parameters.

Inheritance impacts the occurrence of microbes by increasing the number of hosts with at least one colonizing microbe (Fig. 2.2B). The effect is most

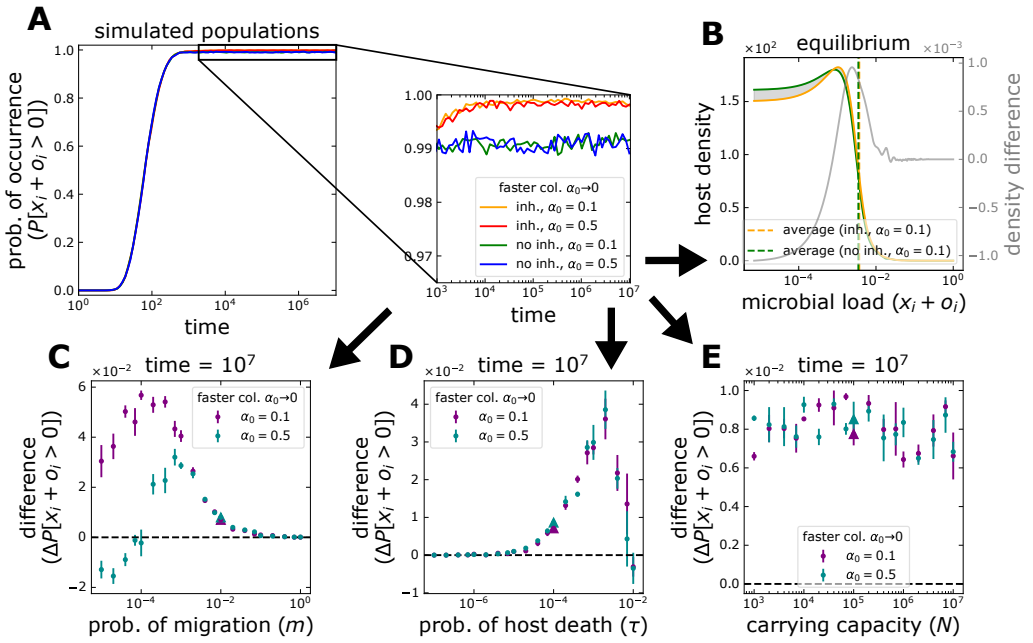

Figure 2.2: **Microbial occurrence in hosts under microbial inheritance.** (**A**) Starting from a condition where all hosts are initially empty, the microbial occurrence increases through time. At first sight, this increase is largely independent of α_0 and the inheritance of microbes. A closer look at equilibrium abundance reveals that inheritance increases the occurrence, in this case, regardless of how rapidly hosts are occupied (α_0). (**B**) The increase results from a distribution of microbial load across the host population where the microbe-free state is less common. A microbial load of 10^{-5} corresponds to 1 microbe per host. In (C-E), single parameters are modified from the case shown in (A-B) (with parameters $m = 10^{-2}$, $\tau = 10^{-4}$, and $N = 10^5$, indicated by the triangles in (C-E)). (**C**) A large migration from the pool of colonizers, $m \to 1$, hinders any effect of inheritance on occurrence as hosts are readily colonized. The change peaks and decreases for smaller m, as for $m \to 0$ hosts are less likely to be colonized. The change can even be negative for slowly occupied hosts where the few colonizing microbes are lost to stochasticity. (**D**) The gain from inheritance is maximal for intermediate values of host death probability, τ. Long living hosts, $\tau \to 0$, are colonized even without inheritance. Short living hosts, $\tau \to 1$, are less likely to be colonized and thus transmit microbes through inheritance. (**E**) The carrying capacity for microbes of a host, N, and α_0 do not alter the gain from inheritance. Points and bars in (C-E) indicate the average and standard deviation of 6 simulation pairs, with vs. without inheritance, with 10^4 hosts each. Offspring receive 9% of their parent's microbiome on average, $a_i = 0$ and $b_i = 9$ in Eq. (2.4). The whole distributions are shown in Fig. Sup. C.2.

prominent in scenarios where without inheritance, most of the hosts are microbe-free. However, the maximum increase occurs at intermediate immigration and host lifespans (Fig. 2.2C-D). For high immigration, $m \to 1$, hosts

are readily occupied by microbes, so inheritance brings no change. Similarly for a long host lifespan, $\tau \to 0$. On the other hand, if immigration is limited, $m \to 0$, or host lifespan short, $\tau \to 1$, microbes never occur in hosts, so parents cannot transmit microbes to their offspring.

Inheritance might decrease the occurrence if the transfer – which splits the parental microbiome between parent and offspring – makes microbes more susceptible to stochastic fluctuations. This occurs if the microbial frequency of the parent is already low – for example when migration is limiting and microbes proliferate slowly (Fig. 2.2C). This phenomenon might be pronounced for individual taxa. Our analyses from the perspective of a single taxon (Fig. Sup. C.1) found multiple instances where inheritance might decrease the occurrence (Fig. Sup. C.1C-F), but also have a larger effect in situations where the occurrence increases. Additionally, the effect on single taxa depends strongly on the carrying capacity for microbes, N (Fig. Sup. C.1F compared to Fig. 2.2E). Competition for space favours taxa according to their frequency in the pool of colonizers, p_i (Fig. Sup. C.1C). Abundant taxa outcompete rare ones as space is more limited, but only until a point, after which there is no benefit – they readily occur without inheritance. In other words, in microbiomes composed by many taxa, the taxon-level effect of inheritance in terms of occurrence is relative to their environmental abundance.

Inheritance can increase the abundance in hosts, but mostly of those abundant in the environment

Modifying the presence of taxa is not the only effect – inheritance also alters the microbiome composition considerably. Using the distribution of microbial frequencies in hosts, we quantified the change in the average frequencies as compared to its microbe-free birth condition,

$$\Delta E[x_i + o_i] = E\left[x_i + o_i\right]_{\text{inh.}} - E\left[x_i + o_i\right]_{\text{no inh.}}. \tag{2.7}$$

Similarly to Eq. (2.6), we quantified this observable for a single microbial taxon, x_i, as well.

When looking at the distribution of microbial loads and frequencies in hosts, the effect of the 'low inheritance' mode (Fig. 2.1B) is two fold – while hosts with small frequencies might experience the largest increase in microbes, hosts with large frequencies can see the largest decrease of microbes (Fig. 2.2B and Fig. Sup. C.2). Thus, at both microbial load and single taxon levels, hosts with small and large frequencies become rarer. Inheritance makes hosts resemble each other to a greater extend (see the reduced spread of the distributions in Fig. Sup. C.2 and Fig. Sup. C.3). This is equivalent to the

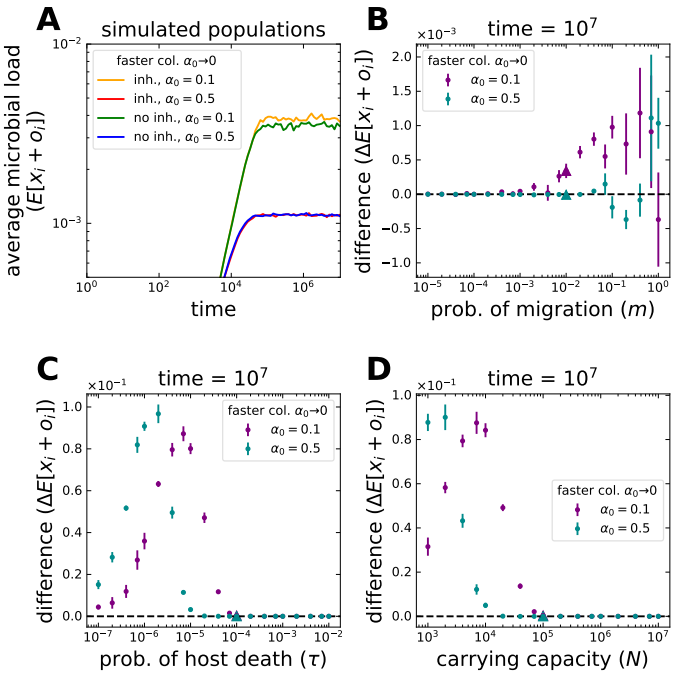

Figure 2.3: **Average microbial load in hosts under microbial inheritance.** (**A**) Starting from a condition where all hosts are initially empty, the average frequency of microbes in hosts increases through time before reaching an equilibrium. In this particular case, inheritance makes such equilibrium abundance larger only when hosts are occupied rapidly, $\alpha_0 \to 0$. This increase results from a host distribution where higher microbial loads are more common (Fig. 2.2B). The cases shown in (A), with parameters $m = 10^{-2}$, $\tau = 10^{-4}$, and $N = 10^5$, are indicated by the triangles in (B-D). A single parameter is varying in (B-D). (**B**) Changes of migration from the pool of colonizers, m, have minimal effect (notice the scale). As $m \to 1$, more microbes colonize the hosts. Still the average microbial load only increases if the loss of microbes to inheritance is less than the gain from proliferation. (**C**) The effect of changes to host death probability, τ, are much larger and maximal at intermediate τ. A faster occupation of hosts makes the effect of inheritance larger for shorter living hosts, $\tau \to 1$. (**D**) In contrast to the occurrence (Fig. 2.2E), changes in the carrying capacity for microbes, N, have a larger intermediate effect. Faster occupation of hosts makes the effect peak for larger N. Points and bars in (B-D) indicate the average and standard deviation of 6 simulation pairs, with vs. without inheritance, with 10^4 hosts each. Offspring receive 9% of their parent's microbiome on average, $a_i = 0$ and $b_i = 9$ in Eq. (2.4). The whole distributions are shown in Fig. Sup. C.2.

effect of increased immigration, which also tends to make microbiomes similar to one another, but increased inheritance does not favour the preservation of the diversity from the pool of colonizers – in contrast to immigration.

An increase in the average load is observed for some conditions (Fig. Sup. C.2). Analogously to the occurrence, such increase peaks at intermediate host death probabilities τ; but also at intermediate carrying capacities N

(Fig. 2.3C-D). The limited time for host colonization impedes any inheritance ($\tau \to 1$), while for $\tau \to 0$ or small N, hosts are fully occupied even without it. The relative effect of inheritance is less for large N. A faster occupation of available space ($\alpha_0 \to 0$) displaces the effect to larger host death probabilities and capacities for microbes. Finally, because the main limitation is the short host lifespan (τ), the influence of immigration (m) is minimal (see the scale in Fig. 2.3B and Fig. Sup. C.4C).

Although higher loads might be reached with inheritance if space is limited (Fig. Sup. C.2C), abundant taxa might increase at the expense of rare ones (Fig. Sup. C.3D and Fig. Sup. C.4D-E). Such reduction is exacerbated by the fast occupation of available space $\alpha_0 \to 0$. Interestingly, this might happen as a result of longer host lifespans as well, if hosts are rapidly occupied by inherited microbes. Such condition favours abundant taxa in the pool of colonizers. Instead, if the occupation is slower, rare taxa increase in frequency, derived from the added benefits of inheritance and a more influential immigration (m).

A particularly relevant question is whether the frequency of a taxon in a specific host (x_i) can be larger than in the pool of colonizers (p_i) – i.e. a benefit is obtained from the host association. We observe this even in the absence of inheritance (Fig. Sup. C.3), where stochastic colonization results in some host containing frequencies larger than in the pool (p_i). The average frequency across hosts, however, can be larger only when the space limitation increases the competition. In this context, inheritance may, in fact, decrease the chances of such outcome, by relating the hosts to each other (Fig. Sup. C.3C-D).

Preferential inheritance can temporally lead to specific taxa overrepresentation

A potential mechanism to increase the average frequency of taxa beyond their frequency in the pool of colonizers (p_i), is preferential inheritance. The asymmetry in inheritance could stem from differences in microbial properties, but also a host's direct or indirect influence. We studied such possibility by manipulating the distribution of the sample inherited, Eq. (2.4). Focusing on a 'low inheritance' mode, we decreased the inheritance of other taxa relative to taxon i, from equal if offspring receive 9% of every taxa on average, to preferential if they receive 9% of taxon i but 1% of others.

For the same parameters as before (Fig. 2.4), we observe no effect if the host lifespan is limiting. In this case, regardless of the frequency in the pool of colonizers (p_i), preferential inheritance does not alter the average frequency of the *i-th* taxon in hosts (Fig. 2.4A), similarly for the probability of immigration

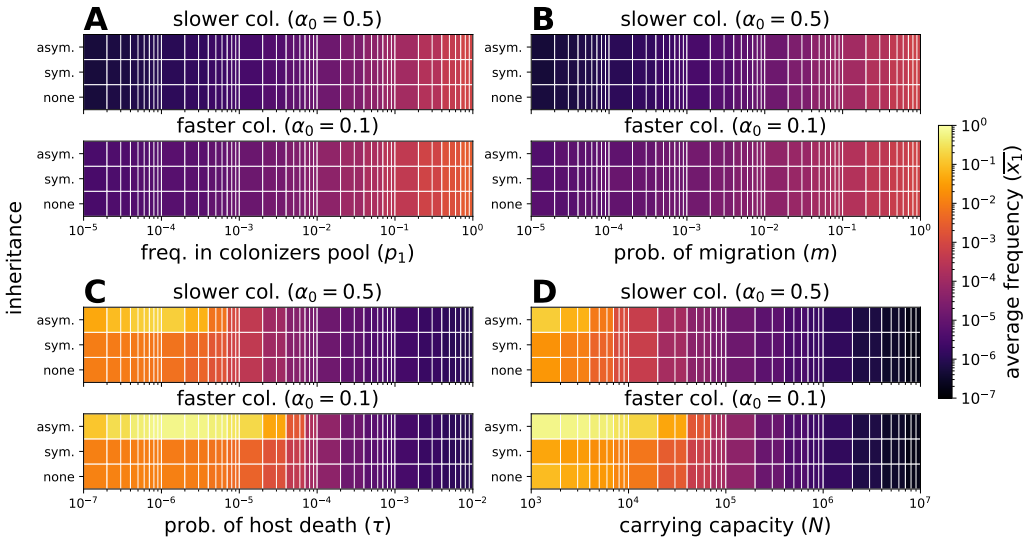

Figure 2.4: **Effect of asymmetric inheritance on the average frequency of a taxon in hosts.** Cases without inheritance and inheritance are compared. Inheritance is symmetric if offspring receive 9% of their parent's microbiome on average ($a_i = 0$ and $b_i = 9$). Inheritance is asymmetric if offspring receive 9% of taxon 1 and 1% of other taxa ($a_i = 0$ and $b_1 = 9$, $b_{i\neq 1} = 99$ in Eq. (2.4)). Available space within hosts is occupied more easily for $\alpha_0 \to 0$. Single parameters are modified from the condition $p_1 = 10^{-2}$, $m = 10^{-2}$, $\tau = 10^{-4}$, and $N = 10^5$. (**A-B**) The average frequency increases for larger abundances in the pool of colonizers (p_1), immigration (m), and $\alpha_0 \to 0$. An asymmetric inheritance has no effect, as hosts are not fully occupied within their lifetime (Fig. Sup. C.2 and Fig. Sup. C.3). (**C**) Longer host lifespans, $\tau \to 0$, increase the average frequency and effect of asymmetric inheritance. The gain is maximal at intermediate τ. Inheritance has more influence before hosts are fully occupied. After this, hosts resemble the colonizers pool. (**D**) The average frequency increases with competition for space (smaller N). While the symmetry of inheritance decreases the average frequency as a result of the reduced initial microbiome variability, asymmetry increases it. Each simulation included 10^4 hosts.

m (Fig. 2.4B). This holds even for fast occupation of available space, $\alpha_0 \to 0$. Only for longer host lifespan, $\tau \to 0$, preferential inheritance leads to an increase (Fig. 2.4C). Besides the almost exclusive occupation of hosts by the *i-th* taxon ($\overline{x_i} \to 1$), the maximum effect is constrained to intermediate τ. This is because the effect of preferential inheritance is transitory for longer living hosts, after which they continue approaching their long term equilibrium, $\overline{x_i} \to p_i$. For faster occupation of available space the gain spans a wider range and shorter host lifespans ($\tau \to 1$). For hosts with short lifespan and limited immigration (in our example $\tau = 10^{-4}$ and $m = 10^{-2}$), the gain from preferential inheritance is largest for decreasing carrying capacity for microbes, N (Fig. 2.4D).

As shown in Fig. 2.4D, inheritance itself might not benefit all microbial

taxa. For some taxa, only preferential inheritance can lead to larger frequencies than without inheritance.

Persistence of lineage taxa in hosts

An extreme case of reliance on microbial inheritance are microbes present in hosts but absent from the environment ($p_i = 0$) (Thompson et al., 2017; Almeida et al., 2019). We refer to these as lineage taxa. We investigated the conditions allowing their persistence under different life-history scenarios (Fig. 2.5).

Within a host, lineage taxa go through the stages sketched in Fig. 2.5A. Depending on the context, after host birth, their frequency might either decrease or increase. If decrease occurs, in a neutral context this trend will not change during the host life. In fact, events of microbiome inheritance will further decrease the frequency in the parent. We found that on average, lineage taxa increase while the inequality

$$x_i + o_i < 1 - \frac{m}{1 - \alpha_0} \tag{2.8}$$

holds (Fig. 2.5C and Appendix C.1). Therefore, lineage taxa increase before reaching carrying capacity, favoured by their fast proliferation ($\alpha_0 \to 1$), but restricted by migration (m). Because the microbial load increases through time ($x_i + o_i \to 1$), alongside the initial state, Eq. (2.8) limits the time of increase. Note that on average, the maximum frequency of lineage taxa is $1-m/(1-\alpha_0)$. From this point on, a decrease driven by the immigration of environmentally present microbes (m) and stochasticity follows. For sufficiently long time, such decrease may lead to their extinction (Fig. 2.5B).

There is a trade-off between the duration of the increase and the maximum frequency of lineage taxa. While small initial microbial loads lead to long durations but small frequencies (as a result of immigration, Eq. 2.8), the opposite is true for high initial loads abundant in lineage taxa. Once increase stops, the time to extinction is proportional to the lineage taxa frequency, Fig. 2.5B. Putting these two times together, the extra time from the increase is behind the subtle effect of the initial microbial load on the total extinction time, Fig. 2.5D. A reduced migration ($m \to 0$) and fast occupation of available space ($\alpha_0 \to 0$) simultaneously increase the frequency and time before extinction.

Looking at the population level, a condition for persistence emerges – namely, an increase of frequency in each host followed by transfer to offspring of a frequency at least equal to that received at birth. This is possible only while the frequency in the parent is larger than initially, Fig. 2.5A. The largest frequencies are expected at intermediate time. In this context, host lifespan,

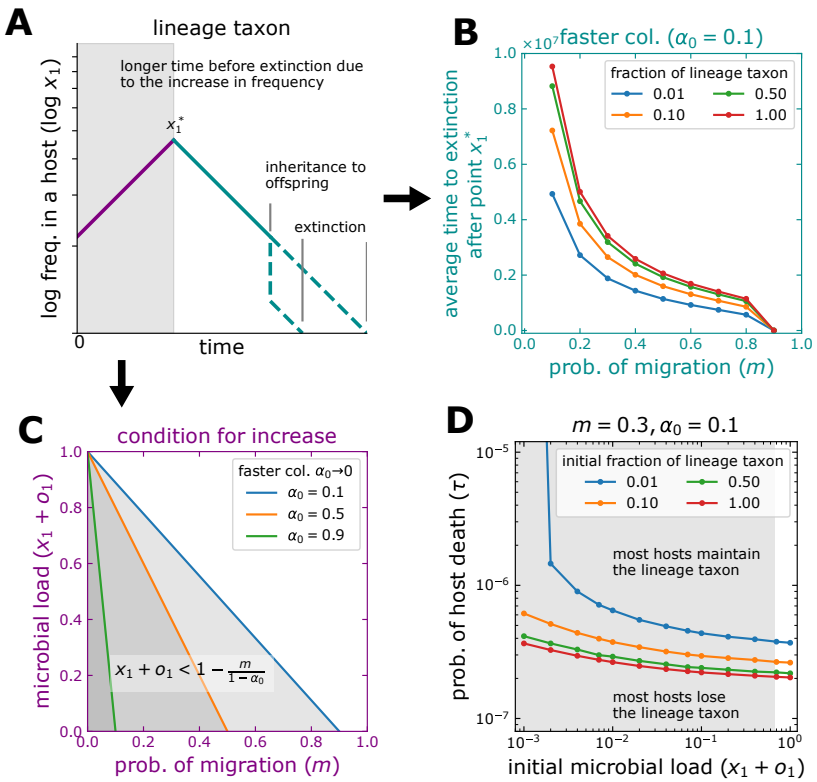

Figure 2.5: **Persistence of lineage taxa in hosts.** A microbial taxon is initially present in hosts $x_1(0) > 0$, but not in the pool of colonizers, $p_1 = 0$. (**A**) The frequency within a host decreases through time. For some conditions, Eq. (2.8), there is a period of increase. If the taxon is transmitted to offspring before the gain is lost, this might persist in the host population (although extinction within the parent occurs sooner). (**C**) Low immigration ($m \to 0$) and fast occupation of available space ($\alpha_0 \to 0$) allow increase and prolong the time before extinction, Eq. (2.8). Large initial available space ($x_i + o_i \to 0$) and lineage taxon fractions ($x_1/(x_1 + o_1) \to 1$) also prolong this time. (**B**) After the increase stops (x_1^*), the average time to extinction is shorter for large immigration ($m \to 1$) and a smaller fraction of the taxon. (**D**) At the host population level, lines indicate the death probability after which most hosts lose the lineage taxon ($\tau_{0.5}$), Eq. (2.9). The early increase shown in (A) only occurs within the darkened area. The distribution of microbes inherited, Fig. 2.1B and Eq. (2.4), affects the initial load and fraction of lineage taxa in offspring. Asymmetric inheritance in low microbial loads might preserve lineage taxa as well as symmetric inheritance in high loads. We set $N = 10^5$. Each point corresponds to 10^4 simulated hosts.

and thereafter the probability of host death (τ) become fundamental. From the distribution of host death events, $\tau e^{-\tau t}$, we see most hosts die early on, potentially while lineage taxa are still abundant; $\tau \to 0$ results in longer living hosts – those more likely to lose lineage taxa. We estimated the probability

of host death at which a fraction z of hosts loses the taxa,

$$\tau_z = -\frac{1}{t_z}\ln(1-z) \tag{2.9}$$

where t_z, the time at which lineage taxa remain present in a fraction z of the host population, is obtained from the distribution of extinction times. Based on the former observations (Fig. 2.5D), our model predicts that regardless of the distribution of inherited microbes (Fig. 2.1B), preferential inheritance of lineage taxa in small microbial loads might favour their persistence as effectively as large but non-preferential microbial loads.

When the distribution of inherited microbes matters

We proposed that a finite set of shapes captures most of the possible microbial inheritance distributions (Fig. 2.1B) – low, high, and seed-like inheritance – all characterized by the most likely microbiome fraction transferred to the offspring. So far, we have focused on the impact of low inheritance on the microbiome (Fig. 2.2-2.5). As mentioned before, because we enforce the conservation of microbes in our model, i.e. the microbes are transferred from the parent host to the offspring, the outcome of low and high inheritance is equivalent: although the parental microbiome is distributed differently, the outcome is indistinguishable at the host population level, because hosts are indistinguishable.

When referring to certain life-histories, other distribution shapes may alter the impact of inheritance. To find out differences between the effect seed-like inheritance and our former results (where we assumed low inheritance) we compared the occurrence and average microbial frequencies.

We found most changes are minimal, however, differences appear for extreme parameters. A seed-like inheritance might better guarantee the occurrence of microbes in extremely adverse life-histories – e.g. rare environmental migration ($m \to 0$) and short host lifespan ($\tau \to 1$) simultaneously (vertical axis on Fig. Sup. C.5A-B). Exceptions could arise for a slower occupation of available space (α_0). For individual microbial taxa, changes are greater in occurrence as well (Fig. Sup. C.6); however, derived from the competition for limited space (N), the effect of a seed-like inheritance is case-specific. Moreover, both maximum increase and maximum decrease occur at intermediate m (Fig. Sup. C.6B) and τ (Fig. Sup. C.6C). In microbiomes composed of taxa with different environmental frequencies (p_i), while some taxa gain, others lose from inheritance (Fig. Sup. C.6A).

Under less adverse conditions, seed-like inheritance might allow larger microbial loads. That is the case when either host lifespan (horizontal axis on Fig. Sup. C.5A) or migration (Fig. Sup. C.5B) is limiting. The gain from

a seed-like inheritance can be large, especially for a small carrying capacity for microbes N (Fig. Sup. C.5C). The consistent microbial transfer and reduced variation are beneficial. Nonetheless, at the single taxon level, gains are minimal (Fig. Sup. C.6). At this level, a limiting carrying capacity for microbes, where competition increases, might even lead to a decrease (Fig. Sup. C.6D). In this case, the variation provided by the low inheritance mode is more beneficial.

In summary, regardless of the distribution of microbes inherited (Fig. 2.1B), life-history seems intrinsically linked to the effect of microbial inheritance on the microbiome composition.

2.5 Discussion and conclusion

The impact of microbial inheritance on host-associated microbial communities is largely unknown. In this work, we explored its potential effects under diverse life-history scenarios, including multiple distributions of microbes inherited (Fig. 2.1). Using a model free of selection – i.e. without microbial fitness differences or effect on host fitness – we shed light on the conditions where microbial inheritance may influence the microbiome composition, showing its impact but also its limits.

Our work emphasizes the role of life-history over host-microbe associations (Fig. 2.2-2.3). Even without symbiotic benefits, the inheritance process itself might alter the microbiome composition (Leftwich et al., 2020). Using a discrete generation model, Zeng et al. considered microbial inheritance in neutral associations over evolutionary timescales – specifically, its effect on the microbial diversity and the distribution of frequencies (Zeng et al., 2015) . Our results, however, highlight the relevance of within-generation probabilistic events – environmental colonization, host lifespan, or carrying capacity for microbes – as ecological drivers to constrain inheritance.

A crucial constraint is the host lifespan. Similarly to Van Vliet and Doebeli, but without any impact on the host fitness, we observe that the environmental acquisition of microbes makes the effects of inheritance transient (Fig. 2.2D, 2.3C and 2.4C) (Van Vliet and Doebeli, 2019). Short-living hosts (relative to the microbial timescale) could influence their commensal microbiome over their whole lives, while long-living hosts only during the first stages of development. The rapid proliferation of inherited microbes or isolation from the environment might prolong the period of influence. This is in contrast to Van Vliet and Doebeli, where selection within isolated hosts acts against costly symbiosis, reducing the period of mutualists presence.

We observed that the effect of inheritance may differ between taxa. Microbiomes assembled entirely from the environment are prone to variation

when migration between hosts is rare (Zapien-Campos et al., 2020; Bruijning et al., 2020). Inheritance might increase the presence of certain microbes, but in contrast to environmental migration, inheritance reduces the variation between hosts and potentially their microbial diversity. This reduction, which especially affects rare taxa, is more pronounced if the carrying capacity is limited (Fig. Sup. C.1 and Sup. C.4), where competition is larger. Bruijning et al. have observed that under selection, such decreased variation and diversity could be detrimental for adaptation to changing environments (Bruijning et al., 2020).

Initially, we assumed no distinction between microbial taxa – only their frequency determined the population dynamics (Eq. 2.1). This could be modified in at least two ways. First, fitness differences could influence the birth and death rates of microbes. Although this is certainly relevant, it diverts from our focus on inheritance. Instead, we addressed a possibility crucial for inheritance – the asymmetric transfer of microbes (Fig. 2.4). Such asymmetry could emerge from differences in microbial capabilities at play during the transfer process, including oxygen tolerance (Moeller et al., 2018) (obligate anaerobes tend to be transmitted vertically) and sporulation (Browne et al., 2016) (spores might allow the transfer of oxygen-sensitive bacteria). Alternatively, hosts could selectively transfer certain microbes to their offspring (Bright and Bulgheresi, 2010). Interestingly, we observe that inheritance alone is not always beneficial; some taxa might only benefit when transferred asymmetrically (Fig. 2.4).

We have emphasized the importance of looking at rare taxa. Such is the case of lineage taxa (Fig. 2.5), microbes absent from the environment that only propagate by inheritance. Our results indicate the importance of modelling the stochasticity and conservation of microbes – only in this way did we appreciate that inheritance can lead to stochastic loss (Fig. 2.2-2.3) and that persistence of lineage taxa may be prolonged by asymmetric inheritance (Fig. 2.5D). Because microbial frequencies are often small, the omission of stochastic effects from models could lead to misestimate the impact of inheritance.

Vertical transfer of microbes might occur in the most diverse host species (Funkhouser and Bordenstein, 2013; Bruijning et al., 2020), with only a few exceptions (Hammer et al., 2019). A great diversity of reproduction and delivery modes might, in turn, determine the distribution of their inheritance – namely the number of microbes transferred and its probability. A comparison of two qualitatively distinct distributions (low and seed-like inheritance in Fig. 2.1B), indicates they might influence the presence and frequency of microbes differently (Fig. Sup. C.5). A consistent cargo in seeds might guarantee the presence of certain microbes in plants (Özkurt et al., 2020), who

might sometimes benefit from being the first colonizers (Zapien-Campos et al., 2020). In contrast, greater variation might be expected for mammals, where changing amounts of microbes are obtained from epithelia during delivery (Perez-Muñoz et al., 2017; Funkhouser and Bordenstein, 2013). Overall, these intrinsic differences might affect the ecological and evolutionary dynamics of hosts and microbes.

We found that microbial inheritance is effective only for some life-histories. While it has been shown that symbiosis (Van Vliet and Doebeli, 2019) and fidelity of inheritance (Bruijning et al., 2020) can evolve driven by selection, our results suggest the evolution of life-history traits itself, independent of symbiosis, can impact the relevance of microbial inheritance. Interestingly, the emergence of symbiosis could lead to selection acting on the more evolvable and impactful traits – not only the fidelity of inheritance (Bruijning et al., 2020).

Investigating microbial inheritance experimentally poses technical challenges (Perez-Muñoz et al., 2017). However, developments using diverse host species (Moeller et al., 2018; Björk et al., 2019; Renelies-Hamilton et al., 2021; Özkurt et al., 2020), suggest that our predictions could be tested experimentally. Firstly, that inheritance is more influential at intermediate host lifespan, environmental migration, or carrying capacity (Fig. 2.2-2.3). Related host species with diverse life histories could be compared (Song et al., 2020); alternatively, control could be increased using model organisms amenable to manipulate such traits (Bosch et al., 2019). Secondly, that the maximum lineage taxa frequency changes with life-history (Eq. 2.8), could be tested using germ-free or gnotobiotic hosts (Özkurt et al., 2020). Finally, the effect of distinct distributions of microbes inherited (Fig. 2.1) could be surveyed.

Our approach simplifies the complexity of natural microbiomes. A natural step forward would be considering fitness differences among microbes. These could interact with inheritance to preserve or out-compete certain microbes. Secondly, the host population structure could be included. In such a scenario, subpopulations characterized by different microbiomes could emerge (Leftwich et al., 2020). Moreover, critical connectivity might be needed for inheritance to be effective. Finally, we did not account for specific reproductive ages (or development). This might be particularly relevant because, as we have shown, the effect of inheritance erodes through time.

Microbial inheritance can influence the occurrence and abundance of microbes within the host-associated commensal microbiome. Even the persistence of microbes absent from the environment could be facilitated in some cases. These findings extend to diverse scenarios of inheritance representative of different host species. However, inheritance is not a silver bullet, instead life-history in terms of environmental immigration, early microbial prolifera-

tion, and host lifespan limit its magnitude and temporal extent. Only certain naturally occurring host-microbiome pairs might meet such conditions to exploit its benefits.

CHAPTER 3

General death-birth models with immigration

In Chapters 1 and 2 I assumed that microbes are neutrally adapted to their host habitat. Although in this way, I focused on a single ecological driver, neutrality is undoubtedly a strong assumption. In this Chapter, I relax the neutral assumption for the growth and death rates to investigate its effect on the occurrence and abundance of microbes in a simpler model.

This Chapter, available as a preprint (Zapién-Campos et al., 2021b), has been submitted for peer-review under the title *The effect of microbial selection on the occurrence-abundance patterns of microbiomes*, coauthored by Michael Sieber and Arne Traulsen. The authors' contributions are detailed at the end of the thesis.

3.1 Abstract

Theoretical models are useful to investigate the drivers of community dynamics. Notable are models that consider the events of death, birth, and immigration of individuals assuming they only depend on their abundance – thus, all types share the same parameters. The community level expectations arising from these simple models and their agreement to empirical data have been discussed extensively, often suggesting that in nature, rates might indeed be neutral or their differences not important. But, how robust are these model predictions to type-specific rates? And, what are the consequences at the level of types? Here, we address these questions moving from simple to diverse communities. For this, we build a model where types are differently adapted to the environment. We adapt a computational method from the literature to compute equilibrium distributions of the abundance. Then, we look into the occurrence-abundance pattern often reported in microbial communities. We observe that large immigration and biodiversity – common in microbial systems – lead to such patterns, regardless of whether the rates are neutral or non-neutral. We conclude by discussing the implications to interpret and

test empirical data.

3.2 Introduction

Theoretical models have been instrumental in understanding ecological systems. Historically, a handful of puzzling natural observations have motivated their development – from the limits of exponential growth by Malthus (Malthus, 1798) to the competition of species by Lotka and Volterra (Lotka, 1932; Volterra, 1928).

The stark difference of the frequencies of species within communities is one such observation. While few species are very abundant, many others barely appear in community surveys (McGill et al., 2007). Two hypotheses have dominated the scientific discussions. On one hand, it is proposed that biotic interactions and environmental filtering make trophically similar species occupy different niches, which allows differences in abundance while preserving diversity. This is known as niche theory (Chase et al., 2003). Alternatively, Hubbell and others (Rosindell et al., 2012) have emphasized that even if niche differences are discounted, so only species' abundances matter, random fluctuations can lead to the patterns of abundance and diversity observed in nature. This is known as neutral ecological theory (Hubbell, 2001).

Despite their stringent assumptions, neutral models often predict patterns observed in communities as different as the tropical rainforest of Barro Colorado island (Hubbell, 2001) and host-associated microbiomes (Burns et al., 2016; Adair et al., 2018; Sieber et al., 2019). With time, neutral models have become null hypotheses used to discard the need for complex mechanistic explanations in data at the community level (Rosindell et al., 2012).

But how does a neutral model work? In a neutral model the death and birth of individuals account for changes in community composition. However, because each rate is identical for all types, after some time, stochastic drift leads to the extinction of all but one type (Vellend, 2010). Thus, to preserve diversity, an external source of individuals by immigration or speciation is needed. Here, neutral theory builds upon island biogeography. In this theory, MacArthur and Wilson (Macarthur and Wilson, 1967) have modelled the community composition of small habitats ("islands") connected by migration to a larger habitat ("mainland"). In neutral models, a local community commonly receives individuals from an external and larger community (Kessler and Shnerb, 2015). Such community can itself undergo internal changes or, by separation of time scales, assumed to be constant (Hubbell, 2001; Kessler and Shnerb, 2015).

Early on, neutral models have been used in macroecology to address the patterns of diversity and abundance of species (Hubbell, 2001; Rosindell et al.,

2012). More recently, driven by developments in sequencing technologies, the study of patterns of occurrence and mean frequency in microbial communities has become possible (Nemergut et al., 2013). At this scale, ecological drift also seems to greatly influence the community dynamics, leading to hypothesize that many microbial taxa could be classified as neutral (Sloan et al., 2006; Sieber et al., 2019). However, few taxa, referred to as non-neutral, have occurrences and frequencies different than neutrally expected. It has been suggested that the last group might include, among others, pathogens and symbionts (Sieber et al., 2019).

At least two possibilities could lead to deviations from neutrality. Either different processes from those in the neutral model are necessary, or, alternatively, not all the parameters of the model are actually neutral. Both of these lead to develop models of selection (Vellend, 2010). Although many such models have been developed from niche theory assumptions, fewer have been developed from a neutral theory basis (Rosindell et al., 2012; Zhou and Ning, 2017). A direct connection from neutral to selective models would allow to comparing their patterns while acknowledging that both might be operating simultaneously. Indeed, the role of non-neutral processes can only be rejected after ensuring that they can not produce "neutral" patterns (Zhou and Ning, 2017), especially in data.

Neutral and niche models have been connected in several ways (Gravel et al., 2006; Haegeman and Loreau, 2011; Kessler and Shnerb, 2015). Some authors have assumed that the rate of types are solely determined by the environment, finding that neutrality might overshadow the niche structure effect (Chisholm and Pacala, 2010), depending on diversity, dispersal, and niche overlap (Gravel et al., 2006). Alternatively, using Lotka-Volterra models with immigration, the effect of competitive interactions has been studied. Early models focused on intraspecific (Volkov et al., 2005) or interspecific (Allouche and Kadmon, 2009) competition. Later on, both were considered simultaneously. Haegeman and Loreau tuned the niche overlap using symmetric interactions to investigate the success behind the neutral assumption (Haegeman and Loreau, 2011). Kessler and Shnerb classified the dynamics emerging from interspecific interactions, finding that the neutral case links all classes (Kessler and Shnerb, 2015). Focusing on intraspecific interactions, Gravel et al. studied the influence of immigration, suggesting a continuum from competitive to stochastic exclusion (Gravel et al., 2006). Throughout these studies, diversity, community size, and environmental fluctuations seem to have great relevance, as pointed out by Chisholm and Pacala and Fisher and Mehta.

This previous research has proven useful to bridge neutral and selective theories. The link has been instrumental to consider migration, speciation,

and stochastic demography key components in ecology. Along this line and motivated by the particularities of microbial communities, large community size and taxa diversity (Sloan et al., 2006; Costello et al., 2012; Nemergut et al., 2013), here we investigate the commonly observed occurrence-abundance pattern in neutral and non-neutral contexts. Similarly to Sloan et al. and Allouche and Kadmon, we model death, birth, and immigration within a community, but in contrast to these neutral models, type-specific growth and death rates are determined by the environment.

3.3 Results

A spatially-implicit death-birth model with immigration

We consider a set of local communities connected by immigration to a larger community which contains multiple types of individuals. While local communities change as a result of the death, birth, and immigration of individuals, the larger community changes on a much longer time-scale – so immigration to local communities can be assumed to be constant. To derive a dynamical equation of a local community composition, we account for the events that change the frequency x_i of each type $i = 1, ..., S$ within each local community. Individuals die with a rate proportional to the product $x_i\phi_i$ of their frequency and their death rate ϕ_i. Additionally, they are born proportional to the product $x_i f_i$ of their frequency and their growth rate f_i – or arrive with a fraction of the immigration rate m that reflects their frequency p_i in the external environment. Combining these processes, we obtain

$$\frac{dx_i}{dt} = f_i x_i - \phi_i x_i + m p_i. \tag{3.1}$$

Assume for now an equal death rate for all types, $\phi_i = \phi$, so only f_i, m, and p_i are free parameters. To hold the community size constant, we use $\sum_i dx_i/dt = 0$ to find $\phi = \bar{f} + m$, where $\bar{f} = \sum_j x_j f_j$ is the average growth rate of a randomly selected individual. In this way

$$\frac{dx_i}{dt} = x_i(f_i - \bar{f}) + m(p_i - x_i). \tag{3.2}$$

Without immigration, $m = 0$, Eq. (3.2) shows that only types whose growth rate is larger than the average increase. After sufficient time, only the type with the largest growth rate remains. Coexistence is only possible in the neutral case, where all types have the same growth rate, $f_i = \bar{f}$. There, the initial frequencies remain unchanged. Immigration, $m > 0$, creates an

equilibrium that resembles the external composition, p_i, that for sufficiently large immigration might promote coexistence, especially if types with small growth rate migrate more. Similar results are obtained if we assume equal growth rate for all types, $f_i = f$ in Eq. (3.1) instead. In the general case, growth and death rates have opposing effects.

Eq. (3.1) provides useful insights about the dynamics and equilibria; however, only a stochastic model would allow us to compute observables such as the occurrence frequency and the variance. To develop such a model, we track the vector of absolute abundances instead, $\mathbf{n}$, and list the transition rates that change it. The increase of type i by one individual occurs at the expense of the decrease of type j,

$$R(\mathbf{n} \to \mathbf{n} + \mathbf{e}_i - \mathbf{e}_j) = \phi_j \frac{n_j}{N}(f_i n_i + m p_i). \tag{3.3}$$

Here $\mathbf{e}_i$ and $\mathbf{e}_j$ are vectors whose i-th or j-th element equals one and zero elsewhere. The carrying capacity of the community is given by N. The master equation accounts for changes in the probability of observing the community composition $\mathbf{n}$ through time,

$$\begin{aligned} \frac{\partial P(\mathbf{n},t)}{\partial t} = &- \underbrace{\sum_{\substack{i,j \\ i\neq j}} P(\mathbf{n},t) R(\mathbf{n} \to \mathbf{n} + \mathbf{e}_i - \mathbf{e}_j)}_{\text{Probability outflux}} \\ &+ \underbrace{\sum_{\substack{i,j \\ i\neq j}} P(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j, t) R(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j \to \mathbf{n})}_{\text{Probability influx}}, \end{aligned} \tag{3.4}$$

where $P(\mathbf{n}, t)$ is the probability density of community composition $\mathbf{n}$ at time t.

In this work, we investigate the probability distribution at equilibrium, i.e. the state where the master equation equals zero. In this case, the influx and outflux to each state balance each other, ending up with a system of equations that can be solved to find $P(\mathbf{n})$. For communities composed of two types ($S = 2$), a detailed balance analysis (Gardiner, 2004) leads to a recurrence equation of the arbitrarily denoted type 1

$$P(n_1) = P(0) \prod_{\mathbf{n}=(0,N)}^{(n_1-1,1)} \frac{R(\mathbf{n} \to \mathbf{n} + \mathbf{e}_1 - \mathbf{e}_2)}{R(\mathbf{n} + \mathbf{e}_1 - \mathbf{e}_2 \to \mathbf{n})}, \tag{3.5}$$

satisfying $\sum_0^N P(n_1) = 1$. In this case, the transition rates are simplified to the single variable n_1, using $n_2 = N - n_1$ and $p_2 = 1 - p_1$.

For communities with more than two types ($S > 2$) analyses are more challenging, as all possible compositions must be considered. This is particularly true for microbial communities, where many types interact (10^1 to 10^4 taxa are common) in large communities (10^3 to 10^{14} individuals). Although a recurrence equation exists (Allouche and Kadmon, 2009), the exponential increase in the number of states and transitions with S and N, make its computation unfeasible. This is a problem common to microscopic and even mesoscopic descriptions, which has been deemed "the curse of dimensionality" (Xu and Chou, 2018). In neutral models, the equality of rates allows to reduce analyses to a single dimension – that of a focal type (Sloan et al., 2006). However, unless density dependence is neglected, non-neutral models are inherently multidimensional, as transitions depend on the current community composition.

A potential way forward is to acknowledge that, typically, rather than being interested in the probability of every possible community, we are interested in marginal probabilities. In other words, the added probabilities over various dimensions. Methods of model reduction have been developed towards this aim. Based on various assumptions, these methods sacrifice "microscopic" information in the interest of specific observables. Jahnke introduced the *model reduction by conditional expectations* (MRCE), where, while selected types are described stochastically, others are modeled using a mean-field approximation (Jahnke, 2011). The MRCE is derived from the Bayes theorem, by which $P(\mathbf{n}, t)$ is given by the product of two probabilities, one for some chosen types and one for the conditional probability of the others. Then, the probabilities of the others are replaced by expected abundances. Because of the last point, the method is particularly suited to systems where types have peaked distributions and large populations – a situation that can be akin to some microbial communities.

In this paper we combine the MRCE method (Jahnke, 2011) with a detailed balance analysis (Gardiner, 2004) to compute the marginal probability distribution of types within a microbial community. For each distribution at equilibrium, we extract the probability of occurrence, $P(n_i \geq 1)$, the mean frequency $E(n_i)/N$, and compare them in situations of neutrality versus non-neutrality.

To apply the MRCE method, we adapt our model to the convention in (Jahnke, 2011). First, we split the vector of abundances $\mathbf{n} \in \mathbb{Z}^S$ into a focal type i, n_i, and the set of others, $\mathbf{n}_j \in \mathbb{Z}^{S-1}, j \neq i$, for which the marginal probability, $\tilde{P}(n_i) \approx P(n_i)$, and the expected abundance conditioned on the focal type, $(\tilde{\mathbf{n}}_j|n_i) \approx \sum_{\mathbf{n}_j} \mathbf{n}_j P(\mathbf{n}_j|n_i)$, are approximated. Then, each transition rate is factored as the product of rates of the focal type and other

types,

$$R(\mathbf{n} \to \mathbf{n} + \mathbf{e}_i - \mathbf{e}_j) = R_i^+(n_i) R_j^-(\tilde{n}_j|n_i) \tag{3.6a}$$

$$R(\mathbf{n} \to \mathbf{n} - \mathbf{e}_i + \mathbf{e}_j) = R_i^-(n_i) R_j^+(\tilde{n}_j|n_i). \tag{3.6b}$$

In our model, $R_i^+(n_i) = f_i n_i + m p_i$, $R_j^-(\tilde{n}_j|n_i) = \phi_j \frac{\tilde{n}_j}{N}$, $R_i^-(n_i) = \phi_i \frac{n_i}{N}$ and $R_j^+(\tilde{n}_j|n_i) = f_j \tilde{n}_j + m p_j$. With these transformations, the equilibrium is given by the simplified master equation of the focal type i,

$$\begin{aligned} 0 = &- \tilde{P}(n_i) \left(R_i^-(n_i) \sum_{j \neq i}^{S} R_j^+(\tilde{n}_j|n_i) + R_i^+(n_i) \sum_{j \neq i}^{S} R_j^-(\tilde{n}_j|n_i) \right) \\ &+ \tilde{P}(n_i + 1) R_i^-(n_i + 1) \sum_{j \neq i}^{S} R_j^+(\tilde{n}_j|n_i + 1) \\ &+ \tilde{P}(n_i - 1) R_i^+(n_i - 1) \sum_{j \neq i}^{S} R_j^-(\tilde{n}_j|n_i - 1), \end{aligned} \tag{3.7a}$$

and a set of equations for the expected abundance of the others conditioned on the abundance of the focal type $(\tilde{\mathbf{n}}_j|n_i)$,

$$\begin{aligned} \mathbf{0} = &- (\tilde{\mathbf{n}}_j|n_i) \tilde{P}(n_i) \left(R_i^-(n_i) \sum_{j \neq i}^{S} R_j^+(\tilde{n}_j|n_i) + R_i^+(n_i) \sum_{j \neq i}^{S} R_j^-(\tilde{n}_j|n_i) \right) \\ &+ \tilde{P}(n_i + 1) R_i^-(n_i + 1) \left(\sum_{j \neq i}^{S} R_j^+(\tilde{n}_j|n_i + 1)((\tilde{\mathbf{n}}_j|n_i + 1) + \mathbf{e}_j) \right) \\ &+ \tilde{P}(n_i - 1) R_i^+(n_i - 1) \left(\sum_{j \neq i}^{S} R_j^-(\tilde{n}_j|n_i - 1)((\tilde{\mathbf{n}}_j|n_i - 1) - \mathbf{e}_j) \right). \end{aligned} \tag{3.7b}$$

We solve this system of equations in the range of $n_i = 0, \ldots, N$, starting from $n_i = N$. By definition $\tilde{P}(N + 1) = 0$, so no probability flux to or from $N + 1$ occurs. Then, the influx from $n_i = N$ implies $R_j^-(\tilde{n}_j|N) = 0$, specifically $(\tilde{\mathbf{n}}_j|N) = \mathbf{0}$. We end up with a simplified system of equations for $n_i = N$. To compute $\tilde{P}(N - 1)$ and $(\tilde{\mathbf{n}}_j|N - 1)$ from this, we assume without loss of generality $\tilde{P}(N) = c_p$, where c_p is a positive constant. Consecutive $\tilde{P}(n_i - 1)$ and $(\tilde{\mathbf{n}}_j|n_i - 1)$ are computed iteratively. Finally, the normalization $\sum_0^N \tilde{P}(n_i) = 1$ is enforced.

A reliable numerical method is needed to solve Eq. (3.7a-3.7b). The large difference between the magnitudes of $\tilde{P}(n_i - 1)$ and $(\tilde{\mathbf{n}}_j|n_i - 1)$ can cause

numerical problems. To avoid them, we extract $\tilde{P}(n_i - 1)$ from Eq. (3.7a) and substitute it in Eq. (3.7b) – note that all else are known values. The resulting system of equations is solved for $(\tilde{\mathbf{n}}_j | n_i - 1)$, and these substituted in Eq. (3.7a) to compute $\tilde{P}(n_i - 1)$. Caution is needed in cases that lead to a normalized $\tilde{P}(N) \approx 0$, especially if computations are performed in a machine with limited float representation. In this case, we find the $n_i = n_i^*$ closest to $n_i = N$ that while declaring $P(n_i > n_i^*) = 0$ and $P(n_i^*) = c_p$ allows for the iterative solution.

Compared to the fully stochastic model that scales with N^S, here, we solve $N(S-1)$ equations for the marginal probability of each type, i.e. $N(S^2 - S)$ equations for the community. This model reduction allows us to approximate the equilibrium of large communities with many interacting types more rapidly.

The neutral expectation

We start by considering the neutral case – a situation where the rates of all types are equal ($f_i = \phi_i = 1$ for all i in $\{1, ..., S\}$). In contrast to the deterministic model at equilibrium, Eq. (3.2), the frequencies of single stochastic realizations change through time, driven by the probabilistic nature of events. As a result, a distribution of frequencies centered at the value set by the source of immigrants (p_i) emerges. The spread of this distribution inversely depends on the magnitude of the immigration, m.

As shown in Fig. 3.1A, large immigration drives the equilibrium distribution towards its mean value, p_i. On the contrary, without or little immigration, the distribution splits. Thus, the frequencies zero (no individuals of the i-th type) and one (only individuals of the i-th type) are the most probable, decaying towards intermediate frequencies. This is a consequence of noisy fluctuations that, for a single realization, lead to the extinction of all but one type. Whether the frequency one or zero is most likely depends on the proximity of the initial state.

The mean frequency of the stochastic model identically corresponds to the frequency of the deterministic model. As shown in Fig. 3.1B, regardless of the total immigration, the mean frequency of a type increases linearly with the fraction of migrants of its kind.

Besides the mean frequency, one of the simplest, but most informative observables is the occurrence frequency of individuals of a given type in the community. In other words, the probability of observing at least one individual of that type, $P(n_i \geq 1)$. Immigration increases this probability up to the point where the type is always observed in the community (Fig. 3.1C). Importantly, this probability does not increase linearly with the fraction of migrants. Instead, an S-shaped curve is observed, where changes of immigration of rare

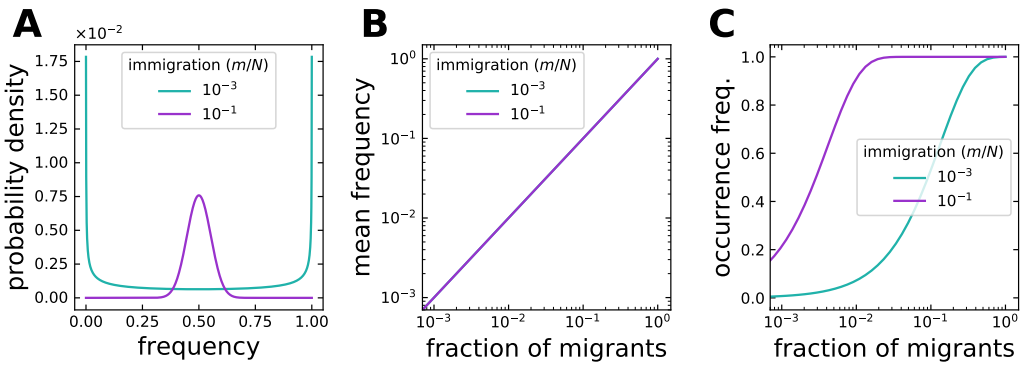

Figure 3.1: **Expected equilibrium of a type if rates in the community are neutral.** (**A**) If the immigration is very small, the population either goes extinct or reaches fixation. A larger immigration reduces the variation in frequency, centered at its fraction of immigrants (here $p_1 = 0.5$). (**B**) The mean frequency increases with the fraction of immigrants p_1, but is independent of the immigration rate m. (**C**) Also the occurrence frequency increase with the fraction of immigrants (p_1), but in an S-shaped manner that depends on m. Deviations from these patterns have been suggested to indicate non-neutral rates (Sieber et al., 2019). The community size is $N = 10^3$.

or abundant types do not modify their occurrence.

Using two simple observables, the mean frequency and the occurrence frequency, we can describe the state of types within a community. In the following, we relax the assumption of neutrality – not enforcing equal growth and death rates. Then, we contrast both observables to their neutral expectation.

Immigration lessens the effect of growth and death differences

To understand the effect of non-neutral rates, we start from a community composed of only two types. Furthermore, we assume only one of them has a non-neutral rate, either f_i or ϕ_i. In this way, we aim to see the effect in the neutral and non-neutral fractions of the community.

For a growth rate below one ($f_1 < 1$) or a death rate above one ($\phi_1 > 1$), the non-neutral type has a reduced mean frequency that preserves its linear relationship to the fraction of immigrants (Fig. 3.2A-B and Fig. 3.3A-B). However, in contrast to the neutral expectation, immigration does play a role, as large migration can reduce the changes occurring in the internal community dynamics (compare panels A to B in Fig. 3.2-3.3). In this context, the neutral type ($f_2 = \phi_2 = 1$) benefits from the reduced proliferation of its partner, thus, gaining in frequency, especially if most immigrants belong to the neutral type.

A similar picture arises for the occurrence pattern. While the non-neutral

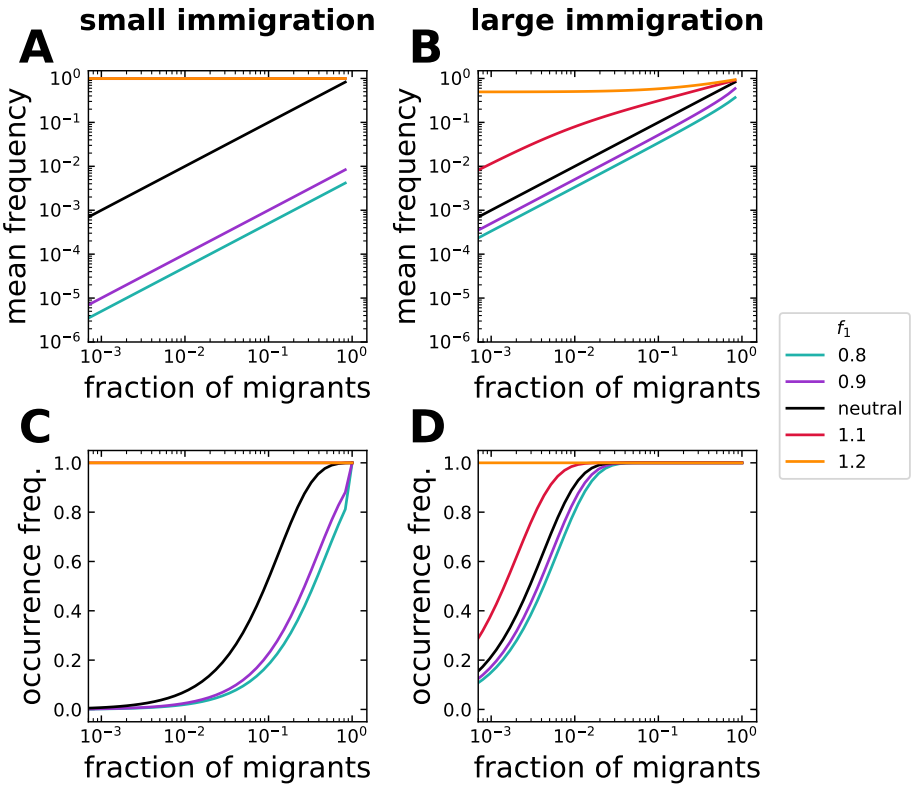

Figure 3.2: **Effect of non-neutral growth rates on the equilibrium of a community with two types.** One of two types has non-neutral growth rate ($f_1 \neq f_2 = 1$), but the death rate is neutral ($\phi_1 = \phi_2 = 1$). In contrast to its all-neutral ($f_1 = f_2 = 1$) expectation, a lower growth rate of the non-neutral type ($f_1 < f_2$) reduces its mean frequency and occurrence. The change can be of several orders of magnitude. Inversely, a larger growth rate of the non-neutral type ($f_1 > f_2$) increases its mean frequency and occurrence. The effect of growth rate differences on the internal dynamics is reduced if immigration is larger, especially for slowly growing types. Immigration is (**A**, **C**) $m/N = 10^{-3}$ and (**B**, **D**) $m/N = 10^{-1}$, with community size $N = 10^3$.

type occurs less frequently, the neutral type thrives, occurring more often than when both types are neutral (Fig. 3.2C-D and Fig. 3.3C-D). The change can be as severe as losing all non-neutral individuals from the community (panel C in Fig. 3.2-3.3). Crucially, large total immigration can prevent this (compare panels C to D in Fig. 3.2-3.3), even if most migrants are of the neutral type.

Once the roles are reversed, so the non-neutral growth rate is above one ($f_1 > 1$) or the death rate below one ($\phi_1 < 1$), the mean frequency and occurrence patterns mirror the previous results (Fig. 3.2 and Fig. 3.3). Although changes produced by non-neutrality in growth ($f_1 \neq 1$) or death

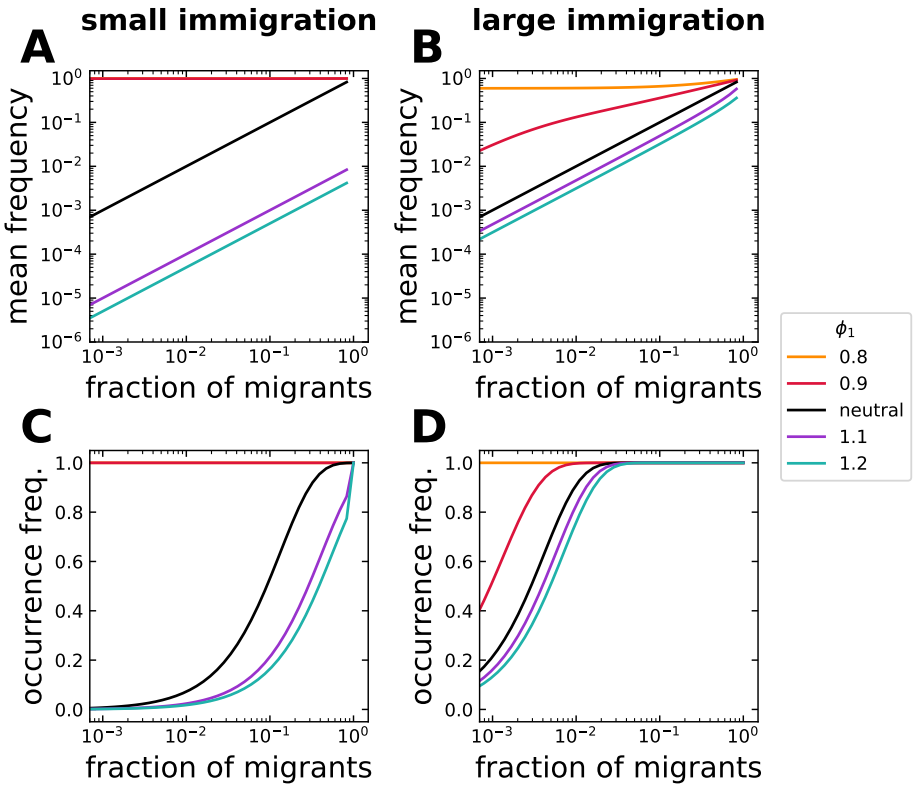

Figure 3.3: **Effect of non-neutral death rates on the equilibrium of a community with two types.** One of two types has non-neutral death rate ($\phi_1 \neq \phi_2 = 1$), but neutral growth rate ($f_1 = f_2 = 1$). Differences in death rates modify the mean frequency and occurrence of both types. A larger immigration reduces differences to the all-neutral ($\phi_1 = \phi_2 = 1$) expectation in a similar fashion to differences in growth rate (Fig. 3.2). Immigration is (**A**, **C**) $m/N = 10^{-3}$ and (**B**, **D**) $m/N = 10^{-1}$, with community size $N = 10^3$.

($\phi_1 \neq 1$) rates are qualitatively similar, they show quantitative differences.

We conclude that even for the simplest community (one with two types), just one non-neutral rate is enough to change the community occurrences and abundances substantially from their all-neutral expectation. This is more visible through the mean frequency (as changes of several orders of magnitude are possible) and for communities with little external migration – where the internal dynamics is more important.

Neutral and non-neutral patterns are similar at the community level but full of differences at the level of types

Communities with two types might occur *in vitro*. However, in nature, communities are much more diverse, especially for microbes. We have pro-

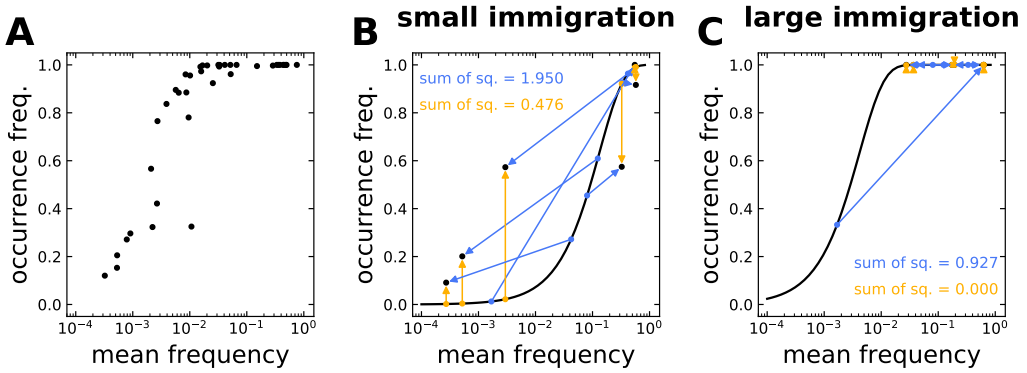

Figure 3.4: **Occurrence-abundance pattern in general non-neutral communities.** (**A**) The non-neutral pattern of a diverse community largely resembles neutral patterns, see Fig. 3.1C. (**B**) However, the change from neutrality of each type can be large (blue arrows), shown here for $m/N = 10^{-3}$. In general, the mean frequency does not equal the fraction of immigrants p_i, assuming otherwise underestimates the change from neutrality (yellow arrows). (**C**) Similar to a community with two types, Fig. 3.2-3.3, the overlap to the neutral expectation increases when immigration, m, is increased to $m/N = 10^{-1}$. The growth and death rates, f_i and ϕ_i, were sampled from a normal distribution with mean 1 and standard deviation 0.1, where $P(f_i < 0.8) = P(f_i > 1.2) \approx 0.023$ and $P(\phi_i < 0.8) = P(\phi_i > 1.2) \approx 0.023$. The fractions of migrants p_i range from 10^{-4} to 10^{-1} and have a $G \approx 0.6$, Eq. (3.8), indicating intermediate immigration asymmetry. Except from the immigration rate m, all rates in (**B**-**C**) are equal. The community size is $N = 10^3$.

duced random instances of such diverse communities, sampling growth and death rates, f_i and ϕ_i, from a normal distribution with mean one and a desired standard deviation. Similarly, we have produced random fractions of migrants, p_i, just conditioned on the Gini index of the community,

$$\mathrm{G} = \frac{1}{S-1} \sum_{i,j} |p_i - p_j|. \tag{3.8}$$

This number that indicates the asymmetry in immigration between types from zero to one, allow us to compare communities quantitatively, regardless of their number of types S. As an example, for $G = 0$ the fractions of migrants are identical for each type, while for $G = 1$ the source pool only contains a single type.

Using these parameters, we have computed the occurrence and abundance frequency of all types in a certain community. Interestingly, the community patterns that we observe are very similar to those expected from neutrality (Fig. 3.4A compared to Fig. 3.1C) – even if asymmetries of growth, death, and immigration increase (Fig. 3.5). In particular, large immigration together with high biodiversity consistently lead to these patterns (Fig. 3.5). This

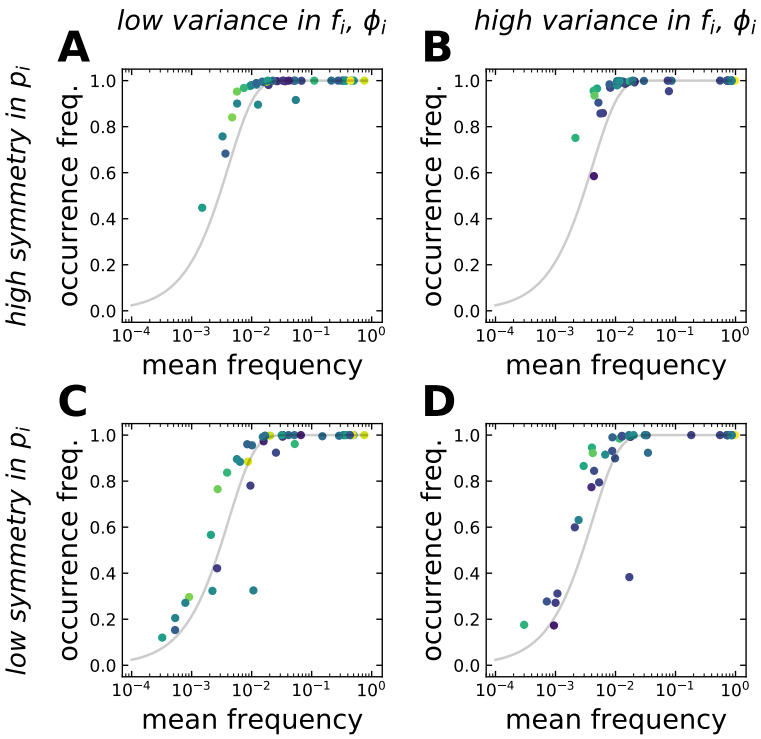

Figure 3.5: **Occurrence-abundance pattern for different levels of asymmetry in the parameters.** A pattern is robust to various asymmetries in the immigration, p_i, and growth and death rates, f_i and ϕ_i. Each community has forty types. For low symmetry in immigration, the types span the range more widely. Colors from dark to light indicate how non-neutral a type is, quantified as the geometric distance from $(f_i, \phi_i) = (1, 1)$. Types overlap regardless of their non-neutrality. The fractions of immigrants, p_i, have a $G \approx 0.3$ (**A**-**B**) or $G \approx 0.6$ (**C**-**D**). The growth and death rates, f_i and ϕ_i, were sampled from a normal distribution with mean 1 and standard deviations 0.1 (**A**, **C**) or 0.2 (**B**, **D**). In the last case, $P(f_i < 0.8) = P(f_i > 1.2) \approx 0.159$ and $P(\phi_i < 0.8) = P(\phi_i > 1.2) \approx 0.159$. Immigration is $m/N = 10^{-1}$, with community size $N = 10^3$.

indicates that neither neutrality nor non-neutrality, but large immigration and biodiversity are behind these patterns.

Even when neutral and non-neutral patterns are similar at the community level, we observe large differences at the level of types. While in the "all-neutral" case, the mean frequency equals the fraction of migrants, $E(n_i)/N = p_i$, this is not the case in a non-neutral scenario (Fig. 3.4B-C). Neither is for the occurrence frequency. The distance from the neutral expectation of each type is not simply related to the level of non-neutrality of its own parameters. Rather, neutral and non-neutral types fall on, above, or below the neutral expectation (Fig. 3.5), highlighting the inherent multidimensionality determining the equilibrium of these communities.

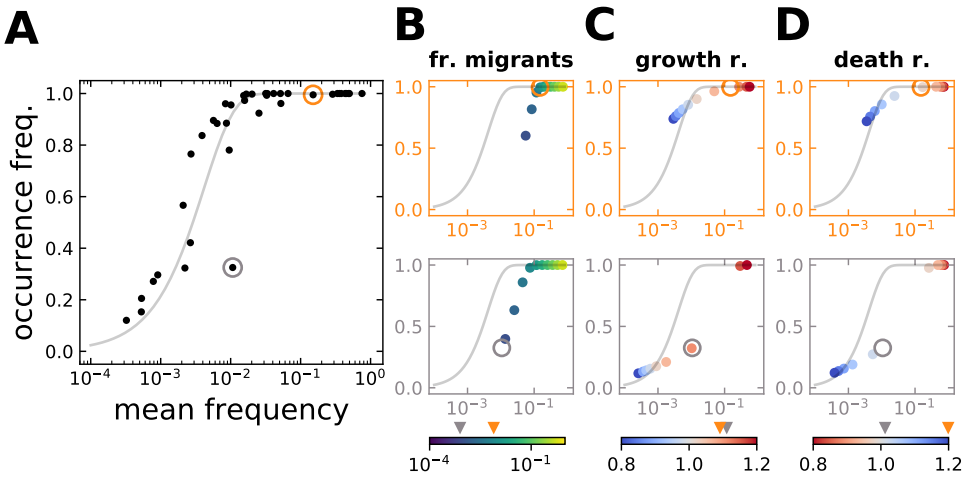

Figure 3.6: **Effect of growth, death, and immigration at the level of types.** (**A**) The community shown corresponds to Fig. 3.5C, with $G \approx 0.6$ for p_i, and f_i and ϕ_i drawn from $\mathcal{N}(1, 0.1)$. Two types are spotted by circles, one that falls on the neutral expectation and the other distant from it. Single parameters are modified in (**B-D**) for both types. Arrows in the colorbars indicate their original values. (**B**) For large fractions of migrants, p_i, non-neutral types are indistinguishable from the neutral expectation; only for small fractions they are below it. (**C**) Different growth rates, f_i, lead non-neutral types to fall on, above, or below the neutral expectation. Changes are especially abrupt for the type with less immigration. (**D**) Different death rates, ϕ_i, mirror the effect of changing growth rates qualitatively. Immigration is $m/N = 10^{-1}$, with community size $N = 10^3$.

To investigate the effect of single parameters at the level of types, we chose two representative types – one close to the neutral expectation, and another one distant from it (Fig. 3.6A). Our results show that types do not remain on or far from the neutral expectation. Rather, the relative magnitude of their growth and death rate, f_i and ϕ_i, is crucial to observe simultaneous decrease or increase in occurrence and mean frequency (Fig. 3.6C-D). In particular, types with a smaller fraction of immigrants, p_i, experience more abrupt changes. Only large fractions of immigrants allow to overcome the effect of growth and death rate differences, leading to large occurrence and mean frequency at the level of types (Fig. 3.6B).

To test neutrality the niche structure must be known first

So far we have used our model to compute observables based on known parameters. However, we can invert this process to infer parameters from simulations or experimental data.

Particularly relevant is the possibility of testing niche structure in data

(Sloan et al., 2006; Burns et al., 2016; Adair et al., 2018; Sieber et al., 2019). Our model indicates care is needed to quantify the true difference from neutrality (Fig. 3.4B-C). In fact, the comparison of the selective case to the neutral case can only be inferred after fitting all parameters of the general model (m, p_i, f_i, and ϕ_i for all i). This is in contrast to the – often used – method by Sloan et al. for neutral conditions, where only the immigration rate m is fitted, while all growth and death rates are assumed $f_i = \phi_i = 1$, and the fraction of migrants p_i equalled to the mean frequency $E(n_i)/N$. Our results indicate these assumptions on the data are unfounded and lead to underestimate niche structure (Fig. 3.4B-C), especially in large communities with many types. Moreover, the consistent occurrence-abundance pattern that we observe (Fig. 3.5), and often reported in data (Burns et al., 2016; Adair et al., 2018; Sieber et al., 2019), emerges from a general death-birth processes with immigration, Eq. (3.3), not just from a neutral process (where $f_i = \phi_i = 1$ for all i). Niche structure – and thus neutrality – can not be discarded or confirmed if certain parameters are fixed *a priori* (Sloan et al., 2006).

The large number of parameters to be fitted requires large datasets. For a community with S types, $3S + 1$ parameters must be fitted, thus requiring at least $3S + 1$ data points. The $2S$ data points obtained from the occurrence and mean frequencies are not sufficient. We propose to include additional observables that can be readily computed from data (Grilli, 2020). These might include, but not be limited to, raw and central moments of the frequency. From this set of observables, available Bayesian methods (Gelman et al., 2013) can be used to infer the parameters using Eq. (3.7a-3.7b).

In Fig. 3.7, we show two potential observables, the variance and the second moment of the distribution. In a community with two types, $S = 2$, both observables reflect the differences in growth, f_i, and death rates, ϕ_i. Only some variances overlap for distinct rates. In this sense, the second raw moment might provide more information to discriminate them. A set of similar observables could allow to characterize the rates of empirical communities.

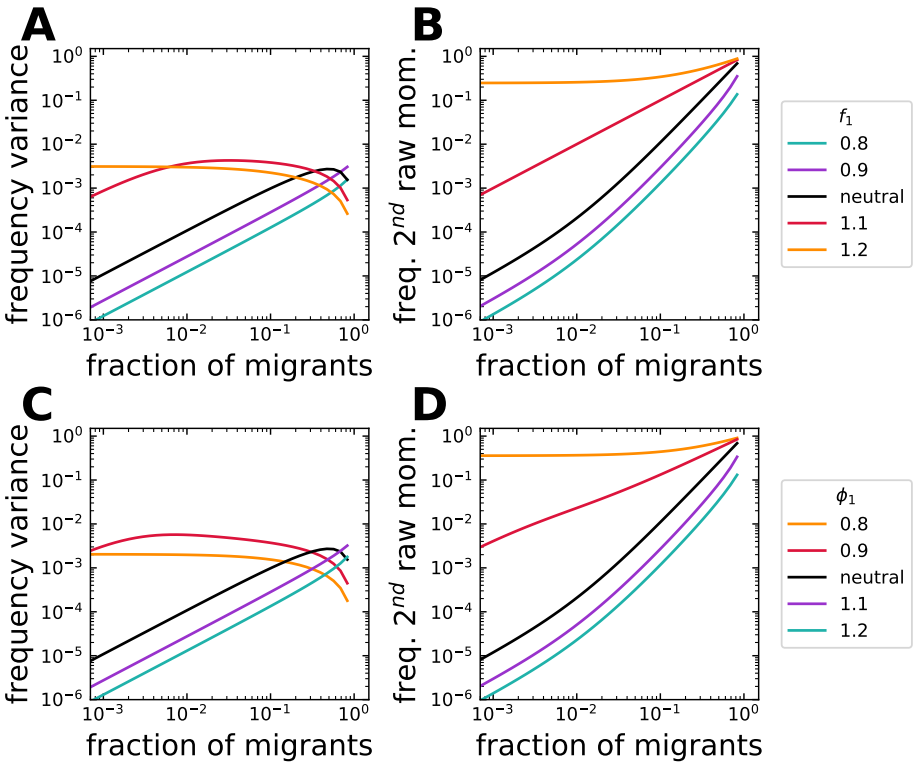

Figure 3.7: **Variance and second raw moment of the frequency.** A community with two types is considered. (**A**-**B**) One type has non-neutral growth rate ($f_1 \neq f_2 = 1$) but neutral death rate ($\phi_1 = \phi_2 = 1$), or (**C**-**D**) a neutral growth rate ($f_1 = f_2 = 1$) but non-neutral death rate ($\phi_1 \neq \phi_2 = 1$). (**B**, **D**) $f_1 > 1$ and $f_1 < 1$ lead to a second raw moment above or below the neutral expectation, respectively. This moment increases continuously with the fraction of migrants, p_i, while the variance reaches a maximum at intermediate p_i (**A**, **C**). In contrast to the second raw moment, the variance of different growth and death rates overlaps. Differences in death rates mirror the effect of growth rate differences qualitatively. Immigration is $m/N = 10^{-1}$, with community size $N = 10^3$.

3.4 Discussion

Understanding the drivers of communities is one of the main objectives of ecological research. In this work, we have used a stochastic death-birth model with immigration to investigate the equilibrium distribution of communities. Comparing cases where changes only depend on the abundances to cases where types have different birth or death rates, we have identified conditions leading to a robust occurrence-abundance pattern – often reported empirically.

Our approach acknowledges the intrinsic density dependence of communities, Eq. (3.3), but simultaneously allow us to compute the equilibrium distribution of large and diverse communities, Eq. (3.7a-3.7b). Combining a

method of model reduction (Jahnke, 2011) and a detailed balance analysis (Gardiner, 2004), we asked questions directly linked to empirical observations. In contrast to studies emphasizing biotic interactions (Volkov et al., 2005; Allouche and Kadmon, 2009; Haegeman and Loreau, 2011; Kessler and Shnerb, 2015), our model can be classified with studies that focus on the differential adaptation to the environment (Gravel et al., 2006; Chisholm and Pacala, 2010). As some of these studies, our results highlight the central role of immigration and biodiversity in community ecology (Chisholm and Pacala, 2010; Fisher and Mehta, 2014).

We tested the reliability of our approach by reproducing known results of neutral adaptation (Sloan et al., 2006). Namely, that the mean frequency of a type equals its immigration and that the occurrence frequency increases in an S-shaped manner with the mean frequency, Fig. 3.1. These results already capture the important role of immigration but discard the frequency dependent effects of other types – for which biodiversity might be important.

The match between community level patterns of neutral models and empirical data has been documented extensively (Rosindell et al., 2012; Burns et al., 2016; Adair et al., 2018; Sieber et al., 2019). Still, some empirical evidence is at odds with neutral theory (O'Dwyer et al., 2015; Davies et al., 2011). The mismatch with evolutionary history – including phylogenetic trees (O'Dwyer et al., 2015; Davies et al., 2011), is one of them. It has been observed that mild differences in adaptation lead to full agreement (Rosindell et al., 2015) – indicating the need to consider models with differential adaptation, even if this is mild.

Here, we considered a general death-birth model where large immigration consistently led to a robust occurrence-abundance pattern. Interestingly, evidence suggests that large immigration might indeed be common in various environmental and host-associated microbiomes (Sieber et al., 2019). Others that deviate from the occurrence-abundance pattern have small immigration (Sieber et al., 2019). Such seems to be the case in *Caenorhabditis elegans*, where active destruction of microbes during feeding results in reduced immigration to the gut microbiome (Vega and Gore, 2017).

A second observation is that with differential adaptation, biodiversity takes a central role. In contrast to the simplest community of two interacting types (Fig. 3.2-3.3), diverse communities promote an occurrence-abundance pattern that resembles the neutral case (Fig. 3.5). With biodiversity, less extreme occurrences and mean frequencies are observed (compare Fig. 3.2-3.3 to Fig. 3.5). Our results agree with research showing that in the limit of high biodiversity, various neutral and non-neutral patterns converge at the community level (Chisholm and Pacala, 2010).

Previous research has speculated about the ecological role of types based

on their location in the occurrence-abundance curve (Sieber et al., 2019) – the motivation being the possibility to identify microbial taxa actively involved in biotic interactions. Our results indicate that such direct identification from occurrence-abundance curves remains challenging, mainly because neutral and non-neutral types can overlap (Fig. 3.6). We propose a way forward, based on the inclusion of new observables computed from data (Grilli, 2020) (Fig. 3.7) combined with robust fitting approaches (Gelman et al., 2013).

Our focus at the level of types revealed the difficulty of assessing niche structure and neutrality from empirical data. While niche and neutral patterns can be indistinguishable at the community level, at the level of types, big differences are observed (Fig. 3.5-3.6). Commonly, in microbial ecology, models have been tested at the community level, where, embraced by a principle of parsimony, neutral interpretations have been suggested (Burns et al., 2016; Adair et al., 2018; Sieber et al., 2019). Our model suggests this is indeed sensible for community level questions. However, for questions at the level of types – including that of ecological roles – general models including differential adaptation can not be avoided. In this case, no parsimonious preference can be given to neutral hypotheses.

The last observation calls for a broader discussion on terminology. As defined by Fisher and Mehta, a community is "statistically neutral" if its distribution can not be distinguished from a distribution constructed under the assumption of ecological neutrality. We must note, however, that ecological neutrality implies statistical neutrality, but statistical neutrality does not necessarily imply ecological neutrality (Fisher and Mehta, 2014). As our results indicate, a reference to large immigration and biodiversity, rather than neutrality, is more accurate and prevents misleading interpretations, that in their worst form, could lead to unfounded generalizations or hold research questions back. On the contrary, our results suggest that numerous questions about neutrality, adaptation, and ecological roles, in microbial ecology and elsewhere are yet to be answered.

Although we mainly focused on microbial communities, our work can be framed in the larger macro-ecological literature. There, a substantial number of models have linked neutral and niche theories (Volkov et al., 2005; Allouche and Kadmon, 2009; Haegeman and Loreau, 2011; Kessler and Shnerb, 2015; Gravel et al., 2006; Chisholm and Pacala, 2010). Heated debates have occurred; however, they have benefited from a close revision of the assumptions on the models and a careful discussion of their implications (Chisholm and Pacala, 2010; Rosindell et al., 2012, 2015). The observation of asymptotically equivalent patterns for neutral and non-neutral rates is one of their main results (Chisholm and Pacala, 2010). We believe microbial research can be guided along this line while offering powerful methods to investigate

general ecological questions (Grilli, 2020). In particular, the possibility to work, *in vivo* and *in vitro*, with large and diverse communities in much shorter time scales (Prosser et al., 2007).

Finally, we should mention some limitations of our work. A limitation of origin is that we considered a differential adaptation to the environment as the sole source of non-neutrality. Certainly, this is not true in nature, where types take part in numerous symbiotic interactions (Kessler and Shnerb, 2015). Therefore, any empirical application of our model should be preceded by evidence of little to no symbiosis. A technical limitation is that we have only approximated the stochastic dynamics (Jahnke, 2011). Our results should be more robust in large communities where types have limited variance (Jahnke, 2011). Interestingly, large immigration – which appears to be common in microbial communities (Sieber et al., 2019) – might lead to satisfying this condition.

Although we provided a focused analysis of the occurrence-abundance pattern at equilibrium, future work could study its dynamics (Zapien-Campos et al., 2020) and derive exact equations for these and other observables (Grilli, 2020). In addition, identifying neutral and non-neutral types remains an open problem. The development of methods for parameter inference from data (Grilli, 2020) seems the way forward.

3.5 Conclusion

Here, we presented a general death-birth model with immigration. Using a method of reduction for the stochastic model, we analysed the equilibrium distribution of abundances for communities equally or differently adapted to the environment. We observe that the community pattern of occurrence-abundance, often reported empirically, is consistently observed in conditions of large immigration and high diversity, regardless of the adaptation to the environment. However, at the level of types, differences in adaptation still lead to large changes.

Code availability

The data generated and analysed in this chapter can be simulated from the Python code available via GitHub at
https://github.com/romanzapien/occurrence-abundance.git.

CHAPTER 4

Closing remarks

4.1 Summary

This thesis compiles my work on the topic of host-associated microbiomes. Each chapter presented a theoretical model where the microbiome composition was analysed. Throughout, I relied on the birth, death, and immigration of microbes as the main ecological drivers. Then, in each chapter, a certain ecological driver, characteristic of host association, was investigated:

- In Chapter 1, I showed the effect of host lifespan – first focusing on the early colonization of microbe-free newborns and then at equilibrium. We observed a large influence of host lifespan in short living hosts and in those with limited environmental immigration – where host subpopulations with distinct microbiome compositions can coexist.

- In Chapter 2, I showed the effect of parental transfer of microbes to newborns. We observed that the life-history traits – microbial immigration, host lifespan, microbial carrying capacity, and the composition of the source of migrants – all determine the effect of the parental transfer of microbes. Still, the "microbial inheritance" could be a reliable mechanism of influencing the microbiome for many host species.

- Finally, in Chapter 3, I relaxed the assumption of neutral adaptation in a simpler death birth-immigration model. We observed that large immigration and high biodiversity – both common in microbial communities – lead to a robust occurrence-abundance pattern at the community level. This pattern is almost indifferent to the differential adaptation of microbes to their habitat, so external drivers outplay local drivers.

Up to this point, each chapter has addressed ecological questions in isolation. In the upcoming paragraphs, I turn isolated results into general

statements – some of these were hypotheses that became asseverations, while others were surprising results that challenge the intuition.

4.2 Overarching conclusions

Some ecological processes are specific to host-associated microbes

A pressing question in microbial ecology is whether theories developed in a macroecological context can be applied to microbes. Similarly to Grilli (2020), this thesis shows that applying them is possible. Nonetheless, other drivers – prominent in the microbial context – will often be needed. I have illustrated this in Chapters 1 and 2, where the host lifespan and the parental transfer of microbes can greatly influence the microbiome composition.

In addition, it has been argued that the microbiome could sometimes become a host trait (Van Vliet and Doebeli, 2019; Bruijning et al., 2020). Interestingly, compared to genetically encoded traits, the microbiome composition could be modified during the life of the host. A possible way of modifying it is what I called in Chapter 2 "microbial inheritance". Our results indicate that host manipulation is possible. By transferring microbes to newborns, hosts can influence not only the presence, Fig. 2.2, but also the abundance of certain microbes, Fig. 2.3 – promoting diversity and preventing extinctions, Fig. 2.5. Nonetheless, not all host species are equally capable of doing this, as life-history traits play a crucial role, Fig. 2.2-2.3. In changing environments, this plastic microbiome composition could be beneficial for the hosts (Bruijning et al., 2020).

Although I investigated salient ecological drivers of a host association – host lifespan and microbial inheritance – many others are yet to be investigated. Future work could dwell upon comparisons between environmental and host-associated habitats in the context of homeostasis, rhythms, and directional migration – e.g., in the digestive tract.

Stochasticity without selection can give rise to empirical patterns

Possibly the most remarkable result of all models in this thesis is how besides the strong assumption of neutral adaptation, these can lead to diverse ecological dynamics and equilibria. As I highlight in the discussion within each chapter, the predictions even match empirical observations.

The first observation is the emergence of distinct microbiome compositions within the host population. In Chapter 1, this was a consequence of the microbiome changing in two time-scales: a fast time-scale of colonization

where individual microbiomes greatly differ, and a second time-scale, where the microbiomes slowly converge to the composition of the source of immigrants, Fig. 1.2.

A second observation is the relevance of bottlenecks. In my models, the immigration of microbes is the first bottleneck. This can differ greatly between host species (Sieber et al., 2019), creating ample variation of the microbiome composition, Fig. 1.2, 1.3, and 1.5. The second bottleneck is the transfer of microbes from parents to newborns. This could also differ greatly between host species, Fig. 2.1, potentially influencing the long-term composition of the microbiome. In the absence of microbial "inheritance", or mechanisms like immunity, targeted nutrients, or physiological adaptations, the microbiome dynamics is largely self-determined. Bottlenecks can be a less sophisticated way to tighten or loosen control over the microbiome.

A final observation is that, interestingly, ecological drift is relevant in large communities. As my results show, bottlenecks can have long-lasting consequences, even if a community is large at present. Moreover, the timing of bottlenecks can be extremely important – e.g., the moment of host death in Chapter 1 and the moment of parenting in Chapter 2.

Maps of life-history traits to identify patterns across host species

I have argued previously that some ecological drivers are specific to the microbiome ecology. The added complexity might seem overwhelming; however, identifying general patterns in nature might still be possible.

One way is to create maps of possible life-history traits. This idea is similar to that of a morphospace – where existing organisms are located upon a space where axes indicate the value of traits (Mitteroecker and Huttegger, 2009). I have used this approach in Chapter 1, where axes indicated immigration and host lifespan, Fig. 1.1. For microbial immigration, hosts are separated into those highly connected to the environment and those highly isolated. For host lifespan, hosts are separated in long-living, like humans, and those with a short life, like *D. melanogaster* and *C. elegans*. Based on empirical observations (Sieber et al., 2019; Gibson et al., 2018), I have located some representative species and related them to their predicted microbiome composition, Fig. 1.3-1.6.

In Chapter 2 the microbiome size emerged as another relevant trait. This is because, commonly, individual microbial taxa have a low frequency – thus, their presence is always challenged. On one extreme, limited available space leads to extreme competition; on the other, enough available space allows microbial coexistence, Fig. Sup. C.1.

The composition of the source of migrating microbes was another trait.

Balanced immigration can alleviate the fierce competition for space – preventing extinctions within the hosts, Fig. 3.5. In all chapters, I assumed a constant source of immigrants. An interesting possibility is to investigate the coupling between the microbiomes of hosts and the source of migrants.

The life-history traits mentioned above are just a set of those that could be potentially important. Future studies could be guided by a trait map, allowing the identification of general microbiome patterns across host species.

Distinct ecological processes can lead to the same patterns

A major finding is that care is needed to attribute results to specific ecological drivers. First, in Chapter 1 I showed how multiple values of microbial immigration and host lifespan could lead to the same qualitative and quantitative results, Fig. 1.3-1.6. Then, in Chapter 3 I showed how large microbial immigration and high biodiversity could dominate the effect of the growth and death rates, Fig. 3.5.

Although in this thesis I emphasized the level of populations, other authors have focused on the community level – even reaching different conclusions with the same data. One example is the occurrence-abundance pattern in Chapter 3, which in my case indicates the importance of large immigration and biodiversity, not that adaptation is neutral (Sloan et al., 2006). What might be clear at one ecological level might not be at another.

In a context where multiple ecological drivers are at play, this careful analysis will be important – especially if not only models but also empirical data are analysed. In fact, whether the quality of empirical data can be sufficient to distinguish between the ecological drivers is an open question. Extra control could be gained in laboratory settings.

Models are useful to investigate otherwise infeasible questions

At the moment, various questions about the ecology and evolution of host-associated microbiomes are not experimentally feasible (Koskella et al., 2017). The work in this thesis shows how mathematical models offer an option. The constructive approach of modeling leads to question assumptions, as well as the interaction between drivers of the dynamics. Nonetheless, solid theories in biology can only be built together with empirical evidence. As laboratory technologies improve, the results in this thesis could be tested.

The rich ecological theory has come to be thanks to the dialogue between mathematical and empirical studies. I see this as the way forward for microbiome research. Which, as might be clear at this point, still has numerous questions to answer.

4.3 Open questions

The results presented in this thesis include just a couple of the ecological processes that might be relevant in *host-associated* microbiomes. Many processes are yet to be studied, and additional questions have emerged.

Are there analytical results at hand?

An open question is whether analytical results can be obtained for the ecological processes studied here. Although we simplified the complexity of real hosts and microbes – particularly the host lifespan and the parental transfer of microbes – analytical solutions seemed out of reach. Future research could aim to obtain such solutions, that similarly to macro-ecology, could allow an extensive application to empirical studies (Allouche and Kadmon, 2009; Ofiţeru et al., 2010).

What will data inference tell us?

An interesting possibility is applying the models presented here to fit empirical data. For this, not only reliable inference methods are needed, but also rich datasets where the processes here described have been characterized. The fitting of data would allow us to contrast the predictions shown here, shed light on the relevant ecological drivers in nature, and inform the development of future microbiome models.

What is the effect of more realistic life-history traits?

Although I considered simple mechanisms of host reproduction, host lifespan, and microbiome inheritance, nature is more complicated. In fact, as argued in Chapters 1 and 2, because the timing of host death and host reproduction had a strong influence, the assumption of how they operate might have a central role. In addition, microbial immigration and the carrying capacity for microbes are two traits that evidence suggests could change throughout the life of a host (Burns et al., 2016).

What is the effect of selection on the more complicated models?

Neutral adaptation is undeniably a strong assumption. As remarked in Chapter 3, different ecological processes can lead to similar patterns. It is, therefore, interesting to consider how different would the findings of Chapters 1 and 2 be without neutrality. Specifically, how influential are the host

lifespan and the parental transfer of microbes to newborns in this context. Moreover, selection does not stop with microbes – as Van Vliet and Doebeli (2019) show, selection acting at the level of hosts can be highly influential.

What is the effect of including evolutionary processes?

The models presented here are ecological. Although I made reference to evolutionary implications, the possibility of including evolutionary processes explicitly – such as mutation and diversification – remains open. Apart from changes in the microbes, an interesting possibility is to allow the host traits to evolve – e.g., the amount of microbial immigration, the host lifespan, and the parental transfer of microbes to newborns. Other models have shown such possibilities in scenarios where selection plays an important role in the multilevel selection of the host-microbiome pair (Van Vliet and Doebeli, 2019; Bruijning et al., 2020).

Towards a theory of host-associated microbiomes?

All in all, the results presented here build upon an eco-evolutionary theory of host-associated microbiomes. To achieve such theory, more processes have to be investigated empirically and mathematically. In doing so, this theory could inform the larger ecological and evolutionary theories, from which microbiology has greatly benefited (Costello et al., 2012).

Is there a better level to address microbiome questions?

Possibly the most recurrent question to empirical and theoretical scientists is if there is a level at which microbiomes can be better understood. In this thesis, I have presented models which build from the bottom-up, assuming hosts as habitats and only tracking the identity of microbes without considering functional traits. Others have suggested a different way, in which a top-down approach emphasizes the collective functioning of microbes giving a secondary role to their identities (Louca et al., 2016). It is yet to be seen how both approaches complement each other and how they can allow a dialogue between empirical and theoretical approaches.

In summary, an exciting future lies ahead for microbiome research – and microbiology at large: a future of revisionism, where our understanding of biological concepts will be constantly tested. This thesis introduced me to the topic – specifically to the ecology of host-associated microbiomes. I look forward to future developments with great excitement.

APPENDIX A

On reproducibility and code availability

The work presented in this thesis is theoretical – thus, the main subject of study is models and equations. However, in addition, source code has been essential to solve them or simulate the ecological processes presented here (Fig. 0). The software pipelines that I developed using open source languages can be readily used by interested readers to reproduce our results.

Rather than being a source of extra work, developing publicly available software guided the progression and organization of the projects presented here, allowing a more accessible and clear technical communication.

The source code to generate data and produce figures is available via the following public online repositories:

1. **Chapter 1**. Stochastic colonization of microbe-free hosts.
   https://github.com/romanzapien/microbiome-hostspan.git
2. **Chapter 2**. Parental transfer of microbes to newborns.
   https://github.com/romanzapien/microbiome-inheritance.git
3. **Chapter 3**. General death-birth models with immigration.
   https://github.com/romanzapien/occurrence-abundance.git

Each repository can be pulled from its source. Within each of them, the directories are organized in *simulations*, *numerics*, and *code for figures*. For *simulations* and *numerics*, the source code is separated in the following way:

1. *Source code.* Low-level code constructively leads to top-level functions.
2. *Parameters.* Data will be generated based on the values specified here.
3. *Execute.* Top level functions produce the main results. Roughly, a top level function is equivalent to a figure or figure panel.

For the *code for figures* the data generated by *simulations* or *numerics* is gathered to produce the final figures presented in this thesis.

Each repository provides a detailed annotation of the code and an indication of the programming language versions used to run the code.

APPENDIX B

Supplementary material of Chapter 1

B.1 Supplementary methods

Derivation of the Fokker-Planck approximation with resetting

Define a Markov process in the discrete state space $\{0, 1, ..., N\}$, where N is the maximum capacity for microbes within a single host. Let n_i be the number of individuals of the *i-th* microbial taxon for $i = \{1, ..., M\}$ and n_0 the amount of unoccupied space. The probability of being in state n_i at time $t + 1$ is given by the sum of the probabilities to go in, out and remain in n_i at time t,

$$\Phi_i[n_i, t+1] = \sum_{o_i=0}^{N} P_i[o_i \to n_i]\Phi_i[o_i, t],$$

where $\Phi_i[o_i, t]$ is the probability of being in state o_i at time t, and $P_i[o_i \to n_i]$ is the probability of transitioning from state o_i to n_i. Taking out the probability of remaining at n_i we get

$$\Phi_i[n_i, t+1] = P_i[n_i \to n_i]\Phi_i[n_i, t] + \sum_{o_i \neq n_i} P_i[o_i \to n_i]\Phi_i[o_i, t]$$

However $P_i[n_i \to n_i] = 1 - \sum_{o_i \neq n_i} P_i[n_i \to o_i]$. Therefore

$$\Phi_i[n_i, t+1] - \Phi_i[n_i, t] = -\sum_{o_i \neq n_i} P_i[n_i \to o_i]\Phi_i[n_i, t] + \sum_{o_i \neq n_i} P_i[o_i \to n_i]\Phi_i[o_i, t]$$

Dividing by Δt and taking the continuous time limit, we find the time continuous master equation

$$\frac{\partial \Phi_i[n_i, t]}{\partial t} = -\sum_{o_i \neq n_i} T_i[n_i \to o_i]\Phi_i[n_i, t] + \sum_{o_i \neq n_i} T_i[o_i \to n_i]\Phi_i[o_i, t],$$

where $T_i[n_i \to o_i]$ and $T_i[o_i \to n_i]$ are transition rates. This equation contains transitions to and from neighbouring states of n_i, i.e. $n_i - 1$ and $n_i + 1$, but also to and from non-neighbouring states due to resetting events.

The next derivation steps focus on approximating the neighbouring transitions locally around n_i. However, because the resetting events are non-neighbouring transitions, we have to treat them separately. To do this, we assume that each $T_i[n_i \to o_i]$ and $T_i[o_i \to n_i]$ can be linearly separated in contributions due to resetting and non-resetting events.

Define R_i as the only state towards which the resetting occurs, so the master equation for this state is

$$\begin{aligned}\frac{\partial \Phi_i[R_i,t]}{\partial t} =& -\sum_{o_i \neq R_i} T_i[R_i \to o_i]\Phi_i[R_i,t] + \sum_{o_i \neq R_i} T_i^*[o_i \to R_i]\Phi_i[o_i,t] \\ &+ \sum_{o_i \neq R_i} \tau \Phi_i[o_i,t],\end{aligned}$$

where $T_i^*[o_i \to R_i]$ are transition rates from neighbouring o_i to R_i, and τ is the transition rate from non-neighbouring o_i to R_i (which we assume to be independent of o_i). We can rewrite this as

$$\begin{aligned}\frac{\partial \Phi_i[R_i,t]}{\partial t} =& -\sum_{o_i \neq R_i} T_i[R_i \to o_i]\Phi_i[R_i,t] + \sum_{o_i \neq R_i} T_i^*[o_i \to R_i]\Phi_i[o_i,t] \\ &+ \tau(1 - \Phi_i[R_i,t])\end{aligned}$$

For any other state $n_i^* := n_i \neq R_i$, the master equation is

$$\frac{\partial \Phi_i[n_i^*,t]}{\partial t} = -\sum_{o_i \neq n_i^*} T_i^*[n_i^* \to o_i]\Phi_i[n_i^*,t] - \tau\Phi_i[n_i^*,t] + \sum_{o_i \neq n_i^*} T_i[o_i \to n_i^*]\Phi_i[o_i,t],$$

where $T_i^*[n_i^* \to o_i]$ are transition rates from n_i^* to neighbouring o_i. We can combine both master equations in the following way

$$\begin{aligned}\frac{\partial \Phi_i[n_i,t]}{\partial t} =& -\sum_{o_i \neq n_i} T_i^*[n_i \to o_i]\Phi_i[n_i,t] + \sum_{o_i \neq n_i} T_i^*[o_i \to n_i]\Phi_i[o_i,t] \\ &+ \tau\left(\delta_{n_i,R_i} - \Phi_i[n_i,t]\right),\end{aligned}$$

where δ_{n_i,R_i} is the Kronecker delta (1 for $n_i = R_i$ and 0 in all other cases), and $\tau\Phi_i[n_i,t]$ and $\tau\delta_{n_i,R_i}$ are the non-neighbouring outflux and influx, respectively.

Now we can approximate the neighbouring transitions contained in the sums. We define $x_i = n_i/N$ and $r_i = R_i/N$, which are approximately

continuous in the large N limit. With this the sums are replaced by integrals

$$\begin{aligned}\frac{\partial \Phi_i[x_i,t]}{\partial t} =& -\int T_i^*[x_i \to x_i + \Delta x_i]\Phi_i[x_i,t]d\Delta x_i \\ &+ \int T_i^*[x_i + \Delta x_i \to x_i]\Phi_i[x_i + \Delta x_i,t]d\Delta x_i \\ &+ \tau\left(\delta_{x_i,r_i} - \Phi_i[x_i,t]\right),\end{aligned}$$

where δ_{x_i,r_i} is 1 for $x_i = r_i$ and 0 otherwise. We focus on the interval Δx_i around x_i to obtain the Taylor expansion of the influx, $\int T_i^*[x_i + \Delta x_i \to x_i]\Phi_i[x_i + \Delta x_i, t]d\Delta x_i$

$$\begin{aligned}=& \int T_i^*[x_i \to x_i - \Delta x_i]\Phi_i[x_i,t]d\Delta x_i \\ &- \int \Delta x_i \frac{\partial}{\partial x_i}(T_i^*[x_i \to x_i - \Delta x_i]\Phi_i[x_i,t])d\Delta x_i \\ &+ \int \frac{(\Delta x_i)^2}{2}\frac{\partial^2}{\partial x_i^2}(T_i^*[x_i \to x_i - \Delta x_i]\Phi_i[x_i,t])d\Delta x_i + ...\end{aligned}$$

Realizing that the magnitude of the zeroth order term equals the outflux, truncating the expansion at the second order and putting terms together, we find

$$\begin{aligned}\frac{\partial \Phi_i[x_i,t]}{\partial t} =& -\frac{\partial}{\partial x_i}\left(\Phi_i[x_i,t]\int \Delta x_i T_i^*[x_i \to x_i - \Delta x_i]d\Delta x_i\right) \\ &+ \frac{1}{2}\frac{\partial^2}{\partial x_i^2}\left(\Phi_i[x_i,t]\int (\Delta x_i)^2 T_i^*[x_i \to x_i - \Delta x_i]d\Delta x_i\right) \\ &+ \tau\left(\delta_{x_i,r_i} - \Phi_i[x_i,t]\right)\end{aligned}$$

Which simplifies to

$$\boxed{\begin{aligned}\frac{\partial}{\partial t}\Phi_i[x_i,t] =& \frac{\partial}{\partial x_i}\left(-a_i[x_i]\Phi_i[x_i,t] + \frac{1}{2}\frac{\partial}{\partial x_i}\left(b_i^2[x_i]\Phi_i[x_i,t]\right)\right) \\ &+ \tau\left(\delta_{x_i,r_i} - \Phi_i[x_i,t]\right),\end{aligned}} \tag{S1.1}$$

where $a_i[x_i]$ and $b_i^2[x_i]$ are the expected change and squared-change of x_i, respectively. These quantities are derived in the main text based on the process of death, birth and immigration of microbes. Although, the change of x_i depends on all $k \neq i$. Derived from our assumptions, $a_i[x_i]$ and $b_i^2[x_i]$ contain all the information of a host microbiome from the *j-th* microbial taxon or unoccupied space ($i = 0$) perspective.

Importantly, host death, expressed by $\tau(\delta_{x_i,r_i} - \Phi_i[x_i,t])$, has opposite effects on microbes and unoccupied space. For a microbial taxon the frequency resets to zero, $r_i = 0$, while for unoccupied space it resets to one, $r_0 = 1$. Eq (S1.1) is the same as Eq (2) in the main text.

A numerical solution of the stationary model, Eq (5) in the main text, can be found using the boundary conditions $\frac{d\Phi_0[0]}{dx_0} = \frac{d\Phi_0[1]}{dx_0} = 0$ and $\frac{d\Phi_i[0]}{dx_i} = \frac{d\Phi_i[1]}{dx_i} = 0$, alongside $\Phi_0[1] = 1$ for unoccupied space and $\Phi_i[0] = 1$ for microbes (Kumar and Narayanan, 2006), enforcing the normalization condition

$$\int_0^1 \Phi_i[x_i]dx_i = 1$$

B.2 Supplementary figures

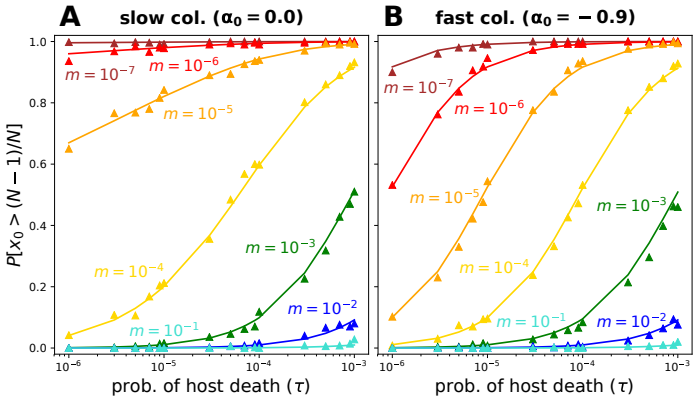

Figure Supplementary B.1: **Comparison between simulations and the model: probability of microbe-free hosts in the stationary distribution.** The $P[x_0 > (N-1)/N]$ is shown, Eq (1.10). Lines show the model prediction, while triangles show the average over the steady state of 500 host samples according to Eq (1.6). The match spans several magnitude orders of migration (m) and probability of host death-birth events (τ). The probability increases for shorter host lifespans (larger τ) and less migration to the hosts (smaller m). The rate of occupation of empty space (α_0) has a larger effect on cases where migration is limited and the host lifespan is long (small τ). Simulations were computed as explained in the Methods. Other parameters: $N = 10^4$. We use Eq (1.5a) where no definition of p_i and a_i is required.

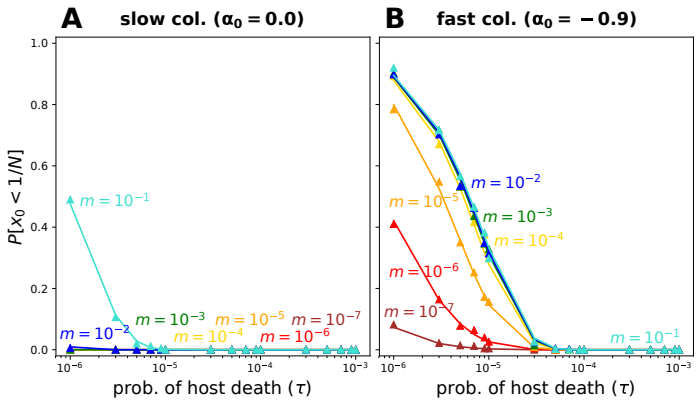

Figure Supplementary B.2: **Comparison between simulations and the model: probability of full colonization in the stationary distribution.** The $P[x_0 < 1/N]$ is shown, Eq (1.9). Lines show the model prediction, while triangles show the average over the steady state of 500 host samples according to Eq (1.6). The match spans several magnitude orders of migration (m) and probability of host death-birth events (τ). The probability increases for longer host lifespans (smaller τ) and larger migration to the hosts (larger m). The rate of occupation of empty space (α_0) has a larger effect on cases where migration is large and the host lifespan is long (small τ). Simulations were computed as explained in the Methods. Other parameters: $N = 10^4$. We use Eq (1.5a) where no definition of p_i and a_i is required.

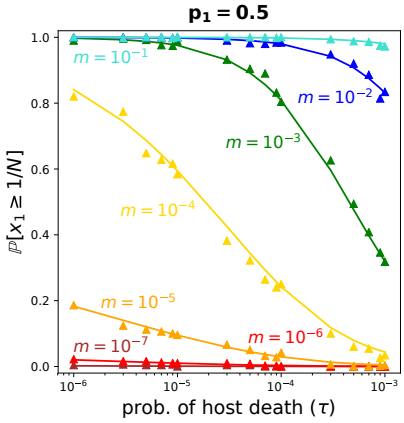

Figure Supplementary B.3: **Comparison between simulations and the model: probability of colonization of microbial taxon 1 in the stationary distribution.** p_1 indicates the frequency of microbial taxon 1 in the pool of colonizers. Lines show the model prediction, while triangles show the average over the steady state of 500 host samples according to Eq (1.6). The match spans several magnitude orders of migration (m) and probability of host death-birth events (τ). The probability increases for longer host lifespans (smaller τ) and larger migration to the hosts (larger m). Simulations were computed as explained in the Methods. Other parameters: $N = 10^4$ and $\alpha_0 = \alpha_1 = 0$.

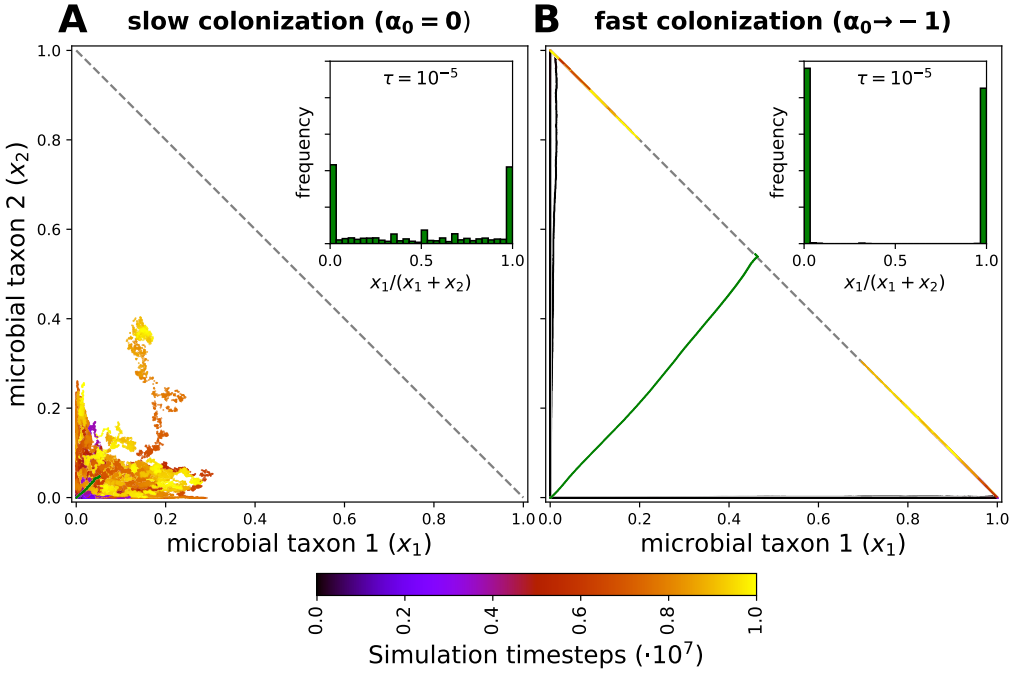

Figure Supplementary B.4: **Individual-based simulations of colonization for two neutral microbial taxa: limited migration.** Except for $m = 10^{-4}$, all parameters are equal to those in Fig 1.2. (**A**) The limited migration and slow empty space occupation impedes the host from being colonized completely. (**B**) When empty space is occupied rapidly, although the mean is conserved, the distribution becomes sharply bimodal as a result of the fast proliferation of the first colonizer, and a slow convergence to the long-term equilibrium, which within the time-range simulated is not reached. For finite host lifespans, such dynamics can produce alternative microbiomes and partial colonization of hosts in the equilibrium.

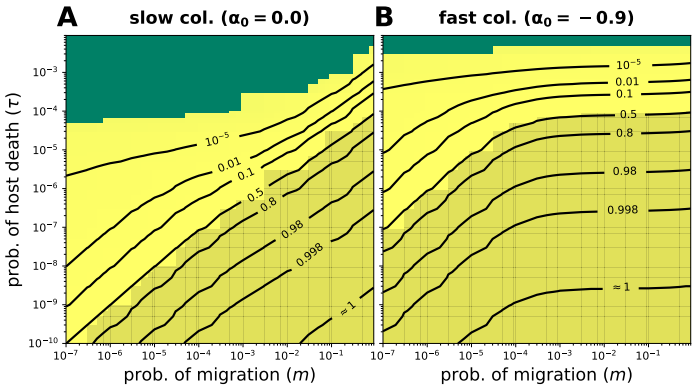

Figure Supplementary B.5: **Probability of full colonization in the stationary distribution: smaller microbiome size.** Except for $N = 10^3$, all parameters are equal to those in Fig 1.5. The smaller capacity for microbes of a host makes full colonization more likely, and migration (m) has increased influence for larger τ.

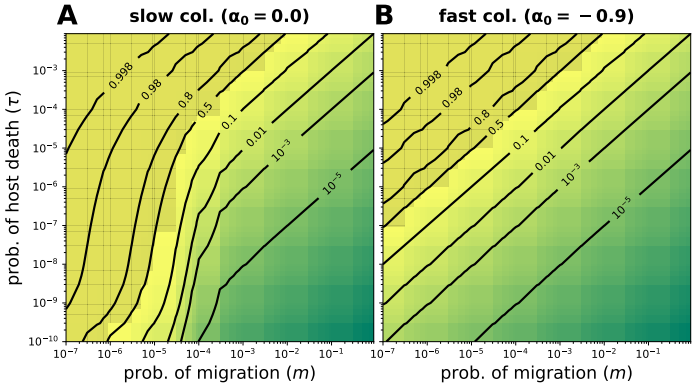

Figure Supplementary B.6: **Probability of microbe-free hosts in the stationary distribution.** The $P[x_0 > (N-1)/N]$ is shown, Eq (1.10). (**A**) Migration (m) is the main driver of the microbe-free state, but still interacting with the probability of host death-birth (τ). The microbe-free state prevails for small m, increasing in the direction of a short host lifespan (large τ). (**B**) Although a faster occupation of empty space decreases its probability, microbe-free hosts are still expected. Moreover the host lifespan (via τ) becomes more relevant. Other parameters: $N = 10^4$. We use Eq (1.5a) where no definition of p_i and a_i is required.

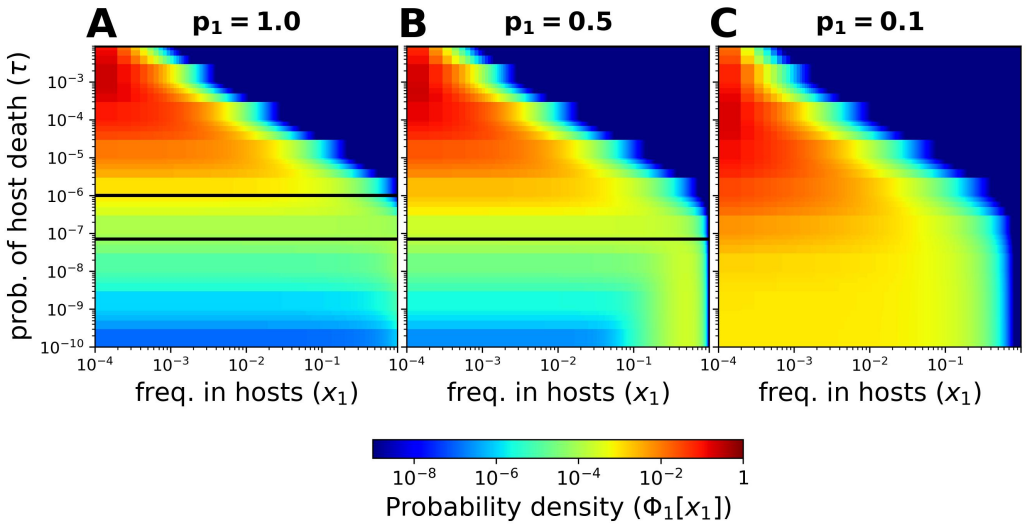

Figure Supplementary B.7: **Probability density of microbial taxon 1 as a function of host death**. The cross-section of Fig 1.4 where $m = 10^{-3}$ is shown. (**A**). If there are only microbes of type 1 in the pool of colonizers ($p_1 = 1$), small τ implies that there is a single maximum at $x_1 = 1$ – the hosts tend to be fully occupied. Bimodality is observed for $7 \cdot 10^{-8} \lesssim \tau \lesssim 10^{-6}$ – some hosts are occupied, but some remain empty. For large τ, hosts tend to remain empty and the distribution has a single maximum at $x_1 = 0$. Black lines indicate the boundaries separating them (see Fig 1.4). (**B**) If the microbe is present in the pool of colonizers at $p_1 = 0.5$, no bimodality is observed. For small τ the frequencies are representative of the pool of colonizers and for large τ most hosts do not contain microbe 1. (**C**) If the microbe is rare in the pool of colonizers, $p_1 = 0.1$, the distribution has a single peak at $x_1 = 0$. This occurs for all values of τ shown here, because there is not enough time in the host to reflect the small number of microbe 1 individuals in the pool of colonizers (any probability smaller than 10^{-9} was considered as zero, $N = 10^4$ and $\alpha_0 = \alpha_1 = 0$).

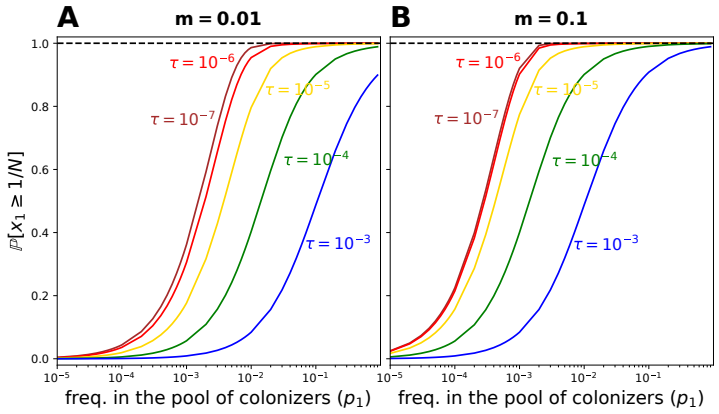

Figure Supplementary B.8: **Probability of colonization of microbial taxon 1 as a function of its frequency in the pool of colonizers**. The results of multiple probabilities of host death-birth events (τ) are shown. Overall, the probability of colonization increases with the frequency in the pool of colonizers (p_1), but decreases as the host lifespan shortens (larger τ). A smaller migration (m) decreases the probability. Other parameters: $N = 10^4$ and $\alpha_0 = \alpha_1 = 0$.

APPENDIX C

Supplementary material of Chapter 2

C.1 Supplementary methods

Deterministic and stochastic components of the model

We have introduced a model of the microbiome dynamics where we track the frequencies of a taxon i, x_i, and the set of other taxa, o_i; together, the vector $\mathbf{x} = \{x_i, o_i\}$. In Eq. (2.5) we expressed the model in the form of a stochastic differential equation – that describes the microbial dynamics within a host during its lifespan– where the deterministic, $\mathbf{A}[\mathbf{x}]$, and stochastic, $B[\mathbf{x}]$, contributions were introduced. Changes have magnitude $\frac{1}{N}$. The deterministic part is given by the expected change of $\mathbf{x}$ that results from the transition probabilities in Eq. (2.1),

$$\mathbf{A}[\mathbf{x}] = \frac{1}{N}\frac{1}{1-\tau}\begin{bmatrix} T^{o_i-}_{x_i+} + T^{o_i}_{x_i+} - T^{o_i+}_{x_i-} - T^{o_i}_{x_i-} \\ T^{o_i+}_{x_i-} + T^{o_i+}_{x_i} - T^{o_i-}_{x_i+} - T^{o_i-}_{x_i} \end{bmatrix}. \tag{S2.1}$$

The stochastic part is related to the matrix of covariant change of $\mathbf{x}$:

$$V[\mathbf{x}] = \frac{1}{N^2}\frac{1}{1-\tau}\begin{bmatrix} T^{o_i-}_{x_i+} + T^{o_i}_{x_i+} + T^{o_i+}_{x_i-} + T^{o_i}_{x_i-} & -(T^{o_i+}_{x_i-} + T^{o_i-}_{x_i+}) \\ -(T^{o_i+}_{x_i-} + T^{o_i-}_{x_i+}) & T^{o_i+}_{x_i-} + T^{o_i+}_{x_i} + T^{o_i-}_{x_i+} + T^{o_i-}_{x_i} \end{bmatrix}. \tag{S2.2}$$

$B[\mathbf{x}]$ is the matrix that satisfies $B[\mathbf{x}]^T B[\mathbf{x}] = V[\mathbf{x}]$. This is calculated analytically (Allen, 2007) after defining the quantities $w = \sqrt{\det(V[\mathbf{x}])}$ and $d = \sqrt{\sum_i V[i,i] + 2w}$,

$$B[\mathbf{x}] = \frac{V[\mathbf{x}] + wI}{d}, \tag{S2.3}$$

where I is the identity matrix.

Note that Eq. (S2.1) and Eq. (S2.2) refer to the lifetime of a single host, therefore we divide by $1-\tau$ to remove it from each transition probability. We had introduced $1-\tau$ in Eq. (2.1) to explain the effect of host death at the population level.

Condition for deterministic increase of lineage taxa

We start from the definition of $\mathbf{A}[1]$, Eq. (S2.1). This equation indicates the deterministic change of frequency of a lineage taxon (x_i) as a function of x_i, other microbes frequency (o_i), and parameters of migration (m), frequency in the pool of colonizers (p_i), and how rapidly available space is occupied (α_0). Asking under which condition $\mathbf{A}[1] > 0$, leads to

$$T^{o_i-}_{x_i+} + T^{o_i}_{x_i+} > T^{o_i+}_{x_i-} + T^{o_i}_{x_i-}$$

Using the definition of the transition probabilities in Eq. (2.1) and simplifying, we find

$$(1-x_i)\left((1-m)\frac{x_i}{\alpha_0 x_0 + x_i + o_i}\right) > x_i\left(m + (1-m)\left(1 - \frac{x_i}{\alpha_0 x_0 + x_i + o_i}\right)\right)$$

where we used the fact that lineage taxa are absent from the pool of colonizers, $p_i = 0$. Simplifying and solving for $x_i + o_i = 1 - x_0$, we find

$$x_i + o_i < 1 - \frac{m}{1-\alpha_0} \tag{S2.4}$$

Thus, the growth of lineage taxa stops before the microbial load, $x_i + o_i$, reaches frequency 1, as this is constrained by migration, m, and how rapidly available space is occupied, α_0.

C.2 Supplementary figures

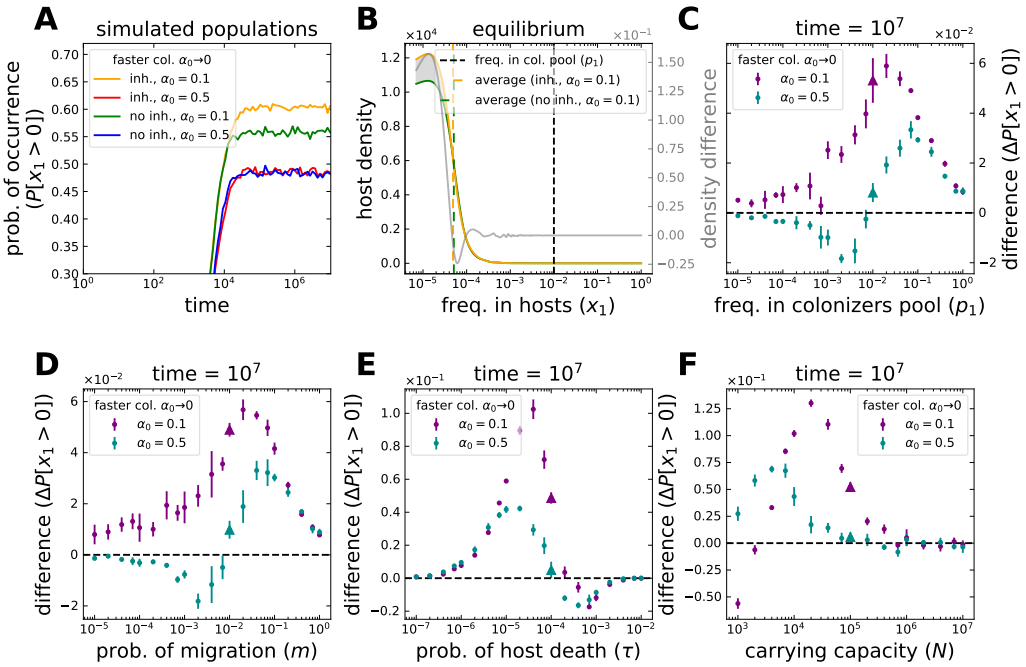

Figure Supplementary C.1: **Occurrence of a microbial taxon in hosts under microbial inheritance.** We repeat the analysis from Fig. 2.2, but instead of load, $x_i + o_i$, we look into a single microbial taxon, x_i. (**A**) Starting from a condition where all hosts are initially empty, the microbial occurrence increases through time. In this particular case, inheritance increases the occurrence if hosts are colonized rapidly, $\alpha_0 \to 0$. (**B**) The hosts now contain the taxon in small frequencies. The cases shown in (A-B), with parameters $p_1 = 10^{-2}$, $m = 10^{-2}$, $\tau = 10^{-4}$, and $N = 10^5$, are indicated by the triangles in (C-F). (**C**) Changes are small for other frequencies in the pool of colonizers, p_1, but those at intermediate values benefit the most from inheritance. (**D**) The maximum change occurs for intermediate migration from the pool of colonizers, m. For $m \to 1$ the taxon colonizes hosts even without inheritance. Instead for $m \to 0$ the taxon does not colonize the hosts. (**E**) Larger changes occur for intermediate host death probabilities, τ, and fast colonization. Long living hosts, $\tau \to 0$, contain the taxon even without inheritance. Short living hosts, $\tau \to 1$, are less likely to be colonized by the taxon within their lifetime. (**F**) In contrast to the microbial load (Fig. 2.2E), for a single taxon the maximum change occurs at intermediate capacities for microbes, N. The change can be negative once inheritance favours more abundant taxa competing for limited space (see C-F). Points and bars in (C-F) indicate the average and standard deviation of 6 simulation pairs, with vs. without inheritance, with 10^4 hosts each. Offspring receive 9% of their parent's microbiome on average, $a_i = 0$ and $b_i = 9$ in Eq. (2.4). The whole distributions are shown in Fig. Sup. C.3.

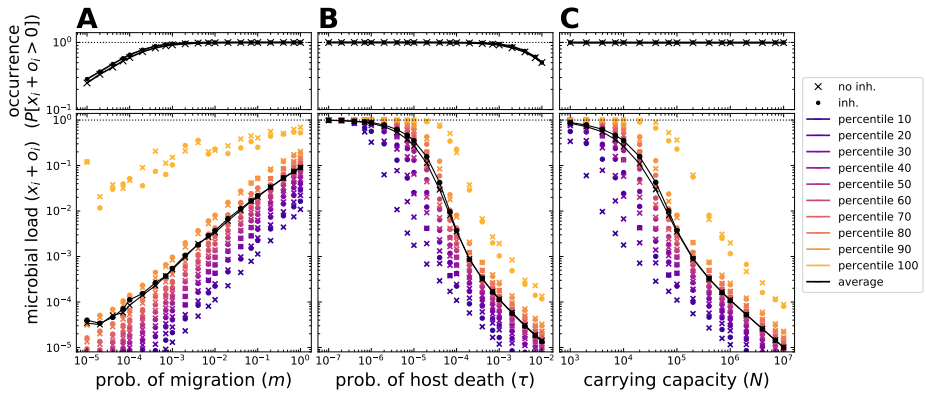

Figure Supplementary C.2: **Microbial load distribution across a host population, with or without microbial inheritance.** The microbial load is the set of all microbes. In contrast to the difference between distributions, Figs. 2.2 and 2.3, here the distributions are shown. The cases without and with inheritance are indicated by $\times$ and $\bullet$, respectively. Single parameters are modified from the condition $m = 10^{-2}$, $\tau = 10^{-4}$, and $N = 10^5$. The probability of occurrence and frequencies within hosts increase for (**A**) larger migration from the pool of colonizers, $m \to 1$, and (**B**) longer host lifespan, $\tau \to 0$. (**C**) While occurrence is constant at 1, frequencies increase for smaller capacities for microbes, N. Inheritance might increase both observables for certain parameter combinations and percentiles of the distribution (compare $\bullet$ to $\times$). The increase is evident for small percentiles. Decrease might occur for large percentiles. Only for $\tau \lesssim 2 \cdot 10^{-7}$ all hosts reach carrying capacity within their lifetime. Each simulation included 10^4 hosts and parameters $a_i = 0$ and $b_i = 9$ for inheritance, Eq. (2.4) – offspring receive 9% of their parent's microbiome on average – and $\alpha_0 = 0.1$ for available space occupation.

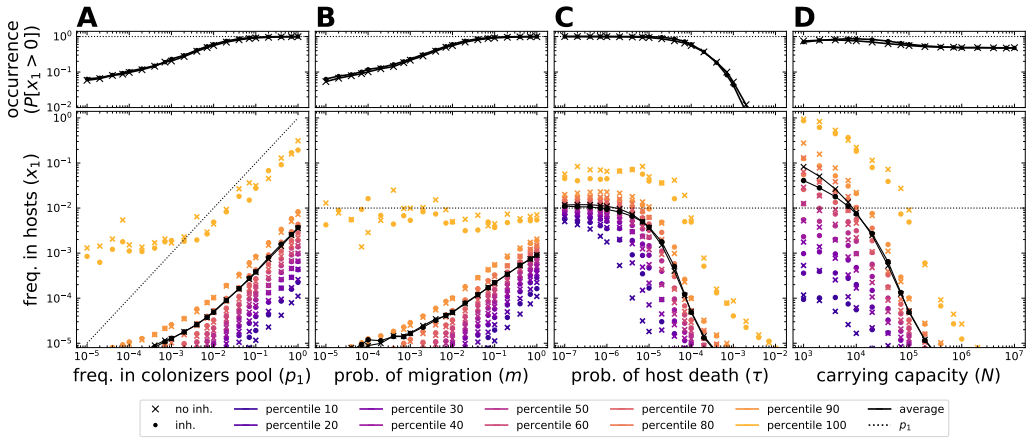

Figure Supplementary C.3: **Frequency of a microbial taxon distribution across the host population, with or without inheritance.** In contrast to the difference between distributions, Figs. Sup. C.1 and C.4, here the distributions are shown. The cases without and with inheritance are indicated by $\times$ and $\bullet$, respectively. Single parameters are modified from the condition $p_1 = 10^{-2}$, $m = 10^{-2}$, $\tau = 10^{-4}$, and $N = 10^5$. (**A**) The probability of occurrence and frequency within hosts increase for higher abundances in the pool of colonizers, $p_1 \to 1$, and (**B**) larger migration from the environment, $m \to 1$. For $p_1 \to 0$, hosts with larger frequencies than in the pool of colonizers ($x_1 > p_1$) might occur stochastically. In contrast to microbial load (Fig. Sup. C.2), inheritance might decrease the frequencies for (**C**) long host lifespans, $\tau \to 0$, and, (**D**) smaller capacities for microbes, N, where hosts are fully colonized. The reduced variability of the early microbiome, makes hosts with initially large frequencies of the microbial taxon less likely. Even if low frequencies increase, the average frequency decreases as a result. Inheritance increases the average frequency for intermediate values of τ and N, where hosts are partially colonized (Fig. Sup. C.2 B-C). Each simulation included 10^4 hosts and parameters $a_i = 0$ and $b_i = 9$ for inheritance, Eq. (2.4) – offspring receive 9% of their parent's microbiome on average – and $\alpha_0 = 0.1$ for the available space occupation.

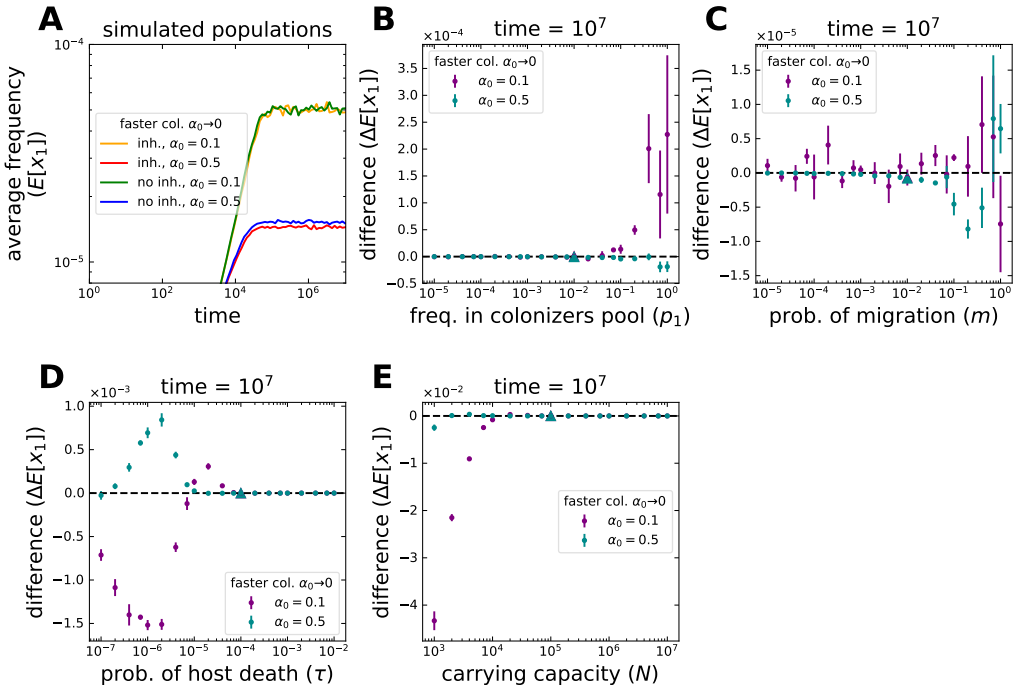

Figure Supplementary C.4: **Average frequency of a microbial taxon in hosts under microbial inheritance.** We repeat the analysis from Fig. 2.3, but instead of load, $x_i + o_i$, we look into a single microbial taxon, x_i. (**A**) Starting from a condition where all hosts are initially empty, the average frequency of microbes in hosts increases through time before reaching an equilibrium. In this particular case, inheritance makes the average slightly larger if hosts are occupied more slowly, $\alpha_0 = 0.5$. Although more hosts harbour the taxon, no change occurs for $\alpha_0 = 0.1$, as inheritance reduces the variability between individuals. The cases shown in (A), with parameters $p_1 = 10^{-2}$, $m = 10^{-2}$, $\tau = 10^{-4}$, and $N = 10^5$, are indicated by the triangles in (B-E). (**B**) No changes occur for multiple frequencies in the pool of colonizers, p_1, and (**C**) migrations from the pool of colonizers, m. (**D**) The largest changes occur for intermediate host death probabilities, τ. For long living hosts, $\tau \to 0$, the change produced by inheritance can be negative. (**E**) Similarly for small capacities for microbes, N, where inheritance causes abundant taxa to outcompete others. Points and bars in (B-E) indicate the average and standard deviation of 6 simulation pairs, with vs. without inheritance, with 10^4 hosts each. Offspring receive 9% of their parent's microbiome on average, $a_i = 0$ and $b_i = 9$ in Eq. (2.4). The whole distributions are shown in Fig. Sup. C.3.

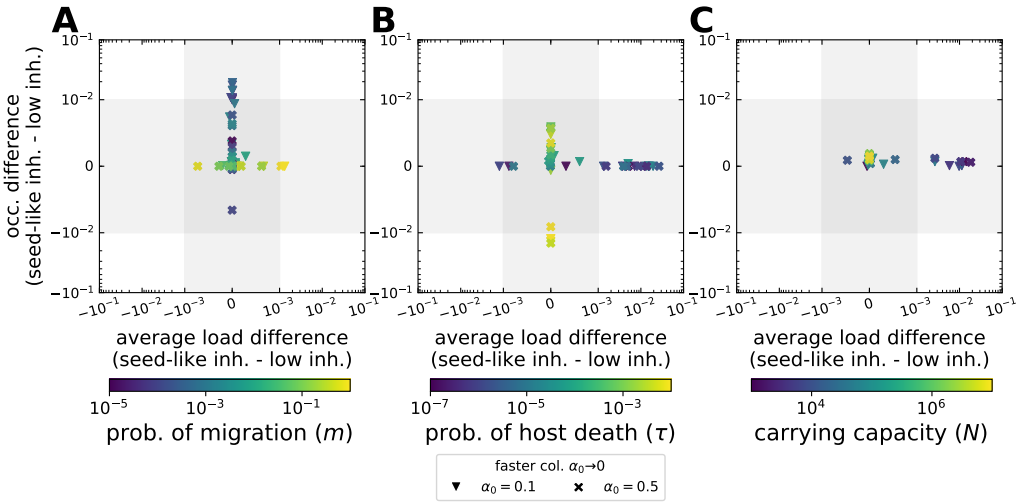

Figure Supplementary C.5: **Difference in microbial load between 'low' and 'seed-like' inheritance.** A positive difference indicates the observable is larger for seed-like inheritance (Fig. 2.1B). For both, low and seed-like inheritance, offspring receive 9% of their parent's microbiome on average ($a_i = 0$ and $b_i = 9$ for low inheritance, and $a_i = 9$ and $b_i = 99$ for seed-like inheritance in Eq. (2.4)). Low inheritance corresponds to data shown in Fig. 2.2 and Fig. 2.3. Single parameters are modified from the condition $m = 10^{-2}$, $\tau = 10^{-4}$, and $N = 10^5$. (**A**) For low migration from the pool colonizers, $m \to 0$, seed-like inheritance increases the microbial occurrence (a exception stems from a slower occupation of available space, $\alpha_0 = 0.5$). For $m \to 1$, it mildly increases the average microbial load. (**B**) For low host death, $\tau \to 0$, this inheritance mode increases the average load importantly. For $\tau \to 1$, it only affects the occurrence, even decreasing it. (**C**) For varying carrying capacity (N), larger average loads are obtained for small N. Each point corresponds to the difference of observables calculated from simulations with 10^4 hosts. The scale of axes is logarithmic, but linear within $\left[-10^{-3}, 10^{-3}\right]$ for the average load, and $\left[-10^{-2}, 10^{-2}\right]$ for the occurrence.

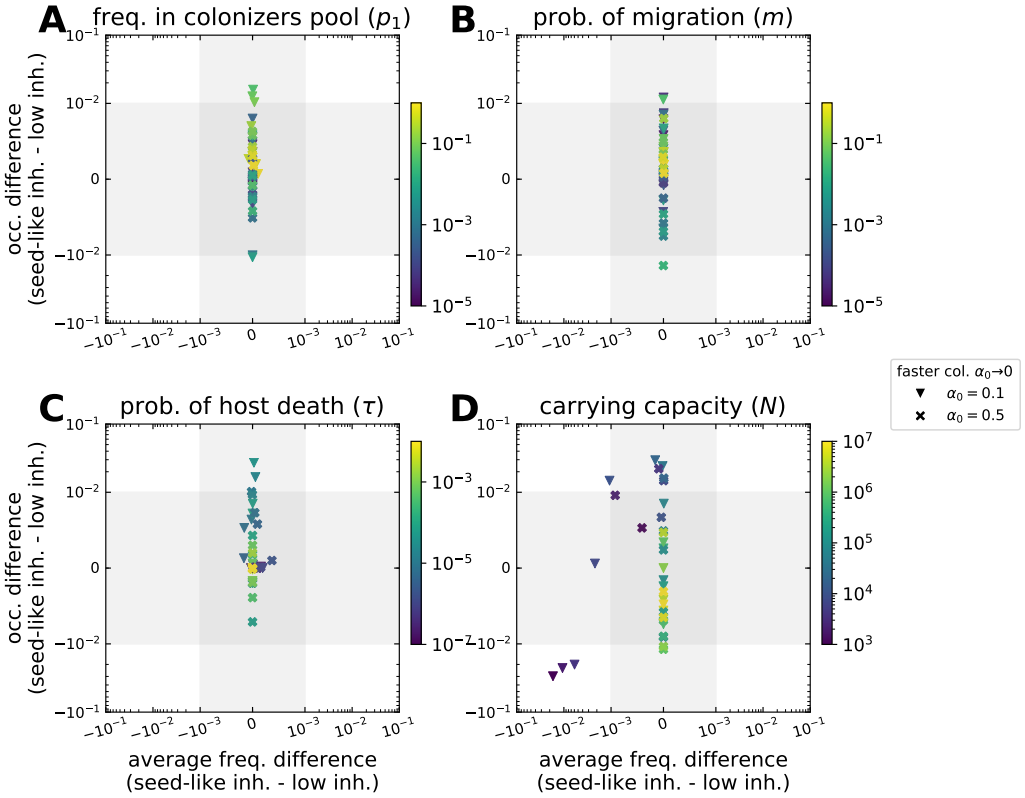

Figure Supplementary C.6: **Difference in the frequency of a microbial taxon between 'low' and 'seed-like' inheritance.** A positive difference indicates the observable is larger for seed-like inheritance (Fig. 2.1B). For both, low and seed-like inheritance, offspring receive 9% of their parent's microbiome on average ($a_i = 0$ and $b_i = 9$ for low inheritance, and $a_i = 9$ and $b_i = 99$ for seed-like inheritance in Eq. (2.4)). Low inheritance corresponds to data shown in Fig. Sup. C.1 and Fig. Sup. C.4. Single parameters are modified from the condition $p_1 = 10^{-2}$, $m = 10^{-2}$, $\tau = 10^{-4}$, and $N = 10^5$. (**A**-**C**) A seed-like inheritance primarily modifies the occurrence for various values of frequency in the pool of colonizers (p_i), migration (m), and host death (τ). (**D**) For varying values of the carrying capacity for microbes (N), the main change is on the occurrence, however, for small N a decrease of average frequency is observed. A decrease or increase of occurrence is not clearly attributable to the rate of host colonization (α_0). Each point corresponds to the difference of simulations with 10^4 hosts. The scale of axes is logarithmic, but linear within $\left[-10^{-3}, 10^{-3}\right]$ for the average frequency, and $\left[-10^{-2}, 10^{-2}\right]$ for the occurrence.

Bibliography

K. L. Adair and A. E. Douglas. Making a microbiome: the many determinants of host-associated microbial community composition. *Current Opinion in Microbiology*, 35:23–29, 2017. (Cited on page 1.)

K. L. Adair, M. Wilson, A. Bost, and A. E. Douglas. Microbial community assembly in wild populations of the fruit fly *Drosophila melanogaster*. *The ISME Journal*, 2018. (Cited on pages 12, 29, 54, 67, 69, and 70.)

E. Allen. *Modeling with Itô Stochastic Differential Equations*. Springer, 2007. (Cited on pages 8 and 91.)

O. Allouche and R. Kadmon. A general framework for neutral models of community dynamics. *Ecology letters*, 12(12):1287–1297, 2009. (Cited on pages 55, 56, 58, 69, 70, and 77.)

A. Almeida, A. L. Mitchell, M. Boland, S. C. Forster, G. B. Gloor, A. Tarkowska, T. D. Lawley, and R. D. Finn. A new genomic blueprint of the human gut microbiota. *Nature*, 568(7753):499–504, 2019. (Cited on pages 34 and 46.)

M. Arumugam, J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, J. Tap, T. Bruls, J.-M. Batto, et al. Enterotypes of the human gut microbiome. *Nature*, 473(7346):174, 2011. (Cited on page 29.)

F. Bansept, N. Obeng, H. Schulenburg, and A. Traulsen. Modeling host-associating microbes under selection. *The ISME Journal*, pages 1–9, 2021. (Cited on page 4.)

Y. M. Bar-On, R. Phillips, and R. Milo. The biomass distribution on earth. *Proceedings of the National Academy of Sciences*, page 201711842, 2018. (Cited on page 28.)

F. U. Battistuzzi, A. Feijao, and S. B. Hedges. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and

the colonization of land. *BMC evolutionary biology*, 4(1):1–14, 2004. (Cited on page 1.)

M. Begon and C. R. Townsend. *Ecology: from individuals to ecosystems*. John Wiley & Sons, 2020. (Cited on pages 2 and 3.)

G. Berg, D. Rybakova, D. Fischer, T. Cernava, M.-C. C. Vergès, T. Charles, X. Chen, L. Cocolin, K. Eversole, G. H. Corral, et al. Microbiome definition re-visited: old concepts and new challenges. *Microbiome*, 8(1):1–22, 2020. (Cited on page 3.)

J. R. Björk, R. B. O'Hara, M. Ribes, R. Coma, and J. M. Montoya. The dynamic core microbiome: Structure, dynamics and stability. *bioRxiv*, 2018. (Cited on page 30.)

J. R. Björk, C. Díez-Vives, C. Astudillo-García, E. A. Archie, and J. M. Montoya. Vertical transmission of sponge microbiota is inconsistent and unfaithful. *Nature Ecology & Evolution*, page 1, 2019. (Cited on pages 28, 34, and 51.)

J. E. Blum, C. N. Fischer, J. Miles, and J. Handelsman. Frequent replenishment sustains the beneficial microbiome of drosophila melanogaster. *MBio*, 4(6), 2013. (Cited on page 36.)

N. A. Bokulich, J. Chung, T. Battaglia, N. Henderson, M. Jay, H. Li, A. D. Lieber, F. Wu, G. I. Perez-Perez, Y. Chen, et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Science translational medicine*, 8(343):343ra82–343ra82, 2016. (Cited on page 31.)

S. R. Bordenstein and K. R. Theis. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biology*, 13(8):e1002226, 2015. (Cited on pages 1 and 30.)

T. C. Bosch, K. Guillemin, and M. McFall-Ngai. Evolutionary "experiments" in symbiosis: The study of model animals provides insights into the mechanisms underlying the diversity of host–microbe interactions. *BioEssays*, 41 (10):1800256, 2019. (Cited on page 51.)

M. Bright and S. Bulgheresi. A complex journey: transmission of microbial symbionts. *Nature Reviews Microbiology*, 8(3):218–230, 2010. (Cited on pages 34 and 50.)

H. P. Browne, S. C. Forster, B. O. Anonye, N. Kumar, B. A. Neville, M. D. Stares, D. Goulding, and T. D. Lawley. Culturing of 'unculturable' human

microbiota reveals novel taxa and extensive sporulation. *Nature*, 533(7604): 543–546, May 2016. (Cited on page 50.)

M. Bruijning, L. P. Henry, S. K. Forsberg, C. J. E. Metcalf, and J. F. Ayroles. When the microbiome defines the host phenotype: selection on vertical transmission in varying environments. *bioRxiv*, 2020. (Cited on pages 34, 35, 50, 51, 74, and 78.)

A. R. Burns, W. Z. Stephens, K. Stagaman, S. Wong, J. F. Rawls, K. Guillemin, and B. J. Bohannan. Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *The ISME Journal*, 10(3):655–664, 2016. (Cited on pages 12, 29, 54, 67, 69, 70, and 77.)

A. R. Burns, E. Miller, M. Agarwal, A. S. Rolig, K. Milligan-Myhre, S. Seredick, K. Guillemin, and B. J. Bohannan. Interhost dispersal alters microbiome assembly and can overwhelm host innate immunity in an experimental zebrafish model. *Proceedings of the National Academy of Sciences*, 114(42):11181–11186, 2017. (Cited on page 28.)

J. M. Cancino and N. Rodger. An ecological overview of cloning in metazoa. *Population biology and evolution of clonal organisms*, pages 153–186, 1985. (Cited on pages 36 and 38.)

H. Caswell. *Matrix population models*. Sinauer Associates, Sunderland MA, 2nd edition, 2001. (Cited on page 8.)

J. A. Chandler, J. M. Lang, S. Bhatnagar, J. A. Eisen, and A. Kopp. Bacterial communities of diverse drosophila species: ecological context of a host–microbe model system. *PLoS Genetics*, 7(9):e1002272, 2011. (Cited on page 13.)

X. Chao and Y. Zheng. Transient analysis of immigration birth–death processes with total catastrophes. *Probability in the Engineering and Informational Sciences*, 17(1):83–106, 2003. (Cited on page 17.)

J. Chase, M. Leibold, and U. of Chicago Press. *Ecological Niches: Linking Classical and Contemporary Approaches*. Interspecific Interactions. University of Chicago Press, 2003. (Cited on pages 4 and 54.)

R. A. Chisholm and S. W. Pacala. Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *Proceedings of the National Academy of Sciences*, 107(36): 15821–15825, 2010. (Cited on pages 55, 69, and 70.)

T. J. Colston and C. R. Jackson. Microbiome evolution along divergent branches of the vertebrate tree of life: what is known and unknown. *Molecular Ecology*, 25(16):3776–3800, 2016. (Cited on page 34.)

P. I. Costea, F. Hildebrand, A. Manimozhiyan, F. Bäckhed, M. J. Blaser, F. D. Bushman, W. M. De Vos, S. D. Ehrlich, C. M. Fraser, M. Hattori, et al. Enterotypes in the landscape of gut microbial community composition. *Nature Microbiology*, 3(1):8, 2018. (Cited on pages 29 and 30.)

E. K. Costello, K. Stagaman, L. Dethlefsen, B. J. Bohannan, and D. A. Relman. The application of ecological theory toward an understanding of the human microbiome. *Science*, 336(6086):1255–1262, 2012. (Cited on pages 56 and 78.)

T. J. Davies, A. P. Allen, L. Borda-de Água, J. Regetz, and C. J. Melián. Neutral biodiversity theory can explain the imbalance of phylogenetic trees but not the tempo of their diversification. *Evolution*, 65(7):1841–1850, 2011. (Cited on page 69.)

P. De Kruif. *Microbe hunters*. Houghton Mifflin Harcourt, 1926. (Cited on page 1.)

M. Dworkin and D. Gutnick. Sergei winogradsky: a founder of modern microbiology and the first microbial ecologist. *FEMS microbiology reviews*, 36(2):364–379, 2012. (Cited on page 1.)

R. Eisenhofer, J. J. Minich, C. Marotz, A. Cooper, R. Knight, and L. S. Weyrich. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends in microbiology*, 27(2):105–117, 2019. (Cited on page 34.)

M. R. Evans and S. N. Majumdar. Diffusion with stochastic resetting. *Phys. Rev. Lett.*, 106(16):160601, 2011. (Cited on page 17.)

C. K. Fisher and P. Mehta. The transition between the niche and neutral regimes in ecology. *Proceedings of the National Academy of Sciences*, 111 (36):13111–13116, 2014. (Cited on pages 55, 69, and 70.)

W. Frey and H. Kürschner. Asexual reproduction, habitat colonization and habitat maintenance in bryophytes. *Flora*, 206(3):173–184, 2011. (Cited on pages 36 and 38.)

L. J. Funkhouser and S. R. Bordenstein. Mom knows best: the universality of maternal microbial transmission. *PLoS Biol*, 11(8):e1001631, 2013. (Cited on pages 2, 34, 50, and 51.)

C. W. Gardiner. *Handbook of Stochastic Methods*. Springer, NY, third edition, 2004. (Cited on pages 5, 6, 7, 8, 9, 16, 17, 31, 39, 57, 58, and 69.)

S. Gavrilets. Perspective: models of speciation: what have we learned in 40 years? *Evolution*, 57(10):2197–2215, 2003. (Cited on page 4.)

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. CRC Press, 2013. (Cited on pages 67 and 70.)

B. Gibson, D. J. Wilson, E. Feil, and A. Eyre-Walker. The distribution of bacterial doubling times in the wild. *Proceedings of the Royal Society B: Biological Sciences*, 285(1880):20180789, 2018. (Cited on pages 28 and 75.)

T. E. Gibson, A. Bashan, H.-T. Cao, S. T. Weiss, and Y.-Y. Liu. On the origins and control of community types in the human microbiome. *PLoS Computational Biology*, 12(2):e1004688, 2016. (Cited on page 29.)

D. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976. (Cited on pages 5 and 6.)

D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of chemical physics*, 115(4):1716–1733, 2001. (Cited on page 6.)

D. Gonze, L. Lahti, J. Raes, and K. Faust. Multi-stability and the origin of microbial community types. *The ISME Journal*, 11(10):2159, 2017. (Cited on page 29.)

I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series and Products*. Academic Press, London, 1994. (Cited on page 38.)

D. Gravel, C. D. Canham, M. Beaudet, and C. Messier. Reconciling niche and neutrality: the continuum hypothesis. *Ecology Letters*, 9(4):399–409, 2006. (Cited on pages 4, 55, 69, and 70.)

J. Grilli. Macroecological laws describe variation and diversity in microbial communities. *Nature Communications*, 11(1):1–11, 2020. (Cited on pages 67, 70, 71, and 74.)

B. Haegeman and M. Loreau. A mathematical synthesis of niche and neutral theories in community ecology. *Journal of theoretical biology*, 269(1):150–165, 2011. (Cited on pages 55, 69, and 70.)

T. J. Hammer, D. H. Janzen, W. Hallwachs, S. P. Jaffe, and N. Fierer. Caterpillars lack a resident gut microbiome. *Proceedings of the National Academy of Sciences*, 114(36):9641–9646, 2017. (Cited on pages 25, 29, and 30.)

T. J. Hammer, J. G. Sanders, and N. Fierer. Not all animals need a microbiome. *FEMS Microbiology Letters*, 366(10):fnz117, 2019. (Cited on pages 2, 29, 30, 34, and 50.)

S. P. Hubbell. *The unified neutral theory of biodiversity and biogeography*. Monographs in Population Biology. Princeton University Press, Princeton, NJ, USA, 2001. (Cited on pages 2, 4, 12, and 54.)

T. Jahnke. On reduced models for the chemical master equation. *Multiscale Modeling & Simulation*, 9(4):1646–1676, 2011. (Cited on pages 58, 69, and 71.)

J. Johnke, P. Dirksen, and H. Schulenburg. Community assembly of the native c. elegans microbiome is influenced by time, substrate and individual bacterial taxa. *Environmental Microbiology*, 22(4):1265–1279, 2020. (Cited on page 34.)

D. A. Kessler and N. M. Shnerb. Generalized model of island biodiversity. *Physical Review E*, 91(4):042705, 2015. (Cited on pages 54, 55, 69, 70, and 71.)

J. E. Koenig, A. Spor, N. Scalfone, A. D. Fricker, J. Stombaugh, R. Knight, L. T. Angenent, and R. E. Ley. Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences*, 108(Supplement 1):4578–4585, 2011. (Cited on pages 30 and 31.)

B. Koskella, L. J. Hall, and C. J. E. Metcalf. The microbiome beyond the horizon of ecological and evolutionary theory. *Nature Ecology and Evolution*, 1(11):1606, 2017. (Cited on pages 3 and 76.)

P. Kumar and S. Narayanan. Solution of fokker-planck equation by finite element and finite difference methods for nonlinear systems. *Sadhana*, 31 (4):445–461, 2006. (Cited on pages 18 and 84.)

E. Kyriakidis. Stationary probabilities for a simple immigration-birth-death process under the influence of total catastrophes. *Statistics & Probability Letters*, 20(3):239–240, 1994. (Cited on page 17.)

A. V. Leeuwenhoek. Observations, communicated to the publisher by mr. antony van leewenhoeck, in a dutch letter of the 9th octob. 1676. here english'd: concerning little animals by him observed in rain-well-sea-and snow water; as also in water wherein pepper had lain infused. *Philosophical Transactions of the Royal Society of London*, 12(133):821–831, 1677. (Cited on page 1.)

P. T. Leftwich, M. P. Edgington, and T. Chapman. Transmission efficiency drives host–microbe associations. *Proceedings of the Royal Society B*, 287 (1934):20200820, 2020. (Cited on pages 34, 35, 49, and 51.)

L. Li and Z. S. Ma. Testing the neutral theory of biodiversity with human microbiome datasets. *Scientific Reports*, 6:31448, 2016. (Cited on page 31.)

S. J. Lim and S. R. Bordenstein. An introduction to phylosymbiosis. *Proceedings of the Royal Society B*, 287(1922):20192900, 2020. (Cited on page 35.)

A. J. Lotka. The growth of mixed populations: two species competing for a common food supply. *Journal of the Washington Academy of Sciences*, 22: 461– 469, 1932. (Cited on page 54.)

S. Louca, S. M. Jacques, A. P. Pires, J. S. Leal, D. S. Srivastava, L. W. Parfrey, V. F. Farjalla, and M. Doebeli. High taxonomic variability despite stable functional structure across microbial communities. *Nature Ecology & Evolution*, 1(1):0015, 2016. (Cited on page 78.)

R. Macarthur and E. O. Wilson. *The theory of island biogeography*. Princeton University Press, Princeton, NJ, 1967. (Cited on pages 4 and 54.)

T. Malthus. An essay on the principle of population, as it affects the future improvement of society with remarks on the speculations of mr. godwin, m. condorcet, and other writers. *J. Johnson, London*, pages 1–126, 1798. (Cited on page 54.)

I. Martínez, M. X. Maldonado-Gomez, J. C. Gomes-Neto, H. Kittana, H. Ding, R. Schmaltz, P. Joglekar, R. J. Cardona, N. L. Marsteller, S. W. Kembel, et al. Experimental evaluation of the importance of colonization history in early-life gut microbiota assembly. *eLife*, 7:e36521, 2018. (Cited on page 29.)

B. McDonald and K. D. McCoy. Maternal microbiota in pregnancy and early life. *Science*, 365(6457):984–985, 2019. (Cited on pages 34 and 38.)

M. McFall-Ngai, M. G. Hadfield, T. C. G. Bosch, H. V. Carey, T. Domazet-Lošo, A. E. Douglas, N. Dubilier, G. Eberl, T. Fukami, S. F. Gilbert, U. Hentschel, N. King, S. Kjelleberg, A. H. Knoll, N. Kremer, S. K. Mazmanian, J. L. Metcalf, K. Nealson, N. E. Pierce, J. F. Rawls, A. Reid, E. G. Ruby, M. Rumpho, J. G. Sanders, D. Tautz, and J. J. Wernegreen. Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences*, 110(9):3229–3236, 2013. (Cited on page 2.)

B. J. McGill, R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, B. J. Enquist, J. L. Green, F. He, et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology letters*, 10(10):995–1015, 2007. (Cited on page 54.)

J. M. Meylahn, S. Sabhapandit, and H. Touchette. Large deviations for markov processes with resetting. *Physical Review E*, 92(6):062148, 2015. (Cited on page 17.)

E. T. Miller and B. J. M. Bohannan. Life Between Patches: Incorporating Microbiome Biology Alters the Predictions of Metacommunity Models. *Frontiers in Ecology and Evolution*, 7, 2019. (Cited on page 34.)

E. T. Miller, R. Svanbäck, and B. J. Bohannan. Microbiomes as metacommunities: Understanding host-associated microbes through metacommunity ecology. *Trends in Ecology and Evolution*, 2018. (Cited on pages 2 and 28.)

P. Mitteroecker and S. M. Huttegger. The concept of morphospaces in evolutionary and developmental biology: mathematics and metaphors. *Biological Theory*, 4(1):54–67, 2009. (Cited on page 75.)

A. H. Moeller, T. A. Suzuki, M. Phifer-Rixey, and M. W. Nachman. Transmission modes of the mammalian gut microbiota. *Science*, 362(6413):453–457, 2018. (Cited on pages 34, 50, and 51.)

N. A. Moran and D. B. Sloan. The hologenome concept: helpful or hollow? *PLoS Biology*, 13(12):e1002311, 2015. (Cited on page 30.)

P. A. P. Moran. *The Statistical Processes of Evolutionary Theory*. Clarendon Press, Oxford, 1962. (Cited on page 9.)

D. R. Nemergut, S. K. Schmidt, T. Fukami, S. P. O'Neill, T. M. Bilinski, L. F. Stanish, J. E. Knelman, J. L. Darcy, R. C. Lynch, P. Wickey, et al. Patterns and processes of microbial community assembly. *Microbiology*

and Molecular Biology Reviews, 77(3):342–356, 2013. (Cited on pages 55 and 56.)

S. V. Nyholm and M. McFall-Ngai. The winnowing: establishing the squid—[ndash]—vibrio symbiosis. *Nature Reviews Microbiology*, 2(8):632–642, 2004. (Cited on page 31.)

B. Obadia, Z. Güvener, V. Zhang, J. A. Ceja-Navarro, E. L. Brodie, W. J. William, and W. B. Ludington. Probabilistic invasion underlies natural gut microbiome stability. *Current Biology*, 27(13):1999–2006, 2017. (Cited on pages 25, 28, 29, and 30.)

J. P. O'Dwyer, S. W. Kembel, and T. J. Sharpton. Backbones of evolutionary history test biodiversity theory for microbes. *Proceedings of the National Academy of Sciences*, 112(27):8356–8361, 2015. (Cited on page 69.)

I. D. Ofiţeru, M. Lunn, T. P. Curtis, G. F. Wells, C. S. Criddle, C. A. Francis, and W. T. Sloan. Combined niche and neutral effects in a microbial wastewater treatment community. *Proceedings of the National Academy of Sciences*, 107(35):15345–15350, 2010. (Cited on page 77.)

S. P. Otto and T. Day. *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton Univ. Press, Princeton, NJ, 2007. (Cited on pages 5, 6, and 9.)

E. Özkurt, M. A. Hassani, U. Sesiz, S. Künzel, T. Dagan, H. Özkan, and E. H. Stukenbrock. Seed-derived microbial colonization of wild emmer and domesticated bread wheat (triticum dicoccoides and t. aestivum) seedlings shows pronounced differences in overall diversity and composition. *Mbio*, 11(6):e02637–20, 2020. (Cited on pages 34, 36, 38, 50, and 51.)

M. E. Perez-Muñoz, M.-C. Arrieta, A. E. Ramer-Tait, and J. Walter. A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. *Microbiome*, 5 (1):48, 2017. (Cited on pages 13, 28, 34, 36, 38, and 51.)

J. I. Prosser, B. J. M. Bohannan, T. P. Curtis, R. J. Ellis, M. K. Firestone, R. P. Freckleton, J. L. Green, L. E. Green, K. Killham, J. J. Lennon, A. M. Osborn, M. Solan, C. J. van der Gast, and J. P. W. Young. The role of ecological theory in microbial ecology. *Nature Reviews Microbiology*, 5(5): 384–392, 2007. (Cited on pages 2 and 71.)

J. Renelies-Hamilton, K. Germer, D. Sillam-Dussès, K. H. Bodawatta, and M. Poulsen. Disentangling the relative roles of vertical transmission, subsequent colonizations, and diet on cockroach microbiome assembly. *Msphere*, 6(1), 2021. (Cited on pages 34, 35, and 51.)

H. Risken. *The Fokker-Planck Equation*. Springer, 1996. (Cited on page 7.)

J. Rosindell, S. P. Hubbell, F. He, L. J. Harmon, and R. S. Etienne. The case for ecological neutral theory. *Trends in Ecology & Evolution*, 27(4): 203–208, 2012. (Cited on pages 54, 55, 69, and 70.)

J. Rosindell, L. J. Harmon, and R. S. Etienne. Unifying ecology and macroevolution with individual-based theory. *Ecology letters*, 18(5):472–482, 2015. (Cited on pages 69 and 70.)

J. Roughgarden. Holobiont evolution: Mathematical model with vertical vs. horizontal microbiome transmission. *Philosophy, Theory, and Practice in Biology*, 12(2), 2020. (Cited on pages 12, 30, and 34.)

J. Roughgarden, S. F. Gilbert, E. Rosenberg, I. Zilber-Rosenberg, and E. A. Lloyd. Holobionts as units of selection and a model of their population dynamics and evolution. *Biological Theory*, 13(1):44–65, 2018. (Cited on page 31.)

S. L. Russell. Transmission mode is associated with environment type and taxa across bacteria-eukaryote symbioses: a systematic review and meta-analysis. *FEMS microbiology letters*, 366(3):fnz013, 2019. (Cited on page 34.)

C. Sala, S. Vitali, E. Giampieri, Ì. F. do Valle, D. Remondini, P. Garagnani, M. Bersanelli, E. Mosca, L. Milanesi, and G. Castellani. Stochastic neutral modelling of the gut microbiota's relative species abundance from next generation sequencing data. *BMC Bioinformatics*, 17(Suppl 2):S16, 2016. (Cited on page 31.)

R. Sender, S. Fuchs, and R. Milo. Revised estimates for the number of human and bacteria cells in the body. *PLoS Biology*, 14(8):e1002533, 2016. (Cited on page 28.)

M. Sieber, L. Pita, N. Weiland-Brauer, P. Dirksen, J. Wang, B. Mortzfeld, S. Franzenburg, R. A. Schmitz, J. Baines, S. Fraune, U. Hentschel, H. Schulenburg, T. C. G. Bosch, and A. Traulsen. Neutrality in the metaorganism. *PLoS Biology*, 17(6):e3000298, 2019. (Cited on pages 2, 12, 13, 29, 54, 55, 61, 67, 69, 70, 71, and 75.)

M. Sieber, A. Traulsen, H. Schulenburg, and A. E. Douglas. On the evolutionary origins of host-microbe associations. *Proceeding of the National Academy of Sciences*, 118(9):e2016487118, 2021. (Cited on pages 4 and 34.)

W. T. Sloan, M. Lunn, S. Woodcock, I. M. Head, S. Nee, and T. P. Curtis. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environmental Microbiology*, 8(4):732–740, 2006. (Cited on pages 2, 7, 12, 13, 15, 16, 17, 29, 55, 56, 58, 67, 69, and 76.)

S. J. Song, J. G. Sanders, F. Delsuc, J. Metcalf, K. Amato, M. W. Taylor, F. Mazel, H. L. Lutz, K. Winker, G. R. Graves, et al. Comparative analyses of vertebrate gut microbiomes reveal convergence between birds and bats. *MBio*, 11(1), 2020. (Cited on page 51.)

D. Sprockett, T. Fukami, and D. A. Relman. Role of priority effects in the early-life assembly of the gut microbiota. *Nature Reviews Gastroenterology and Hepatology*, 2018. (Cited on page 29.)

W. Z. Stephens, T. J. Wiles, E. S. Martinez, M. Jemielita, A. R. Burns, R. Parthasarathy, B. J. Bohannan, and K. Guillemin. Identification of population bottlenecks and colonization factors during assembly of bacterial communities within the zebrafish intestine. *mBio*, 6(6):e01163–15, 2015. (Cited on page 28.)

W. Z. Stephens, A. R. Burns, K. Stagaman, S. Wong, J. F. Rawls, K. Guillemin, and B. J. Bohannan. The composition of the zebrafish intestinal microbial community varies across development. *The ISME Journal*, 10(3):644, 2016. (Cited on pages 13 and 30.)

R. J. Swift. Transient probabilities for a simple birth-death-immigration process under the influence of total catastrophes. *International Journal of Mathematics and Mathematical Sciences*, 25(10):689–692, 2001. (Cited on page 17.)

C. A. Thaiss, D. Zeevi, M. Levy, G. Zilberman-Schapira, J. Suez, A. C. Tengeler, L. Abramson, M. N. Katz, T. Korem, N. Zmora, et al. Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. *Cell*, 159(3):514–529, 2014. (Cited on pages 4 and 31.)

C. A. Thaiss, M. Levy, T. Korem, L. Dohnalová, H. Shapiro, D. A. Jaitin, E. David, D. R. Winter, M. Gury-BenAri, E. Tatirovsky, et al. Microbiota diurnal rhythmicity programs host transcriptome oscillations. *Cell*, 167(6): 1495–1510, 2016a. (Cited on page 4.)

C. A. Thaiss, N. Zmora, M. Levy, and E. Elinav. The microbiome and innate immunity. *Nature*, 535(7610):65–74, 2016b. (Cited on page 4.)

L. R. Thompson, J. G. Sanders, D. McDonald, A. Amir, J. Ladau, K. J. Locey, R. J. Prill, A. Tripathi, S. M. Gibbons, G. Ackermann, et al. A communal catalogue reveals earth's multiscale microbial diversity. *Nature*, 551(7681):457–463, 2017. (Cited on pages 1, 34, and 46.)

E. J. van Opstal and S. R. Bordenstein. Rethinking heritability of the microbiome. *Science*, 349(6253):1172–1173, 2015. (Cited on page 28.)

S. Van Vliet and M. Doebeli. The role of multilevel selection in host microbiome evolution. *PNAS*, 116(41):20591–20597, 2019. (Cited on pages 12, 30, 34, 49, 51, 74, and 78.)

N. M. Vega and J. Gore. Stochastic assembly produces heterogeneous communities in the *Caenorhabditis elegans* intestine. *PLoS Biology*, 15(3): e2000633, 2017. (Cited on pages 12, 26, 28, 29, 30, and 69.)

M. Vellend. Conceptual synthesis in community ecology. *The Quarterly review of biology*, 85(2):183–206, 2010. (Cited on pages 3, 4, 5, 54, and 55.)

I. Volkov, J. R. Banavar, F. He, S. P. Hubbell, and A. Maritan. Density dependence explains tree species abundance and diversity in tropical forests. *Nature*, 438(7068):658–661, 2005. (Cited on pages 55, 69, and 70.)

V. Volterra. Variations and fluctuations of the number of individuals in animal species living together. *Journal du conseil international pour l'exploration de la mer*, 3(1):3–51, 1928. (Cited on page 54.)

E. Watts, U. Hoßfeld, and G. S. Levit. Ecology and evolution: Haeckel's darwinian paradigm. *Trends in ecology & evolution*, 34(8):681–683, 2019. (Cited on page 3.)

S. Winogradsky et al. *Soil microbiology: problems and methods: 50 years of research.* Masson & Cie., 1949. (Cited on page 1.)

S. Woodcock, C. J. Van Der Gast, T. Bell, M. Lunn, T. P. Curtis, I. M. Head, and W. T. Sloan. Neutral assembly of bacterial communities. *FEMS Microbiology Ecology*, 62(2):171–180, 2007. (Cited on page 12.)

S. Xu and T. Chou. Immigration-induced phase transition in a regulated multispecies birth-death process. *Journal of Physics A: Mathematical and Theoretical*, 51(42):425602, 2018. (Cited on page 58.)

R. Zapien-Campos, M. Sieber, and A. Traulsen. Stochastic colonization of hosts with a finite lifespan can drive individual host microbes out of equilibrium. *PLoS Computational Biology*, 16(11):e1008392, 2020. (Cited on pages 11, 38, 40, 50, 51, and 71.)

R. Zapién-Campos, F. Bansept, M. Sieber, and A. Traulsen. On the effect of inheritance of microbes in commensal microbiomes. *bioRxiv*, 2021a. (Cited on page 33.)

R. Zapién-Campos, M. Sieber, and A. Traulsen. The effect of microbial selection on the occurrence-abundance patterns of microbiomes. *bioRxiv*, 2021b. (Cited on page 53.)

Q. Zeng, J. Sukumaran, S. Wu, and A. Rodrigo. Neutral models of microbiome evolution. *PLoS Computational Biology*, 11(7):e1004365, 2015. (Cited on pages 12, 30, 31, 35, and 49.)

Q. Zeng, S. Wu, J. Sukumaran, and A. Rodrigo. Models of microbiome evolution incorporating host and microbial selection. *Microbiome*, 5(1):127, 2017. (Cited on page 31.)

F. Zhang, M. Berg, K. Dierking, M.-A. Félix, M. Shapira, B. S. Samuel, and H. Schulenburg. *Caenorhabditis elegans* as a model for microbiome research. *Frontiers in Microbiology*, 8:485, 2017. (Cited on pages 13 and 36.)

J. Zhou and D. Ning. Stochastic community assembly: does it matter in microbial ecology? *Microbiology and Molecular Biology Reviews*, 81(4), 2017. (Cited on pages 3, 4, 12, and 55.)

Author contributions

Author initials

AT	Arne Traulsen	FB	Florence Bansept
MS	Michael Sieber	RZC	Román Ulises Zapién Campos

Contributor Roles Taxonomy (**CRediT**, http://credit.niso.org)

[**C**] Conceptualization, [**D**] Data curation, [**A**] Formal Analysis, [**I**] Investigation, [**M**] Methodology, [**S**] Software, [**U**] Supervision, [**V**] Validation, [**N**] Visualization, [**O**] Writing - original draft, and [**E**] Writing - review & editing.

Contributions per author

	Author	C	D	A	I	M	S	U	V	N	O	E
Chapter 1	AT	x		x		x		x		x		x
	MS	x		x		x		x		x		x
	RZC	x	x	x	x	x	x		x	x	x	x
Chapter 2	AT	x		x	x	x		x		x		x
	MS	x				x		x		x		x
	FB					x			x	x		x
	RZC	x	x	x	x	x	x		x	x	x	x
Chapter 3	AT	x		x		x		x		x		x
	MS	x						x		x		x
	RZC		x	x	x	x	x		x	x	x	x

Román Ulises Zapién Campos

Acknowledgements

Thank you for reading this thesis. Many people have contributed to making it possible. Words are not enough to express my gratitude to all of them.

Thank you Florence, Michael, and Arne for your feedback to fix earlier versions of this thesis. Thank you to my thesis committee and TAC members, Prof. Dr. Hinrich Schulenburg and Prof. Dr. Ute Hentschel Humeida. Also, thank you to my colleagues in the Max Planck Institute for Evolutionary Biology, the CAU, and the CRC 1182: Origin and Function of Metaorganisms – a special shoutout to the IMPRS program and the Max Planck Society for the funding and a nurturing professional experience.

On a personal note, first, I want to thank my supervisor Arne Traulsen. Thank you, Arne, for this opportunity and for being so caring scientifically and personally. I learned so much from you during this time. Thank you for actively ensuring the well-being of the community before anything else.

Thank you to all the other lovely people I met during these years. Your energy and wisdom have always been stimulating. I am especially grateful to all members of the Theory Department and the Group of Dynamics of Social Behaviour in the MPI. An emphatic thank you to the administration, IT, and technical teams that make our MPI function as a unity.

I had the privilege of collaborating with incredible people. Thank you, Arne and Michael, for your many teachings. Thank you, Florence, for the many occasions I have relied on you as a collaborator, mentor, and friend.

Thank you to Yuanxiao, María, Vandana, Mario, and Prateek, friends with whom I enjoyed sharing an office. Thank you to my PhD partners in crime – Bilal, Elena, Khawla, Eric, Wiola, Onur, Aditi, Lara, and Lizel. Thank you Devika, Anu, Yuriy, Hye Jin, Sedigheh, Christin, Saumil, and Christian – with whom I discovered the world through books. Thank you, Jatin, Zahra, and Nikhil, for those conversions over a cup of tea or snacks – your care always warmed my heart. Thank you to my friends in Latin spirit, Karen, Tania, Ana, Vanesa, Juan, and Loukas. Finally, thank you to the many sports teams (Joanna, Nico, Charlotte, Alejandra, Nikoleta, Sap, and Marta) and the one and only BBQ crew (Chaitanya, Hildegard, and Da).

The greatest thank you is for my family. *Gracias mamá, gracias papá,*

gracias Christian, y gracias Estela: por ser el mejor equipo y darme la posibilidad de cumplir este sueño. Los amo y todo lo hago por ustedes.

Declaration

I hereby declare,

i that apart from my supervisor's guidance, the content and design of the thesis is my own work;

ii that the thesis has not been submitted partly or wholly as part of a doctoral degree to any other examining body, and no other materials have been published or submitted for publication than indicated in the thesis;

iii that the preparation of the thesis has been subjected to the Rules of Good Scientific Practice of the German Research Foundation;

iv that prior to this thesis, I have not attempted and failed to obtain a doctoral degree.

Plön, 2021

Román Ulises Zapién Campos