Department of Mathematical Modelling, Statistics and Bioinformatics

Putting ecological theories to the test: individual-based simulations of synthetic microbial community dynamics

Aisling J. Daly

Thesis submitted in fulfillment of the requirements for the degree of
Doctor (Ph.D.) of Applied Biological Sciences

Academic year 2017-2018

Supervisors:	Prof. dr. Bernard De Baets Department of Mathematical Modelling, Statistics and Bioinformatics Ghent University, Belgium
	Prof. dr. ir. Jan Baetens Department of Mathematical Modelling, Statistics and Bioinformatics Ghent University, Belgium
Examination committee:	Prof. dr. ir. Koen Dewettinck (Chairman)
	Prof. dr. ir. Tom Van de Wiele Prof. dr. ir. Dirk Springael Prof. dr. Dries Bonte Prof. dr. Volker Grimm dr. ir. Frederiek-Maarten Kerckhof
Dean:	Prof. dr. ir. Marc Van Meirvenne
Rector:	Prof. dr. ir. Rik Van de Walle

Aisling Jane Daly

Putting ecological theories to the test: individual-based simulations of synthetic microbial community dynamics

Thesis submitted in fulfilment of the requirements for the degree of

Doctor (Ph.D) of Applied Biological Sciences

Academic year 2017-2018

Dutch translation of the title:
Ecologische principes op de proef gesteld: individu-gebaseerde simulaties van synthetische microbiële gemeenschappen

Please refer to this work as follows:
A.J. Daly (**2017**). Putting ecological theories to the test: individual-based simulations of synthetic microbial community dynamics, PhD Thesis, Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Ghent, Belgium.

Acknowledgements

These acknowledgements should begin with the same people as did my PhD - my two promotors Bernard and Jan. Thank you both for your guidance, support and belief in me through the ups, downs and sideways of the last four years. You were always ready to offer your advice, knowledge and encouragement when I needed it, and I look forward to our future work together. I would also like to thank my colleagues in the MRM project, with particular and grateful mention to Joke for our successful collaboration. My gratitude also goes to the members of my doctoral jury — I'm very appreciative of the time and effort they dedicated to reading and responding to this dissertation, which has certainly helped to improve this work.

I've been very fortunate during these last four years to find myself in a department with such a welcoming and good-natured atmosphere. There are far too many appreciated colleagues to thank all by name in these short paragraphs, but if I should leave you out then feel free to ask me for an impromptu dedication the next time we meet in the coffee room.

My thanks go to my office mates over the years. Guillermo, my office buddy since my very first day, how deeply your bright personality and even your office singing are missed. Jan R, thanks for the football and movie discussions (and I suppose the occasional IT help), overlooking your mocking emails about my beloved Arsenal and their inevitable crushing loss to Barcelona (they will only make our eventual triumph that much sweeter). Jan B, I still dislike your plants but deeply appreciate all the advice, support and strange vegetable-based muffins you've given me over the years. Raúl, the latest addition, who was already much appreciated for his friendship, attempts to improve the Kermix playlist, and his pioneering role in the Beer Club. And to those others who shared an office and brightened my day over the years: Federico, Demir, Gisele, and Dailé.

Within the wider BW10 department, thanks to my colleagues and friends for sharing in the good moments, for empathizing during the difficult ones, and for sharing your friendship across different languages, cultures, and tastes in beer. Michael, the most generous sharer of Roomer. Liselotte, whose sunny disposition is always appreciated. Michiel, for the excellent Netflix recommendations, board game evenings and 3D printed Pokémon. Jim, another excellent board game host. Gang, cheerful and full of puzzling poetic stories. The Wouters, both Good and Bad, brewers of strange beers and magnificent limoncello. Tinne, who along with her baking will forever hold a special place in my stomach and my sweet tooth — next year's Flemish Bake-off is yours for sure. Christina, known for her gentle disposition and iron liver (may I never see tsipouro again). José, the chili master. Peter, mysterious judoka and recommender of great bars. Andreia and Niels, full of good plans even when vegetarian food is involved. Youri, steadfast though the beard comes and goes. Giacomo, pizza chef extraordinaire. Thanks also to my masters students Ward, Zeno and Tim, the last of whom I'm glad decided to stick around for 6 more years of good work. And to the many visitors who have passed through and left good memories: Nubia, Dailé, Bac, Gisele, Shuyun, Wenwen, Yuanyuan, Laura, Hanzel, Giulio, and others I'm sorry to have left out.

Finally my thanks go to those nearest and dearest to me. First, to the friends who became my family in Ghent. Yingjie, Guangchao, Liv and soon-to-arrive Lucas, it's meant a lot to me to share in the happy times as your family grows — not least the delicious Chinese barbecues. Marc and Carlotta, I'm grateful to have shared so many great times with you, and to continue to do so even as you move a little farther away. Thanks for all the pizza, political humour, and happy memories. Guillermo and Max, your presence has been missed so much during the last few months as you left for new adventures — Ghent is a less colourful place without you two. Special thanks must also go to Guillermo: for the important place you've had, both professionally and personally, in my life in Ghent during these last years, for your help and support with my work and for your friendship outside it. Both are sincerely missed, although a small consolation is my relief to see the end of the pepper jenevers.

And then my love and gratitude to my family, always there for me despite distance and time zones. To my parents, for their unwavering love, support, encouragement, and inter-continental deliveries of tea. To my brother Ciarán, always taking interest in my work, now venturing off to embark on his own PhD and to make us all proud. To my sister Aoife for constant support and encouragement, so many TV recommendations, and being an unexpected but welcome roommate during the last months of my PhD, mastering the art of spaghetti carbonara and keeping my spirits up. Finally, a broad and grateful thanks to family and friends, old and new, near and far, too many to name, who've helped me along this road.

Aisling Jane Daly
Ghent, October 2017

Contents

Summary

Microbial communities are critical for the proper functioning of each and every ecosystem on Earth. They play central roles in key ecological, geochemical, medical, manufacturing and industrial processes, as well as being vitally important to human and animal health.

These critical functions are not performed by microbial species in isolation, but rather by communities containing numerous and diverse species. The ability to understand the structure and functioning of these complex communities is crucial to manage natural communities, to protect them from ongoing and significant anthropogenic environmental changes and thus preserve their vital processes, as well as to rationally design engineered microbial communities for important applications ranging from food technology, to medical and pharmaceutical uses, to various industrial and bioindustrial processes.

For these purposes, increasing effort is being dedicated to the development and deployment of tools, techniques and models that allow the prediction of behaviour and functionality in microbial communities, known as *Microbial Resource Management* (MRM). To develop these MRM tools, scientists first require a deeper and more fundamental understanding of the interactions taking place within these microbial communities, since these are the underlying mechanisms driving community behaviour.

An important tool in this respect is the use of mathematical models. Models are abstractions of reality which allow for the testing of hypotheses in a controlled way. Their reduced complexity and inherent manipulability make them the *in silico* counterparts of *in vitro* model (or synthetic) ecosystems constructed in the laboratory. Developing mathematical models related to MRM theories will not only help

to understand the fundamental mechanisms underpinning these theories, but can also help to develop new hypotheses by highlighting interesting or unexpected behaviour, which can then be explored and tested in more focused modelling and experimental studies.

In this thesis, we focus on the use of *in silico* synthetic microbial communities to test important microbial ecological theories relating to community stability, diversity, functionality and productivity. For this purpose, we use an individual-based approach to develop mathematical models representing *in silico* synthetic communities with different levels of microbial interactions, diversity and complexity.

In **Part I**, we gather and synthesize the existing *in vitro* and *in silico* techniques and knowledge relevant to our study of microbial ecological theories.

In Chapter 2, we discuss the important roles played by microbial communities in numerous vital domains of life, then survey the experimental techniques used to study these communities and to develop ecological theories regarding their composition and functionality. We outline why synthetic microbial communities are particularly suited for this type of theory development.

In Chapter 3, we survey the mathematical modelling approaches suitable for microbial communities. We focus on the modelling approaches which have been developed to study key aspects of microbial communities in terms of their functionality: their spatial structures and dynamics, the interactions taking place within the community, and the interactions occurring with the environment.

We discuss in Chapter 4 another key aspect of microbial communities in terms of their structure and behaviour, namely their diversity. We define precisely what is meant by diversity and how it can be quantified through the use of dedicated indices, before providing a comparative survey of selected key indices. We then provide an analogous discussion of evenness, the more complex of the two components of biodiversity.

In **Part II**, we begin our modelling studies by formulating *in silico* synthetic microbial communities designed to interrogate the research questions outlined in Part I. For these simulation studies, we develop an individual-based framework that is progressively extended.

In Chapter 5 we develop an individual-based model of three *in silico* microbial species to examine how evenness and the type of interspecies competition affect the diversity and stability of the community. We describe in detail our model and the processes it incorporates, then explain the set-up of the *in silico* experiments used to investigate our research questions. The results of these simulation studies are presented, and conclusions are drawn regarding the role of initial evenness and competition type.

This model is extended in Chapter 6 to include a fourth species. This increased level of complexity has two important consequences. First, it admits more possi-

ble competition schemes, which we enumerate and study. Second, the increased richness permits us to conduct a second *in silico* experiment in addition to the approach used in Chapter 5, namely an invasion experiment. These two set-ups have different implications for the diversity and functionality of our *in silico* community, which we test through extensive simulation studies.

Another important extension to our modelling framework is developed in Chapter 7, which is the inclusion of resource dynamics. This is achieved through the consideration of an *in silico* environmental substrate to which the demographic processes are linked. The impact of resource dependence of various types, as well as the effects of variable community evenness, are assessed through *in silico* experiments.

In Chapter 8, we assess the effects of resource dependence on the spatial population dynamics of the *in silico* community, and the consequent impacts on community diversity and functionality. Aside from spatial aspects, we also consider more complex forms of resource dependence, as well as their consequences for the stability of the community.

In **Part III**, we incorporate data from *in vitro* synthetic microbial communities in our modelling framework in order to bring it closer to reality. We make use of a dataset related to bioaugmentation of synthetic microbial communities in water treatment sand filters, which we describe along with the experiments designed to obtain this data.

In the first part of Chapter 9, we describe a predictive model developed for the purpose of highlighting interactions between microbial strains which are of interest for bioaugmentation in this setting, and can help to lessen the *in vitro* experimental load.

In the second part of Chapter 9, we apply our individual-based framework. For this purpose, we assess the interactions, first by identifying strain identity effects in the data, then by synthesizing these effects into a competition structure for our model. After formulating our model, we employ it as an *in silico* counterpart of the *in vitro* sand filter community to determine if this is able to reproduce the key behavioural trends occurring *in vitro*.

Finally, in **Part IV** we summarize the modelling studies described in this thesis as well as their conclusions, and discuss their implications for the field of synthetic microbial ecology. We also outline the promising research avenues opened by the work contained in this thesis.

In sum, we employ *in silico* synthetic microbial communities to test key ecological theories. Using an individual-based framework, we assemble various *in silico* communities for the purpose of testing microbial ecological theories relating to community stability, diversity, functionality and productivity.

Our work has implications for the management of natural communities, and the engineering of synthetic communities for various applications. The modelling framework we have developed is flexible, extendable to other avenues of research, and furthermore the modelling techniques described in this thesis are not limited in their applicability to microbial ecology, but can be used in other disciplines and fields, such as the marine, food and agricultural sciences.

Nederlandstalige samenvatting

Microbiële gemeenschappen zijn cruciaal voor de werking van het Aardse ecosysteem. Ze spelen een uitermate belangrijke rol in ecologische, geochemische, medische en industriële processen. Daarnaast zijn van vitaal belang voor de gezondheid van mens en dier.

Het zijn niet de afzonderlijke soorten die deze functies vervullen, maar wel diverse microbiële gemeenschappen. Een goed begrip van de structuur en het functioneren van zulke gemeenschappen is van cruciaal voor een goed management van natuurlijke gemeenschappen, en om deze te beschermen tegen significante veranderingen van hun leefmilieu om zo hun vitale rol te vrijwaren. Bovendien is een dergelijk begrip noodzakelijk voor een rationeel ontwerp van microbiële gemeenschappen met het oog op belangrijke toepassingen binnen onder meer de voedings- en farmaceutische industrie.

Met dit in het achterhoofd wordt er veel aandacht besteed aan de ontwikkeling en implementatie van instrumenten, technieken en modellen die het mogelijk maken om het gedrag en functionaliteit van microbiële gemeenschappen te voorspellen, het zogenaamde Microbial Resource Management. Om deze MRM technieken te ontwikkelen, dienen onderzoekers te beschikken over een grondige en fundamentele kennis van de interacties die plaatsgrijpen binnen microbiële gemeenschappen, vermits deze hun dynamiek sturen.

In dit kader zijn wiskundige modellen steeds belangrijker. Modellen zijn abstracties van de realiteit die kunnen gebruikt worden om op een gecontroleerde manier

hypotheses te verifiëren. Het zijn als het ware de *in silico* tegenhangers van *in vitro* synthetische ecosystemen die in het laboratorium zijn gecreëerd. De ontwikkeling van wiskundige modellen in het kader van MRM kan niet alleen helpen om de fundamentele mechanismen achter de MRM theorieën beter te begrijpen, maar tevens om de ontwikkeling van nieuwe hypotheses te stimuleren.

In dit proefschrift richten we ons op *in silico* synthetische microbiële gemeenschappen om belangrijke microbiële ecologische theorieën met betrekking tot de stabiliteit, diversiteit, functionaliteit en productiviteit van microbiële gemeenschappen te verifiëren. Hiervoor maken we gebruik van een individu-gebaseerde benadering die verschillende niveaus van microbiële interacties, diversiteit en complexiteit toelaten.

In *Deel I* verzamelen en synthetiseren we informatie en kennis aangaande de bestaande *in vitro* en *in silico* technieken die relevant zijn voor ons onderzoek.

In Hoofdstuk 2 bespreken we de belangrijke functies die microbiële gemeenschappen vervullen. Tevens bespreken we de experimentele technieken die gebruikt worden om deze gemeenschappen te bestuderen en ecologische theorieën over hun samenstelling en functionaliteit te ontwikkelen. Vervolgens duiden we aan waarom synthetische microbiële gemeenschappen zich uitermate goed lenen voor deze theorieontwikkeling.

In Hoofdstuk 3 bespreken we de wiskundige modellen die geschikt zijn om de dynamiek van microbiële gemeenschappen te simuleren. We richten ons op de modellen die werden ontwikkeld om belangrijke aspecten van zulke gemeenschappen te bestuderen, zoals hun ruimtelijke structuur en dynamiek, de interacties binnen de gemeenschap en met de omgeving.

In Hoofdstuk 4 bespreken we de diversiteit van microbiële gemeenschappen. Meer in het bijzonder gaan we na wat daarmee precies bedoeld wordt en hoe deze kan gekwantificeerd worden met behulp van indices, alvorens een vergelijkend overzicht van de belangrijkste indices te geven. Daarna geven we volgt een gelijkaardige bespreking voor de gelijkheid (“evenness”) van microbiële gemeenschappen, een van de twee componenten van biodiversiteit.

In **Deel II** beginnen we onze modelleringsstudie door het formuleren van *in silico* synthetische microbiële gemeenschappen die ontworpen zijn om de onderzoeksvragen opgeworpen in Deel I te beantwoorden. Voor deze simulatiestudies ontwikkelen wij een individu-gebaseerd model dat stapsgewijs wordt uitgebreid.

In Hoofdstuk 5 ontwikkelen we een individu-gebaseerd model met drie *in silico* microbiële soorten, om te onderzoeken hoe de gelijkheid en het soort competitie de diversiteit en stabiliteit van de gemeenschap kunnen beïnvloeden. We beschrijven in detail ons model en de processen die erin zijn opgenomen en geven de details over de opzet van *in silico* experimenten die worden gebruikt om onze onderzoeksvragen te onderzoeken. De resultaten van deze simulatiestudies worden

besproken en er worden conclusies getrokken over de rol van gelijkheid en de soort competitie.

Dit model wordt uitgebreid in Hoofdstuk 6 zodat er vier soorten kunnen beschouwd worden. Deze verhoogde complexiteit heeft twee belangrijke gevolgen. Ten eerste laat deze meer interactie regels toe. Ten tweede maakt ze het mogelijk om *in silico* invasie-experimenten uit te voeren. Zulke experimenten hebben een eigen impact op de diversiteit en functionaliteit van de *in silico* gemeenschappen, die we openbaren door uitgebreide simulatiestudies.

Het model wordt verder uitgebreid in Hoofdstuk 7 met het incorporeren van de substraatdynamiek. Dit wordt mogelijk door een *in silico* substraat te beschouwen dat de demografische processen beïnvloedt. Het effect van verschillende soorten hulpbronnen wordt onderzocht via *in silico* experimenten.

In Hoofdstuk 8, beoordelen we de gevolgen van de substraatafhankelijkheid op de ruimtelijke populatiedynamiek van de *in silico* gemeenschap, en de finale effecten op de diversiteit en functionaliteit van de gemeenschap. Naast deze ruimtelijke aspecten, beschouwen we ook meer complexe vormen van substraatafhankelijkheid, evenals de gevolgen voor de stabiliteit van de gemeenschap.

In **Deel III** maken we gebruik van gegevens over een *in vitro* synthetische microbiële gemeenschap om het modelleerkader dichter bij de realiteit te brengen. Meer in het bijzonder gebruiken we gegevens over een synthetische microbiële gemeenschap die een rol speelt bij de bioaugmentatie van zandfilters. We beschrijven deze gegevens evenals de uitgevoerde *in vitro* experimenten.

In de eerste sectie van Hoofdstuk 9 beschrijven we een datagedreven model voor het voorspellen van interacties tussen microbiële soorten die van belang zijn voor bioaugmentatie in zandfilters. Deze benadering maakt het mogelijk om de experimentele last te verlichten.

In de tweede sectie van Hoofdstuk 9, passen we ons individu-gebaseerd model toe. Hiervoor beoordelen we eerst de interacties tussen de microbiële soorten in de beschouwde *in vitro* gemeenschap. Het model gebruiken we uiteindelijk als de *in silico* tegenhanger van de *in vitro* zandfiltergemeenschap en dit om na te gaan of het toelaat om de *in vitro* dynamiek kwalitatief te reproduceren.

In **Deel IV** vatten we tot slot onze modelleerstudie en de belangrijkste conclusies samen en bespreken we de implicaties voor het onderzoeksdomein van de synthetische microbiële ecologie. We geven tevens een overzicht van veelbelovende onderzoeksactiviteiten die uit dit proefschrift voortvloeien.

Samenvatten spitsen we in dit proefschrift toe op *in silico* synthetische microbiële gemeenschappen om ecologische theorieën op de proef stellen. Door het gebruik van een individu-gebaseerd kader, assembleren we diverse *in silico* gemeenschappen om microbiële ecologische theorieën te testen die verband houden met de stabiliteit, diversiteit, functionaliteit en productiviteit van de gemeenschap.

Ons werk heeft gevolgen voor het beheer van natuurlijke gemeenschappen en het samenstellen van synthetische microbiële gemeenschappen. Het ontwikkelde modelleerkader is flexibel en uitbreidbaar. Bovendien is de inzetbaarheid ervan niet beperkt tot toepassingen in microbiële ecologie, maar strekt deze zich tot vele andere wetenschappelijke disciplines.

List of symbols

α	order of an entropy (−)
A	total extent of mineralization (%)
c	endogenous mineralization rate (%h^{-1})
D	diffusion coefficient ($\mu m^2 s^{-1}$)
ϵ	mobility rate (T^{-1})
ϵ_c	critical mobility rate (T^{-1})
E_c	substrate conversion efficiency (g_{mass}/g_{sub})
E_r	reproductive efficiency (g_{mass}/g_{mass})
E_0	initial evenness (−)
E	Simpson evenness (−)
E_H	Heip evenness (−)
λ	lag time (%h^{-1})
H	Shannon diversity
H_α	Rao diversity of order α
H_E	Shannon evenness (−)
H_{GS}	Gini-Simpson diversity (−)

H_{Si}	Simpson diversity (—)
I	identity matrix (—)
I_0	initial substrate concentration (g)
$I^t(i,j)$	biomass at cell (i,j) at time t ($g\ \mu m^{-2} s^{-1}$)
J	Pielou evenness (—)
K_s	Monod half-saturation constant (gL^{-1})
L	side length of lattice (—)
μ	Gompertz maximum mineralization rate (s^{-1})
μ_{max}	Monod maximum growth rate (s^{-1})
M	tournament matrix (—)
m_{max}	maximum biomass of an individual (g)
∇^2	Laplacian operator (—)
N	total number of lattice sites (—)
$\varnothing$	empty lattice site (—)
P	cumulative mineralization (%)
p_i	proportion of community represented by species i (—)
$\mathbf{p}$	vector of species proportions in a community (—)
q	order of a diversity index (—)
${}^qD$	diversity of order q (—)
r	random number drawn from the unit interval (—)
r_I	rate of substrate uptake ($g\ s^{-1}$)
R_{SFI}	richness of SFIs in a co-culture (—)
R_T	total richness in a co-culture (—)
R^2_{CV}	cross-validated R^2 statistic (—)
σ	competition rate (T^{-1})
S	number of species in a community (—)
$s^t(i,j)$	substrate concentration at cell (i,j) at time step t (g)
T	number of generations simulation is evolved (—)

Δt	time discretization step size (—)
τ_i	substrate threshold of species i (g)
x	first space coordinate of a lattice site (—)
Δx	space discretization step size (—)
$\hat{X}$	first-order Sobol index (—)
y	second space coordinate of a lattice site (—)
z	test statistic (—)
z_{crit}	critical test statistic value (—)
Z	similarity matrix (—)

List of acronyms

ABC	Approximate Bayesian computation
ABM	Agent-based model
AOC	Assimilable organic carbon
BAM	2,6-dicholorobenzamide
CA	Cellular automata
DNA	Deoxyribonucleic acid
FBA	Flux balance analysis
GFP	Green fluorescent protein
HPC	High performance computing
MFA	Mean field approximation
μIBE	Microbial individual-based ecology
MMO	Minimal medium ONPG
MRM	Microbial Resource Management
ODD	Overview, Design concepts, and Details
ODE	Ordinary differential equation
OTU	Operational taxonomic unit

PDE	Partial differential equation
PIE	Probability of interspecific encounter
PLM	Population-level model
RMSE	Root-mean-square-error
rRNA	Ribosomal ribonucleic acid
SAD	Species abundance distribution
SEM	Spatially explicit model
SF	Sand filter
SFI	Sand filter isolate
SIM	Spatially implicit model
SPDE	Stochastic partial differential equation
WWTP	Waste water treatment plant

List of Figures

List of Tables

PART I

INTRODUCTION AND BACKGROUND

1

Introduction

1.1 Overview

Microbial communities are both bewilderingly complex and vitally important. They play central roles in key ecological, geochemical, medical and industrial processes. They drive our planet's biogeochemical cycles, including the carbon and nitrogen cycles that circulate these key elements between the Earth's soil, ocean and atmospheric ecosystems, thereby permitting life to continue. In their role as the primary drivers of these element cycles on a planetary level, microbial communities are essential for the functioning of each and every ecosystem on Earth.

Furthermore, while the work of microbial communities allows human, plant and animal life to persist, their initial establishment was also due to the slow and steady efforts of microbial communities. Over millions of years, they transformed the Earth's atmosphere from its unwelcoming primeval state, heaving with corrosive chemical compounds, until it was rich enough in oxygen to permit the development of multicellular life. Their importance continues to the present day, where microbial communities underpin the functioning of every ecosystem on the planet.

1

Despite their singular importance to our existence on this planet, humanity did not begin to study microbial communities until relatively recently. Although their existence was postulated as far back as the 6th century, microbes were not observed and identified until the 16th century, when Antonie van Leeuwenhoek designed and built a microscope with which he observed “animalcules”, and Robert Hooke recorded his observations in moulds of what he named “cells”.

The field of microbiology was solidified as a scientific discipline in the 19th century when Louis Pasteur disproved the theory of spontaneous generation of matter, and showed that microbial life could not arise from non-living materials. Then in the early 20th century, the development of enriched culture techniques permitted the cultivation of microbial species, which began to reveal the incredible extent of microbial biodiversity.

The study of microbial community life has advanced ever since, with an ever increasing understanding of their assembly and functioning allowing microbial communities to be put to work. They drive many industrial processes, from food technology to medicine and drug development, to the production of chemicals and fuels. Even before humanity gained a working understanding of their functionality, we were able to initiate and even steer some of the processes driven by microbial communities, particularly with regards to food manufacturing processes such as fermentation. The manufacture of cheese, pickled items, alcohol and fermented meats is many centuries old, and was developed through trial and error well before scientific knowledge permitted any close study of the microbial communities driving this process.

In the current day, there is more and more concern for sustainability and climate compatibility in our industrial processes, due to the increasing recognition of the dangers of anthropogenic changes to our environment. For this reason, many processes dominated or driven by microbial communities are considered superior to processes driven by chemical or physical reactions, which typically result in emissions harmful to global ecosystems.

Hence stakeholders of various kinds are increasingly interested in the development and availability of tools and techniques that can be used to predict the behaviour and sustainably manage complex microbial communities. Indeed, understanding the ecology of microbial communities — that is, their interactions with each other and with their environments — has been singled out as “*one of the most compelling intellectual challenges facing contemporary ecology*” (Prosser et al., 2007).

Concurrently, the most recent decades have produced incredible advances in various scientific disciplines such as molecular and evolutionary ecology, together with significant progress in technologies such as DNA sequencing, genomics, proteomics, and metabolomics. These advances in theoretical and applied sciences have facilitated improved observation, understanding and prediction of the func-

tioning of microbial communities and the processes they drive. However, much remains to be done.

For example, until very recently it was not practically possible to identify microbes and observe their functions at the level of single cells, neither in simplified laboratory settings nor in more complex natural communities. While this is now changing with the development and applications of new technologies for single-cell microbiology and microbial ecology, it has only raised more questions for these fields.

Most notably, the observation of individual microbes *in situ* has led to one of the most important recent findings of microbiology: that microbes which are genetically identical and situated in a well-mixed environment can nevertheless have different phenotypes (Kreft et al., 2013). This level of individual variation even in homogeneous environmental conditions calls into question the assumption underpinning all population-level experiments, namely that all individuals are roughly the same and thus population-level averages are a sufficiently accurate estimation of their characteristics. With this paradigm now significantly undermined, the implications for the future development of the field are still emerging, but certainly imply a major shift in conceptual and practical approaches.

The current challenge for microbial ecologists is to sustain and protect the Earth's microbial communities and resources, in order to preserve the ecosystems and processes dependent on them. This endeavour has led to major effort being focused, for example, on understanding the effects of biodiversity on ecosystem stability and functioning. The recent and ongoing biodiversity crisis has made clear the links between these processes, but improving our understanding of biodiversity is complicated by its entangled taxonomic, functional, spatial and temporal aspects.

Thus it is not surprising that the large majority of studies attempting to develop theories relating to the links between biodiversity, stability and functioning have been phenomenological rather than conceptual. Until very recently, the role of ecological theory in microbial ecology has been neglected and it is the increasingly strong belief of microbial ecologists that "*advances in microbial ecology are limited by a lack of these conceptual and theoretical approaches*" (De Roy et al., 2014).

Theory, rather than context-specific observations, is necessary to "*classify, interpret and predict the world around us*" (Prosser et al., 2007). We require theory to interpret our scientific observations and to extrapolate our insights to other settings. It can enable a better understanding of the crucial factors steering microbial communities by providing a framework within which to gather, synthesize and understand experimental observations and to validate their implications.

An example of successful theory development at the macro-scale is the development of epidemiological models of the spread of diseases in humans, animals or plants, which have been tested and improved to such an extent that stakeholders use their predictions to steer policy decisions. Such a body of theory and models

would have equally useful applications in improving for example the efficiency and sustainability of waste water treatment strategies or the manufacture of biofuels, both of which have important roles in any sustainable industrial policy.

Hence we expect that increasing interest and effort will be dedicated to the development and deployment of tools, techniques and models that allow for the prediction of behaviour and functionality in microbial communities. This effort, namely the development of tools and techniques that have come to be referred to as Microbial Resource Management (MRM), will permit the management and protection of natural communities, as well as the rational design of engineered communities for important industrial applications. To develop these MRM tools, we first require a deeper and more fundamental understanding of the interactions taking place within these microbial communities, since these are the fundamental mechanisms driving community behaviour.

An important tool in this respect is the use of mathematical models. Models are abstractions of reality that allow for the testing of hypotheses in a controlled way. Their reduced complexity and inherent manipulability make them the *in silico* counterparts of *in vitro* model ecosystems constructed in the laboratory.

Developing mathematical models related to MRM theories can not only help to understand the fundamental mechanisms underpinning these theories, but can also help to develop new hypotheses by highlighting interesting or unexpected behaviour, which can then be explored and tested in more focused modelling and experimental studies. Mathematical models can also be used for predictive purposes. When constructed based on sound ecological theories, models can not only help to understand the fundamental processes underlying these theories, but also to predict under which conditions these theories may no longer hold, or under which conditions these theories are particularly key to community functioning.

Furthermore, if the appropriate data is available, then models can be calibrated and validated in order to make quantitative predictions about community stability and functionality. Technology is now approaching a sufficiently sophisticated stage to allow for the collection of the type of data needed for this purpose, increasing even further the promise of modelling approaches for the purpose of microbial ecology theory development.

1.2 Research questions

In this thesis, we will develop mathematical models to represent *in silico* “synthetic ecosystems”, in order to test microbial ecological theories. In these *in silico* microbial communities, different levels of microbial interactions and complexity will be assembled. This mathematical approach to ecological theory development will allow us to test ecological theories on microbial communities which so far have been

adopted from macro-ecology without much consideration of the appropriateness of this extrapolation.

More specifically, in this thesis we will investigate the roles of (i) community diversity, (ii) community structure, (iii) community architecture, and (iv) microbe to microbe interactions, in ecosystem stability and functionality.

For this purpose, we formulate several research questions to guide our investigations, which will be elaborated and motivated in subsequent chapters:

1. What effect does initial evenness have on maintaining community diversity?
2. Which types of competitive interactions can help to maintain community diversity, and which types can threaten it?
3. What effect does initial evenness have when a community is faced with invasion?
4. What effect does initial evenness have on maintaining community functionality?
5. If interactions within a community are dependent on resource availability and use, how does this affect community diversity and functionality?
6. How does the spatial structure of a community affect its stability and functionality?

These research questions will be studied through the use of *in silico* microbial communities designed to highlight the relevant community processes and mechanisms underlying our research questions.

1.3 Scope of the thesis

In the remainder of Part I, we will gather and synthesize the existing *in vitro* and *in silico* techniques and knowledge relevant to our study of microbial ecological theories.

In Chapter 2, we discuss in more detail the important roles played by microbial communities in numerous vital domains of life, then survey the experimental techniques used to study these communities and to develop ecological theories regarding their composition and functionality. We then outline why synthetic microbial communities are particularly suited for this type of theory development, before discussing one such group of theories in particular, namely those related to the link between biodiversity and functionality.

In Chapter 3, we survey the mathematical modelling approaches suitable for microbial communities. We structure this summary in terms of the basic unit of the models under discussion, namely whether this basic unit is a community, a population, or an individual. We focus on the modelling approaches that have been developed to study several key aspects of microbial communities in terms of their functionality: their spatial structures and dynamics, the interactions taking place within the community, and the interactions occurring with the environment.

We discuss another key aspect of microbial communities in Chapter 4, which has been highlighted as particularly important for understanding their structure and behaviour: their diversity. We define precisely what is meant by diversity (not an easy task, as we shall discover) and how it can be quantified through the use of dedicated indices, before providing a comparative survey of selected key indices. We then focus on evenness, which is the more complex of the two components of diversity, and has attracted less attention. We again provide a comparative survey of the indices that have been proposed to quantify evenness.

In Part II, we begin our modelling studies by formulating *in silico* synthetic microbial communities designed to interrogate the research questions outlined in Section 1.2. For these simulation studies, we develop an individual-based framework that is progressively extended.

We begin in Chapter 5 by developing an individual-based model to study how evenness and the type of interspecies competition will affect the diversity and stability of the community. We describe in detail our model and the processes it incorporates, then explain the set-up of the *in silico* experiments used to investigate our research questions. The results of these simulation studies are presented, and conclusions are drawn.

This model is extended in Chapter 6 to include a fourth species. This increased level of complexity has two important consequences. First, it admits more possible competition schemes, which we enumerate and study. Second, the increased richness permits the use of a second *in silico* experimental set-up in addition to the approach used in Chapter 5, namely an invasion experiment. These two set-ups have different implications for the diversity and functionality of our *in silico* community, which we test through extensive simulation studies.

Another important extension to our modelling framework is developed in Chapter 7, which is the inclusion of resource dynamics. This is achieved through the use of an *in silico* environmental substrate to which the demographic processes are linked. The impact of resource dependence of various types, as well as the effects of variable community evenness, are assessed through *in silico* experiments.

In Chapter 8, we assess the effects of resource dependence on the spatial population dynamics of the *in silico* community, and the consequent impacts on community diversity and functionality. Aside from spatial aspects, we also consider

more complex forms of resource dependence, as well as their consequences for the stability of the community.

In Part III, we incorporate data from *in vitro* synthetic microbial communities in our modelling framework in order to bring it closer to reality. We make use of a dataset related to bioaugmentation of synthetic microbial communities in water treatment sand filters, which we describe along with the experiments designed to obtain these data.

In the first part of Chapter 9, we describe a predictive model developed for the purpose of highlighting interactions between microbial strains that are of interest for bioaugmentation in this setting, and can help to lessen the *in vitro* experimental load. In the second part of Chapter 9, we apply our individual-based framework. For this purpose, we assess the interactions, first by identifying strain identity effects in the data, then by synthesizing these effects into a competition structure for our model. After formulating our model, we employ it as an *in silico* counterpart of the *in vitro* sand filter community to determine if this is able to reproduce the key behavioural trends occurring *in vitro*.

Finally, in Part IV we summarize the modelling studies described in this thesis as well as their conclusions, and discuss their implications for the field of synthetic microbial ecology. We also outline the promising research avenues opened by the work contained in this thesis.

2

Biological background

2.1 Introduction

Microbial communities are ubiquitous on Earth, and are estimated to encompass upwards of 10^{30} individual micro-organisms (Schloss and Handelsman, 2004). Such an incredible number is simply too gigantic for the human mind to picture, but an idea of the staggering abundance of microbial life can be taken from the observation that the current human population of Africa is estimated to be equal in number to the microbes living in a single teaspoon of soil (Editorial, 2011).

These microbial communities play key roles in human and animal health, industrial, medical and pharmaceutical processes, and global ecosystems (Hanemaaijer et al., 2015). They drive the Earth's biogeochemical cycles and have preserved them even through severe environmental disturbances that resulted in mass extinctions of animals and plants (Hallam and Wignall, 1997), leading to the sobering observation that "*[m]icrobial life can easily live without us; we, however, cannot survive without the global catalysis and environmental transformations it provides*" (Falkowski et al., 2008).

These functions, which are so key to our global ecosystem, are not performed by any single microbial species, but rather by diverse communities (Larsen et al., 2012). The ability to understand the structure and functioning of these complex communities is crucial to manage natural communities, to protect them from ongoing and significant anthropogenic environmental changes (Benner et al., 2013) and thus preserve their vital processes, as well as to rationally design engineered microbial communities for important applications ranging from food technology, to medical and pharmaceutical uses, to various industrial and bioindustrial processes (Friedman et al., 2017).

In this chapter, we first discuss in Section 2.2 the importance and uses of microbial communities. Then, in Section 2.3, we summarize the experimental tools and techniques that are typically used to study the key aspects of microbial communities. These aspects can be summarized by posing the following basic questions (Little et al., 2008). First, who is present in the community? This refers to the *structure* of the community, namely which species are present and in what proportions (Section 2.3.1). Second, what are they doing? This corresponds to the *functionality* of the community (Section 2.3.2). Third, how are these functions being accomplished? This encompasses the *interactions* taking place between micro-organisms within the community, as well as with their environment (Section 2.3.3).

To understand the fundamental ecological processes underlying these questions, one type of microbial community in particular has recently gained much attention. This is a *synthetic microbial community*, which we define and discuss in detail in Section 2.4. We then delve deeper into one key property that synthetic communities are particularly suited for investigating, and which will be a focus of this thesis: biodiversity. This property is defined and its important role in proper community functioning is discussed in Section 2.5. We then focus our discussion in Section 2.6 on the two components of biodiversity, namely richness and evenness. Finally, in Section 2.7 we summarize our findings and their implications for our studies.

2.2 The importance of microbial communities

Microbial communities can be put to work in myriad beneficial ways. To name just a few representative examples, they can be employed to: produce foods (Wolfe and Dutton, 2015); treat waste water (Röling et al., 2010); degrade various compounds including cellulose, plastic and heavy metal toxins (Tan et al., 2015); clean contaminated soils (Hairston et al., 1997); leach minerals (Bertrand et al., 2015); drive soil processes such as nitrogen fixation that are key to agriculture (Hanemaaijer et al., 2015); and produce biofuels (Zomorrodi and Segrè, 2016).

Microbial communities also play important roles in human health and disease (Stein et al., 2013). There is a significant and growing body of evidence that numerous and varied diseases (pictured in Figures 2.1 and 2.2), such as inflammatory

bowel disease, obesity and diabetes, are not associated with any single microbial species but rather with altered, unbalanced or malfunctioning microbial communities (Friedman et al., 2017). This is particularly acute in the gut microbiome, where external disturbances or perturbations such as diet changes can shift the composition of the resident microbial communities (Walker et al., 2011). These changes in composition or structure can lead to biodiversity loss, which has been shown to increase the risk of detrimental bacterial infections (Dethlefsen and Relman, 2011).

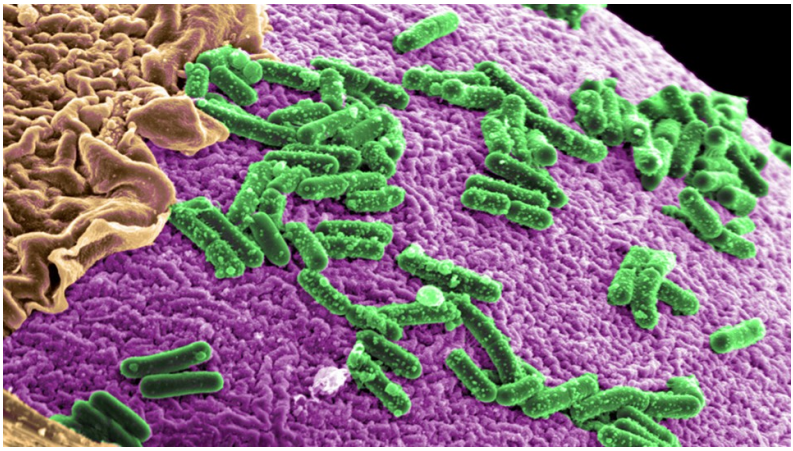


Figure 2.1: Changes to microbial communities in the human gut have been attributed to various diseases including colon cancer, diabetes, and even neurological diseases such as Alzheimer's diseases. Image courtesy of Pacific Northwest National Laboratory (under Creative Commons license).

Once these links between malfunctioning gut microbial communities and ill health were recognized and understood, efforts could be made to reverse these effects. An example is the use of faecal transplants, which involves introducing samples extracted from the gut microbiome of a healthy donor into the patient's intestines in order to prompt the re-establishment or re-orientation of their gut microbiome, and has been shown to be highly effective in certain cases, for example against forms of colitis induced by *Clostridium difficile* (Bakken et al., 2011).

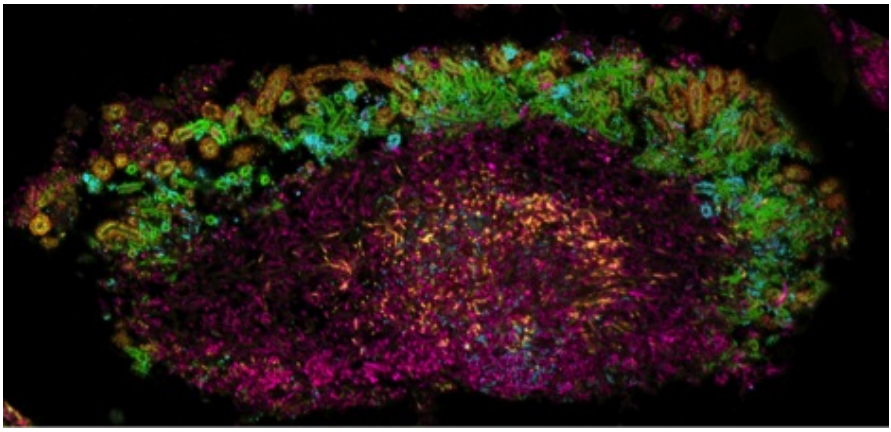


Figure 2.2: Microbial communities form in human dental plaque (Valm et al., 2011).

Microbial communities also play vital roles in food technology, where they drive the bioprocesses that are responsible for the manufacturing of bread, wine, beer, vinegar and cheese, among many other examples (Hanemaaijer et al., 2015). The activity of these microbial communities (pictured in Figure 2.3) are responsible for the distinctive flavour, texture, and aroma of fermented foods (Hutkins, 2006). Humans have developed over many thousands of years a sufficient knowledge of the composition and behaviour of these communities as to be able to control and steer them to produce different types of fermented foods. This can be achieved through the manipulation of environmental conditions such as temperature, salinity, and moisture, which affect the functionality of the microbial communities and alter their effects on the food (Sabra et al., 2010).

Probiotics are defined as microbes that are beneficial to gastrointestinal health, by promoting the proper functioning of microbial communities in the gut (Pham et al., 2009). Probiotic treatments use microbial strains that are identical to those in the human microbiome, and are thus regarded as safer than other treatments (Rastall et al., 2005). Through the use of these probiotic strains, researchers and physicians seek to interfere with the bacteria causing ill gastrointestinal health, in order to retrieve the condition of a healthy gut microbiome that is able to exclude or repel deleterious pathogens (Jenkinson and Lamont, 2005).

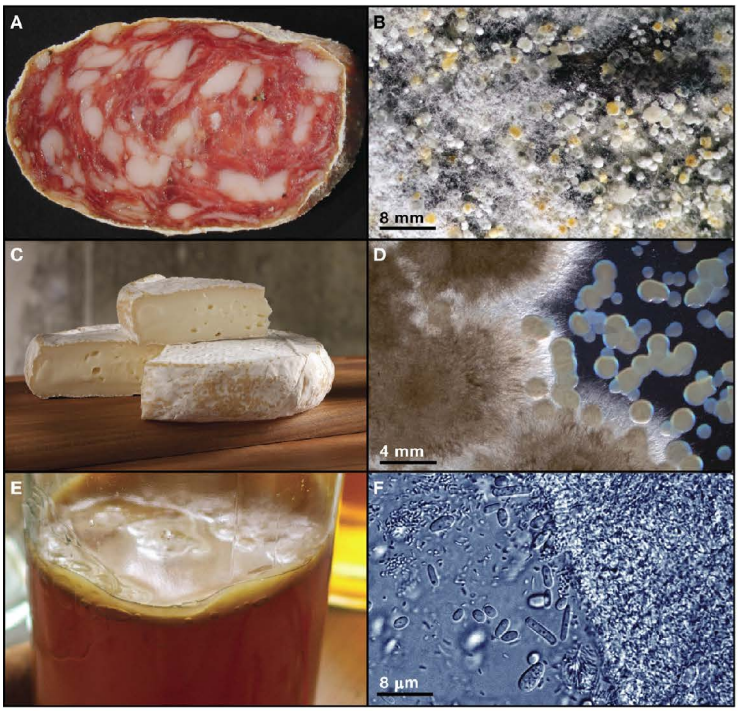


Figure 1. Multi-speci
nities Form during th
mented Foods
(A) Fermented meats, s
duced by fermentation
bacteria.
(B) During the aging proc
colonized by a mixture
visible as white and yell
tous fungi (diffuse wh
Penicillium.
(C) Cheeses, such as the
shown, are made throug
by lactic acid bacteria.
commonly called a rind,
and contributes to the fla
the cheese.
(D) A rind biofilm plated
shows a subset of the m
tous fungi on the left)
bacteria on the right) me
communities.
(E and F) (E) Visible mi
form in liquid fermenta
mented tea, commonly k
microbial cells within th
can be seen in the micr
typically composed of
acetic acid bacteria (sma
involved in the fermenta
ethanol and carbon diox
teria then ferment the et
acid. The intact biofilm
and bacterial cells are sl
All photos by Benjamin
Hill Farm) and (E) (Adam

Figure 2.3: Microbial communities drive the fermentation of food and drink (Wolfe and Dutton, 2015). Fermented meats such as salami (A) are colonized on their surfaces by communities of bacteria and yeast (B). Cheeses (C) are also produced by fermentation, during which microbial communities form a biofilm (D) which is called the rind. Fermented teas such as kombucha (E) involve the formation of a floating biofilm (F) which consists of a community of bacteria and yeasts.

Recent advances in genomic and metagenomic sequencing are providing researchers with catalogs of the bacterial, fungal, and viral diversity in many traditionally produced fermented foods (Table 1; reviewed in Bokulich and Mills, 2012). These communities range in composition from those dominated by bacterial species to those dominated by fungal species, with some communities containing a mix of both bacteria and fungi. Certain bacterial groups such as the lactic acid bacteria (LAB) and acetic acid bacteria (AAB), as well as fungal species such as *Saccharomyces cerevisiae*, have well-established roles in fermentation. However, the increasingly detailed analysis of the microbial diversity of fermented foods is revealing many additional species whose roles have not been characterized extensively, if at all. For example, marine-associated *Pseudoalteromonas* are dominant members of some meat and fish communities (Hultman et al., 2014) (in Table 1, these and all non-LAB/AAB fall under "other bacterial groups").

After characterizing the diversity of a microbial community, one of the biggest challenges in the study of microbial ecosystems is the difficulty in culturing community members in the laboratory. Because MCoFFs have defined starting materials as growth substrates (e.g., milk, grapes, and wheat flour) and known incubation conditions, these same conditions can be replicated in the lab and used as starting conditions for isolation of community members. Indeed, some food-associated microbes are already well-established model organisms, such as *Saccharomyces cerevisiae* and *Lactococcus lactis*.

Experimentation using MCoFFs is also
the fact that they are extremely accessib
tems. The production of fermented foods h
tervals (from daily to seasonally), and com
short timescales (from days to months), all
access to many replicated samples over
periods. Fermented foods are often produ
geography rooms, also increasing the acc
These communities form as part of discrete
of cheese), which allows well-defined s
sampling.

MCoFFs have some potential limitations
Given the short timescales required to form
may be fewer opportunities for species to
horizontal gene transfer between speci
MCoFFs suggests that at least some memb
nities have coexisted long enough to allow
(Chiellini et al., 2014; Rossi et al., 2
tion within communities. For example, fer
(Table 1), providing ample opportunities fo
kombucha consist of a pellicle that contain
in a mixed biofilm (Figures 1E and 1F). Thes
spread all around the world (Marsh et a
geographically separated communities th
from initially identical species and genetic

Because MCoFFs grow on raw food mat
meat, or milk, most nutrients are not limite

This has been accomplished with some success for inflammatory bowel disease, antibiotic-associated diarrhoea and irritable bowel syndrome, among other gut illnesses, as well as for illnesses not specific to the gut such as respiratory infections, vaginitis and hypercholesterolaemia (Pham et al., 2009). For example, experimental studies have shown that probiotics can have therapeutic benefits for patients suffering from irritable bowel syndrome, although both the magnitude of the benefit and the most effective species for probiotic treatment could not be identified with certainty (Moayyedi et al., 2008). The most frequently used probiotics consist of communities of microbial strains from the genus *Lactobacillus*, an example of which is pictured in Figure 2.4 (Pham et al., 2009).

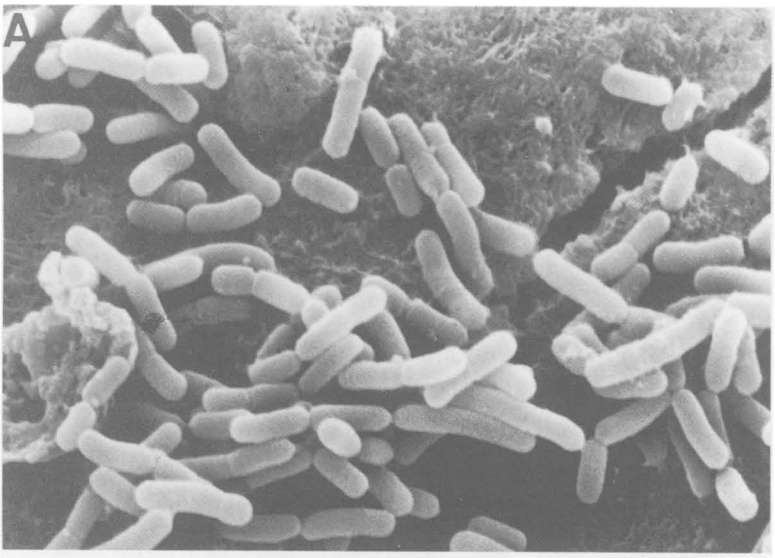


Figure 2.4: An electron microscope image of *Lactobacillus acidophilus*, often used in probiotic products such as yoghurt (Bernet et al., 1994).

Microbial communities also play vital roles in many industrial processes, and are used to produce various chemicals, materials and fuels (De Roy et al., 2014). In this context, the engineering of microbial communities is seen as a keystone technology for sustainable energy technologies, in order to provide alternatives for the current and unsustainable dependence on fossil fuels (Sabra et al., 2010). The biotechnology industry is estimated to represent billions of dollars in the global economy (Stewart et al., 2001), and involves such products as amino acids, organic acids, antibiotics, enzymes, vitamins and pharmaceutics (Sabra et al., 2010). Other industrial bioprocesses driven by microbial communities include the production of biogases, bioethanol and biohydrogen (Bader et al., 2010).

2.3 Experimental techniques for studying microbial communities

After surveying the undoubted importance of microbial communities in our bodies, ecosystems and industries, we now turn to the question of how researchers study these communities. An understanding of the structure and functioning of these communities is vital in order to permit the proper management of natural communities, as well as to enable the design and control of engineered microbial communities.

2.3.1 Censusing the community

The most intuitive avenue with which to begin is the question of who is present in the first place? The *structure* of a community refers to its composition (which strains are present), as well as the abundances of its member species. This touches on two components of biodiversity, namely richness (the number of species present) and evenness (the relative abundances of the different species). The important roles of these components in community diversity, stability and functionality will be discussed in more detail in Section 2.6, and their quantification will be the subject of Chapter 4. Before we can address such quantification, we must first take a step back and address the question of how one can observe, characterize and identify microbes in synthetic communities. Established techniques to census microbial communities can be grouped into two broad approaches: culture-based methods and culture-independent methods.

Culture-based methods rely on the isolation and cultivation of microbes (pictured in Figure 2.5). These studies provide morphological and physiological data that are used to characterize and identify microbial species (Little et al., 2008). However, these techniques present two major drawbacks. First, it is estimated that less than 1% of the strains in the global microbiome can be cultured (Stewart, 2012). Hence, the relatively few strains that are culturable cannot be said to be broadly representative of microbial strains in general. In particular, the choice of medium represents a significant bias in terms of the microbial strains able to be cultured *in vitro*. Many studies use rich media, which provides plentiful nutrition in contrast to the frequently nutrient-poor environments found in natural ecosystems, and hence favours the cultivation of different microbial species (Stefani et al., 2015). This issue can be addressed through the use of multiple and selective media, however, this represents an important and sometimes prohibitive cost in terms of time and labour. Second, the identification of microbial species by their morphological and physiological features requires cultivation for a sufficiently long time that these features (such as their metabolic profile) can emerge and be observed and characterized (Bertrand et al., 2015).

Figure 2.5: Pure cultures of (left to right) *Micrococcus luteus*, *Chromobacter violaceum*, and *Serratia marcascens*. Source: University of Wisconsin-Madison, Virtual Microbiology lab.

Culture-independent methods can address some of these issues, but bring their own drawbacks. One of the most frequently used culture-independent techniques for censusing microbial communities is 16S rRNA amplification (Little et al., 2008). This technique involves amplifying the genes from environmental samples using universal or specific primers, screening the resultant clones for differences in their sequences, and identifying microbial species based on these sequence differences (Sinclair et al., 2015). Since this technique can be applied for both culturable and non-culturable microbial species, it addresses one of the main issues with culture-based methods. However, it also suffers from an analogous issue to media in culture-based methods, namely the choice of primer. Universal primers, despite their name, may not detect all species in the community (Madigan et al., 2008). There is the additional problem of interspecies or horizontal gene transfer, which muddies the question of which microbe belongs to which species or taxa (Hellweger et al., 2016a). To circumvent rather than address this issue, researchers use the term OTU (operational taxonomic unit) to refer to "a group of phylogenetically related micro-organisms", without specifying their actual taxonomy (Bertrand et al., 2015).

The difference in the results that culture-based and culture-independent methods can produce was highlighted by a study that analysed the microbial communities in contaminated soils using both techniques (isolation and cultivation using seven different growth media and DNA pyrosequencing, respectively) (Stefani et al., 2015), representing one of the most comprehensive comparisons of microbial communities from polluted soils using these two approaches. The two resulting datasets only agreed for 2.4% of the bacterial OTUs and 8.2% of the fungal OTUs. This lack of agreement between the two main approaches available to researchers for censusing microbial communities is a powerful illustration of the obstacles facing microbial diversity quantification.

We also note the existence of sensitivity thresholds for the various techniques used to quantify diversity in microbial communities. These thresholds have led researchers to estimate that in practice microbial strains that are present in abun-

dances constituting less than 1% of the community will not be detected by sequencing techniques and therefore will be missing from the community census (Bertrand et al., 2015). Although efforts to remedy this obstacle are ongoing, it can imply that these censusing techniques are susceptible to overlooking rare species.

2.3.2 Understanding community functionality

After identifying the microbes present in a community, the next question typically asked is what these microbes are doing. Associating microbes with the processes they carry out refers to the *functionality* of the community, a multi-faceted aspect encompassing all forms of a community's behaviour, including its metabolic processes, interactions with its environment and responses to disturbances or perturbations (Ogunseitan, 2005). A community's ability to maintain its functionality in the face of such disturbance is referred to as its robustness (Stenuit and Agathos, 2015). Experimental studies interrogating issues of functionality typically address one of the components of robustness, which include a community's temporal stability (ability to maintain its structure over time), its resistance (ability to resist change following perturbations) or its resilience (ability to return to its original or previous structure after disturbances) (Little et al., 2008).

Until recent genetic advances, addressing these questions using culture-based methods has involved isolating microbes from natural communities, determining which strains are culturable and then, after cultivation, inferring from their substrate use the activity of each strain in the natural community (Ogunseitan, 2005). Culture-independent methods are necessary to determine the functional roles of non-culturable microbes, the most common of which is the use of metagenomics, comprising two broad types: sequence-based and functional metagenomics (Röling et al., 2010).

Sequence-based metagenomic methods involve the analysis of genomes from a community of microbes, via the extraction of DNA which is then cloned to a culturable host bacterium and subsequently analyzed, for example to determine the levels of richness or diversity present in the samples (Hanemaaijer et al., 2015). Functional metagenomic methods seek to associate genes with the different functions and processes being carried out in the community. To avoid the complications inherent in identifying gene function based only on sequencing, functional metagenomic techniques involve transfecting the host bacterium with genes that supply a certain function and screening for enzymatic activities, which only requires that the gene be expressed, and not necessarily recognizable by its sequence (Little et al., 2008). Thus a functional metagenomic approach can permit the identification of novel enzymes whose functions would not be recognized based on sequence alone (Lam et al., 2015). An example of such an approach is

the identification of various genes of significant interest for biocatalysis in industrial and pharmaceutical applications (Streit et al., 2004). Hence sequence-based metagenomic methods are suitable for studying the ecology and assembling the genome of a community without revealing the functions associated with these genes, whereas functional metagenomic methods can detect genes which produce functional enzymes, without shedding much light on which microbial species the genetic material actually originated from (Lam et al., 2015).

Additional insights into community functionality can be gained through single-cell analyses, which are targeted at the level of individual microbes rather than at the overall community level (De Roy, 2014). This approach takes into account the importance of individual variation between micro-organisms, since it has recently been affirmed that even in well-mixed environments, individual microbes that are genetically identical can still differ in their phenotypic characteristics (Kreft et al., 2013). This level of individual variation was not previously suspected, and provoked calls for a reconsideration of approaches to studying microbial communities (Hellweger et al., 2016a). Other important examples of individual variation include differences in growth rate and cell division capability between microbes of the same species (*Chlamydomonas reinhardtii*) (Damodaran et al., 2015). Techniques for single-cell analyses such as milli- and micro-fluidics, laser scanning and flow cytometry (Röling et al., 2010) are also useful for analyzing rare species, which can represent significant proportions of community richness (McGill et al., 2007) and perform key community functions (Piper et al., 2015) and, as mentioned in Section 2.3.1, are sometimes missed by censusing techniques due to their sensitivity limits.

2.3.3 Identifying interactions within the community

2.3.3.1 Techniques

Interactions between organisms have, along with metabolism and reproduction, been identified as “one of the most fundamental features of life” (Bertrand et al., 2015). Hence the interactions of microbes with each other and their environment are key to understanding community functionality, since these interactions drive fundamental processes such as metabolite transfer and growth inhibition, and also regulate the size, activity, diversity and productivity of the community (Tan et al., 2015).

Studying these various interactions between microbial populations in natural communities is complex due to their incredible diversity, which results in large numbers of species engaging in multiple different interactions with different partners, which may even be occurring simultaneously (Tan et al., 2015). Thus the fundamental basis of studies of these different types of interactions between cell popu-

lations are co-culture experiments. These are experimental set-ups where "two or more different populations of cells are grown with some degree of contact between them" (Goers et al., 2014). Among the possible motivations for this approach are: the study of interactions between cell populations in nature, the improvement of culturing success for cell populations, and establishing synthetic interactions between populations (Goers et al., 2014). Co-cultures are synthetic systems that have gained particular interest from microbiologists in recent years due to their reduced complexity and increased controllability, which favours them over more complex natural systems for examining ecological theories relating to microbial interactions, their mechanisms and effects (De Roy et al., 2014).

However, there are multiple issues with using co-cultures in the lab. It is not straightforward to determine under which conditions multiple cell types grow; different cell types have different optimal growth conditions, and there are few established protocols for determining how such cell types should be co-cultured (Sabra et al., 2010). If the conditions are not optimized for all constituent cell types, one cell type will typically dominate or outcompete the other cell types. One can then resort to the control of population levels, for which various techniques exist, including auxotrophic cross-feeding and toxin-antitoxin systems (Tanouchi et al., 2012). However, these methods do not allow precise and careful control of the different population ratios. Furthermore, they are not applicable to sufficiently many different cell types (Rollié et al., 2012).

Genetic tools can help to address the deficiencies of co-culture approaches. The essence of this technique for associating genes with functions is to construct random mutations in genetic code, and then search for the resulting mutants (De Roy, 2014). Assessing the phenotypic characteristics of these random mutants is considered to be the "minimally biased" approach to determine which genes are necessary for a certain function to occur (Little et al., 2008). Instead of using random mutations and hoping that these provoke a noticeable change in community functionality, targeted approaches are also possible, however, these so-called gene arrays necessitate some *a priori* knowledge about the genes in question (Madigan et al., 2008). An example of this approach is the identification of genes involved in quorum sensing in the gut of caterpillars (Borlee et al., 2008).

2.3.3.2 Interaction types

Interactions between organisms can be classified into three broad types. *Mutualism* refers to interactions where both individuals benefit, *commensalism* refers to interactions benefiting one individual and having no effect on the other, and *antagonism* refers to competitive interactions (Tanouchi et al., 2012). An extended description of all possible pairwise interactions can be found in the review of Faust and Raes (2012).

Mutualist, or cooperative, interactions facilitate the growth and survival of participating organisms, often through the production and use of shared "public goods", for example in co-cultures of *Pseudomonas aeruginosa* and *Burkholderia cepacia* which must produce molecules called siderophores in order to acquire iron from the environment (Hibbing et al., 2010). Cooperation at species or population level can involve for example the exchange of metabolites essential for growth processes, as was observed in synthetic communities of *Escherichia coli* strains (Wintermute and Silver, 2010).

Commensal interactions are less prevalent, and it has indeed been postulated that "purely commensal relationships may not exist" (Little et al., 2008), although it may instead be the case that it is actually the benefit to the second partner that has not yet been discovered (Brenner et al., 2008; Tan et al., 2015). For example, microflora in the human gut that were previously thought to be neutral partners in commensal interactions were later recognized to in fact be playing critical roles in the proper functioning of the gut microbiome, by initiating host immune defence mechanisms against infections, specifically by activating certain receptors critical for the protection of the gut microbiome against damages due to infection (Rakoff-Nahoum et al., 2004).

Competitive interactions may be the most common and commonly studied interactions occurring in microbial communities (Hibbing et al., 2010). Various forms of competitive interactions between micro-organisms have been observed and described. Perhaps the most common is *exploitative competition*, where microbes compete for a shared and limited resource, for example nutrients, light, water or space (Bertrand et al., 2015). This is an indirect form of competition, since the focus of the interaction is the shared resource. In contrast, *interference competition* constitutes direct warfare between microbes through the use of toxic compounds such as antibiotics (Hibbing et al., 2010). An additional form of interference competition is the disruption of signalling mechanisms, which can interfere with competitive or defensive actions and thus confer a competitive advantage to the disrupter, as has been suggested to be the case in competition between *Pseudomonas aeruginosa* and *Agrobacterium tumefaciens* in synthetic co-cultures (An et al., 2006). However, the importance and prevalence of this interaction mechanism is less established in comparison to the production of toxic compounds (Little et al., 2008).

Two further types of competitive interaction are predation and parasitism, which are similar in that they both involve one species benefiting from the interaction while the other species suffers. They are typically differentiated based on the time scale over which their effects are felt; predation occurs over a brief period, while parasitism continues over a significant period of time (Bertrand et al., 2015). Predation in particular has been singled out as a key stabilizing mechanism in macro-ecological communities, whereby the predator located at the top or apex of the food chain mediates the abundances of the species below it, a mechanism

that can have effects that are outsize compared to the predator population density (Parker and Kamenev, 2010; Saleem et al., 2012; Chu and Adler, 2015). In microbial settings, predation has been shown to be the main mortality mechanism in aquatic communities, and has also been proposed as a regulator of community richness and evenness (Zhang et al., 2007). Predation pressure of this type can also be exploited in synthetic communities as a form of population control (Kunin et al., 2008).

An example of predation in synthetic microbial communities can be found in a community of two *Escherichia coli* strains, where the predators induce the expression of a "suicide protein" in their prey, causing their death, while the predators require a signal from the prey microbes in order to produce a key protein (Balagaddé et al., 2008). Through experimental manipulation of this co-culture, the authors were able to retrieve the dynamics characteristic of predator-prey systems, such as extinction and coexistence. For an even finer classification of predatory competitive interactions, Martin (2002) classified various predation strategies according to their level of specialization, such as pack predation or direct invasion of the cytoplasm.

2.3.4 Making predictions at the community scale

Once a body of fundamental theories and knowledge in microbial ecology is present, this can be used as a solid foundation for predictive modelling techniques. This enables researchers to make predictions about communities that cannot be cultured or studied using current techniques, which as discussed in Section 2.3.1 constitutes the majority of natural communities. The fundamental knowledge required for such predictions lies in the areas outlined in the preceding sections: the composition, functionality and interaction network of the community (Little et al., 2008). This will be discussed in more detail in Section 3.2.4.

To gather the data, insights and knowledge necessary for an understanding of the fundamental ecological processes underlying functionality, synthetic communities have been singled out as particularly promising (Brenner et al., 2008; De Roy et al., 2014; Goers et al., 2014; Stenuit and Agathos, 2015).

2.4 The rise of synthetic microbial communities

2.4.1 Limitations of engineered pure cultures

When using natural microbial communities, the factors and associated mechanisms underlying a community's functionality often cannot be elucidated, and instead remain "black boxes" (Brenner et al., 2008). We may be able to enumerate and measure what goes into the community, and measure the corresponding community output or performance, but what occurs in between remains a mystery.

At the other extreme, there is an extensive and established knowledge base regarding the functionality of pure cultures (cultures containing only one microbial strain), which has been achieved through the use of genomic, transcriptomic, proteomic and metabolomic tools (Jessup et al., 2004). Once researchers were confident in their understanding of the functioning of these pure cultures, the desire grew to apply this knowledge by manipulating and controlling them as engineered pure cultures (De Roy et al., 2014).

Examples of the manipulation of pure cultures include the improvement of their resistance to stress and disturbance, the increase of productivity, the improvement of functionality via redundancy of key traits, the strengthening of toxin degradation capability, and the production of new or different compounds (Benner and Sismour, 2005). These engineered strains have obvious industrial, medical and pharmaceutical applications in the settings described in Section 2.2.

Thus while the study of pure cultures allows researchers to gather information on the genetic, physiological and morphological characteristics of specific and individual microbes (Leonard et al., 2008), they do not permit the study of any factors which influence the functionality of microbial communities. Additionally, only a very small fraction of microbial species are actually culturable (see Section 2.3.1), significantly limiting the representativeness and applicability of the knowledge gained through engineered pure culture studies. Furthermore, significant differences in the behaviour of microbial species have been observed when the species are cultivated in pure cultures compared to when they are embedded in a community (Jessup et al., 2004), pointing to the importance of interspecies interactions which cannot be accounted for in pure cultures and are ubiquitous in community settings (Stenuit and Agathos, 2015).

These factors, combined with the previously discussed importance of microbial communities for ecological, medical, industrial and research applications (Section 2.2) have motivated researchers to consider the engineering of communities rather than pure cultures. Thus techniques also shift to observing and characterizing behaviour in a top-down way, at the level of the community rather than at the

level of a species or strain (Stenuit and Agathos, 2015). Hence the increasing use of metagenomics (the genomic profile of the community), metatranscriptomics (the community transcript profile), and metaproteomics (community protein profile) (Röling et al., 2010) to describe the diversity, structure and composition of the community, and as well as the presence and level of expression of genes (Stenuit and Agathos, 2015).

Here we note the terminology used to describe this field, more specifically the term 'synthetic microbial ecology'. It is sometimes conflated or confused with synthetic biology, which is concerned with the engineering of cells (rather than communities), or with systems biology, which refers to the top-down approach of understanding a system through the characterization of its constituent components (Röling et al., 2010). In contrast, synthetic microbial ecology concerns the "rational design and theory-driven manipulation" of engineered artificial microbial ecosystems (Stenuit and Agathos, 2015).

2.4.2 Synthetic microbial ecology approaches

Synthetic microbial ecology involves the construction of synthetic microbial communities for the purpose of improving the understanding of fundamental microbial ecological principles and theories (Jessup et al., 2004), for example regarding interactions between microbes as well as with their environment, the relationship between diversity and functionality, the mechanisms of metabolic processes and many more (De Roy et al., 2014). This encompasses the design and construction of microbial communities with desired characteristics and functionality, whether for practical or research applications. These communities are designed and constructed "bottom-up" by assembling at least two microbial populations in properly characterized environmental conditions which are controlled by the researcher (Tan et al., 2015).

For the purpose of studying microbial ecological theories, synthetic communities can be seen as a midpoint between mathematical models and natural communities (Song et al., 2014). Synthetic communities are closed systems which can avoid the complex and confounding background processes present in natural communities, but they are also closer to biological reality than mathematical models (Tanouchi et al., 2012). They are necessarily much less complex than natural communities, but it is exactly this comparative simplicity that allows researchers to control and replicate synthetic communities to a sufficient degree as to permit the scientifically sound study of theories and questions that cannot be addressed through observation of natural communities or pure culture experiments (De Roy et al., 2014). Criticisms of their simplicity compared to natural communities misunderstand the purpose of synthetic communities in this endeavour: "*to simplify nature so that it can be more easily understood*" (Jessup et al., 2004).

Synthetic communities also allow researchers to plan ahead and develop techniques to differentiate the different microbial species prior to their culturing, or to ensure that these species can coexist (Tan et al., 2015), hence accounting for significant issues related to culture-based methods (see Section 2.3.1). An example of an important tool in the former case is the use of fluorescent tagging, achieved through the use of green fluorescent protein (GFP). From the expression of GFP, researchers are alerted that the gene linked to GFP has also been expressed, permitting differentiation and identification of different strains (Madigan et al., 2008). In the latter case, microfluidic approaches allow researchers to precisely control the spatial organization of community members, as well as determine their environmental conditions (Stewart, 2012).

Synthetic communities can and have been used to test various theories regarding the factors which affect the behaviour and functionality of natural microbial communities. Only by understanding these fundamental factors can we gain sufficient insight into microbial communities to be able to engineer and steer them, and synthetic microbial ecology is well suited for obtaining these insights. We provide several examples here.

First, synthetic communities can be employed to study the interactions between microbes as discussed in Section 2.3.3, for example by genetically engineering strains in order to ensure the presence of desired interactions and facilitate the tuning of the parameter(s) mediating the interaction (Wintermute and Silver, 2010), which would be far more complex (perhaps prohibitively so) in natural communities due to the presence of other confounding interactions. Synthetic ecology approaches also permit the manipulation of environmental conditions in order to induce or control interactions (Klitgord and Segre, 2011). For example, it has been shown that cooperation can always be induced in all possible pairs of seven microbial species (*Escherichia coli*, *Helicobacter pylori*, *Salmonella typhimurium*, *Bacillus subtilis*, *Shewanella oneidensis*, *Methylobacterium extorquens*, and *Methanosarcina barkeri*), through the appropriate environmental manipulation, more specifically by altering the composition of the media (Klitgord and Segre, 2010).

Second, synthetic communities have been designed to study resilience, which is the presence of functionally redundant microbes that help the community to resist stress and disturbances by ensuring that key functions can continue even in the case of stress-induced extinctions (Vannecke, 2015). For example, it has been shown that in the face of changing community composition due to environmental stress, communities can still maintain stable functionality in terms of their productive output, due to the presence of functionally redundant genes (Kraft et al., 2014). Synthetic community experiments have been used to study this type of biodiversity-productivity relationship in the face of different environmental stresses such as temperature, pH or salinity (De Roy et al., 2014). For example, Wittebolle *et al.* (2009) constructed over 1.000 synthetic communities using the same 18 strains with varying evenness, and investigated the productivity and

functionality of the community when confronted with salinity stress. Their results showed that highly uneven communities were less able to resist stress than more even communities, which could maintain their functionality under the stress conditions. A similar positive relationship with functionality was also demonstrated for richness, using synthetic communities with richness varying up to 72 species (Bell et al., 2005). We will return to the topic of the relationship between functionality, diversity, richness and evenness later in Section 2.5.

A conceptually related study focused on the link between functionality and community structure (rather than richness). Yu *et al.* 2016 studied microbial communities involved in methane oxidation by constructing synthetic communities of 50 species, in order to understand the main drivers of functionality in such communities. This experiment also served as a test of the extent of the differences between natural communities and the synthetic communities constructed to represent them, a topic that is the subject of ongoing discussion in synthetic microbial ecology (Ponomarova and Patil, 2015). Yu *et al.* observed commonalities between the natural and synthetic community dynamics, but also noted significant differences in the identities of the key species (in terms of functionality) in the synthetic communities compared to the natural communities, despite similar environmental conditions. Hence the authors advocate for "intelligent community design" when constructing synthetic communities, rather than using a a random sampling of species, in order to more realistically represent the natural community (Yu et al., 2016).

Third, synthetic communities are well suited for studying the effects of spatial structure, due to their high controllability and bottom-up construction. Techniques for inducing defined spatial structures in synthetic communities by restricting mobility include the use of solid media, inducing the establishment of biofilms, or the use of microfluidics (Tan et al., 2015). A key study in this context was carried out by Kerr *et al.* (2002), who used a synthetic community of three *Escherichia coli* strains. Plating of this community on sold media constrained the mobility of the microbes such that stable spatial structures formed, which permitted coexistence of all three species by localizing their interactions. In contrast, when the same community was cultured in a well-mixed environment, no spatial structure could be induced, which negated the possibility of coexistence and invariably led to extinctions. The same effect was observed through the use of microfluidics, when a different synthetic community of three species (*Azotobacter vinelandii*, *Bacillus licheniformis*, and *Paenibacillus curdlanolyticus*) was cultured in a well-mixed setting and a spatially structured environment (Kim et al., 2008). Once again, only the spatially structured environment could support the coexistence of all three species by localizing their interactions. Furthermore, the microfluidic device (pictured in Figure 2.6) allowed the authors to vary the magnitude of spatial separation between the different species, permitting the study of a spectrum of localization rather than just the two extremes of solid media and well-mixed en-

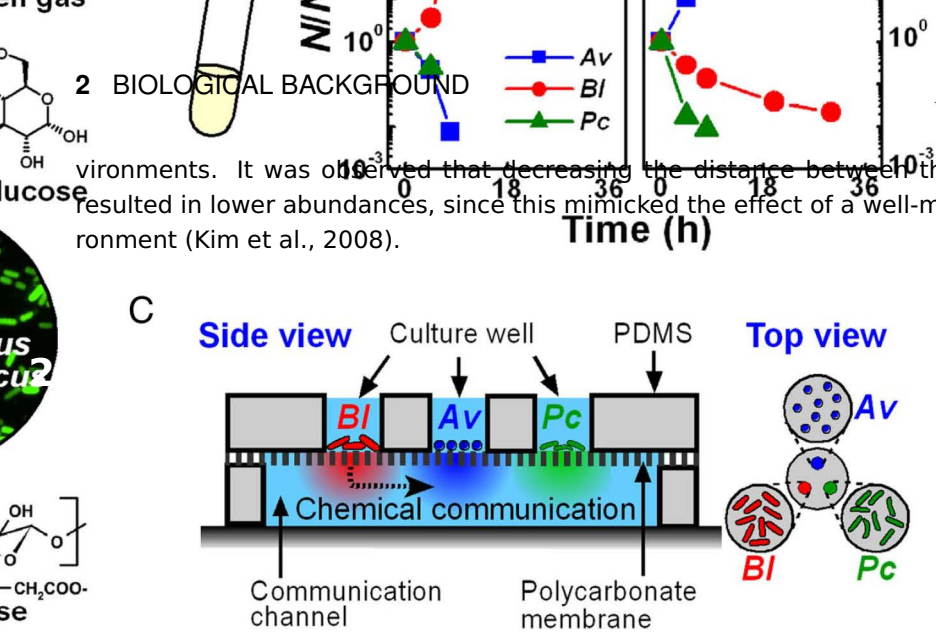


Figure 2.6: Microfluidic device used by Kim *et al.* (2008) to control the spatial structure of a synthetic community with syntrophic interactions. (*B*) Graphs show the survival ratio of each species (N/N_0) as a test tube in nutrient-rich TSB/1771 (*Left*) and nutrient-poor CP (*Right*) media, indicating instability (*C*) A schematic drawing of the microfluidic device used to co-culture the three species stably by communication channel.

ires spatial structure to maintain stable coexistence. (*A*) A schematic drawing of the wild-type soil unity with syntrophic interactions. (*B*) Graphs show the survival ratio of each species (N/N_0) as a est tube in nutrient-rich TSB/1771 (*Left*) and nutrient-poor CP (*Right*) media, indicating instability C) A schematic drawing of the microfluidic device used to co-culture the three species stably by mmunication channel.

Of these different key factors which drive the functionality of microbial communities, we will focus on several in particular: biodiversity, interspecies interactions (in particular competition), and spatial structure. The latter two will be discussed in more detail in Chapter 5, while we now delve deeper into the (microbial) ecological importance of biodiversity.

2.5 Biodiversity: a key factor in community functioning

One of the most important aims in ecology is to identify and comprehend the mechanisms that sustain biodiversity, which is often critically important for the stability of ecosystems (Tilman et al., 2014). Loss of biodiversity can hinder the proper functioning of ecosystems, in drastic cases even leading to mass extinctions and system collapse (Zeeman and Jonsson, 2005). Recently, the nearly unanimous scientific belief has formed that a significant portion of the Earth's species are currently being driven to extinction due to human actions, threatening biodiversity across a wide number of ecosystems (Wilcove et al., 1998). Thus the issue of biodiversity loss has taken on a particular urgency in recent years (Mendes et al., 2008).

Many theoretical models have shown that species in competition can coexist — and hence maintain the ecosystem's biodiversity — if ecological processes such as competition and movement take place over small spatial scales (Laird and

MICROBIOLOGY

CHEMISTRY

Schamp, 2008; Nadell and Bassler, 2011; Frey and Reichenbach, 2011; Adamson and Morozov, 2012; Borenstein et al., 2013). This is also true in the case of communities with non-transitive competition between species — that is, communities where a strict competitive hierarchy does not exist (Laird and Schamp, 2006). The classic example of non-transitive competition in a community of three species is the rock-paper-scissors configuration, also known as cyclic competition. There exist many examples of communities in nature that demonstrate this type of competition, such as invertebrates living in coral reefs (Jackson and Buss, 1975), Arctic lemmings (*Dicrostonyx torquatus*) in Greenland (Gilg et al., 2003), side-blotched lizards (*Uta stansburiana*) in California (Sinervo and Lively, 1996), the Pacific salmon (*Oncorhynchus nerka*) (Guill et al., 2011), certain bacterial species engaging in antibiotic production (Reichenbach et al., 2007), microbial populations of colicinogenic *Escherichia coli* (Kerr et al., 2002), and communities of cryptic species of the nematode *Litoditis marina* (De Meester et al., 2016)

From among the various ecosystems in which cyclic competition has been observed, microbial communities have become model systems for studying the complex interplay between the nonlinear dynamics of evolutionary games, stochastic fluctuations arising from the probabilistic nature of interactions, and spatial organization (Frey and Reichenbach, 2011). Much effort has been dedicated to advancing the qualitative and quantitative understanding of mechanisms that sustain biodiversity and ensure the viability of microbial communities, by allowing for species diversity and social behaviour such as cooperation.

Biological variability between micro-organisms is caused by selection pressure due to environmental heterogeneities (physical, chemical or biological) in conjunction with genomic mutations. Many of these processes can occur over time scales unsuitable for *in vitro* study (Bertrand et al., 2015), complicating efforts to elucidate their functioning. A further complication lies in the incredible diversity of microbial communities (Tan et al., 2015). A single gram of soil has been estimated to contain upwards of hundreds of thousands of bacterial OTUs (van Elsas et al., 2012), which is orders of magnitudes richer than for example the 300.000 plant species estimated to be present on the entire planet (Villenave et al., 2011).

The general approach in such cases is to simply compare different experimental settings and search for correlation between biodiversity and functionality, for example after some disturbance or perturbation to the community (Matthiessen et al., 2010). Such experiments often reveal little or no relationship between biodiversity and functionality, but there have been significant exceptions (Bertrand et al., 2015). A clue to the reasons behind this can be found in recent studies showing that functional changes in microbial communities can be less strongly linked to community diversity than they are to *components* of diversity such as the relative abundances of dominant species (Patra et al., 2006; Attard et al., 2011). For example, it was shown that the methane oxidation functionality of a mixed community of heterotrophs and methanotrophs was stimulated not only by

methanotroph richness, but also by heterotroph richness (Ho et al., 2014). Here the authors were able to untangle the confounding factors to demonstrate that increased heterotroph richness by itself led to increased functionality, however the mechanism underlying this stimulation remains unclear. An additional confounding factor in such diversity-functionality studies in natural communities is the co-variation of environmental factors which may be difficult to disentangle from variation in diversity levels (Bertrand et al., 2015).

Synthetic communities allow for a different approach, since they permit researchers to assemble communities "bottom-up" in order to obtain a specific diversity, rather than working with whatever level of diversity happens to be found in a natural community. This additional manipulation and hence controllability allows the analysis of (possible) causal relationships between functionality and diversity to be investigated in communities with for example equal population abundances of key dominant species (Hellweger et al., 2016a). Such studies generally reveal a positive relationship between community diversity and functionality (Bertrand et al., 2015), but an important caveat is their significantly reduced richness compared to natural communities.

2.6 The components of biodiversity

Richness and evenness, the components of diversity, play different roles in community functioning and have therefore often been treated separately (Hillebrand et al., 2008). Species richness refers to the absolute number of species present in the population of interest, while species evenness refers to the relative abundances of the different species, so that a population is described as completely even if all species are equally abundant (Heip, 1974). Species richness is responsible for the number of functional traits in a community, while evenness may influence the richness effect by controlling the variation of traits present in the community (Lemieux and Cusson, 2014).

Species evenness has been shown to be a key factor in preserving the functional stability of ecosystems (Hillebrand et al., 2008; Wittebolle et al., 2009; De Roy et al., 2013). Evenness is also known to have a positive impact on productivity by increasing the representation of each species' functional traits (Lemieux and Cusson, 2014). Despite this recognition, studies and conservation efforts often focus on restoring or maintaining richness, since the impact of richness on many ecological processes has been well described, see e.g. Crowder *et al.* (2010), Hooper *et al.* (2005), and Isbell *et al.* (2009b). In contrast, much less attention has been paid to the ecological effects of disrupted evenness (Hillebrand et al., 2008), an unfortunate oversight since environmental degradation and damage due to human actions can skew the relative abundance of species, and because uneven

communities are often more susceptible to invasion and less resilient to stresses and disturbances (Wittebolle et al., 2009).

Declining evenness has also been shown in field studies to be an important early warning sign of diversity decline, specifically in response to species invasion (Wilsey et al., 2009). The authors noted that if they had taken species richness as the only index of diversity, they would have falsely concluded that diversity had not changed over the first year of study. Initial drops in evenness preceded the drops in richness that occurred in the second year, an effect the authors singled out as interesting for future study (Wilsey et al., 2009).

In general, the development of theory has outpaced experimental studies concerning evenness as a mechanism promoting maintenance of biodiversity (Isbell et al., 2009b). Empirical evidence was provided by one study showing that maintenance of biodiversity can be promoted by a rare species advantage (Wills et al., 2006) or a common species disadvantage (Harpole and Suding, 2007), both of which are mechanisms leading to a more even community.

As a further example, one experimental study set in a field in Canada varied species evenness and the identity of the dominant plant species in order to test whether plant productivity would increase with increasing evenness, and whether such a relationship would be dependent on species identity (Wilsey and Potvin, 2000). Results showed that biomass production increased linearly with increasing evenness, and was invariant of the identity of the dominant species. These results support the view that a decrease in plant diversity due to human actions would lead to an indirect decrease in productivity (Hillebrand et al., 2008).

Further evidence for the importance of evenness can be found in a global meta-analysis of 54 studies regarding the diversity-productivity relationship in forest ecosystems (Zhang et al., 2012). The authors concluded that the strong positive effects on productivity due to increased evenness provide strong empirical evidence to support the theoretical assertion that evenness affects the relative strength of interspecific and intraspecific interactions within communities, therefore causing an appreciable shift in the diversity-productivity relationship (Zhang et al., 2012). Furthermore, the authors suggest that the lack of attention paid to evenness effects in previous empirical studies can be attributed to the limited levels of evenness found in those experiments, where typically only high and “realistically low” levels of evenness were included; testing so few evenness conditions is unlikely to unveil sufficiently significant behaviour (Polley et al., 2003; Isbell et al., 2009a).

Although ecological studies have only recently begun to examine the mechanisms underlying such evenness effects, studies so far suggest that many of the same processes underlying the impacts of species richness may be at work (Hillebrand et al., 2008). For instance, evenness in bacterial communities promotes resilience to disturbance by ensuring sufficient densities of species in key functional

roles (Wittebolle et al., 2009). This is akin to the “insurance effect” described in the species richness literature (Lemieux and Cusson, 2014). One possibility is that decreasing evenness leads to increasingly underused niches that become fully vacant once species are lost, to the detriment of the ecosystem’s proper functioning (Crowder et al., 2010).

We finally note that diversity has been shown to be partitionable in an additional manner than its decomposition into richness and evenness. It can also be partitioned into alpha diversity and beta diversity (Chao et al., 2012). Alpha diversity refers to the diversity within a community (sometimes called small-scale diversity), while beta diversity refers to the diversity between different communities (sometimes called large-scale diversity) (Jost, 2007). However, this differentiation between small and large-scale diversity is rarely used in microbial settings, since the extreme levels of heterogeneity in micro-environments such as soil make it difficult to define the “local” environment (Bertrand et al., 2015). This is in addition to the difficulty in delineating microbial species due among other factors to frequent horizontal gene transfers (discussed in more detail in Section 4.2.6).

2.7 Conclusions

Microbial communities play vital roles in key ecological, geochemical, medical and industrial processes. Their incredible versatility and functionality have motivated researchers to mimic them by constructing synthetic microbial communities. These engineered communities are not only useful for practical applications in medicine and industry, but can also be used to study fundamental microbial ecological theories and principles that can be difficult to address using natural communities. The knowledge gained through such synthetic microbial ecology studies not only helps researchers to maintain and manage natural communities, and to preserve their vital functions in the face of climate change, but can also be used to carefully design and construct engineered communities with desired characteristics and functions.

While synthetic microbial ecology has already proved its undoubted potential in these areas, it is still a young field and there is much more to be accomplished. A current major drawback is the significantly reduced richness of synthetic communities compared to natural communities. Most synthetic communities contain four species or fewer (De Roy et al., 2014). This limits the applicability of the insights, knowledge and theories developed through their use in synthetic microbial ecology studies, as well as their potential practical applications. This limitation of synthetic communities as model systems is not unique to microbial ecology, but rather is common to all ecological fields where model systems are studied in order to gain insights which are then extrapolated to full-scale ecosystems (Jessup et al.,

2004). This extrapolation has its limits, which requires careful consideration of the effects of scaling up from model systems.

This has motivated calls for a ramping up of the complexity of synthetic communities in order to further bridge the gap to natural communities (Tan et al., 2015). To aid this effort, researchers have highlighted the need to complement synthetic microbial ecology studies with modelling studies, namely by developing mathematical models that can represent the *in silico* counterparts of *in vitro* synthetic communities (Stenuit and Agathos, 2015). This endeavour, and its potential for furthering the already substantial progress in synthetic microbial ecology, forms the discussion in the subsequent chapter.

3

Modelling background

3.1 Introduction

Mathematical models are abstractions of reality which seek to mimic certain behaviours or dynamics of a natural system through their mathematical description (Song et al., 2014). Models can help to further our understanding of the fundamental mechanisms driving microbial community dynamics, as well as aiding in the development of new hypotheses by highlighting interesting or unexpected behaviour. When constructed based on ecological theories, models can not only help to understand the fundamental processes underlying these theories, but also to predict under which conditions these theories may no longer hold, or under which conditions these theories are particularly key to community functioning (Klimenkoa et al., 2016).

Furthermore, if the appropriate data is available then models can be calibrated and validated in order to make testable quantitative predictions about community stability and functionality, which can then be verified using *in vitro* experiments. With recent high-throughput technological advances, particularly related to the

development of -omics tools and techniques, the technology and tools available to researchers is now sufficiently sophisticated to permit the collection of the type of data needed for this purpose (De Roy et al., 2014), increasing even further the potential of mathematical modelling for the purpose of microbial ecology theory development.

Such ecological modelling approaches are well established in the macro-ecology literature (Lewis et al., 2016) but is more complex in microbial ecology. In particular, certain microbial traits and characteristics cannot be found, and are not possible, in macro-scale communities (Bewick et al., 2017). An example is the capacity of microbes to interact over long distances, including competitive interactions using toxic compounds that may diffuse over long distances (Hibbing et al., 2010), and cooperative interactions between community members using mechanisms such as quorum sensing (Pérez-Velázquez et al., 2016). Thus while many fundamental behaviours are similar to those found in macro-ecological communities, and allow for the application of macro-scale ecological modelling knowledge and tools, the distinctive characteristics of microbial organisms also require new and different modelling approaches to more fully capture the dynamics of microbial communities (Bewick et al., 2017).

Therefore in this chapter we gather and synthesize existing modelling knowledge and techniques that are relevant to our study of microbial ecological theories. This chapter is structured as follows. In Section 3.2 we outline the general approaches that have been used to model microbial communities, structured in terms of the different possible basic modelling unit: communities (Section 3.2.1), populations (Section 3.2.2), or individuals (Section 3.2.3). We then focus our survey on the modelling approaches that have been developed specifically to study functionality effects due to community spatial structure and dynamics (Section 3.3) and the interactions taking place within the community (Section 3.4), in particular the interactions occurring with the environment (Section 3.4.1) and between individuals (Section 3.4.2). Finally in Section 3.5 we summarize the gaps remaining in our knowledge of the fundamental mechanisms and processes underlying these features, as well as the suitability of IBMs for addressing these open questions.

3.2 Modelling scales for microbial communities

The modelling approaches typically used to study microbial communities can be grouped according to the basic unit with which they are constructed, in a similar way to macro-ecological models (Lewis et al., 2016). Models seek to capture the characteristics of these different units, as well as the interactions between them and their environment. In broad terms, microbial communities can be modelled

using a basic unit of: (i) a community, (ii) a population, or (iii) an individual.

3.2.1 Communities

At the coarsest scale, several approaches exist to model at the level of entire communities. From this perspective, a microbial community is no longer a consortium of species, but rather a single "super-organism" comprising various genes and reactions (Song et al., 2014). Hence the interactions being captured in such models are not between microbial species or strains, but rather between genes and/or reactions.

The focus of super-organism models is the community's metabolic network, which encapsulates its metabolites, their transport, and relevant intracellular reactions (Orth et al., 2010). The network is typically represented by a list of mass balance equations for the metabolites, which essentially makes an accounting of all inputs and outputs from the network to determine the net gain or loss of a particular metabolite (Hanemaaijer et al., 2015). When combined with appropriate flux boundary conditions, this system of equations is called a stoichiometric model. When using stoichiometric modelling with the super-organism approach, the metabolic network is constructed for the community as a whole (Greenblum et al., 2012).

The main advantage of the super-organism approach is that established methods developed for single-species models are easily applicable (Borenstein, 2012). Such tools include the vast array of genome-based single-species metabolic models which are used to infer the metabolic functionality of the species (Reed and Palsson, 2003), and the "reverse-ecology" framework which aims to develop computational tools for analyzing genome-scale models, in order to permit the characterization of the natural habitat of microbial species, as well as the prediction of the interactions between these species and their environments (Levy and Borenstein, 2012).

An example of the use of community-level modelling to study the functionality of a microbial community is an *in silico* study of the human gut microbiome by Greenblum *et al.* (2012). Using a database of genes and genomes, the authors used metagenomic methods to identify enzymes that were then used to construct the metabolic networks of the community under different environmental conditions, providing insights into the metabolic functionality and stability of the community as a whole.

The main drawback of the community-level approach is that it is unable to provide insights into the population dynamics of the community, since it is by design blind to these different populations, but considers them as one aggregated whole. Hence the structure and composition of the community, as well as the interactions occurring within it, are all neglected. All of these aspects have been highlighted

as important mechanisms in steering community stability and functionality (see Section 2.3.2).

An illustration of the implications of this coarse-grained modelling approach can be found in the concept of keystone species. These are species which are considered crucial for maintaining the functionality and stability of a community, and whose loss typically provokes chains of extinction that in severe cases may lead to total ecosystem collapse (Ebenman and Jonsson, 2005). Community-level models constructed using the super-organism approach may be able to highlight a gene or reaction that is important for the functionality or stability of the community, but by design are unable to identify which particular member of the community is responsible for this critical gene or reaction (Widder et al., 2016). Hence the mechanisms underlying their central role in the community dynamics cannot be elucidated with a community-level modelling approach.

3.2.2 Populations

3.2.2.1 Modelling approach

The next scale available to modellers is at the level of populations, where each population represents a microbial species or taxa, and the model mimics the dynamics of these populations as well as the interactions between them. Hence these population-level models (PLMs) directly model changes in populations, and therefore assume that stochasticity at the individual level can be averaged into a deterministic population-level effect (Zomorrodi and Segrè, 2016). PLM studies in microbial ecology are generally concerned with how the interactions between species, whether they are direct (e.g. competition or cooperation) or indirect (e.g. occurring through the environment), affect the structure, stability and/or functionality of microbial communities (Wade et al., 2016).

PLMs can take the form of difference equations if time is considered discretely, or ordinary differential equations (ODEs) if time is considered continuously (Song et al., 2014). These two approaches track the evolution of the species fractions through time. If the spatial nature of the environment is also taken into account, partial differential equations (PDEs) are typically used. These track the evolution of population densities (biomass per unit area) through time (Zomorrodi and Segrè, 2016).

Generally, PLMs have been described as "*strategic models made to be as simple as possible to reveal general explanations*", and are thus excellent choices for studies aiming to determine general theories applicable to a wide range of organisms or ecosystems (Hellweger et al., 2016a). Hence they are well suited for the type of theory development studies central to synthetic microbial ecology (Prosser et al.,

2007), and for which purpose they have been used with great success in other fields, such as macro-ecology and epidemiology (Adamson and Morozov, 2012).

PLMs are less suitable for prediction purposes, due to their general nature. Synthetic microbial ecology in particular is concerned with the steering of engineered and natural microbial communities, which requires predictions regarding the behaviour of specific communities under specific environmental conditions (De Roy et al., 2014).

On the other hand, the use of population-level models for theory development is already well established in microbial ecology, and is increasing still further due to advances in metagenomic tools that allow researchers to determine the composition of a microbial community, as well as to measure the abundances of the constituent species (Song et al., 2014). These data allow researchers to gain more insights into the population dynamics, so that they can construct models that are both more complex and more realistic. We now outline a few of the most commonly used PLM approaches in this setting.

3.2.2.2 Examples of PLMs in microbial ecology

Spatially implicit PLMs

The most commonly used ODE-based approach is the Lotka-Volterra predator-prey model (Lotka, 1925; Volterra, 1926). This model is a system of coupled ODEs, and is often used in microbial ecology to investigate the population dynamics resulting from competitive and mutualistic interactions between microbial populations (Zomorrodi and Segrè, 2016). The model was originally proposed for a community of two species (one predator and one prey), but can be generalized to an arbitrary number of species, as well as being able to mimic interactions other than predation.

The generalized Lotka-Volterra model represents the population dynamics of species i using the following formulation:

$$\frac{dx_i}{dt} = x_i\left(\mu_i + \sum_{j=1}^{S} a_{ij}x_j\right), \tag{3.1}$$

for $i = 1, \ldots, S$, where μ_i is the growth rate of species i, and a_{ij} is the interaction coefficient for species i and species j (Lotka, 1925; Volterra, 1926). Depending on the type of interaction occurring between the two species, this coefficient can be positive, negative or zero (Faust and Raes, 2012).

Thus the generalized Lotka-Volterra equation directly models the interactions between populations in terms of the effect of one population on the growth of an-

other. However, it does not account for indirect interactions such as exchanges of metabolites or quorum sensing (Hibbing et al., 2010).

Examples of the use of the generalized Lotka-Volterra model in microbial ecology studies include the modelling of: microbial interactions in the human gut (Stein et al., 2013); the competitive interactions between harmful pathogens and the resident communities found in pork products (Cornu et al., 2011); and a study which used simple theories about community assembly to predict the structure of synthetic microbial communities of eight species (Friedman et al., 2017). In the latter study, the generalized Lotka-Volterra model was used to simulate pairwise competitive interactions, to understand why some resulted in coexistence of both species, and others resulted in the competitive exclusion of one species (Figure 3.1).

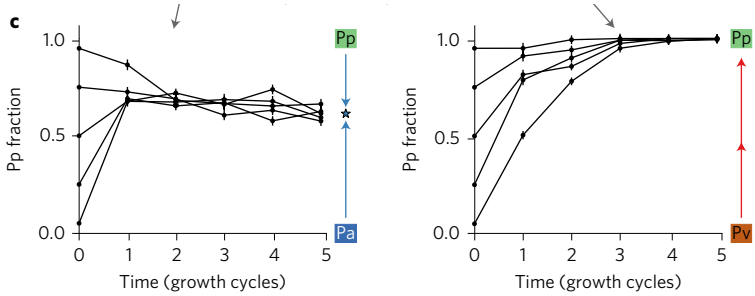


Figure 3.1: Friedman *et al.* (2017) used a generalized Lotka-Volterra model to study community assembly rules in synthetic microbial communities. This figure shows the outcomes of pairwise competitions, by plotting the evolving population fraction of one of the competing pair, which either stabilized at an intermediate value indicating coexistence (left) or increased until the second species was excluded (right), resulting in monoculture.

In the study of Stein et al. (2013) of the human gut microbiome, which focused on the structural population dynamics of the community, the authors extended the generalized Lotka-Volterra model by including additional terms that describe the effect of environmental perturbations. Their ODEs then took the form:

$$\frac{dx_i}{dt} = x_i\left(\mu_i + \sum_{j=1}^{S} a_{ij}x_j + \sum_{k=1}^{S} b_{i,k}c_k\right), \tag{3.2}$$

where the second sum on the right hand side represents the effects of environmental perturbations on species i. Stability analysis of the extended model allowed the authors to explain how the stability of the gut community could be threatened by external perturbations, even for a significant period after the removal of the perturbations (Stein et al., 2013). This extension illustrates how ODEs, and the Lotka-Volterra model in particular, can be modified to simulate other factors that may influence the population dynamics, such as environmental variations or perturbations.

Spatially explicit PLMs

Spatially explicit PLMs, which as their name suggests explicitly account for space in their representations of natural ecosystems, are used in settings where the spatial dynamics are known or suspected to play a key role in mediating the population dynamics. In Section 3.3 we will discuss in more detail why this distinction is important, and the underlying mechanisms driving it.

An example of a setting in microbial ecology where spatial considerations are important is the study of biofilms, since the population growth dynamics of these structured communities (both natural and synthetic) can depend very strongly on the spatial distributions of the microbial species and the environmental substrate(s) (Wang and Zhang, 2010). PDEs are typically used to model the population dynamics of the species in a biofilm community. These equations track the biomass density (or volume density in the case of a three-dimensional model) of the different species through time, and can account for processes such as reproduction, dispersal, attachment to and detachment from the biofilm (shown in Figure 3.2).

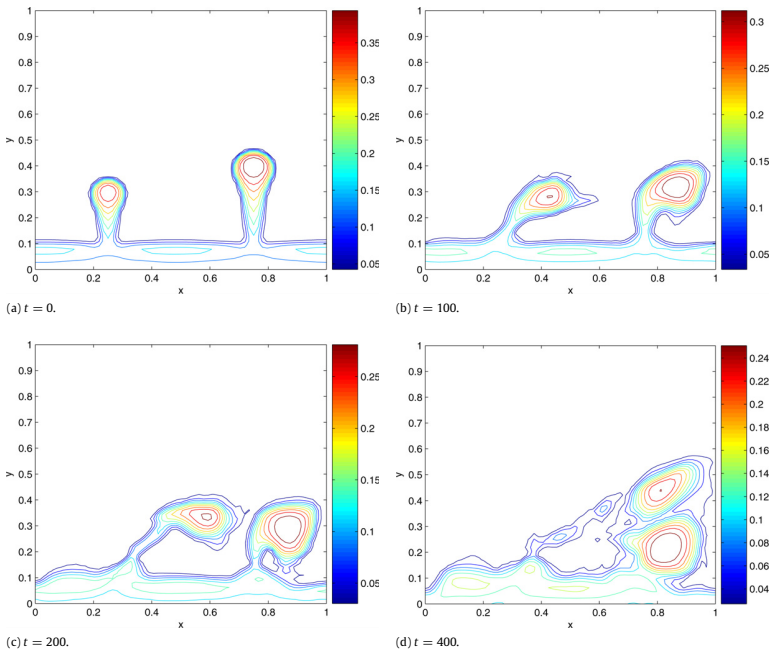


Figure 3.2: Two-dimensional growth and detachment of two mushroom-shaped biofilms, modelled using a PDE approach and shown at four different time points (Wang and Zhang, 2010). The colour bars indicate biomass density.

An example of a simple PDE model of a biofilm is given by:

$$\frac{\partial f_i}{\partial t} = \mu_i f_i - \frac{1}{\rho_i}\frac{\partial g_i}{\partial t} \tag{3.3}$$

where f_i is the volume fraction of species i, μ_i is its growth rate, ρ_i is its constant density, and g_i is the mass flux of species i, which means the biomass of species i displaced per unit time and area (Wanner and Gujer, 1986). This one-dimensional model can be used to study the biofilm's steady-state growth dynamics, such as its structure (thickness, etc.) and the spatial distributions of the microbial species as well as the substrate concentrations.

An extensive body of work exists relating to the modelling of biofilms, and the spatially explicit PLM approaches which have been developed are able to reproduce and explain some of the complex phenomena related to the formation, structure and functionality of biofilms (D'Acunto et al., 2015). Important questions still remain unanswered, such as the effects of incorporating the numerous physical, chemical, biological and ecological processes occurring in a biofilm in a unified analytical or computational model (Wang and Zhang, 2010).

Spatially explicit PLMs, particularly PDE-based models, are also often used in microbial invasion studies (Bewick et al., 2017), since spatial factors can play important roles in the success or failure of invasions. We will discuss the modelling of microbial invasions in more detail in Section 3.4.2.3.

Another example of the use of PDE models in microbial ecology is a study of the functional resilience of communities (König et al., 2017). We recall from Section 2.3.2 that the resilience of a community refers to its ability to return to its original or previous structure after disturbances, and thereby maintain or preserve its functionality. This can be particularly important in synthetic microbial communities, for example if the community has been engineered for biodegradation purposes. In this case, it is important that the community can maintain its biodegrading functions even in the face of disturbances or perturbations, since this could otherwise lead to the failure of the wider bioremediation process (Poggiale et al., 2015).

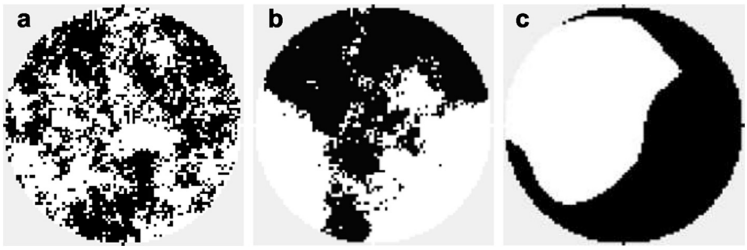


Figure 3.3: Examples of disturbance patterns in an *in silico* microbial community simulated using PDEs, with (a) high, (b) moderate, and (c) no fragmentation (black: disturbed area, white: undisturbed area) (König et al., 2017). The authors used such simulation scenarios to assess biodegradation performance in the face of disturbance, in order to study the system's functional resilience.

In their study of functional resilience, König *et al.* (2017) used a PDE approach, namely reaction-diffusion equations, to simulate the community's population growth dynamics and degradation of substrate. The model incorporates the processes of: substrate uptake by bacteria, uptake allocation, bacterial dispersal, growth, and substrate diffusion. By implementing various scenarios representing different disturbances, the authors could study the spatiotemporal dynamics of the community's recovery in terms of its functionality (Figure 3.3), and found that different local environments were responsible for different phases of the community's functional recovery. These results suggest that spatial dynamics are crucial for the maintenance of biodegradation functionality when the community is confronted with disturbances, and are a good example of the insights into fundamental community processes that can be gained using PLM approaches.

Disadvantages

It is clear that PLMs are well established and undeniably useful, and hence an extensive knowledge base exists for their construction and analysis. However, they have several drawbacks that are especially significant for microbial communities.

First, PLMs consider averages at the population-level of characteristics such as growth rate, and hence ignore the variability and heterogeneity recently discovered to exist between individuals of the same genotype in the same environment (even when well-mixed), which has lately prompted much discussion about reorienting microbial modelling approaches (Kreft et al., 2013). These variations are lost when considering population averages, which can have serious implications for the validity of the modelling approach if these variations are sufficiently extensive or significant, as we shall see in Section 3.2.3.

Second, PLMs cannot capture interactions at the scale of individuals, which are important in spatially structured environments where individual microbes can only interact locally (Daly et al., 2016). The extent to which these interactions are

localised can be crucially important for the stability of the community dynamics, as will be discussed in more detail in Section 3.3, and cannot be captured by PLM approaches.

Third, PLMs cannot account for adaptation, a key and ubiquitous microbial process (Bertrand et al., 2015). It has been noted that "*practically everything [microbes] do is in response to their environment*" (Hellweger et al., 2016a), but PLMs cannot capture these important responses to changes both internal and external to the community at the level of the individual microbes who are driving this process.

Overlooking these three features — intraspecies variability, localized interactions, and adaptive processes — not only affects how realistic a model can be in relation to the natural community it seeks to represent, but also reduces the ability of the model to help researchers understand the roles of these features in community stability and functionality. This is particularly relevant to synthetic microbial ecology, where a key aim is to achieve a deeper understanding of the fundamental processes driving and maintaining community stability and functionality. In this context, processes occurring at the individual level, of which the principals are the three we have highlighted, are increasingly recognized as vitally important.

These deficiencies in the PLM approach can all be addressed by moving to a finer modelling scale, namely the scale of individuals.

3.2.3 Individuals

3.2.3.1 Modelling approach

Individual-based models (IBMs) track through time the characteristics, activities and interactions of each and every individual within a community (Hellweger et al., 2016a). Thus IBMs, in contrast to PLMs, do not describe changes at the population level, but instead describe the activities and properties of individuals and how they respond to their environment. Changes at the population level then emerge automatically from these collective interactions between individuals, a phenomenon named *emergence*. For this reason, IBMs are classified as 'bottom-up models', since they describe the lower organizational level in order to predict the higher organizational level (Mabrouk, 2010).

The terms 'individual-based model' and 'agent-based model' are sometimes used interchangeably. However, the term 'agent' is more general since an agent is not necessarily an individual (Railsback and Grimm, 2012). Agents can cover many scales, from individual cells and organisms, to social groups such as families, or larger social or economic organizations like businesses or public health care systems (Grimm et al., 2005).

Similar to IBMs are cellular automata (CA), which also model time and space discretely (Permogorskiy, 2015). CA models differ from IBMs in that they consider the spatial cells instead of the individuals occupying them (Ferrer et al., 2008). Hence CA approaches are concerned with the global geometric patterns that emerge from the local interactions (Manukyan et al., 2017), while IBMs focus on individual variability and how this affects collective population-level behaviour (Hellweger et al., 2016a).

The bottom-up construction of IBMs takes individuals rather than populations as the basic modelling unit, which admits variation between individuals, allows the spatial heterogeneity of this system to emerge naturally as a result of the localised interactions, and can easily mimic adaptive processes (Kreft et al., 2013). These features address the key limitations of PLMs that we outlined in Section 3.2.2. For these and other reasons which we shall now discuss, the use of IBMs in microbiological and microbial ecology studies has gained increasing favour in recent years (Railsback and Grimm, 2012; Larsen et al., 2012; Song et al., 2014; Widder et al., 2016).

Advantages and disadvantages

The main advantages of IBMs are their "*maximally flexible*" representations of individuals, their characteristics and behaviour (Klimenkoa et al., 2016) and their explicit description of interactions between individuals. This flexibility permits a gradual introduction of complex behaviours into an individual's interactions, so that their population-level effects can be assessed separately (Ferrer et al., 2008). This makes IBMs particularly suitable for studies focused on theory development, since the impacts of different elements of a model and the consequent emergent collective behaviour can be studied in a modular way. This flexibility also permits the construction of models with a desired degree of complexity, to study whether certain population-level effects will still emerge under more specific or more general conditions or settings (Song et al., 2014).

Additionally, by mapping individual interactions to population dynamics in this way, IBMs can use data from both levels: observations of individual behaviour are used as model input and observations of population dynamics are compared with model output (Hellweger et al., 2016a). This ability of IBMs to to incorporate both lower and higher levels of organization is a significant advantage.

At the lower level, IBMs can incorporate sub-models of, for example, intracellular dynamics (Widder et al., 2016). Such models can steer the behavioural dynamics of an individual microbe mechanistically rather than phenomenologically (Kreft et al., 2013). At the higher level, IBMs can model the dynamics of complex communities or ecosystems in a minimally complex way, since they rely on the description of individual actions (Klimenkoa et al., 2016). As will be discussed in

Section 3.3, IBMs can account for space in a straightforward manner since individuals and their interactions are localized. Even indirect interactions between individuals can be captured by IBMs, since these are emergent properties of the direct interactions of individuals with each other and with their environment, and hence do not need to be explicitly accounted for (Kreft et al., 2013).

The main disadvantages of IBMs are that they can require significant amounts of experimental data at the individual scale, and may come with a high computational cost (Kreft et al., 2013). Reducing the computational costs of IBM simulations typically takes two forms: either confining the *in silico* domain to a small representative space, or aggregating individuals via the use of the "super-individual" concept (Song et al., 2014). The latter technique has been highlighted as particularly relevant for microbial communities (Hellweger, 2008).

The disadvantage relating to experimental data is increasingly being addressed through experimental advances such as microfluidics, flow cytometry and microscopy (discussed in Section 2.3.2) that allow for the collection of data at the individual level (Wessel et al., 2013). These data permit the calibration and validation of IBMs which are necessary to make well-supported predictions (Widder et al., 2016). This modelling endeavour has been further catalyzed by the increasing availability of the necessary computing power, which has made it feasible to simulate large numbers of individuals *in silico* (Mabrouk, 2010). An example of this synthesis is the deployment of IBMs to model the complex structures of biofilms in conjunction with confocal microscopy observations (Picioreanu et al., 2004).

Another current limitation of the IBM approach is the young age of the methodological framework for developing, implementing and validating IBMs, analogous to those which have been developed over many decades for PLMs (Wade et al., 2016). Simply by virtue of their youth, IBM approaches do not yet dispose of a similarly well established methodological framework, the development of which is also complicated by the same characteristics of IBMs that make them so attractive as alternatives to PLMs, namely their ability to incorporate complex behaviour and large amounts of intra-species variability (Ferrer et al., 2008). However, important and ongoing efforts have been undertaken to establish such a methodology, including the definition of standard set of terms and enumerating the key steps involved in the procedure (Augusiak et al., 2014) and an outline of good practice documentation for model development and testing (Grimm et al., 2014), so that modellers already dispose of important resources when deciding how best to validate and evaluate their IBMs.

In comparison to other modelling approaches, IBMs are more difficult to analyse, which has also made them more difficult to describe and disseminate (Ferrer et al., 2008). Many IBMs are too extensive and their processes and sub-processes too detailed to be described in one research paper, and a standardized framework for their description is still being elaborated and established (Mabrouk, 2010).

Efforts to address the lack of rigour in describing IBMs include the use of the ODD (Overview, Design concepts, and Details) protocol proposed by Grimm *et al.* in 2006 and updated in 2010. This standard protocol consists of seven elements, grouped into Overview (purpose, state variables and scales, process overview and scheduling), Design concepts, and Details (initialization, input, submodels), and has found increasing favour as a standardized framework for describing IBMs (Hellweger et al., 2016a), an example that we will follow in our own modelling studies later in this thesis.

3.2.3.2 Examples of IBMs in microbial ecology

IBMs have been used to model microbial communities in: wastewater treatment plants (Van Loosdrecht et al., 2002; Picioreanu et al., 2004; Xavier et al., 2005; Laspidou et al., 2010); medical settings (Murphy et al., 2008; Seal et al., 2011); food manufacturing processes (Ginovart et al., 2007); and various other environments such as soil or marine ecosystems (Ginovart et al., 2005; Gras Moreu et al., 2011; Koenigstein et al., 2016).

To help understand the establishment of IBMs in the microbial sciences, Hellweger and Bucci (2009) reviewed 46 published papers related to IBMs of microbial and phytoplankton ecosystems, with a particular focus on why the various authors selected an IBM approach. They broke down the motivations as related to: the importance of the intra-population variability or heterogeneity (46%); the emergence of population level patterns (24%); the discreteness of the individuals (5%); or other reasons (26%).

IBMs can also be of significant use in synthetic microbial ecology in particular, since they permit the simulation and optimization of how individuals interact with each other and the environment before actually constructing them. Therefore they can be an important tool in the rational design of engineered communities, which is a central goal of synthetic microbial ecology (Stenuit and Agathos, 2015).

In this context, IBMs have been extensively used to study microbial interactions, since these can be implemented in a straightforward manner in an individual-based framework (Coyte et al., 2015; Billiard and Smadi, 2015; Centler and Thullner, 2015; Lloyd and Allen, 2015; Germerodt et al., 2016).

For example, Nadell *et al.* (Nadell et al., 2010) used an IBM approach to study how cooperative and cheater microbes can self-organize during biofilm growth, resulting in a clear spatial segregation of the two types. Simulations showed that this spatial segregation permitted the cooperative individuals to interact with other cooperators, avoiding contact with and exploitation by cheaters, which favoured the establishment of cooperation as the dominant interaction in the community. This study is an example of the usefulness of IBMs in theory development, in this case regarding the evolution of cooperation in microbial communities.

Another example is the engineering of *Saccharomyces cerevisiae* strains in a synthetic community so that their growth was dependent on each other through the production of a metabolite required by the other strains (Momeni et al., 2013). The IBM counterpart of this *in vitro* synthetic community revealed that when randomly placed individuals began to merge into colonies, colonies that were engaged in mutual cross-feeding, and that were also by chance located near each other, would grow towards each other to form spatial aggregations of cross-feeders while also excluding cheater strains. This process was identified as one of the main mechanisms maintaining cooperation in the community, and hence its stability and functionality.

3.2.4 Integrative modelling approaches

Due to the particular complexity of microbial communities and their constituent entities and processes, it can be advantageous to combine modelling approaches, so that each modelling component is geared towards the particular facet of microbial communities that it is most suited for.

There exist multiple strategies for such model integration; Song *et al.* (2014) classify them by ascending order of strength as (i) information feedback, (ii) indirect coupling, or (iii) direct coupling. With information feedback, the outputs of one model are used to adjust the assumptions underlying the construction of another, independent model; whereas indirect coupling involves using the outputs of the first model as the inputs of the independent second model (Song et al., 2014). Finally, the strongest form of model integration is direct coupling, when different models are combined into a single simulation system.

For this type of 'multilevel' modelling, IBMs have more promise than other approaches, due to their flexibility and ability to integrate submodels of various kinds (Widder et al., 2016). For example, intracellular dynamics have been incorporated in IBMs by modelling the signalling mechanisms involved in quorum sensing (Pérez-Velázquez et al., 2016) and chemotaxis (Shklarsh et al., 2011).

Integrative modelling approaches have also been highlighted as very promising for predictive purposes (Klimenkoa et al., 2016). Developing such models would, for example, enable researchers to make predictions about microbial communities that cannot be cultured or studied using current techniques, which as discussed in Section 2.3.1 constitute the majority of natural communities.

3.3 Modelling spatial dynamics in microbial communities

The importance of accounting for spatial dynamics in both theory development and predictive studies — the former in particular — is fairly well established in macro-ecology (see e.g. the review by DeAngelis and Yurek (2017)), but is less established in microbial ecology. Possible reasons for this include the difficulties of resolving the very small spatial scales over which microbes disperse and interact, as well as the experimental barriers to obtaining data at these scales (Zomorrodi and Segrè, 2016).

3.3.1 Spatially implicit and spatially explicit models

Two types of mathematical models can be distinguished based on their treatment of space. Spatially implicit models (SIMs) do not account for space in their representation of natural systems, in contrast to spatially explicit models (SEMs) (DeAngelis and Yurek, 2017). This distinction runs across the different modelling scales described in Section 3.2, so that we may speak of spatially implicit or spatially explicit community-level models or PLMs. The majority of IBMs are spatially explicit, although there are some exceptions (Railsback and Grimm, 2012).

SIMs are typically constructed on the basis of the mass action law, which has its roots in the study of chemical reactions (Murray, 2002). This law states that if different particles must collide to initiate a reaction, and the experimental system is well mixed, then the collision and hence reaction rate is proportional to the product of the concentrations of the reactants (Song et al., 2014). This law is used in SIMs to describe the interactions between species as a function of their densities (analogous to the concentrations of chemical reactants). This construction depends on the assumption mentioned above: that the environment is sufficiently well mixed to allow any individual to come in contact with any other. This assumption, known as the mean field assumption, justifies the use of species or population-level averages (Zomorrodi and Segrè, 2016).

However, in many ecological settings in the real world, the mean field assumption of a well-mixed environment does not hold. This is also true for microbial ecosystems in particular, where heterogeneities in for example metabolite, nutrient, and light distributions have been shown to play key roles in community structure (Wimpenny, 1999).

Figure 3.4: A microbial mat from a New England salt marsh, with distinctive spatial structure (Armitage et al., 2012).

Indeed, one of the most important factors for the establishment and stability of a microbial community is the spatial organization of its members (Tan et al., 2015), alongside those discussed in Section 2.3.2. Natural communities often form and maintain a defined spatial structure (an illustration is given in Figure 3.4). Synthetic microbial communities are therefore often engineered to reproduce these spatial structures, so as to stabilize their dynamics and functionality in a similar way (Bertrand et al., 2015).

Therefore, models should take into account heterogeneous space and local interactions in order to obtain a more realistic representation of reality (Hellweger et al., 2016a). This insight is increasingly well recognized in various ecological fields, and has resulted in the ever increasing use of SEMs, which have repeatedly been shown to produce representations and predictions significantly different, and more realistic, than those obtained using mean field models such as SIMs (DeAngelis and Yurek, 2017). In the next section, we ask why this should be the case.

3.3.2 Effects of spatial heterogeneity

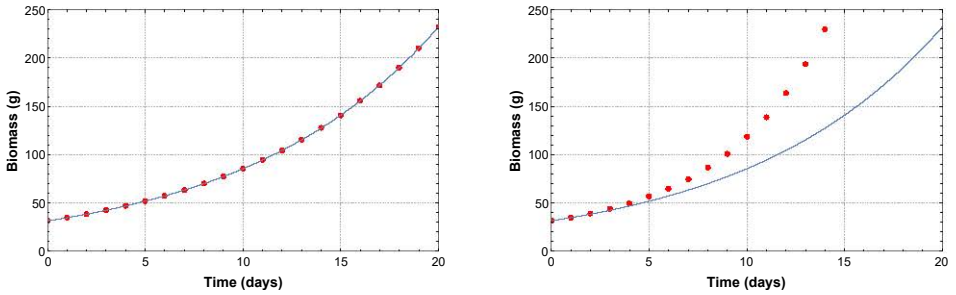
Why are the predictions of spatially explicit models so different from the predictions of their corresponding mean field approximations? A key reason is the

existence of spatial variation in local environments, which is ignored by SIM approaches. When individuals interact with their neighbours, it is the density of these neighbours (necessarily confined to a small region local to the focal individual) that drive the local dynamics, not the density averaged over a larger spatial area. Deviations in local neighbourhood characteristics compared to the population-level average characteristics have been grouped into two types (Dieckmann et al., 2000).

The first type are systematic deviations from the population average, which are typically due to previous interactions between neighbouring individuals (DeAngelis and Yurek, 2017). For example, if species A is preyed upon by species B, then in the local neighbourhoods of individuals of species B, we will find fewer individuals of species A than the population-level average would suggest. SIMs are unable to capture these local interaction effects, and are therefore unable to capture their collective impact on the population-level dynamics (Neuhauser, 2001).

The second type are random deviations from the population average. These are typically due to the effects of finite population sizes, which are not found in the mean field case. Random local deviations can be significant when individuals respond differently to a heterogeneous environment, because the average response of individuals across different environments is not the same as the response of individuals to the average environment (Dieckmann et al., 2000). This phenomenon is known as Jensen's Inequality (Jensen, 1906), and leads to variations in response at the individual level being magnified so that significant population-level changes emerge, leading to disagreements with the mean field prediction.

An example is shown in Figure 3.5, where for illustrative purposes we model the population growth of an *in silico* community using the exponential growth model. The key parameter in this model in terms of individual variability is the growth rate r. If we do not consider variability of this parameter due to the mechanisms described above, then every individual in the community has the same growth rate and there is zero variation in this parameter, i.e. $\Delta r = 0$ (Figure 3.5(a)). If we make the contrary assumption, and allow for a certain level of variability in individiduals' growth rates, then the population-level average growth rate quickly diverges from the mean field prediction which averages across all local environments (Figure 3.5(b)). The use of SEM approaches avoids this pitfall.



(a) $\Delta r = 0$ (b) $\Delta r = 0.1$

Figure 3.5: Illustration of Jensen's inequality for exponential growth with (a) zero variability in individuals' growth rates and (b) variability in individuals' growth rates. Variations in response at the individual level are magnified so that significant population-level changes emerge, leading to disagreements with the mean field prediction.

3.3.3 SEMs in microbial ecology

The *in silico* techniques available to model the effects of mobility and dispersal in microbial communities have been studied in various reviews (see e.g. (Gregorius and Kosman, 2017); (Adamson and Morozov, 2012)). We mention one in particular that has attracted particular attention in the field, and to which we will return in our own modelling study since it has particular importance for the maintenance of *in silico* community diversity.

Reichenbach *et al.* (2007) used a stochastic lattice-based IBM to demonstrate that coexistence of three *in silico* species was mediated by their dispersal. Individuals could move around the lattice by switching places with one of their nearest neighbours, namely, one of the individuals located at an adjacent lattice site. This mobility process occurred at a certain rate ϵ, and led to the formation of spatial structures that were stable in time and permitted the coexistence of all three species (Figure 3.6(a)). But if the mobility rate was increased until it exceeded a critical rate ϵ_c (a function of the lattice size), this provoked extinctions and the loss of the *in silico* community's biodiversity, independent of spatial environment and details of the competitive interactions between individuals (Figure 3.6(b)). This was due to the loss of localization in individuals' interactions, which meant that the interactions approached a well mixed setting and hence the stabilizing spatial structures were unable to form.

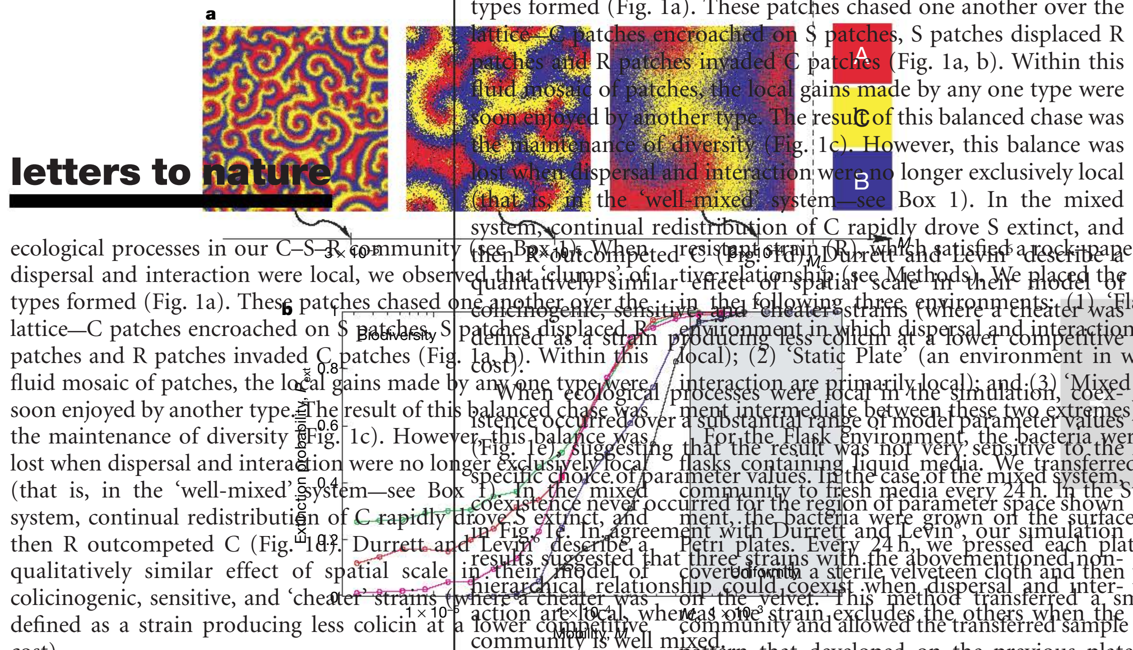


Figure 3.6: Reichenbach *et al.* (2007) used an IBM approach to demonstrate that the diversity of a community of three species was maintained by the localization of interactions between individuals, as determined by their mobility M. Once M exceeds a critical value M_c (dependent on system size) biodiversity is lost and all simulations result in monocultures.

Empirical evidence for this assertion can be found in a study that employed a synthetic community consisting three populations of *E. coli* (Kerr et al., 2002). The study found that diversity was rapidly lost in the synthetic community when dispersal and interactions occurred over relatively large spatial scales, whereas all populations could coexist when the same ecological processes were localized (Figure 3.7).

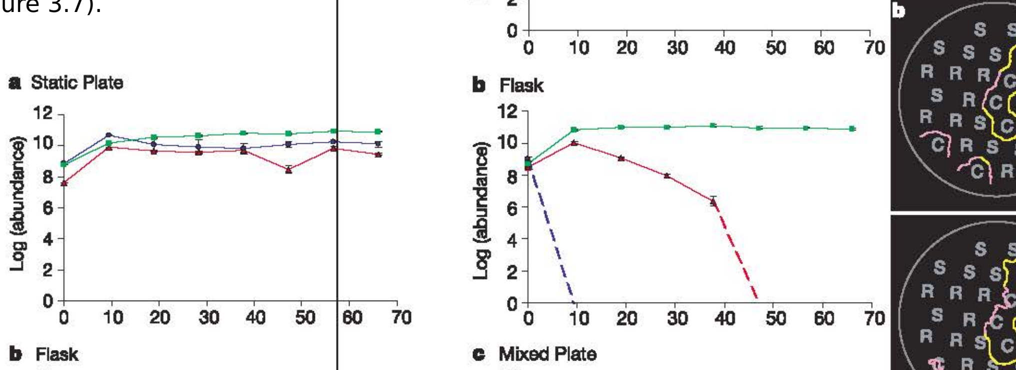


Figure 3.7: Kerr *et al.* (2002) demonstrated that coexistence of a community of three *E. coli* was dependent on the localization of interactions due to spatial heterogeneity. When the species interacted in a spatially structured environment (left), all species could coexist. In contrast, when the species interacted in a well mixed environment (right), coexistence was not possible and only monocultures were obtained.

The lattice-based IBM approach of Reichenbach *et al.* (2007) has inspired numerous studies which employ extensions or modifications of this modelling approach, since it is able to model the emergence of complex dynamics and behaviour from straightforward interactions at the individual level, as is characteristic of IBMs. Some examples of modifications to this model include: reaction rates that vary between species (He et al., 2010), the incorporation of mutations (Mobilia, 2010), and the study of cooperation rather than competition (Helbing and Wenjian, 2008).

3.4 Modelling interactions in microbial communities

The interactions taking place within a community, whether these occur between individuals or between individuals and their environment, have been shown to be key drivers of community stability and functionality (see Section 2.3.3). These interactions can take various forms, from competition or cooperation to predation or commensalism; an in-depth description of the different possibilities can be found in the review of Faust and Raes (2012). Here we focus on the modelling approaches typically used for the several specific types of microbial interactions highlighted in our research questions (Section 1.2).

3.4.1 Interactions with the environment

The interaction between microbe and environment that we have focused on in our research questions is resource limitation. This refers to the presence of a substrate in the *in silico* environment which individuals depend on for growth and other demographic processes, but whose concentration is limited, thus forcing individuals to modulate their behaviour in response to the changing availability of the environmental resource.

In microscopic scale models, the typical approach to modelling resource limitation has been to represent it simply by imposing a constant limit on population size (Nowak, 2006; Riolo et al., 2001). This avoids the necessity of modelling the resource dynamics explicitly, which simplifies model construction and analysis, but reduces the insights this approach can provide into the interactions between individuals and the environment. For this reason, this approach has been adapted to explicitly consider resource fluxes and dynamic population sizes (Requejo and Camacho, 2013; Melbinger et al., 2010; Requejo and Camacho, 2011; Centler and Thullner, 2015).

A similar shift took place in the modelling of biofilm formation, now typically done using PDE approaches where both the growth of cells and the diffusion of nutri-

ents through the bulk liquid are taken into account (Lardon et al., 2011; Ardré et al., 2015; Kragh et al., 2016). This approach has in recent years been extended to IBMs (Centler and Thullner, 2015), where a typical example admits a limiting resource that constrains individuals' reproduction (Requejo and Camacho, 2012).

However, some models of resource-limited reproduction assume that the population in question is well mixed, a typical yet significant simplifying assumption which by design does not permit any effects of spatial structure to emerge. This is despite the fact that spatially structured environments have been acknowledged to result in a significantly different population dynamics than well-mixed environments, as we discussed in Section 3.3.

3.4.2 Competition

3.4.2.1 Pairwise interactions

Much of the modelling literature regarding competitive dynamics has focused on pairwise interactions, most often using ODE-based approaches such as the Lotka-Volterra equations (Levine et al., 2017). As a result, the coexistence of two species in competition has been well explained using the framework of mutual invasion, where each species has a positive growth rate when its density is low and its counterpart is at its carrying capacity (Chesson, 2000). In this way, a species can always rebound from perturbations, and hence maintain the two-species coexistence. The stabilizing mechanism underlying this coexistence criterion of mutual invasion has been shown to be related to niche differences between the two species, which cause intraspecific effects to be more negative than interspecific effects (Adler et al., 2007). Hence when the density of one of the species increases, its growth rate is reduced relative to other species, which wards off its competitive exclusion and helps to maintain coexistence.

Coexistence is then dependent on these niche differences between the two species being more significant than their difference in fitness, otherwise a species with a comparatively weaker fitness would not be able to invade its competitor and thereby maintain coexistences (Adler et al., 2007). Examples of these stabilizing niche difference mechanisms include the limitation of the two species by different resources (in well-mixed settings explored by SIMs) or when the two species prosper in different locations in the landscape (in spatially heterogeneous settings modelled by SEMs) (Chesson, 2000).

3.4.2.2 Higher-order interactions

However, this mutual invasion framework is not transferable to systems of three or more species (Levine et al., 2017). The underlying mechanism is intuitive: when

one species is suppressed in a system of two species, its counterpart will always be able to persist if the mutual invasion criterion is satisfied. But in richer communities, the suppression of one species can allow another species to prosper (and maintain the competitive balance between these two species) while a third species suffers. For example, in the rock–paper-scissors game, the suppression of rock will allow scissors to prosper, but both of these changes will negatively affect paper, which depends on there being a sufficient density of rock to keep scissors at bay. This phenomenon is also known as trophic interaction modification, when referring more generally to the modification of a consumer-resource interaction by a third species (Terry et al., 2017).

Thus the mutual invasion criterion does not hold for systems of three or more species linked in this way. Therefore, a different framework is required to understand the mechanisms underpinning coexistence in richer communities. Improving our understanding of these mechanisms which are present only in diverse communities will also improve our understanding of the stability of these communities' diversity.

Two particular types of competitive dynamics have been shown by theoretical and modelling studies (Levine et al., 2017) to emerge only in systems of three or more species: interaction chains and higher-order interactions.

Interaction chains involve pairwise competitive interactions contained within a network of other pairwise interactions. This allows for indirect interaction effects, for example when the direct interactions of a focal pair are affected by changes in density of a third species that interacts with both species in the focal pair (Hibbing et al., 2010). Most notably, these indirect effects can have positive influences on the system even when the direct interactions have negative effects (Stone and Roberts, 1991). The classical example of an interaction chain is the rock–paper–scissors game. This chain consists of interlinked pairwise competitive interactions, which together stabilize the dynamics of the system (Reichenbach et al., 2007). Terry *et al.* used their framework of trophic interaction modifications to further distinguish between indirect interaction effects, characterizing these as either secondary (when the initial direction interaction leads to a change in densities of the secondary species, which then has a further knock-on effect) or density-mediated (due to the trophic links with the third species) (Terry et al., 2017).

In contrast, higher-order interactions refer to interactions that are not pairwise (Levine et al., 2017). In this case, the effect on any one species does not depend on only one other species, but rather on several. For example, if a predator depresses the population of another species, this can cause effects even further down the food chain (Shurin and Allen, 2001).

The difference between these two types of interaction lies in the cause of the (positive or negative) indirect effect (Levine et al., 2017). If the indirect effect is due to changes in the density of the competitor species, it relates to an interaction

chain. In contrast, a higher-order interaction involves changes in the competitive effects of the competitor species. This difference also results in their effects occurring on different time scales: higher order interactions alter the competitive effects of a species on the dynamics of two (or more) other species, changes which occur promptly (Bruno and Cardinale, 2008). In contrast, interaction chains are altered over a longer time scale. To return to our rock–paper–scissors example, an increase in the density of scissors will not immediately affect the dynamics of the system, since the consequent changes in rock and paper densities will not occur immediately, but rather after a certain time lapse (Reichenbach et al., 2007).

The loss of either of these mechanisms can result in so-called extinction cascades (Ebenman and Jonsson, 2005), where the loss of one species triggers further extinctions, and in the worst case total system collapse. Interaction chains consisting of intransitive cycles or loops are particularly susceptible to this effect, where the loss of one species breaks an intransitive cycle, inducing the further loss of the other species in the cycle (Han et al., 2016).

Both interaction chains and higher-order interactions are important mechanisms in stabilizing dynamics in microbial communities, and must particularly be accounted for in synthetic microbial communities. Since many synthetic communities involve more than two species (De Roy et al., 2014), both interaction chains and higher-order interactions may emerge. Their loss or destabilization can therefore have significant consequences for the stability or functionality of the community.

3.4.2.3 Invasion

A specific type of competition is invasion, where an “alien” species infiltrates a community and establishes itself (Kinnunen et al., 2016). Invasion is a particular focus of synthetic microbial ecology studies, either for the purpose of engineering beneficial invasions in natural and synthetic microbial communities, or for protecting these communities against unwanted invasions (De Roy et al., 2014).

Invasion has been described as a four step process (Mallon et al., 2015) involving: (i) introduction, (ii) establishment, (iii) growth and spread, and (iv) impact. Each of these stages involves different dynamics, for which different modelling approaches can be suitable. Considering invasion using this framework also illustrates the benefits of employing an integrative modelling approach (see Section 3.2.4), in order to match the most appropriate model with each step.

In the introduction phase, the invading microbes are transported to the resident community. This dispersal can be active or passive (Mallon et al., 2015), and can be captured by various spatially explicit modelling techniques, from PDEs to IBMs. Mechanisms of interest include quorum sensing, whereby microbes undertake certain actions once a sufficient density of neighbouring individuals is sensed. For

example, flagellar mobility may be triggered by quorum sensing, propelling an invasive type towards the resident community (Pérez-Velázquez et al., 2016).

Once the invader has successfully infiltrated a community, it must then establish itself. This entails resistance to biotic pressures, including counter-measures from resident community members. These interactions between invader and community members can take various forms, and their effects can be positive, negative or neutral for their participants (see Section 2.3.3), and are typically modelled using the generalized Lotka-Volterra model. Since this model suffers from the typical drawbacks of a PLM (see Section 3.2.2), IBMs are more suited for modelling this invasion stage, since by design they account for interactions between individuals; see for example reviews by (Tan et al., 2015) and (Hellweger and Bucci, 2009).

The growth and spread phase depends on the invader's ability to access and exploit new resources in the resident community. Based on their mechanism for achieving this, invaders can be classified as pathogens or non-pathogens (Mallon et al., 2015). Pathogens directly manipulate their environment to create niches that they can exploit, while in contrast non-pathogens rely on environmental disturbances or changes that are beneficial for their invasion. In order to capture the interactions between pathogens and their environment, IBMs have gained favour due to their high degree of specificity, allowing them to capture the complexities of a specific pathogen's life cycle and interaction with its environment. Two illustrative examples of specific pathogenic invaders that have been studied using this approach include *Pseudomonas aeruginosa* (Seal et al., 2011) and *Aspergillus fumigatus* (Pollmächer and Figge, 2014).

The final stage of an invasion is its impact on the resident community, which can be significantly altered in the long term. A successful invasion can drastically alter the resident community's composition and functionality, most notably by inducing changes in its diversity (Acosta et al., 2015). To compare a community before and after invasion, whole-community (or super-organism) models may be useful to gain a "bird's eye" view of changes in community functionality in terms of its metabolic network. However, due to their low resolution such models cannot provide explanations of the mechanisms underlying such shifts, and since they are typically not dynamical (Orth et al., 2010), they can only be used to compare "snapshots" of the community's composition and functionality before and after an invasion event, although such an approach can still yield important observations to motivate further, more pointed studies. Further impacts of invasion include the integration of new mutant organisms into the community, either from the invader's side or from the resident community's (Zomorrodi and Segrè, 2016). These adaptive changes, and their impact on community composition and functionality, can be modelled in several ways. PLMs can employ evolutionary game theory to answer the question of whether a certain mutant type can establish itself in a community of an existing phenotype (see review by (Hummert et al., 2014)), while IBMs are more suited for mechanistic studies of mutation, since they can account for such adaptive proces-

ses at the individual scale (MacPherson and Gras, 2016).

3.5 Conclusions

The complexity of microbial community dynamics continues to drive the development of a vast ensemble of modelling techniques for their simulation and study. The parallel increase in the volume and complexity of experimentally obtained data has further encouraged microbial ecologists to turn to modelling to help them understand their observations, as well as to provide frameworks within which they may develop theories to explain these observations, and eventually to make predictions for unobserved conditions.

Closer collaboration between modellers and experimentalists has already yielded impressive results in the study of microbial community dynamics, and will only increase in the coming years. Both *in silico* and *in vitro* methods will continue to be developed and refined, approaching the ultimate goal of engineering synthetic microbial communities to allow for their management and control.

For this purpose, one modelling paradigm has emerged as particularly promising, namely individual-based modelling. The flexibility of IBMs lends itself well to theory development, since their bottom-up construction allows for a step-wise increase in the complexity of the behaviour or dynamics under consideration. IBMs are also particularly suitable for simulating microbial communities, where multitudes of interactions at the level of individual microbes combine to drive population and community-level effects. The most important mechanisms underlying community stability and functionality — namely the interactions between microbes and with the environment, as well as their spatial dynamics — are straightforward to incorporate in an individual-based framework. Equally straightforward to include in this modelling framework are spatial dynamics, since in IBMs the individuals and their interactions are localized. These models are thus located in the overlap between two classes of model (IBMs and SEMs) and bring the advantages of both approaches.

However, the field of microbial individual-based ecology (μIBE) is young, and much work remains to be accomplished. Before IBMs can be applied for predictive purposes, as is the central aim of synthetic microbial ecology, there remains much more to be achieved in terms of developing the tools and framework for an individual-based approach to theory development.

In Section 1.2, we have formulated several research questions relating to the stability, diversity and functionality of microbial communities. Our goal in Part II of this thesis is to build on the existing μIBE techniques discussed in this chapter, in order to develop an IBM framework which we can employ to address these research questions. Before we begin this endeavour, we bring Part I of this thesis

to a close in the following chapter, where we focus more narrowly on the techniques which can be used to simulate and represent one key aspect of microbial communities, namely their diversity.

4

Assessing diversity of microbial communities

4.1 Introduction

The use of microbial systems to test ecological theories has increased dramatically in recent years, driven in part by the increasing use of high-throughput sequencing technologies (Fulthorpe et al., 2008). In particular, theories about the relationship between ecosystem stability and biodiversity have been tested in various ways using natural and synthetic microbial ecosystems (see Section 2.5 for a more in-depth discussion). As discussed in Section 2.3, when studying competitive interactions using such microbial systems, microbial ecologists often use classical macro-ecological methods to analyze their data, one such tool being diversity indices (Veresoglou et al., 2014). But this approach sometimes proves to be complicated due to key differences in the type of data produced by microbiological or microbial ecological studies (Hill et al., 2003).

Microbial communities often contain organisms of wildly different types, meaning that any diversity quantification approach must be applicable across different do-

mains of life (Mills and Wassel, 1980). An additional difficulty lies in the extension of the notion of "species" to microbial organisms. Classical measures of diversity typically require a clear differentiation between species, which can often be difficult to achieve in microbial communities due to features such as nonhomologous recombination and a lack of sexual reproduction (Doll et al., 2013). The difficulties in directly observing microbes and their distinguishing characteristics are a further impediment to classifying microbes for the purpose of diversity quantification (Hill et al., 2003).

These particular issues are specific to microbial ecology, but the field must also confront the same issues as classical ecology when studying diversity, starting with the most basic question of all: what is diversity?

This fundamental question, and the difficulties in answering it, is addressed in Section 4.2. Several families of diversity indices particularly useful for microbiological applications are described in Sections 4.2.4 and 4.2.5. A further improvement, which can address the difficulties of differentiating species in microbial communities, is discussed in Section 4.2.6. Next, in Section 4.2.7 examples of the use of diversity indices in studies of microbial communities are discussed.

In Section 4.3, we focus more narrowly on evenness, the more complex component of diversity. The most commonly used evenness indices are described in Section 4.3.1, then we discuss the desired biological and mathematical properties of an evenness index in Sections 4.3.2.1 and 4.3.2.2, respectively. Finally, in Section 4.4 we summarize the conclusions that may be drawn from this survey of the diversity literature.

4.2 Diversity

4.2.1 Defining diversity

It is generally understood that species diversity can be split into two components: species richness and species evenness. Species richness refers to the absolute number of species present in the population of interest, while species evenness refers to the relative abundances of the different species — if a population is completely even, all species are equally abundant. However, this is where agreement ends. Dozens upon dozens of different diversity indices can be found in the literature. Such an abundance of diversity indices and their sometimes discrepant behaviour has led to so much confusion that some authors have concluded that the concept of diversity is meaningless. Even as far back as 1971, Hurlbert was moved to declare that *"the term "species diversity" has been defined in such various and disparate ways that it now conveys no information other than "something to do*

with community structure"; species diversity has become a nonconcept" (Hurlbert, 1971). Since then, the picture has only become busier, with a plethora of diversity indices to be found in the literature and more being proposed each year.

It is important to emphasize that none of these numerous and varied diversity indices are wrong. On the contrary, each index has its own unique properties that are useful for specific applications. The key point of Hurlbert's criticism of diversity as a unified concept is that since raw diversity indices exhibit such a wide variety of mathematical behaviours, they cannot all give reasonable results when directly inserted into a general diversity equation or formula (Hurlbert, 1971).

A key problem is the fact that the concept of diversity is often confounded with the indices that measure it. Jost (2006) gives the analogous example of the radius of a sphere being an index of the volume, but obviously itself not the volume. In a similar way, the most commonly-used diversity measure, the Shannon index (Ricotta and Szeidl, 2006), is actually an entropy.

Entropies, which are often confounded with diversity indices, are characterized in several ways (Mora Villarrubia and Ruiz-Castillo, 2010). First, they are continuous measures, so that any small changes in the information probabilities (the equivalent of the species proportions) have proportionately small effects on the entropy value. Second, these measures are symmetric since the ordering of the probabilities does not affect the entropy value. Third, an entropy's maximum is attained when all probabilities are equal (intuitively, this corresponds to the highest uncertainty occurring when all events have equal probability of occurring). Finally, an entropy should have an additive property which implies that the entropy value does not depend on how the sample is divided into different groups or parts (Volij, 2014).

Entropy measures disorder or uncertainty in information, and hence shares important conceptual similarities with diversity. Thus they are reasonable and frequently used indices of diversity, but this does not mean that entropy *is* diversity. Similar arguments can be made regarding many other diversity indices.

Throughout this chapter, we will consider a fully-censused community of S species, with relative abundances denoted by $p_1, \ldots, p_S$; thus, $p_i \geq 0$ and $\sum_{i=1}^{S} p_i = 1$. For convenience, we write $\mathbf{p} = (p_1, \ldots, p_S)$. First, we will briefly describe the most commonly-used diversity measures.

4.2.2 Selected diversity indices

The previously mentioned *Shannon diversity index*, also known as the Shannon–Wiener index, the Shannon–Weaver index and the Shannon entropy (Eliazar and Sokolov, 2010), measures the uncertainty in the outcome of a sampling process. When calculated using base two logarithms, it represents the minimum number

of yes/no questions that are on average required to determine the identity of a sampled species. It is given by:

$$H(\mathbf{p}) = -\sum_{i=1}^{S} p_i \ln(p_i). \tag{4.1}$$

The *Simpson diversity index* represents the probability that two individuals taken at random from the community of interest (with replacement) represent the same species (Keylock, 2005). It is given by:

$$H_{Si}(\mathbf{p}) = \sum_{i=1}^{S} p_i^2. \tag{4.2}$$

On the other hand, the *Gini–Simpson diversity index*, also called the probability of interspecific encounter (PIE), represents the probability that the two individuals represent different species (Jost and Chao, 2008):

$$H_{GS}(\mathbf{p}) = 1 - H_{Si} = 1 - \sum_{i=1}^{S} p_i^2. \tag{4.3}$$

Rao's quadratic diversity index (often called quadratic entropy although it is in fact not an entropy) is defined as the expected dissimilarity between two individuals of a given species assemblage selected at random (with replacement) (Ricotta and Szeidl, 2009), and is given by:

$$H_R(\mathbf{p}) = \sum_{i,j=1}^{S} d_{ij} p_i p_j, \tag{4.4}$$

where d_{ij} is the dissimilarity between species i and j (not necessarily a metric distance) (Ricotta and Szeidl, 2009). Note that H_R reduces to the Gini–Simpson diversity index in the case where $d_{ij} = 1$ for all $i \neq j$, and $d_{ii} = 0$ for all i.

The *Rényi entropy* generalizes several other entropies including the Shannon entropy and the standard Boltzmann–Gibbs entropy, the latter being given by $S = -k_B \sum_i p_i \ln(p_i)$ where k_B is a physical constant known as Boltzmann's constant (Eliazar, 2011). The Rényi entropy of order α for $\alpha \geq 0, \alpha \neq 1$ is given by:

$$H_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \log \sum_{i=1}^{S} p_i^\alpha \tag{4.5}$$

The Shannon entropy is the limiting case of this entropy as $\alpha \to 1$.

The *Tsallis entropy* also generalizes the Shannon and Boltzmann–Gibbs entropies (Hoffmann, 2008). The Shannon entropy is recovered as $q \to 1$. It is given by:

$$H_T(\mathbf{p}) = \frac{1}{q-1}\left(1 - \sum_{i=1}^{S} p_i^q\right). \tag{4.6}$$

While the Rényi and Tsallis entropies both generalize the standard Boltzmann-Gibbs entropy, only the Rényi entropy has an additivity property which implies that the Rényi entropy of a system composed of m independent sub-systems which are governed, respectively, by the probability vectors $\mathbf{p}_1, \ldots, \mathbf{p}_m$, is equal to the sum of the Rényi entropies of its sub-systems:

$$H_\alpha(\mathbf{p}_1 \otimes \ldots \otimes \mathbf{p}_m) = H_\alpha(\mathbf{p}_1) + \ldots + H_\alpha(\mathbf{p}_m). \tag{4.7}$$

4.2.3 Comparisons: a common problem

The indices we have mentioned comprise the most commonly used diversity measures across the various scientific fields concerned with diversity and entropy measurement; unsurprisingly, they represent quite different formulations of the same concept. However, most of them share a common problem: they are ill-suited for both relative and absolute comparisons. Given that one of their main uses in microbiology is for assessing changes in community diversity following perturbations (see Section 2.3), this is a significant issue.

To illustrate this problem, let us consider the simplest possible case of diversity: a community consisting of S equally-common species. In virtually any biological context, it seems reasonable to say that a community C_1 with ten equally-common species is twice as diverse as a community C_2 with five equally-common species. But calculating for example the Shannon entropy using the natural logarithm (as is typical), we arrive at a diversity of 2.30 for the first community and 1.61 for the second.

The first question involves a relative comparison: how should we understand the difference in diversity between these two communities? The diversity of the first community is not twice that of the second, although our intuition tell us otherwise. Second, it is also unclear what these values mean in absolute terms: should we consider a diversity of 2.3 to be high, low or something in between?

As a further example, consider a perfectly even community of one million species. The Gini–Simpson index of this community is 0.999999. We can now imagine that some catastrophe befalls this community - a meteor for example - which wipes out all but 100 species. The Gini–Simpson index of the new community is 0.99. So

despite the fact that more than 99% of the species of the pre-catastrophe community have been wiped out, the Gini–Simpson diversity index only drops by 1%. This extreme non-linearity is illustrated in Figure 4.1. Anyone directly equating the Gini–Simpson index with diversity would conclude that the community's diversity was not greatly affected by the catastrophe, while it is clear that the opposite is true. The Shannon entropy demonstrates the same problem, but to a lesser degree.

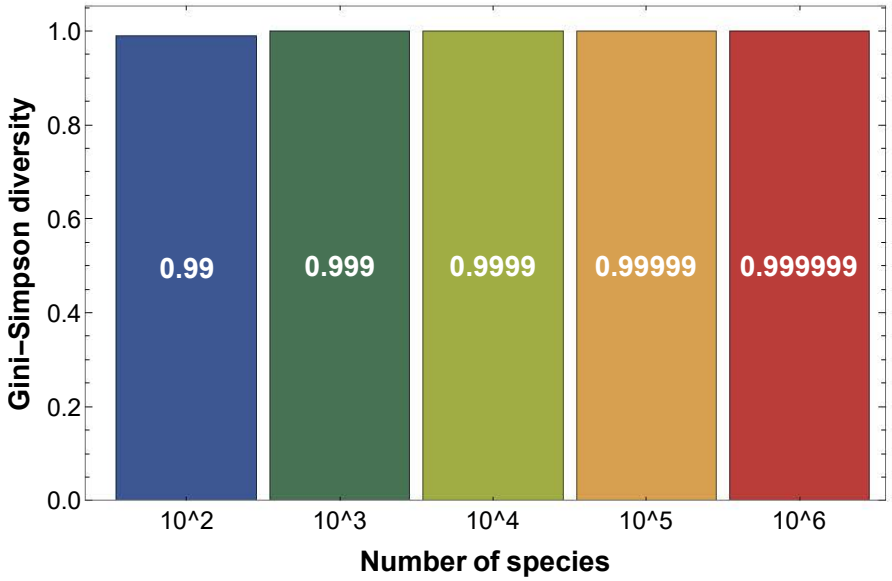


Figure 4.1: Comparison of Gini-Simpson diversity for communities with different richness.

In practice, most ecologists do not seem to be too concerned that diversity indices give results that are difficult to interpret or counter-intuitive (Pallmann et al., 2012). In their view, the actual values of the indices are unimportant, so as long as they can be used to calculate the statistical significance of the drop in diversity following an event (Jost, 2009). Subscribing to this view (which is fairly common in the literature concerning diversity in applied ecological or biological contexts), the conclusions of a study are based on the statistical significance of the result. In many cases, this is not reasonable. The statistical significance of a change in the diversity index often has little to do with the actual magnitude or biological significance of the change. Using the classical example of tossing a coin many times to see if it is biased, a highly significant *p*-value will prove that the coin is biased but will not shed any light on the size or practical importance of the bias.

Other researchers were not content with this state of affairs, and proposed a solution: the use of effective numbers.

4.2.4 Effective numbers

"In physics, economics, information theory, and other sciences, the distinction between the entropy of a system and the effective number of elements of a system is fundamental. It is this latter number, not the entropy, that is at the core of the concept of diversity in biology ... Conversion of these [diversity indices] to effective number of species is the key to a unified and intuitive interpretation of diversity. Effective numbers of species derived from standard diversity indices share a common set of intuitive mathematical properties and behave as one would expect of a diversity, while raw indices do not" (Jost, 2006).

Each diversity index creates equivalence classes among the communities it is applied to. If we apply, for example, the Shannon entropy to a set of different communities, then those communities that share a particular value of Shannon entropy are (according to this index) equivalent with respect to their diversity. In each of these equivalence classes there will be one particular community (call it C) whose species are all equally common (i.e. a perfectly even community). If we return to the intuitive definition of diversity described above - that a community of S equally-common species should have a diversity of S - then all other communities in the same equivalence class as C must also have this same diversity. Thus the problem of determining the diversity of a community reduces to finding an equivalent community (one that has the same value of the diversity index as the community in question) that is perfectly even. For example, if a community is assigned a diversity of 18.2, that means that it is slightly more diverse than a community of 18 totally dissimilar equally abundant species - there are "effectively" 18.2 species.

This problem is straightforward algebra: we need only to calculate the diversity index for D equally-common species (each species therefore having a relative abundance of $1/D$), set this equal to the actual value of the diversity index, and solve that equation for D. This value of D is the diversity of the community according to the chosen diversity index. The number D has been called the *"effective number of species"* by MacArthur (MacArthur, 1965). Other fields have recognized the importance of the effective number of a diversity index since many years ago, though the concept goes by different names depending on the discipline. In physics it is known as the number of states associated with a given entropy, and in economics is called the "numbers equivalent" of a diversity measure (Patil, 2013).

As an example of this conversion algorithm, consider a community whose species abundance distribution is given by $\mathbf{p} = (0.41, 0.21, 0.08, 0.25, 0.04, 0.01)$. The Simpson diversity of this community is $H_{Si} = 0.2828$. To convert this diversity to its effective number equivalent, we need to find a community of D equally abundant species that also has a Simpson diversity of 0.2828. We therefore have that $p_i = \frac{1}{D}$ for $i \in \{1, ..., 6\}$ and that $H_{Si} = 0.2828 = \sum_{i=1}^{6} p_i^2$. It only remains to solve for D. We obtain $D = 3.54$, implying that our six-species community is "effectively as

diverse" as a community of 3.54 equally abundant species.

Thus any diversity index can be converted into an effective number in a few straightforward steps. Converting "raw" indices to effective numbers of species in this way gives them a set of common behaviours and properties. After conversion, diversity is always measured in units of number of species, allowing for easy comparison and interpretation. It also lets us avoid the serious misinterpretations spawned by the non-linearity of most diversity indices (cfr. Section 4.2.3).

Examples of the use of effective number diversity indices to assess changes in microbiological communities are given in Section 4.2.7. We will first address further techniques to clarify and improve the choice and use of diversity indices.

4.2.5 Hill numbers

Most of the diversity indices used in the sciences, including all generalized entropies used in biology and mentioned above, are monotonic functions of $\sum_{i=1}^{S} p_i^q$, or limits of such functions as q approaches unity (Ricotta, 2003). Such indices include: species richness, Shannon entropy, all Simpson measures, all Rényi entropies (Rényi, 1961; Patil, 2002), all Tsallis entropies (Keylock, 2005; Czachor and Naudts, 2002), and many others. All such measures yield a single expression for diversity when the algorithm described in Section 4.2.4 is applied in order to transform the indices into effective numbers (Jost, 2006):

$$ {}^{q}D(\mathbf{p}) \equiv \left(\sum_{i=1}^{S} p_i^q \right)^{1/(1-q)}, \tag{4.8} $$

where the exponent and superscript q is known as the *order* of the diversity. These are often called *Hill numbers* (Hill, 1973). For all indices that are functions of $\sum_{i=1}^{S} p_i^q$, the true diversity depends only on the value of q and the relative species abundances, and not on the functional form of the index. This means that when calculating the diversity of a single community, it does not matter whether one uses the Simpson diversity index, inverse Simpson diversity index, the Gini–Simpson index, etc.; all give the same effective number diversity (or Hill number):

$$ {}^{2}D(\mathbf{p}) = 1/\left(\sum_{i=1}^{S} p_i^2 \right). \tag{4.9} $$

This diversity index depends not only on the species abundance distribution $\mathbf{p}$, but also on q, and hence the index is not univariate, unlike the classical indices described in Section 4.2.1. The parameter q gives the order of the diversity (in the equation above, the order is 2) which indicates its sensitivity to common and

rare species. The diversity of order zero ($q = 0$) is completely insensitive to relative species abundances and is in fact just species richness, which is clearly unaffected by evenness.

All values of q less than one result in diversities that disproportionately favour rare species, while all values of q greater than one lead to diversities that disproportionately favour the most common species (Keylock, 2005). The critical point that weighs all species by their frequency, without favouring either common or rare species, occurs when $q = 1$. Note that ${}^{q}D$ is undefined at $q = 1$ (Hill, 1973), but the limit exists and equals

$$ {}^{1}D = \exp\left(-\sum_{i=1}^{S} p_i \ln p_i\right) = \exp\left(H_{Sh}\right), \tag{4.10} $$

which is the exponential Shannon entropy. This quantity plays a central role in biology, information theory, physics, and mathematics (Lin, 1991; Jost and Chao, 2008; Tuomisto, 2011), and this "*is not a matter of definition, prejudice, or fashion (as some biologists have claimed) but rather a consequence of its unique ability to weigh elements precisely by their frequency, without disproportionately favouring either rare or common elements*" (Jost, 2006).

The diversity of order one (${}^{1}D$) has the properties we would intuitively expect of a diversity index. It always gives exactly S when applied to a perfectly even community with S species. It also possesses the "doubling" property introduced by Hill (1973): suppose we have a community of S species with arbitrary species frequencies $p_1, \ldots, p_S$ with diversity ${}^{q}D$. Suppose further that we divide each species into two equal groups, say males and females, and we treat each group as a separate "species". Intuitively, we have doubled the diversity of the community by this reclassification, and indeed the diversity of the doubled community is always $2 \times {}^{q}D$ regardless of the values of the p_i (Hill, 1973).

4.2.6 Similarity-sensitive measures

4.2.6.1 The problem with microbial species

Converting the established diversity indices into effective numbers is already one improvement. But there is another issue: all of the most commonly-used indices are based on the assumption that distinct species are assumed to have nothing in common. From the definitions of all the indices described in Section 4.2.1, it is clear that they take into account only the number S of different species in a community, and what proportion p_i of the community each species represents. No allowance is made for whether for example species S_i is more similar to S_j than it is

to S_k. This would imply for instance that a community of six dramatically different species is considered to be no more diverse than a community of six species of butterflies.

Diversity indices depend too much on the notion of species, a concept based upon the division of living organisms into different classes that is notoriously problematic (Leinster and Cobbold, 2012). Conventional indices such as Shannon's, Simpson's and species richness depend entirely on this division, and behave badly in the face of taxonomic reclassification. Microbial ecologists in particular have long recognized the need for similarity or distance measures in the quantification of diversity, because of the complexities of microbial taxonomy (Mills and Wassel, 1980).

Ogunseitan (2005) outlines three broad and interrelated causes for the lack of a coherent solution to this issue in microbial ecology:

1. incomplete information on the number of existing microbial species;
2. non-operational definition of a microbial niche;
3. loose definition of microbial strains and species.

The first point has to do with the deficiencies of the currently available technologies and techniques for finding, isolating and culturing microbes. These barriers prevent an accurate estimate of microbiome richness. The second point refers to the difficulties in applying the macro-ecological concept of a niche to microbial ecology. Niches delineate under which environmental conditions a species may persist (Faust and Raes, 2012). This is often very difficult to determine for microbial communities, due to the obstacles in identifying, quantifying and explaining the wide "geographical, geological and ecological" ranges of conditions under which specific microbial species persist (Ogunseitan, 2005). Finally, the third point refers to the difficulty in neatly delineating microbial species, due to such life history features as frequent genetic exchanges, nonhomologous recombination and the lack of sexual reproduction (Doll et al., 2013).

As a result, there is a lack of good diversity measures that reflect the more realistic and nuanced view of varying dissimilarities between species, or at least a lack of understanding of how to use them. Such measures are termed *similarity-sensitive* (Leinster and Cobbold, 2012). The best-known similarity-sensitive diversity measure is Rao's quadratic entropy, H_R (Ricotta and Szeidl, 2009). This measure has been receiving increasing attention, but is still used much less than either the Shannon entropy or Simpson index. The reason for the dearth of similarity-sensitive measures may well be the fact that "theoretical ecologists have been hesitant to introduce new diversity indices when the profusion of similarity *in*sensitive indices is already perceived to form an impenetrable jungle." (Leinster and Cobbold, 2012)

While similarity-sensitive measures already represent an improvement over measures insensitive to species similarity, we can move a step further. Similarity-sensitive measures that are effective numbers have an advantage over earlier similarity-sensitive indices, such as Rao's (De Bello et al., 2010) and Ricotta and Szeidl's (Ricotta and Szeidl, 2006), for the reasons discussed in Section 4.2.4. One such measure was described by Chao *et al.* (2010), who defined a family of effective number similarity-sensitive measures tailored specifically to phylogenetic diversity. These were quickly adopted by the phylogenetic community.

Another family of similarity-sensitive diversity measures was described by Leinster and Cobbold (2012). This family includes — either directly or upon applying a simple transformation — Rao's quadratic entropy, species richness, Shannon entropy, the Gini–Simpson index, the Berger–Parker index, the Hill numbers, the Tsallis entropies, and the entropies of Ricotta and Szeidl. Thus almost all of the measures discussed in Section 4.2.1 can be subsumed in one family of measures that are both effective numbers and similarity-sensitive, as we describe in detail in the next subsection.

4.2.6.2 Leinster & Cobbold

The diversity index proposed by Leinster and Cobbold (2012) takes two inputs:

- **Relative abundance data**: the proportions in which the different species are present, where "species" can represent any biologically meaningful unit, such as species, genus, phylogenetic taxa, etc.
- **Similarity data**: for each pair of species, a number specifying how similar they are, where "similar" can also be used in any biologically meaningful way. For example, a genetic notion of similarity will lead to a measure of genetic diversity, a functional notion of similarity will lead to a measure of functional diversity, and so on. The traditional "naive" model - in which similarities between species are ignored - implicitly takes all similarities between distinct species to be zero. This leads to the so-called naive measure of diversity represented by the Hill numbers (Section 4.2.5).

The diversity index also involves a parameter q ranging from 0 to ∞, which as for the Hill numbers determines how much significance is attached to species abundance. Again, for $q = 0$ species richness attaches as much significance to rare species as common ones. At the other extreme, i.e. $q = \infty$, the index, which corresponds to the one described by Berger and Parker (Berger and Parker, 1970), depends only on the most abundant species; rare species are ignored altogether.

The similarities between S species are encoded in an $S \times S$ matrix $Z = (Z_{ij})$, where Z_{ij} is a measure of the similarity between the i^{th} and j^{th} species. We assume that

$0 \leq Z_{ij} \leq 1$, with 0 indicating total dissimilarity and 1 indicating identical species; therefore $Z_{ii} = 1$. Genetic measures of similarity (often used in microbial ecology and microbiology) are often expressed as percentages, which gives us similarity coefficients Z_{ij} scaled to the unit interval (Lande, 1996). Other typical measures of inter-species distance d_{ij} range instead between 0 and infinity, but these can be scaled to the unit interval through various transformations. The simplest uses the formula $Z_{ij} = e^{-ud_{ij}}$, where u is a constant (Nei, 1972).

Various methods for determining a similarity matrix Z have already been developed, most in connection with Rao's quadratic entropy, which as mentioned in Section 4.2.6 is a similarity-sensitive measure, but is not an effective number. Some are genetic (Hughes et al., 2008), others are functional (Botta-Dukát, 2005; Petchey and Gaston, 2006), taxonomic (Vane-Wright et al., 1991; Warwick et al., 1995; Shimatani, 2001), morphological (Pavoine et al., 2005), or phylogenetic (Faith, 1992; Hardy and Senterre, 2007). They generally associate with each focal species some data concerning the characteristics deemed to be important, such as a list of functional traits, a DNA sequence, a location on a phylogenetic tree, etc. The similarity coefficients Z_{ij} are then computed in terms of some notion of difference between the associated data, depending on its particular characteristics.

Although we might assume that similarity matrices are always symmetric (i.e. $Z_{ij} = Z_{ji}$) since this seems intuitive, this is not necessarily the case (Leinster and Cobbold, 2012). The definition of similarity matrix does not require symmetry, and there are useful non-symmetric similarity matrices; the most relevant to our interests are those corresponding to certain existing measures of phylogenetic diversity (Chao et al., 2010).

The inclusion of a similarity matrix Z is what differentiates the Leinster-Cobbold index from the Hill numbers (see Eq. (4.8)), since the Leinster-Cobbold index also includes a sensitivity parameter q ranging from 0 to ∞. Then for $1 \neq q \neq \infty$ the *Leinster-Cobbold diversity of order q* of the community is given by

$$ {}^{q}D^{Z}(\mathbf{p}) = \left(\sum_{i=1}^{S} p_i (Zp)_i^{q-1} \right)^{1/(1-q)} \tag{4.11} $$

for $i \in \{1, \ldots, S\}$ such that $p_i \neq 0$, i.e. accounting for all species that are actually present.

The cases $q = 1$ and $q = \infty$ are excluded because ${}^{q}D^{Z}(\mathbf{p})$ is not valid for these values. At these values, the index does, however, converge to

$$ {}^{1}D^{Z}(\mathbf{p}) = \frac{1}{(Zp)_1^{p_1} (Zp)_2^{p_2} \cdots (Zp)_S^{p_S}}, \tag{4.12} $$

as $q \to 1$, and to

$$ {}^{\infty}D^{Z}(\mathbf{p}) = \frac{1}{\max_{i \in \{1,\ldots,S | p_i \neq 0\}} (Zp)_i}, \tag{4.13} $$

as $q \to \infty$.

The oldest measure of diversity is species richness. In our notation, this is given by the number $s \leq S$ of values of i such that $p_i \neq 0$. This measure clearly takes no account of unequal similarities between species; it uses the naive model of a community, in which the similarity coefficient Z_{ij} is taken to be 0 (total dissimilarity) if $i \neq j$, and 1 (total similarity) if $i = j$, so Z is the identity matrix I, and $(Zp)_i = p_i$. Hence species richness is the naive diversity of order zero: ${}^{q}D(\mathbf{p}) = {}^{q}D^{\mathbf{I}}(\mathbf{p})$ and thus ${}^{0}D(\mathbf{p}) = s$.

The diversity of order 2 is

$$ {}^{2}D^{Z}(\mathbf{p}) = \frac{1}{\sum_{i,j} p_i Z_{ij} p_j} = \frac{1}{\mu^2} \tag{4.14} $$

where μ^2 is the expected similarity between two individuals chosen at random. This quantity is closely related to a common measure of genetic diversity, which we shall discuss shortly. In the naive model where $Z = I$, Eq. (4.14) represents the inverse Simpson index $1/\sum p_i^2$.

More generally, if we consider any integer $q \geq 2$, and q individuals of respective species $i_1, i_2, \ldots, i_q$, then the product

$$ Z_{i_1 i_2} Z_{i_1 i_3} \cdots Z_{i_1 i_q} \tag{4.15} $$

is a measure of their group similarity. If we now let μ_q be the expected similarity of a randomly chosen group of q individuals (sampled with replacement), then

$$ {}^{q}D^{Z}(\mathbf{p}) = \mu_q^{1/(1-q)}. \tag{4.16} $$

This implies that diversity ${}^{q}D^{Z}(\mathbf{p})$ increases as the mean group similarity μ_q decreases. Equation (4.16) can be applied in situations where many diversity indices are not applicable. The most interesting for our purposes is its application to the estimation of the diversity of a community of microbes, where the notion of similarity can be fairly well-defined, but the question of what constitutes a microbial species is highly problematic (Johnson, 1973; Watve and Gangal, 1996).

The advantage of Eq. 4.16 is that we do not need to know how to differentiate between microbial species. It is sufficient to have a measure of similarity be-

tween two microbial strains. An estimate for μ_q and therefore ${}^{q}D^{Z}(\mathbf{p})$ can therefore be obtained by repeatedly taking q samples from the community, calculating the group similarity for each, and then taking the mean (Leinster and Cobbold, 2012).

4.2.6.3 Diversity profiles

We now have a family of similarity-sensitive diversity measures ${}^{q}D^{Z}(\mathbf{p})$ for each value of the sensitivity parameter q, as opposed to a single measure. The *diversity profile* of a community is defined as the graph of ${}^{q}D^{Z}(\mathbf{p})$ against q. Different communities can thus be compared by means of their diversity profiles as opposed to comparing any single statistic (Patil, 2002).

The region of a diversity profile where q is small gives information about species richness and rare species, since ${}^{q}D^{Z}(\mathbf{p})$ is affected almost as much by rare species as common ones. The tail where q is large gives information about dominance and common species, since here ${}^{q}D^{Z}(\mathbf{p})$ is almost entirely unaffected by rare species. As the sensitivity parameter q increases, the perceived diversity ${}^{q}D^{Z}(\mathbf{p})$ drops. More precisely, the diversity profile is always a decreasing continuous curve (Leinster and Cobbold, 2012).

We can illustrate the usefulness of a diversity profile with an example using a dataset that will be discussed in more detail in Chapter 9. Amongst other information, this dataset includes time series of the species abundance distribution (SAD) for a community of 13 microbial strains, under different experimental conditions. An example of the SAD for a single time point is shown in Figure 4.2, thus representing a static snapshot of the community at that moment. At this time point, the different microbial strains are present in unequal proportions: one strain dominates the community (although not excessively), several strains are present in roughly equal proportions, and there are also several rare strains.

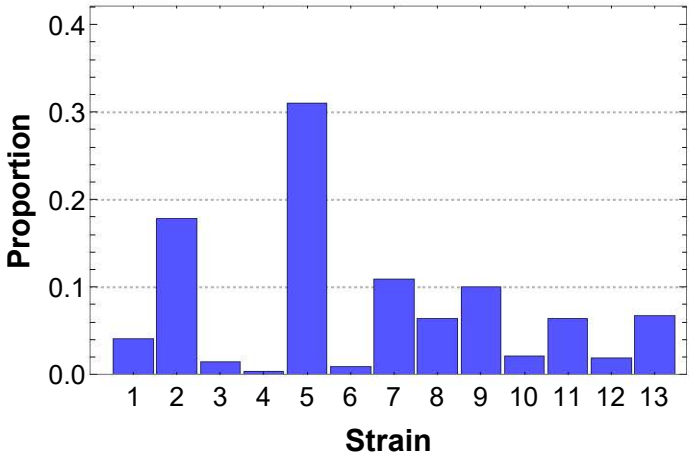


Figure 4.2: Example of a species abundance distribution for a community of 13 microbial strains (described in detail in Chapter 9).

To compare how different diversity indices represent this SAD, we can compute the diversity value for univariate indices and the diversity profile for multivariate indices. Recall that univariate indices do not account for sensitivity to rare or common species, hence they do not vary with q, unlike multivariate indices. We plot in Figure 4.3 three univariate indices, and two multivariate indices. One of the multivariate indices is similarity-sensitive (the Leinster-Cobbold index) and the other is not (the Hill numbers).

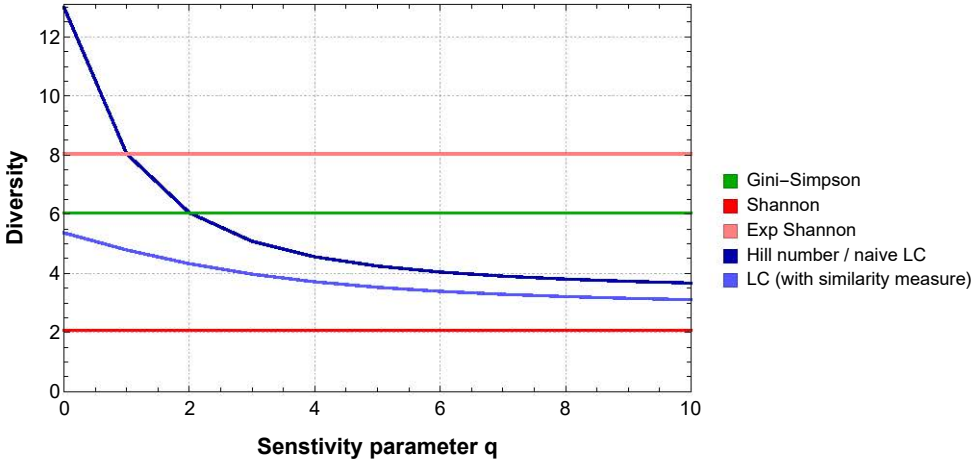


Figure 4.3: Comparison of univariate and multivariate diversity indices applied to the community in Figure 4.2. Multivariate indices (Hill numbers and Leinster-Cobbold index) result in profiles, while univariate indices do not account for sensitivity to rare species and hence do not vary with q.

We can first note the differences between the univariate indices, due to the fact

that they are not effective number indices. Hence they do not have a common unit, and the diversity values they produce for the same community can be significantly different. For the multivariate indices, we notice the difference that results from the inclusion of a similarity measure. For all values of q, the similarity-sensitive index gives a lower value of diversity than the naive index. This is to be expected since the naive index treats all strains as equally different, whereas the similarity-sensitive index considers some strains to be less distinct than others. Thus for example at $q = 1$, where rare and common species are given the same weight, the naive index calculates there to be 8 effective strains in the community, while the similarity-sensitive index calculates there to be 6.

The steepness of the left-hand end of the profiles, where q is small, give us information about the rare species in the community. As q increases, these rare species are given less weight by the index, and therefore the steeper the drop of the profile, the more rare species there are in the community. Again we notice that the naive index considers there to be more rare species than the similarity-sensitive index, since the slope of the former is steeper. In fact, the slope of the similarity-sensitive profile is so small that we can surmise that the similarity measure considers the rare species to be very similar.

As a further example of the application of diversity profiles to microbiologal studies, Turnbaugh *et al.* (2009) conducted an experiment comparing the microbial communities in the guts of lean and overweight humans. The diversity profiles for two particular test subjects from that study were compared, one subject being overweight and the other not. Since only a fraction of microbial species have been isolated and given taxonomic classifications, it was not possible for the authors to partition the microbes into species. Instead, they turned directly to DNA sequencing data. Using the naive similarity matrix (Figure 4.4(a)), the diversity profiles cross at $q \approx 1$. This suggests that the microbiome in the gut of the lean child has greater variety (higher richness), but is less evenly distributed, than that of the overweight subject. However, using a similarity matrix based on genetic measures (Figure 4.4(b)), the diversity in the lean subject is seen to be greater for all values of q, a conclusion supported by results of Turnbaugh *et al.* (2009).

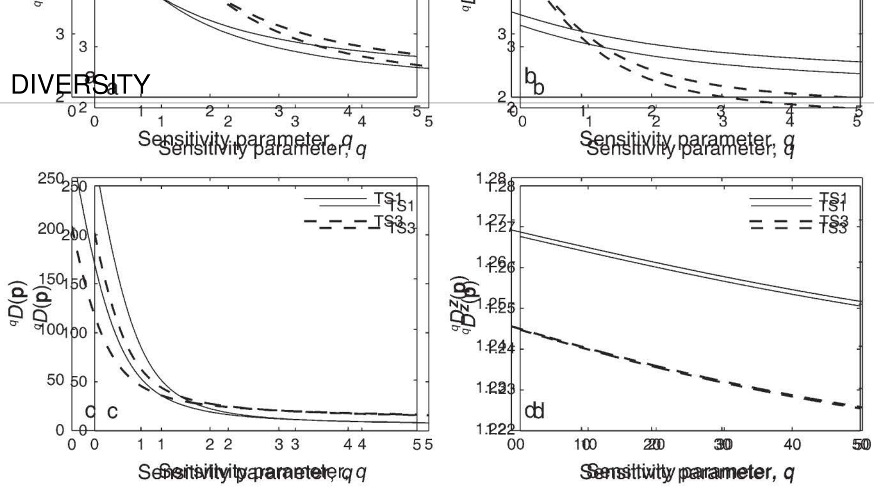


Figure 4.4: Comparison of diversity profiles using (a) a naive similarity measure and (b) a genetic similarity measure (Leinster and Cobbold, 2012) for the datasets of Turnbaugh *et al.* (2009) representing two human gut microbiomes (TS1 and TS3). Using the naive measure, the profiles in (a) cross at a low q value, suggests that the TS1 microbiome has higher richness but is less evenly distributed than that of TS3. But using a non-naive similarity measure in (b) suggests that diversity in the TS1 microbiome is greater for all values of q.

Another study used the Leinster-Cobbold index to calculate the diversity profiles of four microbial datasets, in order to investigate whether these profiles altered the interpretation of said datasets (Doll et al., 2013). For the sensitivity matrix, the authors tested both naive and similarity-sensitive matrices, the latter being based on phylogenetic similarity. The authors found that the results using the latter diversity measure provided insights into the microbial datasets that were not detectable with classical (univariate) diversity metrics (Doll et al., 2013). In particular, they found that the similarity-based and naive diversity profiles only agreed for approximately 50% of cases in their classification of which microbial sample was most diverse, a strong argument for incorporating similarity information into diversity quantification.

A third study used the same index to re-analyze four datasets of fungal microbial communities to determine whether the diversity quantification was different from the classical approach (Veresoglou et al., 2014). Using a phylogenetic similarity measure, the authors noted that the similarity-sensitive measure not only reproduced the patterns obtained using classical measures, but also revealed additional patterns, leading the authors to conclude that their approach was “more likely to uncover subtle treatment effects” (Veresoglou et al., 2014).

4.2.7 Diversity indices in microbiological studies

In contrast to the approaches discussed in Section 4.2.6.2, diversity analysis and quantification of experimental microbial community data are generally still performed using classical univariate, similarity-insensitive diversity measures, al-

4

though recognition of the benefits of other approaches to diversity is beginning to spread (Doll et al., 2013; Acosta et al., 2015; Presley et al., 2014). In such studies, diversity indices are typically employed to demonstrate a change in community size and composition following a disturbance, such as a change in environmental conditions or an invasion event (De Roy et al., 2013). This generally takes the form of a straightforward comparison, using statistical analysis to prove the significance of the change. Studies may focus on short-term or long-term impacts of the disturbance, depending on how frequently diversity is measured following the event (Acosta et al., 2015).

By far the most commonly employed diversity indices for this purpose are taxonomic richness (the naive diversity measure) and the Shannon diversity index (Eq. 4.1) (Horňák and Corno, 2012; Acosta et al., 2015; Bonanomi et al., 2014; Vivant et al., 2013; Van Elsas et al., 2007; Matos et al., 2005).

Where taxonomic richness information is unavailable or unobtainable, for any of the reasons discussed in Section 4.2.6, functional richness levels can be assessed using different approaches (Matos et al., 2005; Liu et al., 2012). Occasionally, functional richness is used to complement taxonomic richness information (Eisenhauer et al., 2012, 2013). Functional diversity levels can be assessed using substrate or carbon source utilization patterns, or dedicated indices (Petchey and Gaston, 2006).

While these approaches can often clearly demonstrate for example an inverse diversity/invasibility relationship, little insight can typically be drawn about the true mechanism of the effect, since experimental designs generally cannot separate individualistic biodiversity effects from synergistic ones (Matos et al., 2005).

This piecemeal approach to diversity measurement can be significantly improved by the use of measures that synthesize the different facets of diversity, an advance that is already being advocated for by microbial ecologists (Armitage et al., 2012; Hodgson et al., 2002; Scheiner, 2012; Presley et al., 2014; Doll et al., 2013). These works illustrate the potential of similarity-sensitive and effective number indices for use in microbiological studies, where they may lead to additional insights that are not detectable with classical diversity measures (Veresoglou et al., 2014).

4.3 Evenness

Diversity is a key mediator of microbial dynamics, yet as we have seen in Section 4.2, there is little consensus about precisely how to measure it. Practically the only agreed upon characteristic of diversity is that it is composed of two components: richness and evenness (Tuomisto, 2012). If we now move our review of diversity to this level, we find that one of these components is fairly straightforward. Richness is (generally) easy to define and measure. So long as there is

an accepted delineation of species in a community, richness is simply a matter of counting. The picture is much cloudier for evenness, the second component of diversity.

Evenness is generally agreed to reflect the equitability of the species abundance distribution in a community (Peet, 1974). It is at this point that the consensus ends. Most likely due to the difficulties in merely defining evenness, much less measuring or modelling it, studies focused on this characteristic are far outnumbered by studies relating to community richness (Hillebrand et al., 2008). However, species evenness has been shown to be a key factor in preserving the functional stability of ecosystems (De Roy et al., 2013). This is particularly true for microbial systems, where it has been shown to promote the resilience of communities to stressors and disturbances (Wittebolle et al., 2009; Yachi and Loreau, 1999). This was discussed in more detail in Section 2.6.

4.3.1 Defining evenness

Before we can address how evenness affects community dynamics, our natural first question is how to measure evenness. However, at this point there is again no consensus on any single evenness index (Smith and Wilson, 1996). Consequently, as was also the case with diversity, there are dozens of different evenness indices to be found in the scientific literature, with more being proposed every year. The high number of evenness indices stems in large part from the lack of a rigorous definition of the concept of "evenness" itself. As Tuomisto (2012) points out, "*because different indices can quantify conceptually different things, two studies whose stated purpose is to document 'evenness' can actually focus on entirely different phenomena*". The huge number of indices can make the field seem impenetrable, but researchers have proposed several approaches to help scientists choose which index is most suited to their needs (Alatalo, 1981; Smith and Wilson, 1996; Stirling and Wilsey, 2001; Kvalseth, 2015), as we shall see in the following sections.

First, we provide a brief list of the most commonly used evenness indices in the microbiological literature. Descriptions of more indices, including those typically used in other disciplines such as macro-ecology, economics and the social sciences, can be found in reviews by Maignan *et al.* (2003), Ginebra and Puig (2010), Eliazar and Sokolov (2012), and Ricotta (2003) among others. Again we use $\mathbf{p} = (p_1, ..., p_S)$ to denote a vector of species proportions, where S is the number of species in the community.

The **Simpson evenness index** is based on the probability that two individuals taken at random from the dataset of interest represent the same type (Hill, 1973). The index ranges from $1/S$ (perfect inequality) to 1 (perfect equality), and is defined as

$$\mathrm{E}(\mathbf{p}) = \frac{1}{S \sum_{i=1}^{S} p_i^2}.$$

This index has for instance been used in studies on the impact of soil invertebrates on grassland diversity (De Deyn et al., 2003) and to assess the positive effects of parasites on biodiversity in animal communities (Mouritsen and Poulin, 2005).

The **Gini evenness index** is based on Lorenz curves, where the cumulative proportion of species is plotted against the cumulative proportion of individuals (Rousseau and Van Hecke, 1999). Aside from its ecological applications, this index is also used as a measure of equality in economic and social studies (Eliazar and Sokolov, 2010). The index ranges from zero (perfect inequality) to 1 (perfect equality) — note that this is typically reversed in economic applications (Eliazar and Sokolov, 2012). The Gini index is defined as

$$\mathrm{G}(\mathbf{p}) = \frac{2}{S-1}\left(S - \frac{\sum_{i=1}^{S} i\,p_i}{\sum_{i=1}^{S} p_i}\right), \tag{4.17}$$

where the p_i are sorted such that $p_i \leq p_{i+1}$. Amongst other applications, this index has been used in ecological studies focusing on communities of moss-dwelling fauna (Vincke et al., 2006) and bacterial soil communities (Harch et al., 1997). It is also frequently used in economic and sociological applications (Lambert and Aronson, 1993).

The **Shannon evenness index** is based on the Shannon-Weaver information entropy H (Shannon, 1948), which was described in Section 4.2.1. The evenness index quantifies the uncertainty in predicting the species identity of an individual taken at random from the dataset of interest (Tuomisto, 2012). The index ranges from zero (perfect inequality) to 1 (perfect equality) and is calculated as

$$\mathrm{H}_E(\mathbf{p}) = -\frac{\sum_{i=1}^{S} p_i \ln(p_i)}{\ln(S)}. \tag{4.18}$$

In ecological studies, it has been used to investigate phenomena such as functional diversity in contaminated soil communities (Derry et al., 1998) and heterogeneous soil communities grazed upon by sheep (Gibson, 1988). This index is also referred to as the **Pielou evenness index**, where it is formulated as

$$\mathrm{J}(\mathbf{p}) = \frac{\mathrm{H}(\mathbf{p})}{\max_S \mathrm{H}(\mathbf{p})} \tag{4.19}$$

where $\mathrm{H}(\mathbf{p})$ is the Shannon entropy (see Eq. (4.1)), and $\max_S \mathrm{H}(\mathbf{p})$ is the maximum value H can take for a community of S species. Since this maximum value is $\ln(S)$, Pielou (1966) framed the index as the Shannon entropy normalized by its maximum value for a given community.

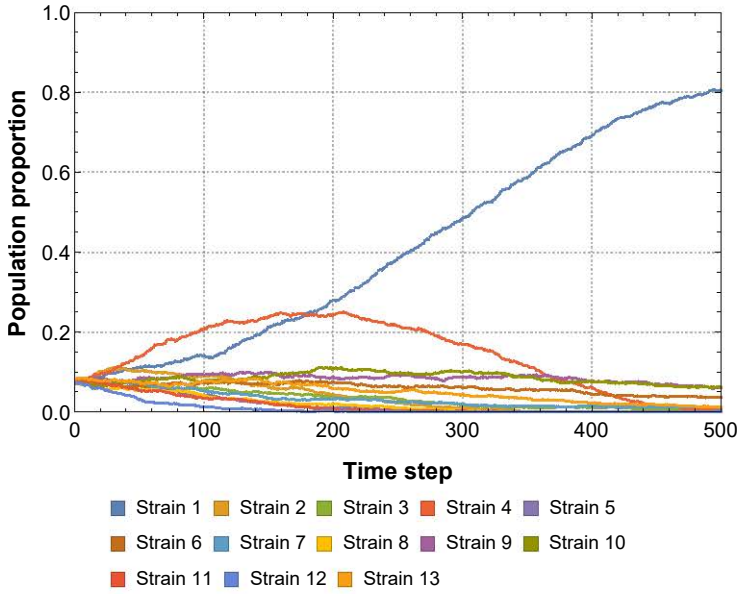
The **Heip evenness index** is again based on the Shannon entropy (Heip, 1974) , and is given by

$$\mathrm{E}_{\mathrm{H}}(\mathbf{p}) = \frac{e^{\mathrm{H}(\mathbf{p})} - 1}{S - 1}, \tag{4.20}$$

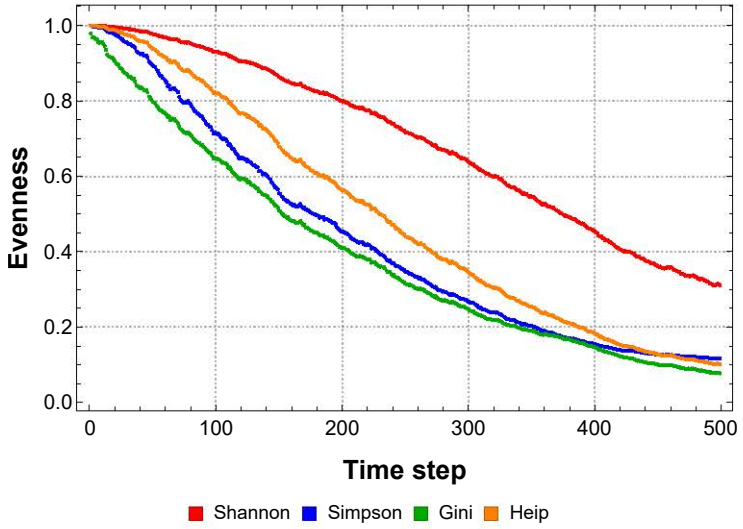
where H is the Shannon entropy. This index was proposed to address the tendency of other indices to overly depend on richness S and their failure to attain sufficiently low values when community evenness should be low (Heip, 1974). This index has for instance been used in the ecological literature to compare the evenness of insect communities subsisting on different vegetation types (Sanderson, 1992), and to study the evenness of marine meiofauna along pollution gradients (Heip et al., 1988).

To illustrate the differences between these evenness indices, we again provide an example with the same microbial dataset used in Section 4.2.6.3 (described in detail in Chapter 9). Now, rather than examine the SAD at a single time point, we can follow it through time by calculating the community evenness at each time point. In Figure 4.5(a) we show the evolution through time of the community's SAD by plotting at each time point the proportion of the population each strain represents. In Figure 4.5(b), we plot the corresponding community evenness at each time point for the four indices described above.

The community is initially very even, but over time evenness decreases sharply as one species begins to dominate the community. All four evenness indices reproduce this behaviour qualitatively, but quantify it differently. Note for example the difference in the curves produced by the Shannon and the Gini evenness indices, which represent the extremes in behaviour of these four indices. The Shannon index is more sensitive to rare species, while the Gini index is more sensitive to common species, which in this example is reflected in the curves the indices produce: the curve of the Shannon index is concave while the curve of the Gini index is convex. These characteristics are sometimes used to classify evenness indices as Type I or Type II, respectively (Peet, 1974). More specifically, Type I indices are characterized by having a second derivative with respect to p_i that increases as p_i tends to zero, whereas Type II indices produce curves with second derivatives that are constant or decreasing for values of p_i tending to zero. The Simpson and Heip indices fall in between these two cases. We can also note that this ordering is not strict: the curves of the Simpson, Gini and Heip indices cross at low evenness.



(a) Evolution through time of the community's species abundance distribution



(b) Evolution through time of community evenness as calculated using different indices

Figure 4.5: Comparison of different evenness indices for the same time series of species abundance distributions.

Many more evenness indices have been proposed in the literature, and to varying extents applied in ecological or theoretical studies. Among others, Smith and Wil-

son (1996), Magurran (2004), Tuomisto (2012) and Kvålseth (2015) provide tables listing lesser known evenness indices than those described here.

4.3.2 Desired properties of an evenness index

4.3.2.1 Biological properties

Many researchers have attempted to address the problem of choosing one evenness index from the plethora available by listing desirable criteria for an index, and then assessing how well these are satisfied by candidate indices. In their influential review paper on the subject, Smith and Wilson (1996) listed 14 criteria for evenness indices, and subdivided these into four essential requirements and 10 desirable features. The four essential requirements have a long history; they can be traced back at least as far as Dalton (1920). The requirements are as follows:

1. Evenness should be invariant under replication: it should not change when a dataset is replicated so that each of the species gives rise to n new species of the same absolute abundance as the original one.
2. Evenness should decrease when abundance shifts from a less abundant species to a more abundant one.
3. Evenness should decrease when a very rare species is added to the population.
4. Evenness should be invariant to scale, so that it depends on the proportional (not absolute) species abundances.

Similarly, Beisel *et al.* (2003) conducted a comparative analysis of 15 different evenness indices, and concluded that the measure should be chosen based on both the type of data to be analysed, and the index properties desired by the users. More recently, but in the same manner, Tuomisto (2012) compared 19 evenness indices, with more indices dismissed on conceptual grounds, and reached similar conclusions.

Another widely agreed upon criterion for an index is known as the Lorenz criterion. This criterion is based on Lorenz curves, which plot the cumulative proportion of species against the cumulative proportion of individuals (Rousseau and Van Hecke, 1999). Hence a perfectly even community has a Lorenz curve that is simply a straight line along the xy-diagonal. As the community becomes more and more uneven, its Lorenz curve falls farther from the diagonal. Then to satisfy the Lorenz criterion, an index must correctly reflect whether the Lorenz curve of one community is below the Lorenz curve of another community (Rousseau, 2011). This

determines a partial order for the indices; only partial since there is a possibility that Lorenz curves of two communities may cross, and therefore it is not possible to strictly order them in this sense. In the case of crossing Lorenz curves, different indices may rank the two community differently in terms of their evenness, and still satisfy the Lorenz criterion (Gosselin, 2001).

Tuomisto (2012) went further, and proposed four basic characteristics that an evenness index must possess if it is to agree with the definition that diversity consists of two independent components, namely richness and evenness. From this statement, which is perhaps the only statement relating to evenness that is nearly universally agreed upon, Tuomisto (2012) inferred the following four characteristics:

1. "Independent" refers to conceptual independence, hence each term (diversity, richness, evenness) refers to a different phenomenon, rather than the same phenomenon measured for different parts of the community.
2. "Independent" also refers to numerical independence, so that richness and evenness can vary independently of each other.
3. Diversity can be partitioned into two components, and there is no need for more; i.e. diversity can be expressed as a function of richness and evenness only.
4. When richness and evenness are combined, the result is a single value: diversity. For the units of conceptually different phenomena such as richness and evenness to combine in this way, they must be combined using multiplication rather than addition.

Tuomisto found these inferences sufficient to derive the formulation:

$$\text{Diversity} = \text{Richness} \times \text{Evenness},$$

and hence defined evenness as

$$\text{Evenness} = \text{Diversity}/\text{Richness}.$$

She then went on to apply this conceptual framework to various indices in the literature in an attempt to come up with a conceptually coherent assessment of the different established indices, and in this way was able to compile a table of the most commonly used indices using a unified notation to describe their calculation.

A different approach was taken by Kvålseth (2015), who instead surveyed how evenness indices are typically used, and therefore how an index should behave in order to make these applications justifiable. He stated that the general purpose of

using an evenness index was to compare the evenness values of different species abundance distributions, in order to draw some conclusions about the similarity or differences between these distributions (as was discussed in Section 4.2.3). For various evenness values $e_1, e_2, e_3, \ldots$ of the same evenness index, he distinguished the following comparisons:

- Size/order comparisons, where the goal is to compare e_1 with e_2 to determine if it is greater: $e_1 > e_2$
- Difference comparisons, where the goal is to compare changes in evenness rather than the values themselves: $e_1 - e_2 > e_3 - e_4$
- Proportional difference comparisons: $e_1 - e_2 > c\,(e_3 - e_4)$

Using this framework, Kvålseth (2015) found that for an index to behave as intuitively expected when used for such comparisons, it must take values throughout its range that are "*accurate, true or valid representations*". He therefore proposed a further index requirement, namely the value validity test, to ensure that all potential numerical values of an index must be reasonable with respect to some general criterion. This essentially ensures that an index measures what it is supposed to measure, and is best illustrated with a numerical example.

Consider, as Kvålseth suggests, the abundance distributions $\mathbf{p} = (0.75, 0.25)$ and $\mathbf{q} = (0.60, 0.10, 0.10, 0.10, 0.10)$. Using the Pielou evenness index (Eq. 4.19), we find $J(\mathbf{p}) = 0.81$ and $J(\mathbf{q}) = 0.76$. On the basis of this index, we would conclude that both abundance distributions have high levels of evenness, with **p** being slightly more even. However, when comparing **p** and **q** to their two extreme distributions (entirely even and entirely uneven), the components of **p** and **q** are equally far from the corresponding components of the extreme distributions, so that in this sense we would conclude that the "reasonable" evenness values for both **p** and **q** would be 0.5.

Thus while introducing an additional numerical requirement may seem unnecessarily complex, Kvålseth argues that an index may otherwise provide only limited information about an abundance distribution's evenness, for example by significantly overestimating the distribution's evenness, or demonstrating discontinuously large increases in evenness when the number of species in the community is increased. These problems have previously been noted elsewhere in the literature (Bulla, 1994; Smith and Wilson, 1996), and can lead to misleading interpretations and comparisons.

4.3.2.2 Mathematical properties

In their influential review paper, Smith and Wilson proposed 10 desirable features that an evenness index should possess in order to behave as would "mathematically" be expected (Smith and Wilson, 1996):

1. The index is maximal when the species abundances are equal.
2. The maximum value of the index is 1.0.
3. The index is minimal, for any number of species, when the species abundances are as unequal as possible.
4. Unrealistically uneven communities should not be necessary before the index value is low (arbitrarily defined as 0.05).
5. The minimum value of the index is 0.
6. The minimum is attainable with any number of species.
7. The index should show a value in the middle of the scale for values we intuitively consider intermediate (arbitrarily defined as 0.25 to 0.75).
8. The index should respond in a reasonable way to a series of communities that changes in evenness (using a series proposed by Alatalo (1981); "reasonable" is taken to mean a convex curve.
9. The index should be symmetric with respect to minor and abundant species, i.e. a community with several abundant species and one minor species should have the same evenness value as one with several minor species and one abundant one.
10. Species abundance distributions that are more skewed should give a lower value of the index.

Many of these are taken from the extensive literature on the subject of differentiating and choosing between evenness indices. These features are not as widely accepted as the four essential biological requirements, in part because they contradict the Lorenz criterion, and even conflict with each other (Ricotta, 2004). The list has also been modified and/or expanded by others, e.g. Eliazar and Sokolov (2012), Ricotta (2004), Ginebra and Puig (2010), Jost (2010) and Mendes *et al.* (2008).

Thus the question of an index's desirable mathematical behaviour is much less settled. However, this can be seen as a consequence of the fact that the mathematical behaviour of an index is often of lesser importance to researchers in ecological or microbial ecological fields. Most reviews on the topic conclude by stating that there is no universal way to measure evenness, and thus researchers must choose the index most suited for their particular needs (Alatalo, 1981; Smith and Wilson, 1996; Ricotta, 2004; Tuomisto, 2012; Kvalseth, 2015). This subjectivity can be regarded instead as flexibility: depending on the particular research question or data type, an index can be selected that is optimal for those specificities. While this limits the comparability of different studies, it also reflects the reality that studies are generally interested in different facets and aspects of evenness and diversity, which consequently can be optimally described by different indices.

4.4 Conclusions

The mathematical consensus is that the ideal diversity measure is one that is an effective number — this will allow for easy interpretation and comparison of the diversities of different communities. As outlined in Section 4.2.4, any diversity index may be converted to its effective number equivalent via straightforward algebra. For this reason, the choice of diversity index need not be constrained by the desire to work with effective numbers, since any shortlist of candidates is not reduced by this requirement, and researchers may choose freely from any of the established indices.

The next choice to be made is whether to work with a diversity measure that is similarity-sensitive or not. The traditional, similarity-insensitive measures (also called the naive measures) are most commonly used in the literature, and the easiest to work with. However, they incorporate no similarity measures due to their key assumption that all species are equally dissimilar. This is clearly not always the case, but it may be that in some cases this simplifying assumption presents no significant drawbacks — for example, if the community under investigation happens to be composed of wildly different species. In the context of the microbial communities that we will be studying and modelling throughout this thesis, this may or may not be the case. In particular, the concept of a “species” in microbial ecology is not always evident, as discussed in Section 4.2.6.

In cases where microbial strains cannot easily be separated into distinct classes, a similarity measure enables researchers to study the community’s diversity without confronting the sometimes tricky issue of species. Several such similarity measures can be found in the literature, based on notions ranging from functionality to phylogenetics. Ultimately, the choice of similarity measure can be based on the type of data being generated.

PART II

MAINTAINING DIVERSITY IN COMPETITIVE COMMUNITIES

5

The impact of initial evenness on biodiversity in a three-species *in silico* microbial community

5.1 Introduction

In Chapter 2, we have discussed the importance of maintaining biodiversity in natural ecosystems, and particularly in microbial systems. Several key mechanisms underpinning biodiversity were highlighted, most notably non-transitive competition (see Section 2.5) and high community evenness (Section 2.6). However, we have also seen in Chapter 3 that empirical and, in particular, modelling studies of microbial community biodiversity often overlook one if not both of these mechanisms (see Section 3.4), despite their significant and well-recognized ecological role.

More specifically, variable evenness in microbial ecosystems has not yet been investigated in computational studies concerning non-transitive competition (see for example Case *et al.* (2010), Cheng *et al.* (2014), and Frachebourg *et al.* (1996). Additionally, there is substantial evidence to suggest that perfectly even communities are rarely found in nature (Wilsey and Polley, 2004; Huston, 1997; Grime, 1998; Smith and Knapp, 2003). Thus it may well be dangerous to assume, as previous studies have done (Hillebrand et al., 2008), that community evenness is always maximal. To address this gap, we develop an individual-based model incorporating both of these two key mechanisms, in order to better understand their roles in maintaining biodiversity at the microscopic scale. In this way, we can address research questions 1 and 2 (see Section 1.2):

- What effect does initial evenness have on maintaining community diversity?
- Which types of competitive interactions can help maintain community diversity, and which types can threaten it?

This chapter is structured as follows. In Section 5.2, we introduce the fundamental processes that underpin ecosystem functioning, and discuss how they are typically modelled. Then in Section 5.3, we describe our model and the set-up of the simulation studies, and describe the computing resources in Section 5.4. The results of these simulation studies are presented and discussed in Section 5.5. In the final section, some conclusions are drawn.

5.2 Fundamental ecosystem processes

As discussed in Chapter 3, recent efforts to understand non-transitive competition as a mechanism maintaining biodiversity have focused on microscopic models describing the interactions at the individual rather than the population level (Adamson and Morozov, 2012). The individuals — whether they represent microbes, lizards or humans — are referred to as *agents*. In such microscopic models, agents are typically subject to three key demographic processes: reproduction, competition and mobility, which we denote as occurring at rates μ, σ and ϵ $\left[\mathrm{T}^{-1}\right]$, respectively, and which in our setting are not dependent on the particular species in the modelled community.

5.2.1 Rate equations

In these settings, reproduction can occur when an agent finds itself adjacent to an empty space — “a space” can be defined in various ways, depending on the

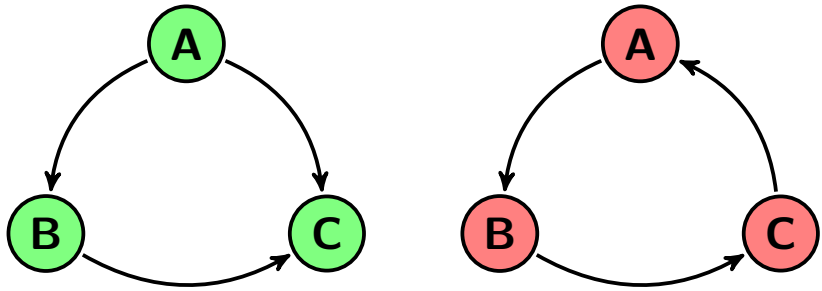
particular model (see Section 3.3) — which is then filled with a new agent of the same species. In order to provide a form of mobility, all agents can at some rate ϵ exchange their position with a nearby neighbour (which can again be defined in several ways) or move to a neighbouring empty space. For a community of three species A, B and C, these processes are represented by the rate equations (5.1), in which Ø represents an empty space:

$$\left\{\begin{array}{lcl} A\emptyset & \xrightarrow{\mu} & AA \\ B\emptyset & \xrightarrow{\mu} & BB \\ C\emptyset & \xrightarrow{\mu} & CC \end{array}\right., \quad \left\{\begin{array}{lcl} AB & \xrightarrow{\sigma} & A\emptyset \\ BC & \xrightarrow{\sigma} & B\emptyset \\ CA & \xrightarrow{\sigma} & C\emptyset \end{array}\right., \quad \left\{\begin{array}{lcl} A\emptyset & \xrightarrow{\epsilon} & \emptyset A \\ B\emptyset & \xrightarrow{\epsilon} & \emptyset B \\ C\emptyset & \xrightarrow{\epsilon} & \emptyset C \end{array}\right. . \tag{5.1}$$

5.2.2 Competition

When competition between the species in a community is fully connected — that is, each species interacts with every other species — this competition structure is called a tournament graph. In the terminology of graph theory, a tournament graph is a directed graph in which every pair of distinct vertices is connected by a single directed edge (Dutton and Brigham, 1983). Such a graph can be interpreted as the outcome of a "round-robin tournament" where every player competes with every other player exactly once; the vertices correspond to the players and the edge between each pair of players is oriented from the winner to the loser (Laird and Schamp, 2015).

Between three species, there are two possible tournament graphs (Figure 5.1). In the first case, known as hierarchical competition, species A beats both species B and species C, while species B beats species C. This leads to a simple chain. In the second case, known as cyclic competition, the three species are engaged in a rock-paper-scissors game.



(a) (b)

Figure 5.1: Competition between three species: (a) hierarchical (transitive) and (b) cyclic (non-transitive).

5.2.3 Mean field analysis

System (5.1) of rate equations gives rise to System (5.2) of ODEs, which model the population densities A, B and C of the three species through time. These equations hold for a well-mixed system with a large number of agents (Reichenbach et al., 2008), two significant simplifying assumptions. As is the convention in such models, the equations include a reproduction rate μ, competition rate σ and $\rho = A + B + C$:

$$\begin{cases} \dfrac{dA}{dt} = A[\mu(1-\rho) - \sigma C] \\ \dfrac{dB}{dt} = B[\mu(1-\rho) - \sigma A] \\ \dfrac{dC}{dt} = C[\mu(1-\rho) - \sigma B] \end{cases} \tag{5.2}$$

However, these equations ignore a key characteristic of interactions between agents, namely the spatial component. This oversight can be addressed by shifting to PDEs. The population densities then become functions of both space and time. In addition, diffusion is introduced to describe spatial movement.

As a further extension, in order to mimic the stochasticity of the system, Gaussian white noise terms $\xi(\mathbf{r}, t)$ with a spatio-temporal dependence are also included. The

resulting system of stochastic PDEs (SPDEs) is given by:

$$\begin{cases} \frac{\partial A}{\partial t}(\mathbf{r},t) = \delta\nabla^2 A(\mathbf{r},t) + \mu A(\mathbf{r},t)[1-\rho(\mathbf{r},t)] - \sigma A(\mathbf{r},t)C(\mathbf{r},t) + c_A\xi_A(\mathbf{r},t) \\ \frac{\partial B}{\partial t}(\mathbf{r},t) = \delta\nabla^2 B(\mathbf{r},t) + \mu B(\mathbf{r},t)[1-\rho(\mathbf{r},t)] - \sigma B(\mathbf{r},t)A(\mathbf{r},t) + c_B\xi_B(\mathbf{r},t) \\ \frac{\partial C}{\partial t}(\mathbf{r},t) = \delta\nabla^2 C(\mathbf{r},t) + \mu C(\mathbf{r},t)[1-\rho(\mathbf{r},t)] - \sigma C(\mathbf{r},t)B(\mathbf{r},t) + c_C\xi_C(\mathbf{r},t), \end{cases} \tag{5.3}$$

where ∇^2 is the Laplacian operator, $\mathbf{r} = (r_1, ..., r_n)$ is a vector in n-dimensional space and δ is the diffusion coefficient (with units of area per time).

By ignoring the noise terms in System (5.3), Reichenbach *et al.* (2007) were able to cast the deterministic equations into the form of a complex Ginzburg-Landau equation:

$$\frac{\partial z}{\partial t} = M\frac{\partial^2 z}{\partial r^2} + c_1 z - (1 - ic_3)|z|^2 z, \tag{5.4}$$

where z is a complex variable, $M = 2\epsilon N^{-1}$ is the mobility assigned to the individuals, and c_1 and c_3 are constants dependent on σ and μ.

This equation gives rise to the formation of dynamic spirals and allows for the derivation of analytic results relating to their wavelength (the distance over which the spiral's shape repeats) and frequency (how often the spiral's shape repeats over a distance relative to its point of origin). When the spirals exceed a certain critical wavelength (which can be associated with a critical mobility value), the spirals exceed the system size, resulting in their destruction and the loss of system biodiversity — the same type of behaviour seen in the deterministic three-species system. Thus by casting System (5.3) into the form of Eq. (5.4), Reichenbach *et al.* (2007) were able to analytically predict whether biodiversity would be preserved or lost, given the position in the parameter space.

While this PDE-based approach allowed for an improved mechanistic understanding of the system's route to extinction, it also comes with the usual drawbacks of population-level models that were discussed in detail in Section 3.2.1. Given the important role that individual variability plays in microbial systems (see Section 2.3.2), and the recent pronounced shift towards individual-based models in the microbiological and microbial ecological literature (see Section 3.2.3), this represents an important gap deserving of further investigation.

5.3 Model description

Our goal is to study the effects of both variable evenness and non-transitive competition on the maintenance of biodiversity. Given the limitations of the mean

field approach discussed in Section 5.2.3, we turn to computational methods to study this system. To this end, we consider a system of three interacting bacterial species, and present a stochastic, spatial individual-based model simulating the system dynamics and allowing for the assessment of the effect of variable initial evenness. As a foundational framework, we use the model discussed in Section 3.3, proposed by Reichenbach *et al.* (Reichenbach et al., 2007). We describe our IBM using the ODD (Overview, Design Concepts and Details) protocol described in Section 3.2.3.

5.3.1 Overview

5.3.1.1 Purpose

The aim of the model is to investigate how variable initial evenness and non-transitive competition between individuals affect the maintenance of community biodiversity.

5.3.1.2 State variables and scales

The model is a two-dimensional representation of an experimental domain divided into a regular grid of size $L \times L = N$, and populated by a community of three species denoted by A, B and C. Each grid site is either occupied by a single individual, or is empty. Individuals are characterized by two state variables: grid position (x, y) and species identity $s \in \{A, B, C\}$.

5.3.1.3 Process overview

We consider an *in silico* microbial community that is initially placed on the grid with a random spatial distribution. The community's initial species abundance distribution is constructed to obtain a desired level of evenness.

An individual can interact with its nearest neighbours, defined as those individuals in its von Neumann neighbourhood (the four grid cells with which it shares an edge). Three possible interactions can occur, representing the three key demographic processes discussed in Section 3.4: reproduction, competition and mobility.

The mechanisms of these interactions are illustrated in Figure 5.2. Reproduction can occur when an individual is located adjacent to an empty grid site, which is then filled with a new individual of the same species. In order to provide a form of mobility, all individuals can exchange their position with a nearest neighbour or

move to a neighbouring empty site. Competition can occur between two neighbouring individuals of different species. The outcome of the competition event is determined by the governing cyclic competition scheme; the defeated individual is removed from the grid and the grid site becomes empty.

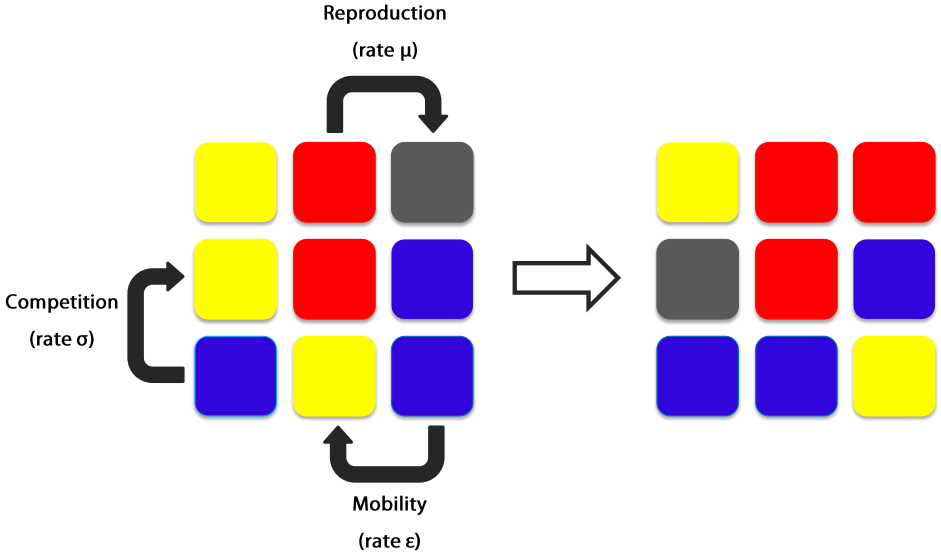


Figure 5.2: Mechanisms of demographic processes, for individual *in silico* bacteria of three species.

5.3.1.4 Scheduling

The IBM proceeds in discrete time steps. Within each time step, the dynamics of the IBM are governed by reproduction, competition and mobility. To simulate the evolution of the *in silico* community, we must specify which type of interaction event will occur and which individual will be the focus of the interaction. For this purpose, we used a modified version of a procedure called the Gillespie algorithm, which is often used in models of biological or chemical systems (Gillespie, 1976). The procedure involves an asynchronous random execution of the interaction events, and assumes that one event occurs at a time. It iterates over the following steps:

1. Set time to $t = 0$ and set the event rates:
   (a) reproduction with rate μ
   (b) competition with rate σ
   (c) mobility with rate ϵ

2. Calculate the overall rate of events $r = \mu + \sigma + \epsilon$
3. Select an individual at random
4. Select one of the focal individual's nearest neighbours at random
5. Select an interaction event with the following probabilities, by drawing a random number from the interval $[0, r]$:
   (a) reproduction with a probability $\frac{\mu}{r}$
   (b) competition with a probability $\frac{\sigma}{r}$
   (c) mobility with a probability $\frac{\epsilon}{r}$
6. Execute the selected interaction event on the selected individual (if permitted) and determine the outcome according to the governing rules
7. Update the grid according to the outcome of step 6
8. Update the time step to $t = t + 1$
9. Return to step 3 and continue until $t = t_{\text{end}}$

Thus the algorithm advances by use of a Monte Carlo step, where random numbers are generated to determine the next process to occur. We aggregate these time steps into generations: a generation is defined as the number of time steps for, on average, each individual to be the subject of one interaction event, i.e. N Monte Carlo steps for a grid of size $L \times L = N$. The length of the simulation is then defined by the number of generations for which the model is evolved (Reichenbach et al., 2007).

5.3.2 Design concepts

- **Emergence:** the spatial patterns and population-level dynamics of the community emerge naturally from the interactions occurring between individuals.
- **Interactions:** individuals interact with each other and their environment by reproducing if located next to an empty site, exchanging sites with their neighbours, or competing with their neighbours.
- **Stochasticity:** the stochasticity in the model arises from the initial spatial distribution of the grid; the interactions between individuals and the environment (reproduction); and the interactions between individuals (mobility, competition).
- **Sensing:** if selected for reproduction, individuals can sense whether their neighbouring site is empty; if so, they will reproduce. If the site is occupied by an individual, no reproduction will occur.

- **Observation:** the data collected from the IBM includes the population count of each species, the community evenness, the spatial distribution of individuals, and their time to extinction. These are tracked for each time step.

5.3.3 Details

5.3.3.1 Initialization

The model is initialized with a random spatial distribution of individuals and empty sites. Initially, a certain proportion of grid sites are left empty; thus the system is initially below carrying capacity. The initial species abundance distribution is determined by a selected evenness value, which is used to generate a vector of population abundances. Individuals are then randomly placed in the remaining grid sites according to these proportions.

To determine the relative proportions of each species in an initial community configuration, we must generate a vector of population proportions satisfying a given initial evenness, using a particular index. For this purpose, we must be able to rely on dedicated sampling algorithms. These algorithms must sample from the space of all possible vectors as uniformly, accurately and inexpensively as possible. In the three-species setting, the space of all possible population vectors is the simplex (Figure 5.3).

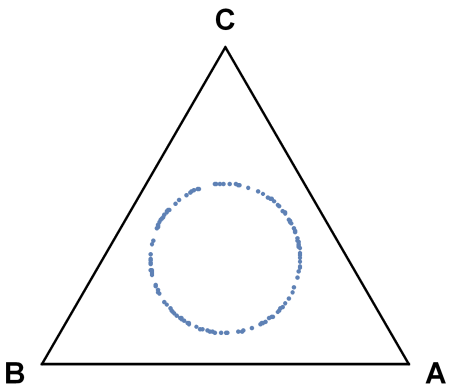


Figure 5.3: Examples of population vectors for a three-species communities corresponding to an evenness value of 0.8 (Gini), plotted in the simplex. In the simplex, points closer to the centre are more even, so that the point in the centre represents the population vector $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, while points closer to the vertices represent more uneven population vectors, so that for example a point at vertex A represents the population vector $(1, 0, 0)$.

Three evenness indices were chosen for implementation: the Simpson, Gini and Shannon indices. As discussed in Section 4.3.2.2, these indices satisfy Smith and Wilson's axioms and are among the most frequently used in the literature.

Recall that the Simpson index ranges from $1/S$ (perfect inequality) to 1 (perfect equality), and is defined as

$$\mathrm{E}(\mathbf{p}) = \frac{1}{S \sum_{i=1}^{S} p_i^2}$$

where $\mathbf{p} = (p_1, ..., p_S)$ is the vector of species proportions and S is the number of species.

For the Simpson index, sampling is straightforward because it is possible to obtain a geometrical description of the sampling region in the space of proportions p_i; that is, solving the expression for the Simpson index to obtain the limiting proportions p_i in terms of evenness E.

Thus for a given evenness value E, species proportions are sampled from the two-dimensional region enclosed by the p-axis and the curves $p = \frac{1}{2} - \frac{\sqrt{\frac{2-3E}{E}}}{2\sqrt{3}}$; $p = \frac{1}{2} + \frac{\sqrt{\frac{2-3E}{E}}}{2\sqrt{3}}$; $p = \frac{1}{3} - \frac{\sqrt{2}}{3}\sqrt{-\frac{E-1}{E}}$; and $p = \frac{1}{3} + \frac{\sqrt{2}}{3}\sqrt{-\frac{E-1}{E}}$.

The first species proportion is selected randomly along the line segments given by the intersection between the sampling region and the line $p = E$. Then the second species proportion is selected randomly along the same segments, subject to constraints imposed by the value of the first species proportion; and so on. The final species proportion is then determined by the proportions of the rest of the species, since these proportions must sum to one.

5

The Gini index ranges from zero (perfect inequality) to 1 (perfect equality) and is defined as

$$\mathrm{G}(\mathbf{p}) = \frac{2}{S-1}\left(S - \frac{\sum_{i=1}^{S} i p_i}{\sum_{i=1}^{S} p_i}\right)$$

where the p_i are sorted such that $p_i < p_{i+1}$.

The Gini index can be sampled using a consistent estimator (Davidson, 2009). Species proportions are sampled from the two-dimensional region enclosed by the p-axis and the lines $p = \frac{3}{4}E - \frac{1}{2}$; $p = -\frac{3}{4}E + 1$; $p = -\frac{3}{2}E + 1$; and $p = \frac{3}{2}E$, where E is the given evenness value. The proportion of each species is sampled according to a given evenness value in a similar manner as described above for the Simpson index.

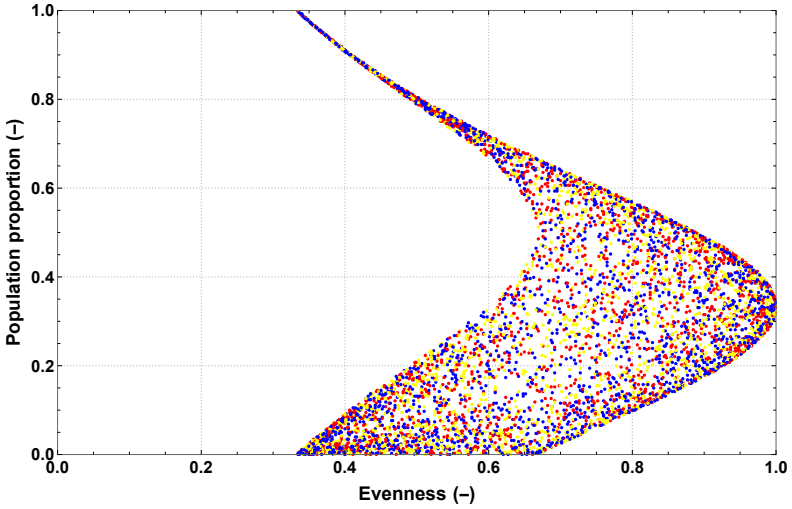
The Shannon index ranges from zero (perfect inequality) to 1 (perfect equality) and is calculated as

$$\mathrm{H}(\mathbf{p}) = -\frac{\sum_{i=1}^{S} p_i \ln(p_i)}{\ln(S)}.$$

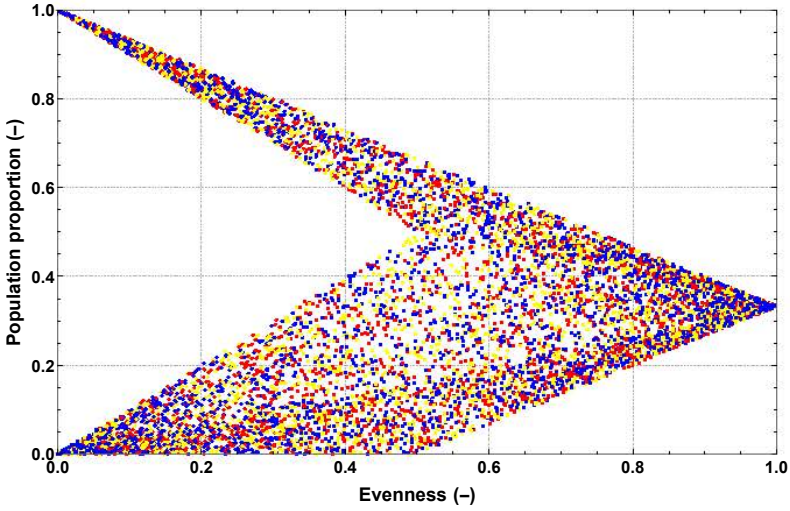
An analytic expression generating population vectors satisfying the Shannon index

is not possible, thus population vectors for this index were sampled using the simulated annealing minimization algorithm, a commonly used algorithm for global optimization (Kirkpatrick et al., 1983). The following parameters were used: a maximum of 50 iterations to stay at a given point; a random jump scale of 2.0; and a tolerance of 0.0001.

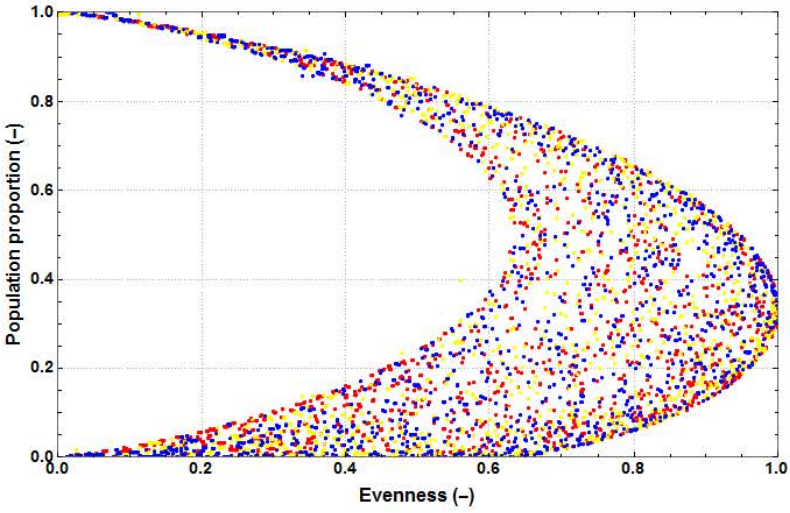
In Figure 5.4 we show the results of these sampling algorithms, for 1.000 different evenness values. Although the three indices measure the same quantity, namely the evenness of the community, we have seen in Section 4.3.1 that they can behave differently due to the differences in their formulation. Accordingly, we can notice in Figure 5.4 differences in their sampling regions. Most notably, the Simpson index ranges between $\frac{1}{3}$ and 1, while the other two indices range between zero and 1. However, they all share the same general shape. For low evenness values, one species is necessarily dominating the community, while the other two species are present in very low proportions. This results in the sampling regions taking the shape of two ”arms“ for low evenness values. In contrast, for high evenness values the three species proportions are all close to $\frac{1}{3}$.



(a) Simpson index



(b) Gini index



(c) Shannon index

A B C

Figure 5.4: Results of sampling algorithms for (a) Simpson evenness index, (b) Gini evenness index, and (c) Shannon evenness index. The evenness value of the population vector is shown on the x-axis, while the three corresponding population proportions are plotted (in different colours) on the y-axis.

Aside from the input variables, all other parameters used to initialize the model are fixed for all simulations, and are shown in Table 5.1.

Table 5.1: Parameters of the individual-based model.

Parameter	Description	Value
L	Grid side length	100
$ø$	Initial proportion empty sites	0.1
μ	Reproduction rate	1
σ	Competition rate	1
ϵ	Mobility rate	4.25
T	Number of generations evolved	500

5.3.3.2 Input

The input variable is the initial community evenness. We check a range of initial evenness values, systematically chosen so as to sample the entire range of possible values: $[0, 1]$ for the Gini and Shannon indices, and $\left[0, \frac{1}{S}\right]$ for the Simpson index, where S is the richness of the community.

5.4 Computing infrastructure

The model was implemented using Mathematica (Version 10, Wolfram Research Inc.). Simulations were executed using the High Performance Computing (HPC) infrastructure at Ghent University[1]. As of 2016, the UGent HPC infrastructure consisted of 568 computing nodes distributed among seven computing clusters. Every node comprises at least eight cores, resulting in a total of 11,328 processor cores available for performing intense computational tasks.

For this work, computing jobs were submitted to the HPC as array jobs, which are useful when confronted with very large numbers of jobs that are largely identical and differ only in the values of parameters they use. Hence, for each competition scheme, an array job was submitted to test a range of initial evenness values. The HPC provides an efficient implementation of array jobs, handling the computations as an array of independent tasks joined into a single job. For the simulations described in this chapter, a typical array job tested 50 different initial evenness values, computing 100 replicates of each condition, for a total of 5000 jobs; each job represented a simulation of 500 generations, or 5 million Monte Carlo steps. This required a total computing time of approximately 200 minutes for the three species model.

[1] https://www.ugent.be/hpc/

5.5 Results and discussion

We present the results of simulation studies of variable initial evenness for the three-species models. For the sake of brevity, in the results shown in the following section we use the Gini evenness index. Simulations using the other two indices (Simpson and Shannon) produce qualitatively similar results, although in the case of the Simpson index, evenness values are contained in the interval $[\frac{1}{S}, 1]$, where S is the number of species, as opposed to the unit interval as in the cases of the Gini and Shannon indices.

5.5.1 Final community configuration

As discussed in Section 5.2.2, there are two possible tournament graphs for three species: hierarchical competition and cyclic competition (Figure 5.1). In Figure 5.5, for each competition scheme we plot the final community configuration of 5.000 simulations on the simplex. Points located at the vertices represent communities with only one surviving species, and two others extinct. Points on the edges of the simplex represent communities with one species extinct (the one represented by the vertex opposite the edge) and two surviving. A point located in the centre of the simplex represents a community where the three species are present in equal proportions, i.e. where evenness is maximal. Thus the closer a point is located to the centre, the higher the evenness of that community. In contrast, a point located close to a vertex (say, species A) represents a community with a very high proportion of species A, and very low proportions of species B and C, i.e. a very uneven community. A point's colour represents the *initial* evenness of the community.

It is then immediately obvious that hierarchical competition — Figure 5.5(a) — results in extremely uneven communities. The final community configurations of almost all simulations are unsurprisingly located at the A-vertex, indicating survival of the top predator and extinction of the two other species. In no cases did the species C at the bottom of the food chain survive. In addition, all points not located at the A-vertex represent communities with with very low initial evenness. These communities were initially dominated by a species other than the top predator, thus extending the time necessary for the system to reach its steady state of complete apex predator domination.

In the case of cyclic competition — Figure 5.5(b) — the final community configurations are distributed much more evenly over the simplex, indicating that a wide variety of final configurations is possible. Most points (45%) are located in the interior of the simplex, indicating that all species persist; 29% are located on an edge,

indicating extinction of one species, whereas 27% of points are located at a vertex, indicating extinction of two species. We note that points located closer to the centre of the simplex (indicating very high final evenness) represent communities with high initial evenness. In contrast, points not in the interior (thus having suffered at least one extinction) represent communities with lower initial evenness. There is thus a positive relationship between initial evenness and coexistence.

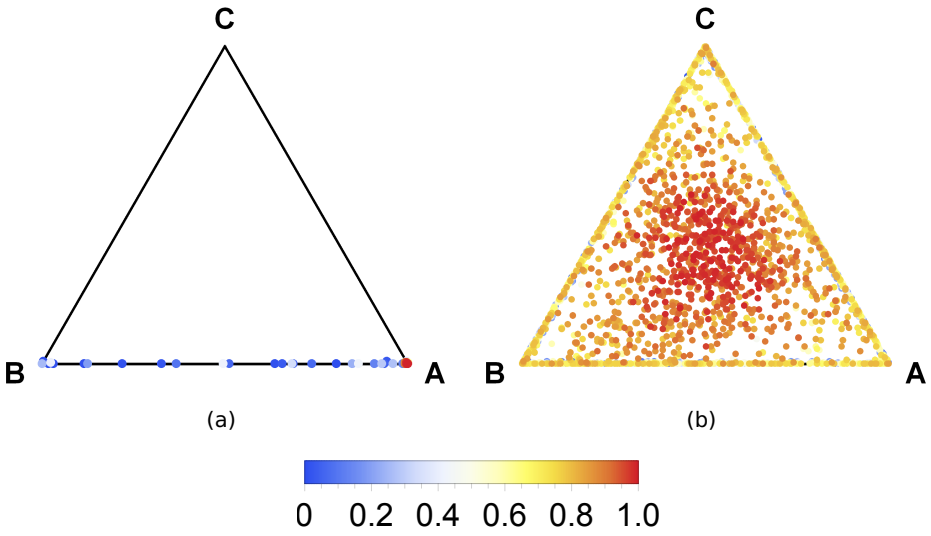


Figure 5.5: Final community configurations of 5.000 simulations for communities of three species mediated by: (a) hierarchical competition and (b) cyclic competition. Colours indicate initial community evenness, according to the colour bar legend.

5.5.2 Extinctions

We can attempt to quantify this positive relationship by looking at the probability of extinction for each species as a function of initial community evenness, shown in Figure 5.6. We firstly note that initial evenness appears to have no effect on the dynamics of communities with hierarchical competition. The dominance of the apex predator is so complete that even in initially extremely uneven communities, which may be dominated by one of the prey species, the probability of extinction is almost zero for the apex predator, and very high for the two prey species lower in the food chain. Thus when considering probability of species extinction, initial evenness does not have any effect on the system dynamics of hierarchical competition.

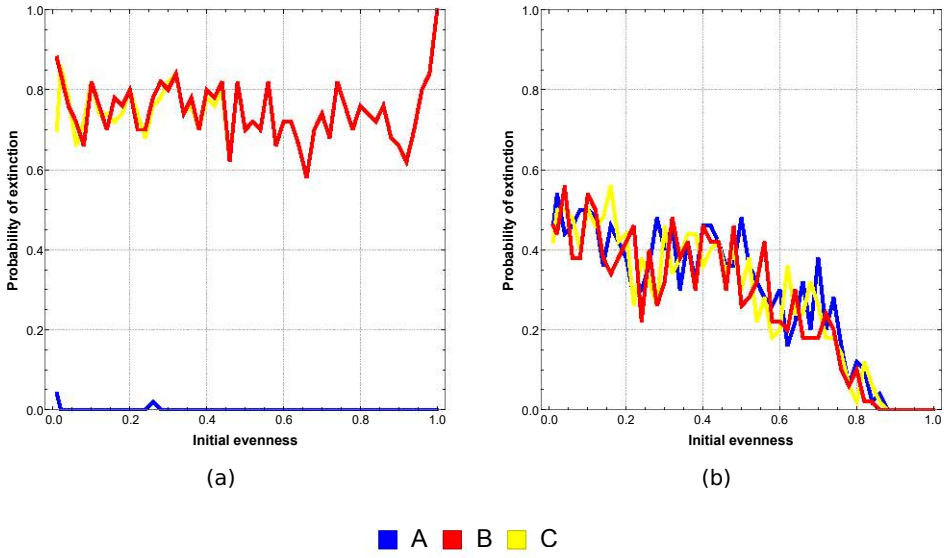


Figure 5.6: For each species, probability of extinction within 500 generations as a function of initial evenness, based on 5.000 simulations: (a) hierarchical competition and (b) cyclic competition.

In contrast, initial evenness has a marked effect on the extinction probabilities of species in cyclic competition. We first note that, as expected, extinction probabilities are extremely similar between species, due to the symmetrical nature of the competition scheme. Any differences are due to stochastic fluctuations. We also note a significant drop in probability of extinction as initial evenness increases, from as high as 0.6 for very low evenness to zero for complete initial evenness. Note that an extinction probability of zero relates to the finite and fixed simulation period, and does not imply that the species will never go extinct.

These transitions are not difficult to understand; with low initial evenness, one species outnumbers the other two, which are present in only small proportions and are hence more vulnerable to their predator, as well as to stochastic fluctuations. They are thus more easily driven to extinction. As evenness increases, the three species are present in ever more equal proportions, allowing them to stay locked in cyclic competition (thus not especially vulnerable to one predator) and less vulnerable to stochastic fluctuations.

As another measure of system stability, we can examine the time until the first extinction event (Figure 5.7). Here the differences between the two competition schemes are again evident. Hierarchical competition implies extremely rapid extinction events, so that biodiversity cannot be maintained over any significant length of time. The variability of extinction times is significantly larger for lower initial evenness. This is due to the fact that low initial evenness can imply communities where the apex predator is initially present in very low proportions. In such

cases there is an initial transient period while the apex predator grows in number, during which the other species can temporarily persist, though their numbers are declining monotonically due to predation by the rapidly increasing population of apex predators.

We note here that these results are in effect “blind” to different compositions of communities with the same initial evenness. For example, two communities may have an evenness value of say 0.2, but Community 1 has a high proportion of species A and low proportions of species B and C, while Community 2 has a high proportion of species C and low proportions of species A and B. In the majority of cases, these two communities do not evolve to give different final results. For example, in virtually all cases with hierarchical competition, species A (at the top of the chain) will dominate and persist alone, regardless of the composition of the initial uneven community. The varying compositions of the communities in general merely delay the system arriving at its steady state. The important exception to this is the case of extremely uneven communities. In such cases, all but one of the species are initially present in such low proportions that they are vulnerable to extinction due to stochastic fluctuations, regardless of their place in the competition structure. This can be seen in Figure 5.7(a), where the variability in the mean time to extinction is significantly larger for very low evenness.

Thus the time until the first extinction can be extended with lower initial evenness, but is in generally still extremely brief. On the other hand, aside from the region of very low evenness where dynamics are dominated by stochastic fluctuations, the *average* time to extinction increases slightly with initial evenness. In this narrow sense, initial evenness can be said to have a small stabilizing effect on system dynamics.

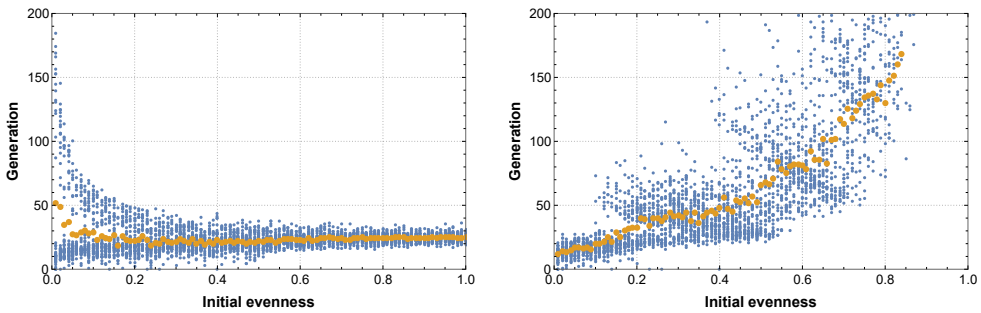


Figure 5.7: Time to first extinction event as a function of initial evenness: (a) hierarchical competition and (b) cyclic competition. Orange points represent the mean time to extinction for each initial evenness value.

In the case of cyclic competition, the behaviour is reversed in that the variability in the time until the first extinction increases with initial evenness. The average

time to extinction also increases exponentially (with growth rate 2.7). In addition, note that very few extinction events occur for evenness higher than 0.8, and none occur for evenness higher than 0.9. Thus a higher initial evenness significantly extends the region of the parameter space where diversity is maintained, supporting experimental observations that shifting to a more even community promotes biodiversity (Isbell et al., 2009b).

This finding is also supported by theoretical results. It was shown for the basic two-species (predator-prey) Lotka-Volterra model that while deterministic dynamics correspond to oscillations about an elliptic fixed point at the center (corresponding here to perfect evenness), extinction occurs via radial diffusion towards the edges (Parker and Kamenev, 2010). In our case, the initial distance from the centre of the simplex corresponds to the initial evenness and hence the larger the initial evenness, the larger the distance to the edges of the simplex, and the longer it takes for the system to reach a homogeneous state.

The mean time to extinction in well-mixed systems for cyclic competition has been well-studied in the literature (Szabó and Fath, 2007). Comparing the results from our IBM to those obtained in (Dobrinevski and Frey, 2012), we do not find a very good agreement: both the qualitative and quantitative behaviour is different. The mean time to extinction, as a function of the initial condition, is much shorter for the IBM than for the mean field approximation. This points to the important role that spatial heterogeneities play in this system, which agrees with previous modelling studies that also found significant differences in behaviour and dynamics between well-mixed and spatially heterogeneous versions of the same experimental system (Adamson and Morozov, 2012; Schreiber and Killingback, 2013; Laird, 2014).

5.6 Conclusions

Two competition schemes have been examined with simulation studies modelling various possible communities, which result in qualitatively different coexistence and extinction scenarios. System behaviour is strongly dependent on initial evenness and competition scheme.

The dynamics induced by the hierarchical competition scheme do not permit coexistence of all species. There are frequent extinction events, which typically occur very rapidly. Varying the community's initial evenness cannot counteract the competitive dynamics which necessarily result in the persistence of a single species, with the other two species quickly collapsing to extinction. Very low initial evenness can only extend the initial transient period before the system settles to its steady state, because in this case the community can be initialized with the apex predator present in initially very low proportions. This effect is lost as initial

evenness increases, after which the community dynamics become insensitive to varying initial evenness.

In the case of the cyclic competition scheme, low initial evenness can counteract the stabilizing dynamics of the competition scheme and provoke extinctions. In contrast, higher initial evenness can have an important stabilizing effect, in the sense that the time until the first species extinction increases significantly as initial evenness increases. By extending the region of biodiversity in this way, there is sufficient time for system behaviour to be affected by other factors such as competition scheme, rates of competition and mobility. These results support experimental observations that biodiversity is promoted by increasing evenness (Isbell et al., 2009b).

Our results demonstrate the danger in overlooking variable community evenness and making the typical assumption that communities are maximally even, despite mounting evidence to the contrary (Wilsey and Polley, 2004; Huston, 1997; Grime, 1998; Smith and Knapp, 2003). This oversight also ignores the fact that damages due to human actions can affect the evenness of natural communities, often making them more vulnerable to invasion, stresses and disturbances (Wittebolle et al., 2009). While theoretical studies such as this one are beginning to increase in number, experimental studies to validate their conclusions are still lacking (Isbell et al., 2009b).

6

The impact of initial evenness and invasion on biodiversity in a four-species *in silico* microbial community

6.1 Introduction

In Chapter 5, we have investigated the effects of variable community evenness on communities' ability to maintain their diversity when there is competition occurring between the community members. We focused in particular on non-transitive competition, since this has been shown to be key in preserving community diversity and functionality (Hillebrand et al., 2008; Wittebolle et al., 2009; De Roy et al., 2013), and has not previously been investigated in conjunction with variable community evenness (Case et al., 2010; Cheng et al., 2014; Frachebourg et al., 1996).

We now shift our focus to communities of four interacting species. This setting is much less studied than the three-species case (Szabó and Fath, 2007). It not only admits more complicated competition schemes, but also allows for different experimental approaches to be considered.

We have previously discussed the different experimental techniques that are used to study interactions in microbial communities, in particular competitive interactions. The first of two broad approaches are co-cultures, where all species are inoculated at the start of the experiment and allowed to evolve together towards a steady state, whatever it might be. This was discussed in detail in Section 2.3.3. The second experimental approach considers invasion, where an alien species infiltrates a stable resident community and must compete with the resident community members in order to establish itself. Invasion theory and experiments were discussed in Section 3.4.2.3.

It has been shown that these two different experimental set-ups can result in very different outcomes even when using the same group of microbial species (Tan et al., 2015; Sekhar et al., 2016; Gilbert et al., 2003; Yoshida et al., 2009) (Horemans *et al.*, 2017, in prep).

With only three species, it is not possible to investigate invasion when the species are engaged in non-transitive competition, since the initial absence of one species results in unbalanced dynamics and leads to extinctions. In contrast, with four species we may begin an experiment with a stable coexisting community of three species, and later add a fourth species.

We therefore develop an IBM of four interacting species, in order to study the effects of non-transitive competition and variable community evenness on the maintenance of community diversity, under the two different experimental set-ups defined above. In this way, we can address research questions 1 and 2 (see Section 1.2) in a more complex setting than in Chapter 5, while also considering question 3:

- What effect does initial evenness have on maintaining community diversity?
- Which types of competitive interactions can help maintain community diversity, and which types can threaten it?
- What effect does initial evenness have when a community is faced with invasion?

This chapter is structured as follows. In Section 6.2, we study non-transitive competition and variable evenness via the use of *in silico* co-culture experiments, in order to understand their effects on community biodiversity. Thus in Section 6.2.1 we discuss non-transitive competition in the case of four species, and in Section 6.2.2 outline the predictions of the mean field approximation in this case.

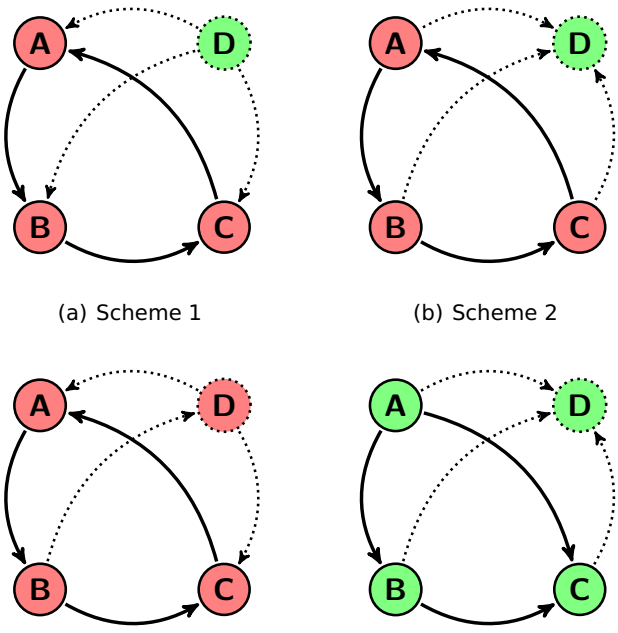
In Section 6.2.3, we present our IBM incorporating cyclic competition and variable evenness, and describe the set-up of the co-culture simulation studies, the results of which are presented and discussed in Section 6.2.4.

Then in Section 6.3 we study the same phenomena of cyclic competition and variable evenness in the context of invasion, again to understand the effects on community diversity and stability. The corresponding IBM and *in silico* experimental set-up are described in Section 6.3.1, and the results of these experiments are presented and discussed in Section 6.3.2. Finally, in Section 6.4 conclusions regarding the two different set-ups are drawn.

6.2 Co-culture experiments

6.2.1 Competition

Competition in the four-species setting is more complex than in the three-species case, since there are now four possible tournament graphs (up to a re-labelling of species) for fully connected competition between four species (Figure 6.1).



(a) Scheme 1 (b) Scheme 2

(c) Scheme 3 (d) Scheme 4

Figure 6.1: Competition between four species.

Schemes 1, 2 and 3 involve three species engaged in cyclic competition between themselves, with a fourth species that interacts with the others in various ways. In Scheme 1, the fourth species dominates all others (and is referred to as the top, or apex, predator); in Scheme 2 it is preyed upon by all others; and in Scheme 3 it dominates two species and is preyed upon by the third. This is shown in Figure 6.1, where those species coloured in red are members of a sub-cycle of three species, while those in green are not. Note that Scheme 3 contains two sub-cycles of three species (A-B-C and A-B-D), while Schemes 1 and 2 each contain only one. Finally, Scheme 4 is the only competition scheme that does not contain a sub-cycle of three species, but rather is a strict competitive hierarchy with the ranking A, B, C, D.

If we relax our requirement that the competition structure be fully connected, we can find many simpler cases. We mention in particular the case of a cycle with neutral pairs (Figure 6.2), which has been extensively studied in the literature (see e.g. Szabó and Szanaider (2004), Intoy and Pleiming (2013), Lütz *et al.* (2013), and Durney *et al.* (2012)). In this case, each species interacts with only two other species (one prey, one predator), and ignores the third. Therefore there are two pairs of non-interacting species, A paired with C and B paired with D, whence the term neutral pairs (Durney et al., 2011), and we are left with a simple cycle of four species.

This four-species neutral pairs scheme produces behaviour qualitatively different from the corresponding three-species case (Figure 5.1(a)), which typically results in either sustained coexistence of all three species or (in the case of oscillations that approach too closely the edge of the simplex) extinction of two species and persistence of the third. In contrast, the four-species neutral pairs case typically results in communities of two non-interacting species (A and C, or B and D) or communities of all four species that are located in the simplex along a gradient connecting the two non-interacting steady states (Dobrinevski and Frey, 2012).

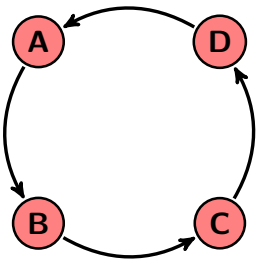


Figure 6.2: Cyclic competition with neutral pairs.

6.2.2 Mean field analysis

In the setting of four species, the mean field approach can be applied to the case of a simple cycle with neutral pairs (Figure 6.2), yielding the following system of ODEs modelling the population densities A, B, C and D of the four species through time:

$$\begin{cases} \frac{dA}{dt} = A[\mu(1-\Omega)-\sigma D] \\ \frac{dB}{dt} = B[\mu(1-\Omega)-\sigma A] \\ \frac{dC}{dt} = C[\mu(1-\Omega)-\sigma B] \\ \frac{dD}{dt} = D[\mu(1-\Omega)-\sigma C] \end{cases} \tag{6.1}$$

where $\Omega = A + B + C + D$.

However, even in this simpler case, the mean field approach fails to yield insight into extinctions (Case et al., 2010). This is because the mean field assumption is only suitable for large populations; the fractions in the mean field equations will never vanish in finite time and hence the approximation breaks down near the extinction of one or more species (Durney et al., 2011). Thus the mean field equations cannot be used to analytically determine either the probability of extinction for each species or the average time to extinction.

Thus for more precise insights into the dynamics of the system, we must turn to other approaches. One study uses a PDE, known as the Fokker-Planck equation, to approximate the dynamics of the stochastic system; the Fokker-Planck equation can then in turn be reformulated as an SPDE called the Langevin equation (Dobrinevski and Frey, 2012). Stochastic simulations of this equation can then be used to obtain mean extinction times and extinction probabilities as a function of the system's parameter values (Dobrinevski and Frey, 2012). However, it should be noted that this approximation of the stochastic system can result in exponentially large errors (Doering et al., 2005).

In the more complex case of fully connected competition (Figure 6.1), we show System (6.2) as an example of the system of ODEs one obtains by applying the mean field approach to Scheme 2 (Figure 6.1b).

$$\begin{cases} \dfrac{dA}{dt} = A[\mu(1-P) - \sigma C] \\ \dfrac{dB}{dt} = B[\mu(1-P) - \sigma A] \\ \dfrac{dC}{dt} = C[\mu(1-P) - \sigma B] \\ \dfrac{dD}{dt} = D[\mu(1-P) - \sigma(A+B+C)] \end{cases} \tag{6.2}$$

Similar systems can easily be obtained for the other fully connected competition schemes. These equations admit several equilibria, whose linear stability can be determined from the governing Jacobian, more specifically the sign of the real part of its eigenvalues. A fixed point is unstable if at least one of the corresponding eigenvalues has a positive real part; otherwise it is stable (Glendinning, 1994).

The system of ODEs representing Scheme 1 (Figure 6.1a) has five equilibrium points: $x_1^* = (1,0,0,0)$, $x_2^* = (0,1,0,0)$, $x_3^* = (0,0,1,0)$, $x_4^* = (0,0,0,1)$ and $x_5^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0)$. The first four equilibria are homogeneous states representing the persistence of a single species, while the fifth equilibrium is the only one permitting a level of coexistence. Of these five equilibria, all are unstable except for x_4^*, which indicates the persistence of species D alone. This is unsurprising given that species D is the apex predator in this competition scheme, and hence dominates the competitive interactions.

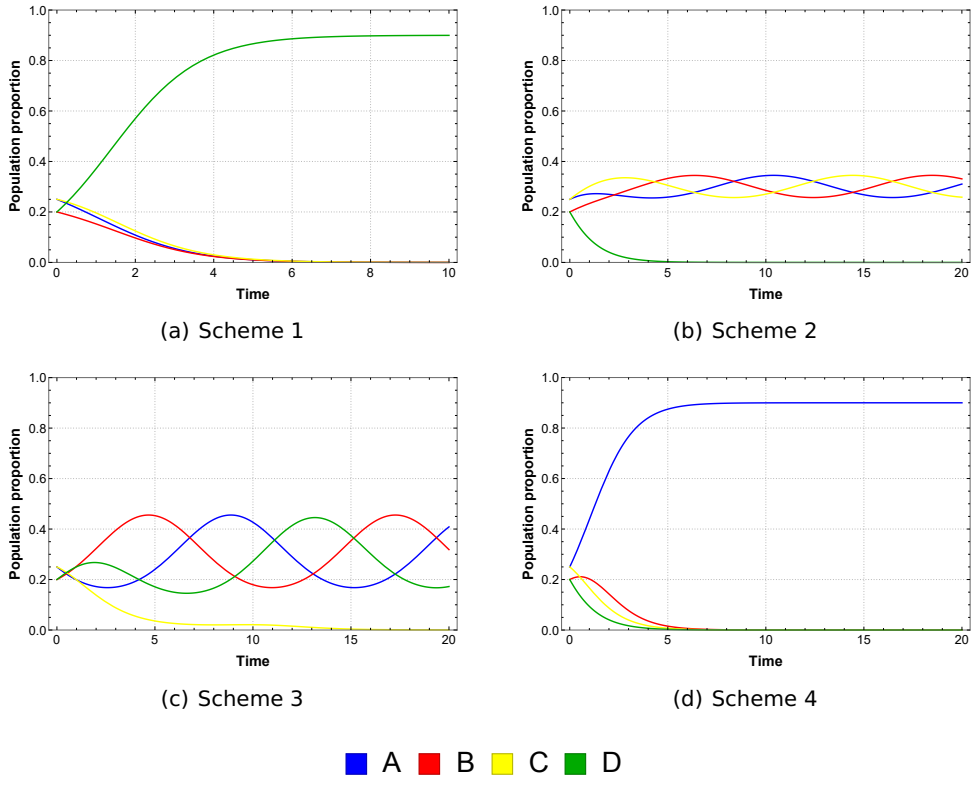
Scheme 2 (Figure 6.1b) admits the same five equilibrium points as Scheme 1, but in this case the only stable state is x_5^*: coexistence of the three species in cyclic competition and extinction of the fourth, species D. All other equilibria are unstable. Again this reflects what we would intuitively expect from this competition scheme — the species that is preyed upon by all others collapses to extinction, while the three remaining species persist in cyclic competition identical to System (5.2), the case of three species in cyclic competition.

Scheme 3 (Figure 6.1c) admits an additional coexistence equilibrium $x_6^* = (\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3})$, which is unstable. In this case, the stable states are x_2^* and x_4^*, indicating the sole persistence of either species B or species D.

Scheme 4 (Figure 6.1d) has only four equilibrium points — it does not admit any steady state permitting coexistence, only the four equilibrium points x_1^* to x_4^*. Of these, only x_1^* is stable, representing the homogeneous state of species A persistence.

In Figure 6.3 below we show the evolution of the species fractions for each competition scheme, starting from an initial condition with fractions $(0.25, 0.2, 0.25, 0.2)$ for species A - D respectively. The differences between the competition schemes in

the mean field approximation are clear: Schemes 1 and 4 quickly settle to a homogeneous steady state, with a single species persisting and all others extinct, while Schemes 2 and 3 result in a fast extinction of one species, with the other three persisting in oscillating fractions. The amplitude of the oscillations is greater, and the period shorter, for Scheme 3 than for Scheme 2.



(a) Scheme 1 (b) Scheme 2

(c) Scheme 3 (d) Scheme 4

Figure 6.3: Mean field evolution of the species fractions over time, for each competition scheme.

For the fully connected cases, no invariant of motion exists that involves all four species fractions. In contrast, the neutral pairs case (Figure 6.2) admits two such invariants, which therefore prevent any species collapsing to extinction (Case et al., 2010). This is why coexistence equilibria involving all four species can be found for the neutral pairs case, where the species fractions oscillate indefinitely along a fixed, closed orbit in the phase space (Dobrinevski and Frey, 2012). This does not hold for the fully connected cases. We can find an invariant of motion if we allow one species to go extinct. For Scheme 2, if $D = 0$ then the algebraic product ABC is an invariant, while for Scheme 3, if $C = 0$ then the algebraic product ABD is an invariant (Lütz et al., 2013). Thus the mean field equations suggest that extinction events occur rapidly and frequently for these competition schemes.

6.2.3 Model description

Given the limitations of the mean field approach discussed in Section 6.2.2, we again turn to computational methods to study the effects of both variable evenness and non-transitive competition on the maintenance of biodiversity in this system. We therefore extend the model developed in Chapter 5 to incorporate a fourth species, and the corresponding competition schemes. We describe our IBM using an abbreviated ODD (Overview, Design Concepts and Details) protocol. Since the four-species model is constructed in an analogous way to the three-species model, we address only the sections of the ODD protocol where the four-species model differs from the three-species model.

6.2.3.1 Overview

Purpose

The aim of the model is again to investigate how variable initial evenness and non-transitive competition between individuals affect the maintenance of biodiversity, this time for a community of four species.

State variables and scales

The model is a two-dimensional representation of an experimental domain divided into a regular grid of size $L \times L = N$, and populated by a community of four species denoted by A, B, C and D. Each grid site is either occupied by a single individual, or is empty. Individuals are characterized by two state variables: grid position (x, y) and species identity $s \in \{A, B, C, D\}$.

Process overview

Three possible interactions can occur: reproduction, competition and mobility. The mechanisms of these interactions are implemented in the same way as for the three-species model (see Section 5.3.1.3). These are illustrated in Figure 6.4 for the case of four species.

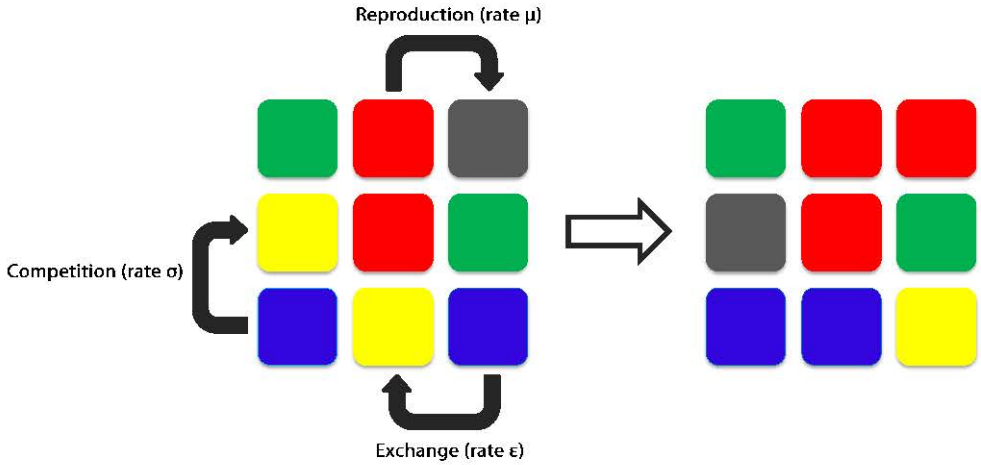


Figure 6.4: Mechanisms of demographic processes, for individual *in silico* bacteria of four species.

Scheduling

The IBM proceeds in the same way as the three-species model, using a modified version of the Gillespie algorithm to determine which interaction occurs at each time step (Section 5.3.1.4). Recall that time steps are aggregated into generations: the number of time steps for, on average, each individual to be the subject of one interaction event, i.e. N Monte Carlo steps for a grid of size $L \times L = N$. The length of the simulation is then defined by the number of generations for which the model is evolved.

6.2.3.2 Design concepts

The design concepts of the four-species IBM are the same as those of the three-species IBM (Section 5.3.2).

6.2.3.3 Details

Initialization

The model is initialized with a random spatial distribution of individuals and empty sites. All species are present in this initial seeding, representing a co-culture experiment of simultaneous inoculation of all species. The initial species abundance distribution is determined by a selected evenness value, which is used to generate

a vector of population abundances. Individuals are then randomly placed in the remaining grid sites according to these proportions.

We again use sampling algorithms to generate vectors of population proportions satisfying a given initial evenness, for our three selected indices: Simpson, Gini and Shannon. In the four-species setting, the space of all possible population vectors is the 3-simplex, also called a tetrahedron (Figure 6.5).

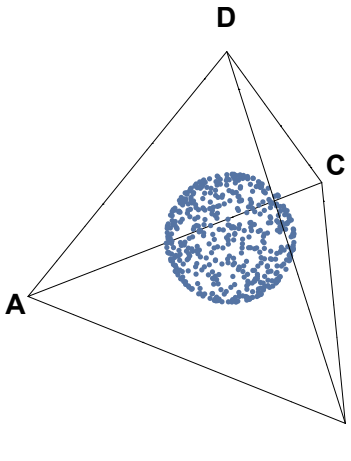


Figure 6.5: Population vectors for four-species communities corresponding to a fixed evenness value of 0.8 (Gini) plotted in the 3-simplex. In this representation, points closer to the centre of the simplex represent population vectors that are more even, while points closer to the vertices represent population vectors which are more uneven. For example, a point at the vertex B represent the population vector $(0, 1, 0, 0)$, while a point at the centre represents the population vector $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.

Aside from the input variables, all other parameters used to initialize the model are fixed for all simulations, and are the same as those for the three-species model (shown in Table 5.1).

6

Input

The input variables are: the competition scheme, and the initial community evenness.

First, the competition scheme is specified. The schemes under investigation are those shown in Figure 6.1. These consist of all possible fully connected competition structures for the richness level S under investigation: four schemes for $S = 4$. The rules of the particular scheme are used to determine the outcome of competitive interactions as described in Section 6.2.3.1.

For each competition scheme, we sample a range of initial evenness values, systematically chosen so as to sample the entire range of possible values: $[0, 1]$ for the Gini and Shannon indices, and $\left[0, \frac{1}{S}\right]$ for the Simpson index, where S is the richness of the community.

6.2.4 Results and discussion

We present the results of simulation studies of variable initial evenness for the four-species model using the co-culture set-up. For the sake of brevity, in the results shown in the following section we use the Gini evenness index. Simulations using the other two indices (Simpson and Shannon) produce qualitatively similar results, although in the case of the Simpson index, evenness values are contained in the interval $[\frac{1}{S}, 1]$, where S is the number of species, as opposed to the unit interval as in the cases of the Gini and Shannon indices.

As discussed in Section 6.2.1, there are four possible tournament graphs in the four-species setting (Figure 6.1). Simulations show that long-term system behaviour is again strongly dependent on both initial community evenness and the details of the competition scheme. High initial evenness extends the region of the parameter space where biodiversity is preserved, in addition to prolonging the time until the first species extinction. Hence there is sufficient time for other factors (such as details of competition scheme, rates of competition and mobility, etc.) to affect the system dynamics. In contrast, if evenness is too low, biodiversity can be lost before other emergent behaviour can be observed.

6.2.4.1 Final community configuration

In Figure 6.6, the final population configurations of 5.000 simulations are plotted on the 3-simplex. Note that when plotting population vectors on the 3-simplex, points located at the vertices correspond to extinction of all but one species, edges correspond to extinction of two out of the four species, and faces correspond to extinction of one species. Points located in the interior of the simplex thus represent communities with coexistence of all four species. Note that the faces of the simplex are “absorbing” in the sense that trajectories into a face are irreversible — once a species has become extinct it cannot reappear.

Echoing the mean field approximations (Figure 6.3), we find that communities governed by Schemes 1 or 4 are almost always dominated by the top predator — all other species quickly collapse to extinction, while the apex predator persists alone.

Table 6.1: Location of final community configurations in the 3-simplex, for simulation length of 200 generations. Points in the interior represent communities with all four species coexisting, those on a face represent communities with three species coexisting, those on an edge represent communities with two species coexisting, and points at a vertex represent monocultures.

Scheme	Interior	Face	Edge	Vertex
1	0%	0%	6%	94%
2	1%	23%	38%	38%
3	0%	1%	15%	84%
4	0%	0%	6%	94%

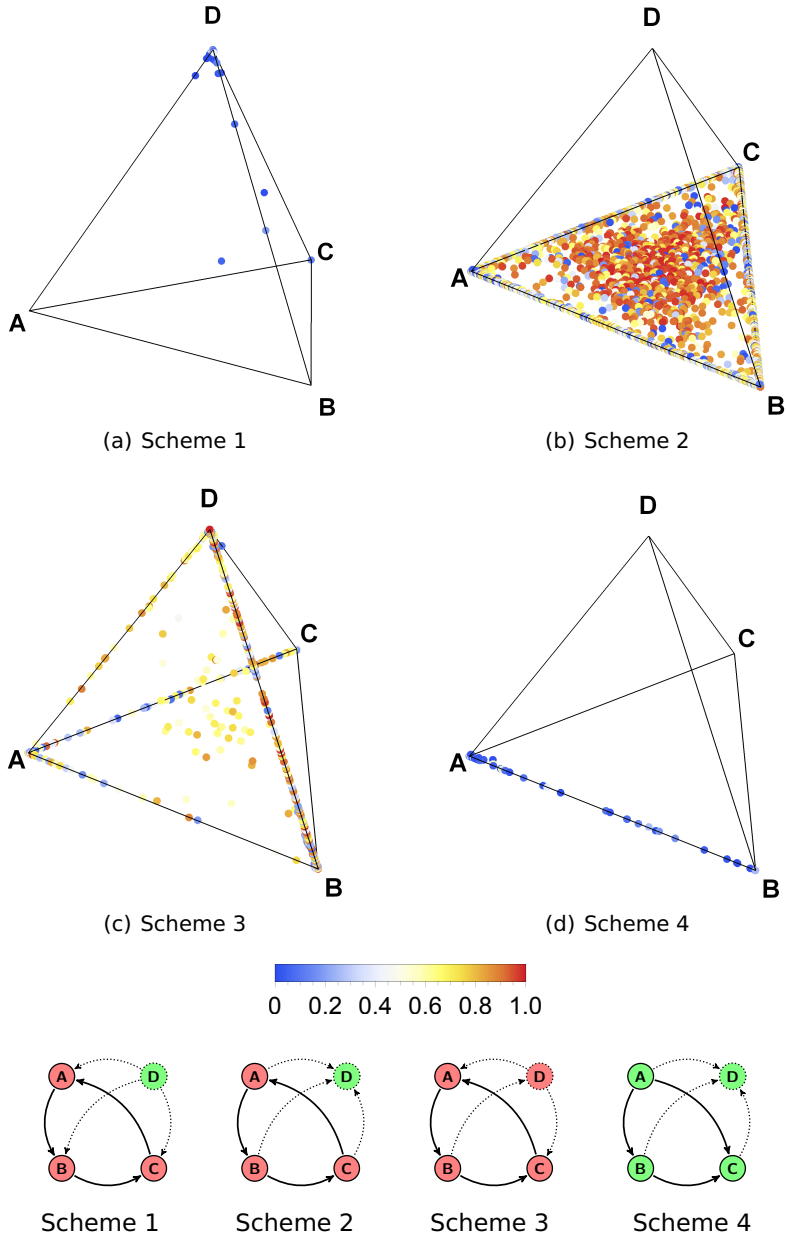


Figure 6.6: Final community configurations for the four different competition schemes. Colours indicate initial community evenness, according to the colour bar legend.

This trajectory is followed in 94% of simulations (see Table 1). The few simulations that do not result in three extinctions instead result in two — not a great difference in terms of biodiversity. Of these, all involve communities that were initially very uneven. This suggests that these communities were initially dominated by a species other than the apex predator, delaying the system's trajectory towards its steady state. This type of behaviour at low evenness has also been noted experimentally, where "idiosyncratic effects" of dominant species at lower evenness initially dampened resource uptake by the community as a whole (Wilsey and Polley, 2004).

In both schemes, there are no points located in the interior of the simplex, indicating that extinction events always occur. Indeed, in the vast majority of cases the maximum of three extinctions occurs. Those points not located at the vertex of the apex predator represent communities with low initial evenness.

In communities subject to Scheme 2, the species preyed upon by all others quickly collapses to extinction, reducing the system dynamics to the cyclic competition seen in the three-species model. Note the similarity between the A-B-C plane in Figure 6.6(b) and the simplex in Figure 5.5(b).

Communities governed by Scheme 3 exhibit richer behaviour, particularly at higher initial evenness. No single species is especially dominant, nor is one species weak to the extent seen under other competition schemes. Extinctions are again ubiquitous — there are almost no points located in the interior of the simplex. Most points are located at a vertex or an edge, indicating at least two extinctions. The few points located on a face are to be found on the A-B-D face, suggesting that when only one species is lost, it is species C; results in the next section regarding extinction further support this observation, as did the mean field approximation in Section 6.2.2. Additionally, these points represent communities with high initial evenness, again suggesting that higher initial evenness supports the preservation of biodiversity.

We note the dependence of some of the results in Table 6.1 on the simulation time — if we extend the simulation time, these probabilities may change. For Schemes 1 and 4, if the simulation time is doubled these probabilities barely change: again approximately 95% of simulations result in lone persistence of the apex predator. This robustness to the maximal simulation time indicates that a certain amount of deviation from the steady state exists due to stochastic fluctuations, which is not mitigated by extending the simulation time. In contrast, for Schemes 2 and 3, doubling the simulation time roughly halves the number of simulations not located at a vertex. This indicates that for these schemes, those simulations that do not complete a trajectory to a steady state are only delayed, and not deflected elsewhere as for Schemes 1 and 4.

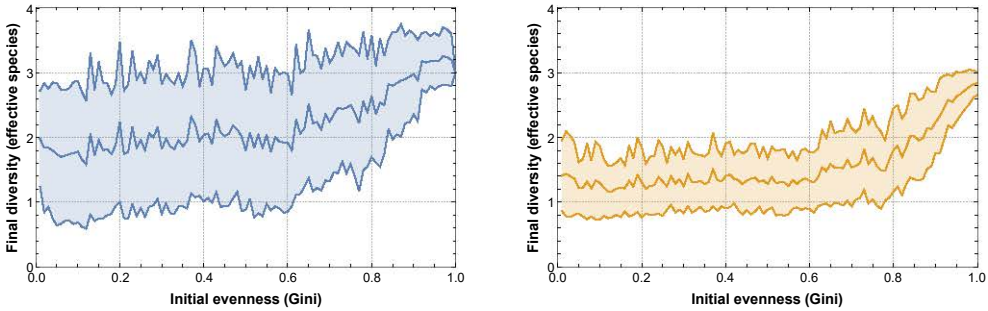
Finally, we can examine the change in community diversity for each competition scheme. For this purpose, we use the Leinster-Cobbold index described in Sec-

tion 4.2.6.2. Since all species in our system are equally distinct, it is most appropriate to use the identity matrix as the similarity matrix. Recall that this index permits the use of diversity profiles, where diversity is plotted as a function of the sensitivity parameter q, which measures the relative weight assigned to rare and common species; $q = 0$ corresponds to species richness, $q = 1$ weighs all species equally, and higher values of q give more weight to common species.

The main purpose of diversity profiles is to study the effects of giving different weights to rare and common species. For the system under consideration, this will not be very informative since there are relatively few species in our community. We instead look at how diversity varies as a function of initial evenness, for different values of the sensitivity parameter q.

For the two schemes that are dominated by a single apex predator (Schemes 1 and 4), final community diversity is extremely low, as expected given the frequent and numerous extinction events. These plots are therefore omitted. For Schemes 2 and 3, we show in Figures 6.7 and 6.8 respectively the mean final community diversity as a function of initial community evenness, for $q = 0$ and $q = 1$.

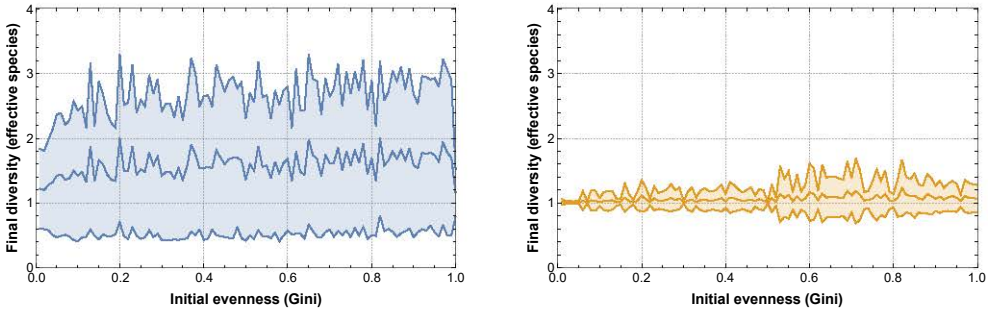
For Scheme 2, the diversity of order zero (equivalent to species richness) is shown in Figure 6.7(a), and demonstrates a positive relationship between initial evenness and final richness, with an increase in the order of one species between low and high initial evenness. However even with very high or complete evenness, coexistence of all four species is not possible. Recall that the typical behaviour for communities under this competition scheme is the rapid extinction of one species; the dynamics is then reduced to the cyclic three-species case. There is also fairly large variability for lower evenness, indicating the possibility of more than one species extinction. For $q = 1$ (Figure 6.7(b)), the index now takes evenness into account, and we find a stronger positive relationship between initial community evenness and final community diversity, with less variability. We note again that diversity higher than three effective species is not possible even for very high initial evenness values.



(a) $q = 0$ (b) $q = 1$

Figure 6.7: Mean final diversity (with standard deviation) as a function of initial evenness, for Scheme 2 with (a) $q = 0$, and (b) $q = 1$.

For Scheme 3, the diversity of order zero is shown in Figure 6.8(a). Here there is no apparent relationship between initial community evenness and final community richness, which remains low for all initial evenness values, indicating the frequency of extinction events. For $q = 1$ (Figure 6.8(b)), there is again no evident relationship between final community diversity and initial community evenness, although the variability in final community diversity does increase with initial evenness. Final community diversity is quite low for all simulations, indicating that communities that do persist with more than one species tend to be fairly uneven.

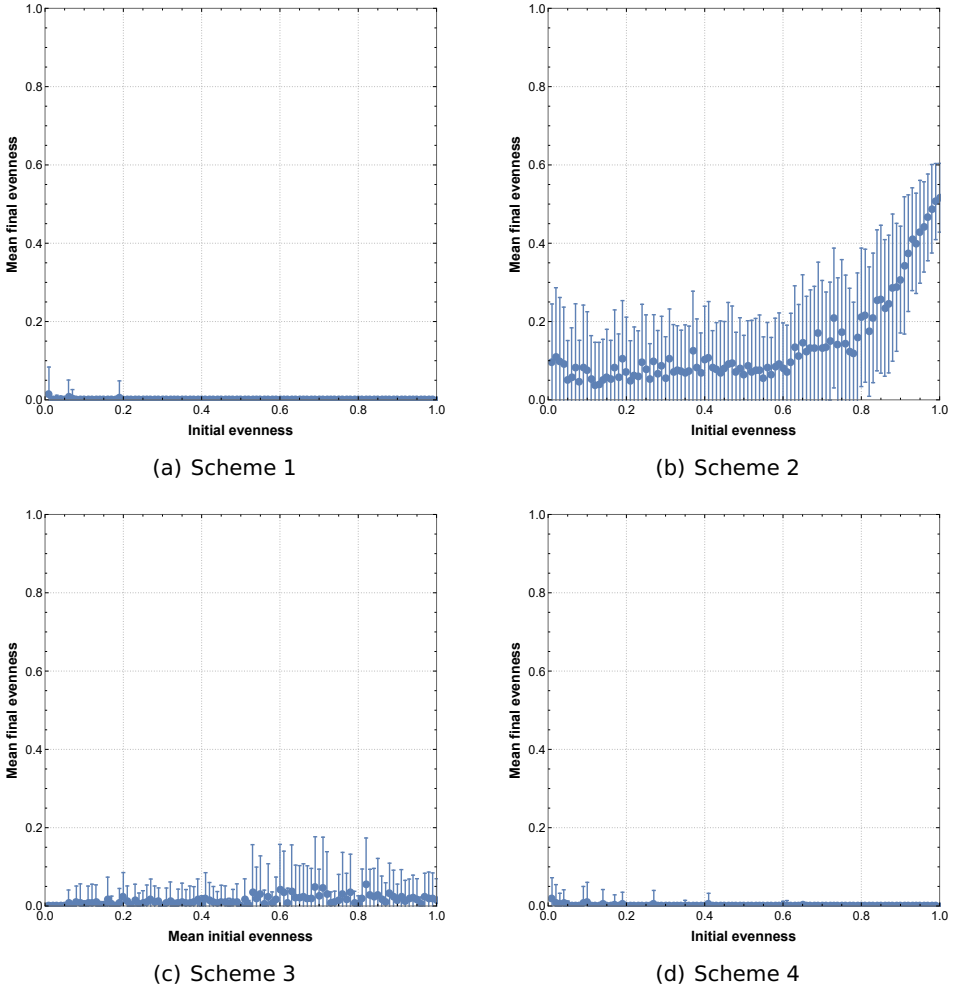


(a) $q = 0$ (b) $q = 1$

Figure 6.8: Mean final diversity (with standard deviation) as a function of initial evenness, for Scheme 3 with (a) $q = 0$, and (b) $q = 1$.

To examine the effect on evenness in isolation, in Figure 6.9 we plot the mean final community evenness as a function of initial community evenness, for the 5.000 simulations. We note the same general relationships as were found for the diversities of order one (Figures 6.7(b) and 6.8(b)) In the case of Scheme 2, the mean final evenness can remain quite high, and generally increases with initial evenness. However, final community evenness always remains under a threshold rep-

resenting the loss of one species. In contrast, in communities subject to Scheme 3 the final community evenness is again generally low, indicating the frequency of extinction events, although the variability in the final evenness increases.



(a) Scheme 1

(b) Scheme 2

(c) Scheme 3

(d) Scheme 4

Figure 6.9: Mean final community evenness (with standard deviation) as a function of initial evenness, for the four competition schemes.

6.2.4.2 Extinctions

Having examined how the final state of communities varies with initial evenness, we now turn our attention to extinction events. We have seen that these occur frequently, and asymmetrically between species. In Figure 6.10 we plot for each competition scheme, the probability of extinction of each species as a function of initial evenness.

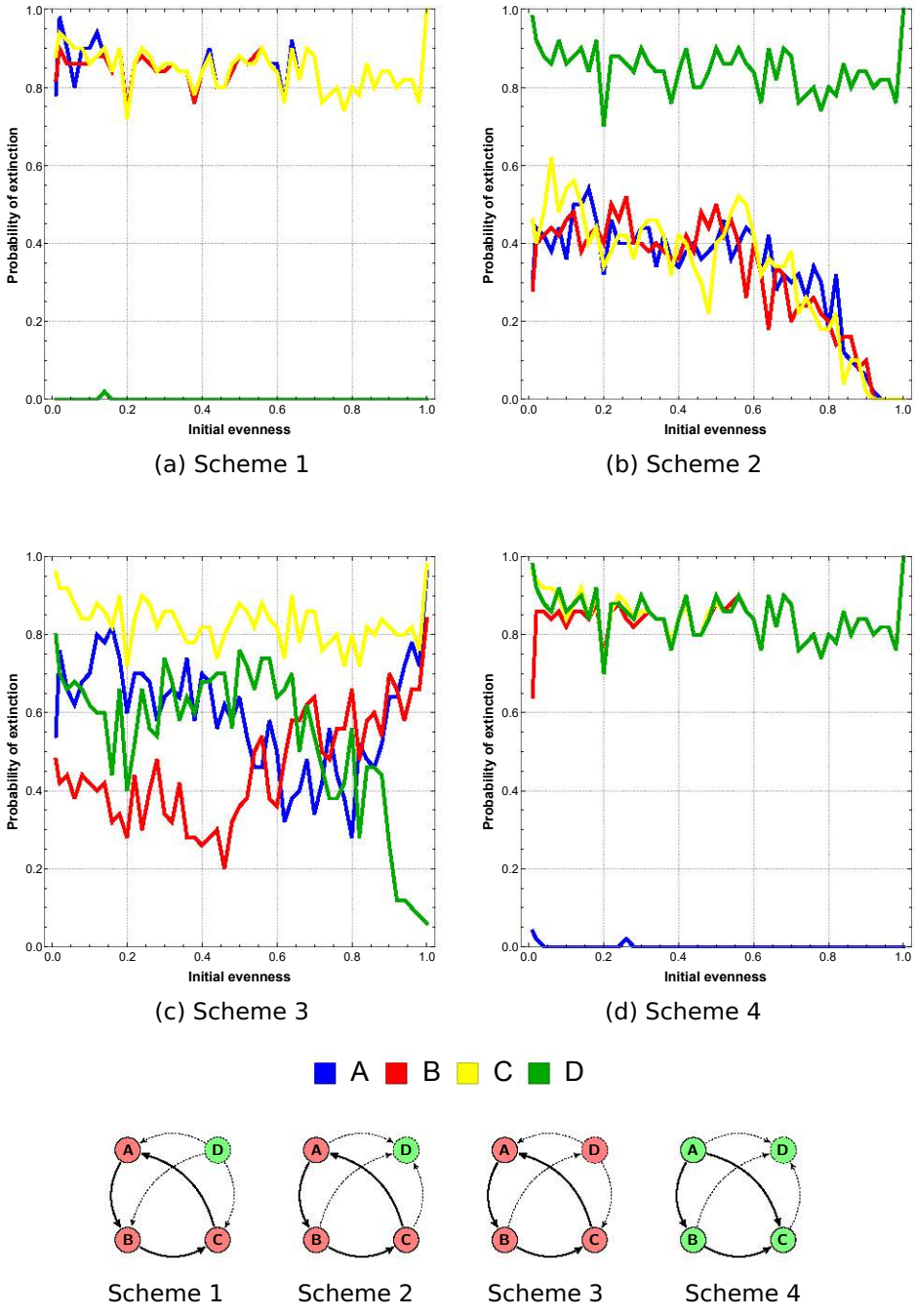


Figure 6.10: Probability of extinction within 500 generations as a function of initial community evenness, defined as the number of simulations per evenness condition that resulted in the extinction of at least one species.

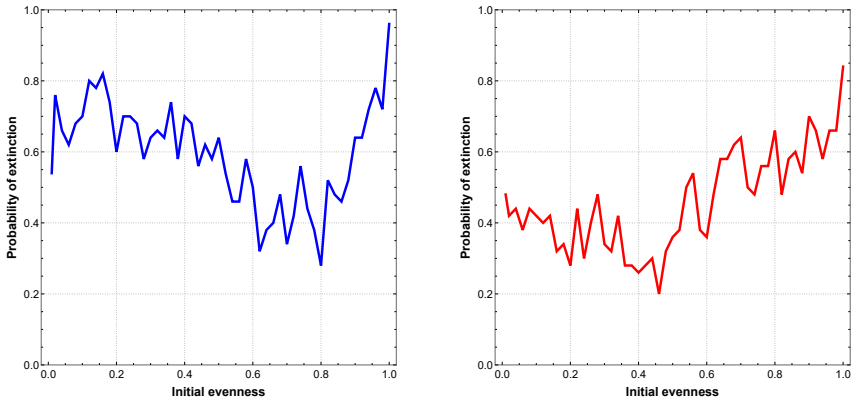
The behaviour of the extinction probabilities reflects the dynamics seen in Chap-

ter 5 for three species, and also mirrors the mean field approximations discussed in Section 6.2.2. The extinction probabilities of species under Schemes 1 and 4 again illustrate the dominance of the apex predator, with the other species very vulnerable to extinction. Increasing initial evenness cannot mitigate the effects of the competition scheme in these cases.

For Scheme 2, we note that the species at the bottom of the food chain cannot be saved from extinction by increasing initial evenness. Rather, species D has a high probability of extinction for all evenness values. However, the other three species (which are engaged in cyclic competition between themselves) benefit from a lower probability of extinction as initial evenness increases. These probabilities decline to nearly zero for sufficiently high initial evenness. Thus high evenness preserves the stabilizing effects of the competition scheme for those species that benefit from it, while it is unable to counteract the effects for the species that suffers.

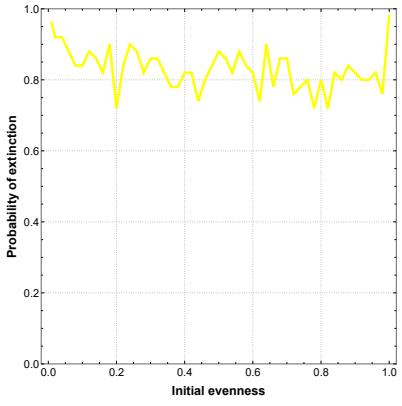
Communities governed by Scheme 3 exhibit richer behaviour, particularly at higher initial evenness, as can be seen in Figure 6.11. No single species is especially dominant, nor is one species weak to the extent seen under other competition schemes. Recall that in Scheme 3, there are two sub-cycles of three species — a cycle A-B-C and a cycle A-B-D (Figure 6.3(d)). Between them, species B can be considered the “stronger” in the sense that it has two preys and only one predator, while species A has one prey and two predators. However, for high evenness, species B has a higher probability of extinction than species A, an analogue of the “survival of the weakest" law seen in the three-species case (Reichenbach et al., 2007) and supported by experimental observations (Berr et al., 2009).

The mechanism that has been suggested to explain this “survival of the weakest" is that at high evenness, species B initially grows quickly as a result of dominating its two prey species. This leads to a sharp drop in species C (which is also preyed on by species D), permitting a rapid increase in the levels of its prey, species A. This in turns leads to a crash in species B, being killed off by the numerous individuals of species A preying on it. This mechanism explains at once the high probability of extinction for species C and species B and the low probability of extinction for species D at high evenness, because species B is the only predator of species D. So as the population levels of species B crash, the population levels of species D may rise unchecked.

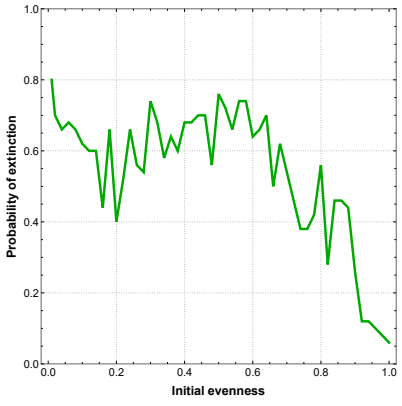


(a) Species A

(b) Species B



(c) Species C



(d) Species D

Figure 6.11: For each species, probability of extinction within 500 generations under Scheme 3.

To attempt to elucidate the relative "strength" of each species in the different competition schemes, we can look at the time before each species goes extinct and rank the species according to longer persistence. Thus the species that persists the longest will be ranked first, while the first species to go extinct is ranked last. We can then follow these rankings as initial evenness is varied (Figure 6.12). Note that the rankings shown are averages of many simulations, and thus are not integer-valued.

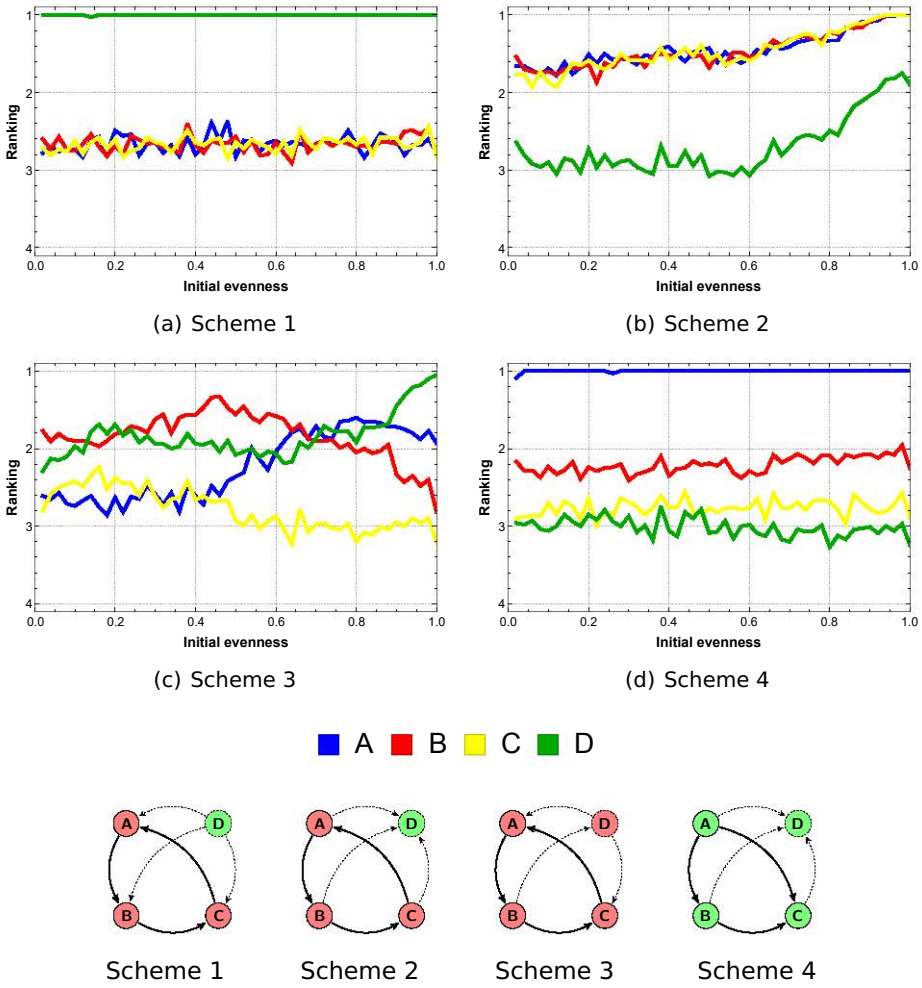


Figure 6.12: For each competition scheme, ranking of species in terms of longer persistence, as a function of initial community evenness.

The values of the rankings are as expected given the probabilities of extinction examined in Figure 6.10. For Schemes 1 and 2, the rankings of the species A, B and C that are in cyclic competition are extremely similar (oscillating slightly due to stochastic events), with species D respectively leading or trailing them in rank. We again observe the positive effect of higher initial evenness on the persistence time for Scheme 2, but not for Scheme 1. For Scheme 4, we see clearly the hierarchy imposed by the competition scheme, with a strict ranking A-B-C-D (we note occasional overlap between species C and D, due to extremely short time between their respective extinctions).

The most interesting feature in Figure 6.12 is that Scheme 3 is the only scheme in which the order of rankings changes with initial evenness. So whereas for low

initial evenness species B is ranked first in terms of persistence, as evenness increases the highest ranked species changes between species A, B and D until finally species D takes first place. The mechanism driving this behaviour is the same one underlying the "weakness" of species C, which also leads to its consistently low ranking across initial evenness values.

As evenness increases, the dynamics of the competition scheme exerts a strong effect on the system behaviour as the species proportions become more equal. Then we find the situation described above for Figure 6.11: species C has the highest probability of extinction (and hence the lowest ranking), while species D has the lowest probability of extinction and thus the highest ranking.

This also explains why the rankings of species A and C differ at higher evenness, whereas for low evenness they are nearly equal: as evenness increases and species C starts to decline in proportion, species A benefits from a decrease in predation due to species C, and its ranking increases accordingly. We note the same trend for species B and D: as one declines in ranking, the other increases.

In contrast, at low initial evenness the "survival of the weakest" mechanism is counteracted. In uneven communities, one species dominates, while the other three are present in small proportions. All species have a similar chance of being drawn as the dominant species in terms of proportion, since this is done randomly in the initialization step; thus all species are equally "strong" in that sense. On the other hand, a species initially present in low proportion has a better chance of persisting longer if it is able to benefit from the dominance of the species with high proportion. There are two ways to benefit from the numerical dominance of a particular species (call it X). The first is direct, as a result of being able to prey upon species X. The second is indirect, as a result of decreased predation from those species preyed upon by species X.

If we look at communities dominated by a single species on a case-by-case basis, we find that species B has the best chances to persist in very unequal communities, as can be seen in Figure 6.12. The reasoning is as follows:

(i) if species A is numerically dominant, this directly and indirectly benefits species C and D;

(ii) if species B is numerically dominant, this directly and indirectly benefits species A;

(iii) if species C is numerically dominant, this directly benefits species B and D and indirectly benefits species B;

(iv) if species D is numerically dominant, this directly and indirectly benefits species B.

Thus in communities that are initially very uneven, the survival of the weakest mechanism is counteracted to a sufficient extent as to reduce the differences between the species that can be noted at higher evenness, a phenomenon reflected in both the rankings and probabilities of extinction.

6.2.4.3 Time to extinction

The system behaviours are not absolute — lower initial evenness can counteract the dominance of species favoured by the competition scheme, allowing for different dynamics to be observed in all four cases. However, in the majority of such cases, low initial evenness only extends the initial transient period before the system settles to a steady state. This can be seen in Figure 6.13.

For three competition schemes, the time until the first extinction event is very brief. Scheme 3 again stands out — there is a much larger variability in the time to the first extinction in communities subject to this competition scheme, and the average is significantly higher than for the other competition schemes. In addition, the average time to the first extinction grows significantly as initial evenness increases. For the other three schemes, increasing initial evenness yields a small increase in the average time until the first extinction event. Thus we see a small stabilizing effect due to initial evenness, as was the case with the three-species model (Figure 5.7).

6.2.4.4 Sensitivity analysis

A global sensitivity analysis of the model was carried out by calculating Sobol indices (Lilburne and Tarantola, 2009). These indices indicate how sensitive a model's outputs are to changes in its inputs, by determining what proportion of the variance of the output is due to each input. In our case, the inputs investigated were reproduction rate μ, competition rate σ, mobility rate ϵ, and initial evenness E_0. The outputs checked were final community evenness, identity of the first species to go extinct, and the time until the first extinction. The Sobol indices for these three outputs were computed for each of the four competition schemes.

This was done by performing a large number of simulation runs with varying values of the inputs. The variance in the outputs is then decomposed into proportions assigned to each input, following a procedure due to Sobol (2001). Computationally, this involves generating two matrices Q and R of dimension $M \times k$, where k is the number of model inputs and M the size of the base sample. The columns of these matrices are samples of the model inputs, selected using Latin Hypercube Sampling (Iman, 2008). Then a matrix Q_i is formed by taking the i-th column from R and all other columns from Q, and a matrix R_i is formed by taking the i-th column from Q and all other columns from R. Then the model outputs are computed for

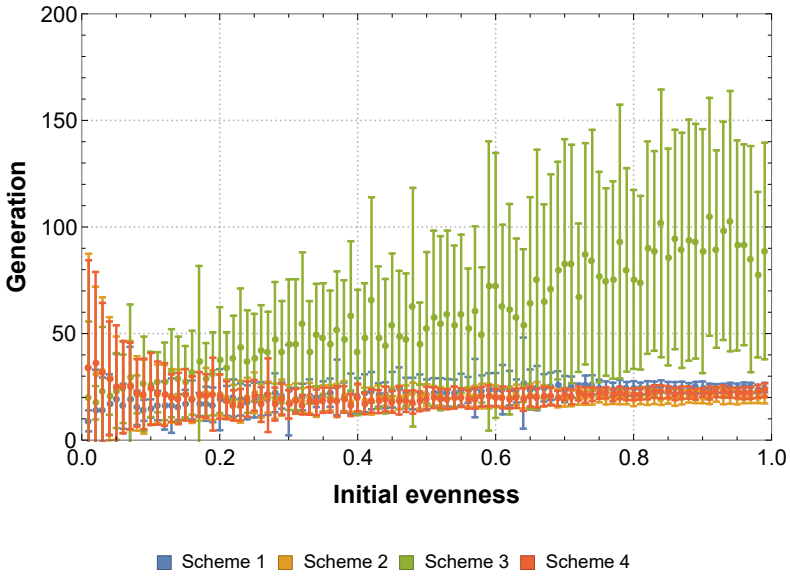


Figure 6.13: Mean time to first extinction event, as function of initial community evenness (plotted with standard deviation).

all the input values in the matrices Q, R, Q_i and R_i, yielding four vectors of model outputs:

$$\mathbf{y}_Q = f(Q), \quad \mathbf{y}_R = f(R), \quad \mathbf{y}_{Q_i} = f(Q_i), \quad \mathbf{y}_{R_i} = f(R_i).$$

The Sobol index for each input is then obtained from these outputs using the following formula:

$$\widehat{X}_i = \frac{\widehat{V}_i}{\widehat{V}} = \frac{\mathbf{y}_Q \cdot \mathbf{y}_{Q_i} - \widehat{f}_0^2}{\mathbf{y}_Q \cdot \mathbf{y}_Q - \widehat{f}_0^2} = \frac{\frac{1}{M}\sum_{j=1}^{M} y_Q^j y_{Q_i}^j - \widehat{f}_0^2}{\frac{1}{M}\sum_{j=1}^{M} \left(y_Q^j\right)^2 - \widehat{f}_0^2},$$

where

$$\widehat{f}_0^2 = \left(\frac{1}{M}\sum_{j=1}^{M} y_Q^j\right)^2,$$

and y_Q^j is the j-th element of the vector $\mathbf{y}_Q$.

This procedure then reveals what proportion of the variance in the selected model outputs can be assigned to each of the selected model inputs. By comparing the

Sobol indices of the different input parameters, we can check which one explains the largest part of the variance in the key model outputs: final evenness, identity of first species extinction, and the time to the first extinction.

The results show that in the large majority of cases (10 out of 12), initial evenness was indeed the input parameter with the largest Sobol index, and hence the input which explained the largest part of the variation in the three selected model outputs. This can be seen in Table 6.2, where we show the Sobol index corresponding to initial evenness, for each of the three selected model outputs, and for each competition schemes. Thus for example, initial evenness explained more than 82% of the variation in the time to first extinction for Scheme 1.

The two entries marked with an asterisk are the only cases where initial evenness did not have the largest Sobol index. This occurred for Schemes 1 and 4, and in both cases the input parameter with the largest Sobol index was mobility. This reflects the findings of Reichenbach *et al.* (2007) who determined (for a system assuming maximal evenness) that coexistence was mediated by mobility.

For all other outputs and all other schemes, initial evenness represented the largest proportion of the variance in final evenness, the identity of the first extinction, and the time until the first extinction. In several cases initial evenness explained more than 80% of these variances. The results in Table 6.2 therefore underline the important role that initial evenness plays in the dynamics of the system.

6.3 Invasion experiments

As discussed in Section 6.1, there are two general experimental set-ups used to investigate microbial competition: co-culture experiments, where all species are inoculated together at the start of the experiment, and invasion experiments, where one species is added at a later stage to a community that has already had a certain amount of time to evolve towards a stable state. In Section 6.2, we have studied *in silico* co-culture experiments addressing non-transitive competition and variable evenness in a community of four species. In this section, we will study these phenomena in the context of an invasion experiment.

Table 6.2: First order Sobol indices for initial evenness, describing what proportion of the variation in the three model outputs can be assigned to changes in this input parameter. The Sobol indices were calculated for each of the competition schemes. The two entries marked with an asterisk are the only cases where initial evenness did not have the highest Sobol index of the tested model input parameters.

Scheme	Final evenness	Identity of first extinction	Time to first extinction
1	0.0709*	0.5186	0.8268
2	0.5726	0.6469	0.8305
3	0.3719	0.4554	0.6408
4	0.4588*	0.8860	0.8351

6.3.1 Model description

To represent invasion of a microbial community, we can use the IBM from Section 6.2. However, we now designate one of our four species as an invader, which will be introduced in the *in silico* community at a later stage than the other three species. To build on the knowledge we have gained in Chapter 5, it is sensible to take species A, B and C as our resident community, since these three species form a stable coexisting community when they engage in cyclic competition, the dynamics of which are well-understood. Thus species D is our invader species, which can interact with the resident community in different ways, depending on the particular competition scheme.

We describe in this section the resulting IBM and experimental set-up using the ODD protocol. Again we outline only those sections where the model or experimental procedure differs from Section 6.2.3.

6.3.1.1 Overview

Purpose

The aim of the model is to investigate how variable initial evenness and non-transitive competition between individuals affect community diversity and stability in the face of invasion, for a stable resident community of three species that is invaded by a fourth alien species.

State variables and scales

The state variables and scales are the same as for the co-culture experiments (described in Section 6.2.3.1).

Process overview

The same demographic interactions as in the co-culture experiments are represented in this invasion experiment: reproduction, competition and mobility. The mechanisms of these interactions are implemented in the same way, as was described in Section 6.2.3.1.

Scheduling

The IBM proceeds in the same way as the co-culture model, using the Gillespie algorithm to determine which interaction occurs at each time step (see Section 6.2.3.1).

6.3.1.2 Design concepts

The design concepts of the invasion IBM are the same as those of the co-culture IBM (Section 6.2.3.2).

6.3.1.3 Details

Initialization

The model is initialized in two stages, representing a first phase of a three-species co-culture, followed by a second phase beginning with an invasion by a fourth species.

The co-culture phase is initialized with a random spatial distribution of empty sites and individuals of species A, B and C. Hence species D is not initially present in the community. The initial species abundance distribution of species A, B and C is determined by a selected evenness value, which is used to generate a vector of population abundances. Individuals are then randomly placed in the remaining grid sites according to these proportions. To generate these vectors of population proportions satisfying a given initial evenness, we use the sampling algorithms described in Section 5.3.3 for the three-species model. Again we use three selected indices: Simpson, Gini and Shannon.

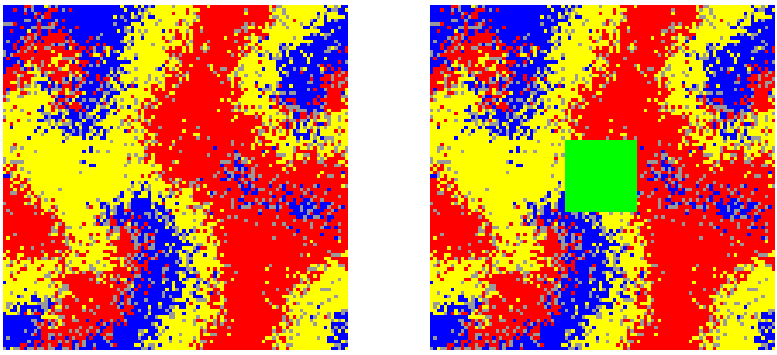
The three-species community is then allowed to evolve for 200 generations. An example of the *in silico* community at this point is shown in Figure 6.14(a) for a high initial evenness condition.

After 200 generations, we model an invasion by species D by emptying a small region in the centre of the grid, and filling these sites with individuals of species D. This is illustrated in Figure 6.14(b). This particular type of invasion geometry has previously been used in the study of biofilm invasions (Merkey et al., 2011) and can represent either artificial inoculation of the invader (in the case of a synthetic community) or an invader colonizing the experimental system via transfer from a bulk liquid or reservoir adjoining the two-dimensional experimental space (Kreft et al., 1998).

After invasion by species D, the system is allowed to evolve as in the co-culture model for another 200 generations. No further invasion events are included, so population growth of species D can occur only through reproduction.

Aside from the input variables, all other parameters used to initialize the model are fixed for all simulations, and are the same as those for the co-culture model (shown in Table 5.1).

Input



(a) (b)

Figure 6.14: A three-species community with high initial evenness (a) after 200 generations which is then (b) invaded by a fourth species (in green).

The input variables are: the competition scheme, and the initial evenness of the resident community.

First, the competition scheme is specified. Starting from a resident community of three species engaged in cyclic competition, there are three possible ways to extend the competition structure to include a fourth species (since we require a fully-connected competition scheme). These three possibilities are represented by Schemes 1, 2 and 3 in Figure 6.1, where the three-species subcycles represent the resident community. In Scheme 1, the invader species preys upon all three species in the resident community, whereas in Scheme 2 the opposite is true: the invader is preyed upon by all three species in the resident community. In Scheme 3, the invader preys upon two species in the resident community, and is itself preyed upon by the third. The rules of the particular scheme are used to determine the outcome of competitive interactions as described in Section 6.3.1.1.

For each competition scheme, we sample 20 initial evenness values, systematically chosen so as to sample the entire range of possible values: $[0, 1]$ for the Gini and Shannon indices, and $\left[0, \frac{1}{S}\right]$ for the Simpson index, where S is the richness of the community.

6.3.2 Results and discussion

In Figure 6.15, we show the final community configurations for each of the three competition schemes. These represent the community compositions after a total of 400 generations. The colours of the points in this figure represent the initial evenness of the three-species co-culture at $T = 0$.

In the case of Scheme 1, the invader is able to defeat all three species in the resident community (see Figure 6.1). Thus we unsurprisingly find in the invasion

experiments that the invader always succeeds in dominating the community to the almost complete exclusion of the other species. Many simulations result in invader monoculture, represented by points located at the vertex D in the 3-simplex in Figure 6.15(a). Some simulations result in different communities, however these communities are extremely uneven and always dominated by the invader. In conclusion, we do not find behaviour significantly different from the corresponding co-culture experiment (compare with Figure 6.6(a)).

For Scheme 2, simulations reveal the opposite situation: the invader is always successfully repelled by the resident community. Thus once again the competitive dynamics are not affected by switching from a co-culture set-up to an invasion set-up, as can be seen when comparing Figure 6.15(b) with Figure 6.6(b). We also note that some simulations result in the extinction of not only the invader, but also one of the resident community members. These simulations represent communities that were initially less even, mirroring the evenness effects seen in Chapter 5.

Table 6.3: Location of final community configurations in the 3-simplex, after a total of 400 generations.

Scheme	Interior	Face	Edge	Vertex
1	0 %	0 %	80 %	20 %
2	0 %	80 %	0 %	20 %
3	20 %	11 %	34 %	35 %

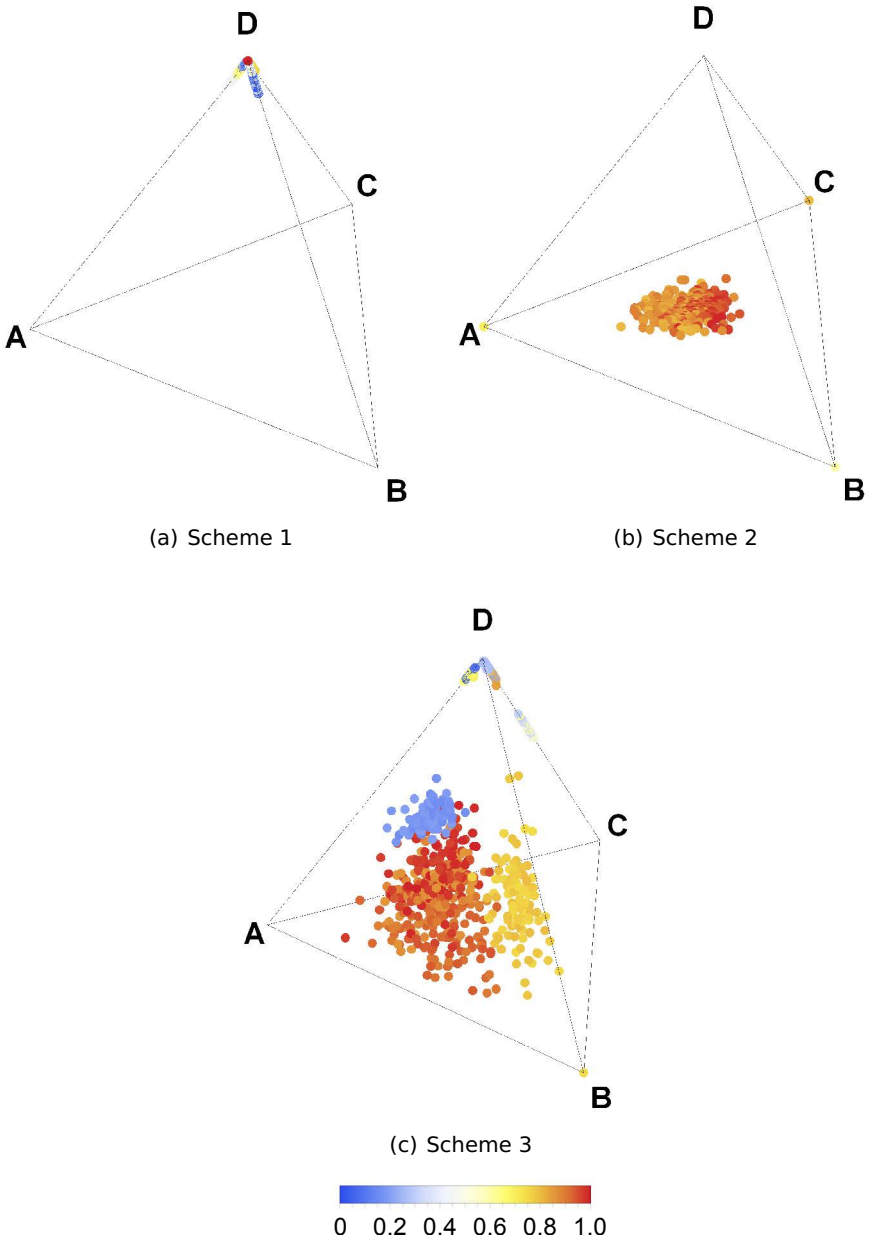
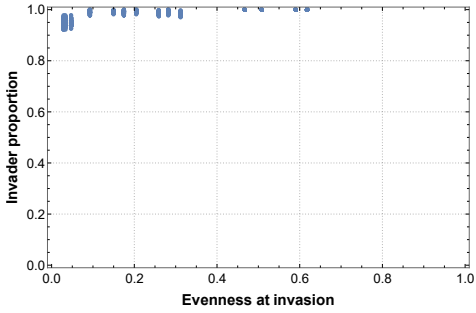


Figure 6.15: Final community configurations (after 400 generations) for each of the three invasion experiments. Colours indicate initial community evenness, according to the colour bar legend. 200 replicates are plotted for each of the 20 initial evenness values.

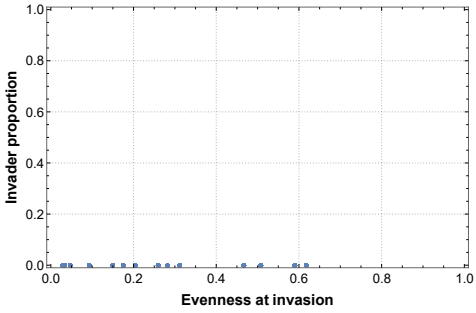
For Scheme 3, we again find richer behaviour since in this competition scheme the invader is not significantly stronger or weaker than the resident community

members. We notice that communities with the highest proportion of the invader (those points closest to the D vertex) generally represent those communities that were initially less even. We also find some resemblance to the co-culture case for high evenness, where most of the final configurations of most communities ended up on the A-B-D face (see Figure 6.6(c)).

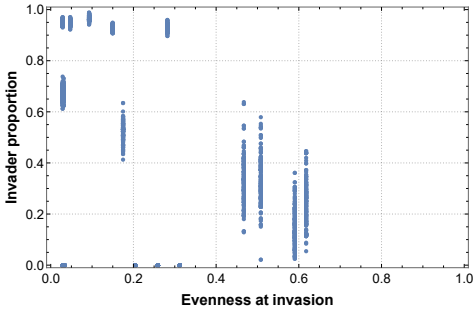
To measure the success of the invasion, we can check how well established in the community the invader was by the end of the simulation period, at $T = 400$, by calculating what proportion of the community it represented. These results are shown in Figure 6.16, as a function of the evenness of the community at the moment of invasion. Using evenness after 200 generations (E_{200}) is more informative than using evenness at $T = 0$ (E_0) since during the intervening time the evenness and configuration of the three-species community can change, as is illustrated in Figure 5.5. This can also be seen in Figure 6.16, which reveals that the sampling coverage of evenness is no longer uniform or equidistant across the entire range of possible evenness values.



(a) Scheme 1



(b) Scheme 2



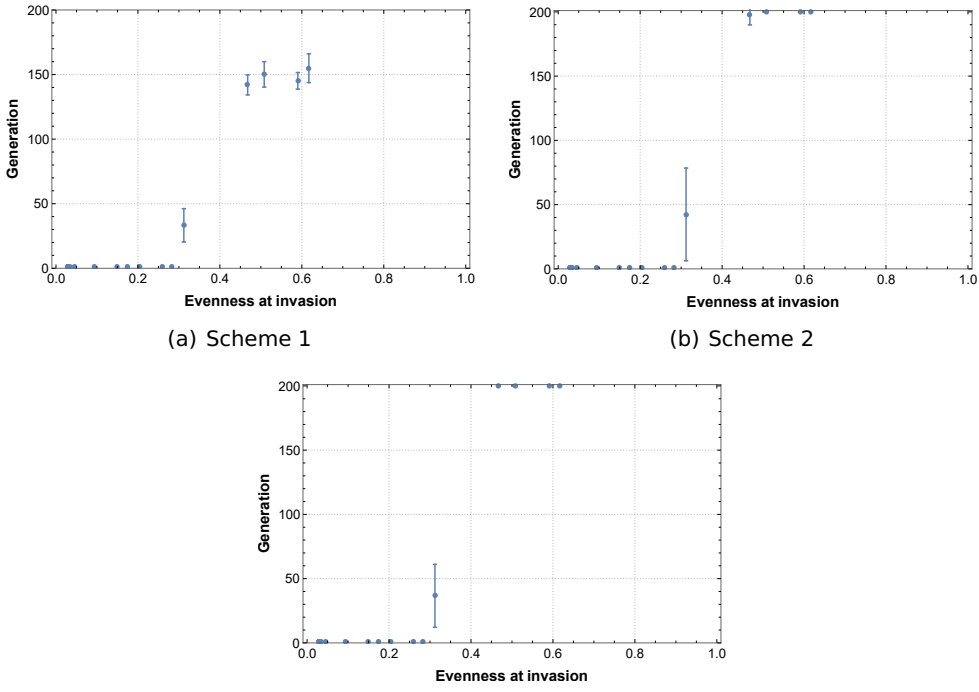
(c) Scheme 3

Figure 6.16: Population proportion representing the invader species (species D) after a total of $T = 400$ generations, as a function of community evenness at the moment of invasion ($T = 200$ generations), for three competition schemes (Figure 6.1). 200 replicates are plotted for each of the 20 initial evenness conditions.

In Figure 6.16, we see that for Scheme 1, the invader always represents the largest

proportion of the final community, to an extensive degree. There is practically no effect of resident community evenness. The same is true for Scheme 2: the invader is always nearly completely excluded, and varying the evenness of the resident community does not change this. In contrast, for Scheme 3 we note a definite evenness effect. We find that a higher resident community evenness before invasion leads to a less successful invasion, in terms of invader proportion at $T = 400$. This agrees with *in vitro* studies which found that less diverse communities were more susceptible to invasion (Wilsey and Polley, 2004; Hillebrand et al., 2008); this is true in particular for synthetic bacterial communities (Hodgson et al., 2002).

As a measure of resident community stability, and its resistance to invasion, we also calculate the average time until the first species extinction within the resident community. These results are shown in Figure 6.17. For Scheme 1, the time to the first extinction within the resident community increases with higher resident community evenness at invasion. However, we have seen that invasions for communities subject to this competition scheme are always almost entirely successful. Hence higher evenness increases the resident community's resistance in terms of time to extinction, but increasing evenness cannot counteract the invader's competitive strength in terms of the competition scheme.



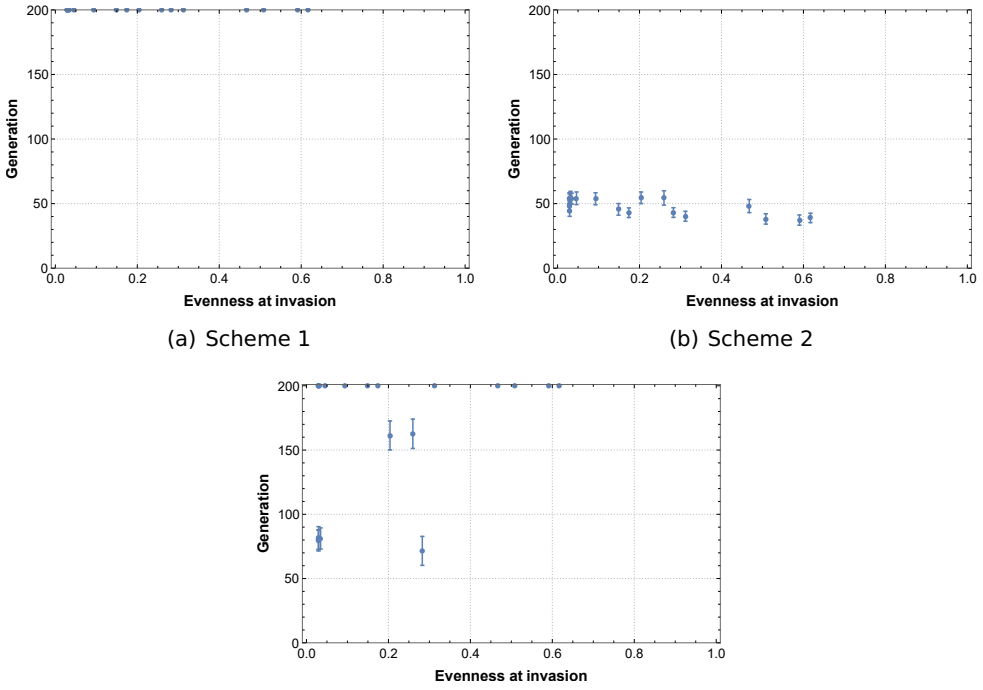
(a) Scheme 1

(b) Scheme 2

(c) Scheme 3

Figure 6.17: Mean time to first species extinction within the resident community as a function of community evenness at the moment of invasion ($T = 200$ generations), for three competition schemes. Mean and standard deviation calculated from 200 replicates per condition.

We find a similar mechanism underlying the invasion dynamics in communities subject to Scheme 2. Increasing evenness of the resident community at the time of invasion results in longer times to extinction for these three species. For high evenness, there are no extinctions within the simulation time (these are plotted as extinction time equal to 201 generations). Scheme 3 exhibits similar behaviour: an abrupt jump in the time to extinction at intermediate resident community evenness, and no resident community extinctions for high resident community evenness.



(a) Scheme 1

(b) Scheme 2

(c) Scheme 3

Figure 6.18: Mean time to invader extinction as a function of community evenness at the moment of invasion ($T = 200$ generations), for three competition schemes. Mean and standard deviation calculated from 200 replicates per condition.

As a further measure of resident community resistance to invasion, we can check the average time until extinction of the invader. These results are shown in Figure 6.18. For Schemes 1 and 2, we note no dependence of average time to invader extinction on the evenness of the resident community at the time of invasion. For Scheme 1, the invader never goes extinct, while for Scheme 2 it always collapses to extinction fairly rapidly, regardless of resident community evenness. For Scheme 3, the picture is less clear. For low resident community evenness, the invader generally persists or has a fairly long time to extinction. The time to extinction drops for intermediate evenness, before increasing again for high resident com-

munity evenness. This may seem to run counter to what we expect, but we should note that it does not disagree with our previous finding, that higher resident community evenness results in lower invader establishment (Figure 6.16).

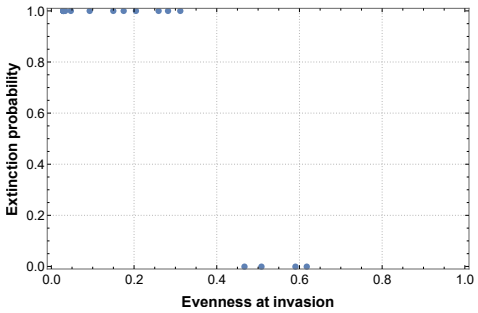


Figure 6.19: Mean extinction probability for the resident community species as a function of community evenness at the moment of invasion ($T = 200$ generations), for Scheme 3. Mean calculated from 200 replicates per condition.

Additionally, in Figure 6.19 we plot the probability of extinction for the resident community species in terms of resident community evenness at the time of invasion. Again find a positive effect of evenness in resisting invasion. For low evenness, at least one member species of the resident community always collapses to extinction in the face of invasion. From the 3-simplex of this experiment (Figure 6.15(c)), we note that this is generally species C. Recall that in co-culture simulations for Scheme 3, this species was highlighted as the weakest in the competition scheme, particularly for low evenness (see Section 6.2.4). Thus our results agree with the co-culture simulations.

6.4 Conclusions

Multiple competition schemes have been examined with simulation studies modelling various possible communities under two different experimental set-ups, which result in diverse coexistence and extinction scenarios. With fully connected competition, the four-species system is generally unstable for all competition schemes under both experimental set-ups. There are frequent extinction events, which typically occur very rapidly relative to the persistence time of stable subpopulations. The dynamics induced by the competition schemes works against the coexistence of all species (Cheng et al., 2014).

System behaviour is strongly dependent on initial evenness and competition scheme. The importance of initial evenness was confirmed by means of a sensitivity analysis. Low initial evenness can counteract the dynamics of the competition scheme in the sense that the identity of the first species to collapse to extinction can

change. But generally, low initial evenness will only extend the initial transient period before the system settles to its steady state. If initial evenness is excessively low, system biodiversity is lost before other emergent behaviours can be noticed.

In contrast, higher initial evenness can have a small stabilizing effect, in the sense that the time until the first species extinction is slightly extended as initial evenness increases. The time until the first extinction is generally quite short for all competition schemes except Scheme 3. In this case, the time until the first species extinction can vary significantly. By extending the region of biodiversity in this way, there is sufficient time for system behaviour to be affected by other factors such as competition scheme, rates of competition and mobility. These results support experimental observations that biodiversity is promoted by increasing evenness (Isbell et al., 2009b).

When considering an invasion experiment, we find similar evenness effects. Higher resident community evenness before invasion leads to a less successful invasion, in terms of invader proportion at the end of the simulation, probability of extinction and time until the first extinction of a resident community member. These results agree with empirical studies from different natural and synthetic systems (Wilsey and Polley, 2004; Hillebrand et al., 2008; Hodgson et al., 2002).

Our results demonstrate the danger in overlooking variable community evenness and making the typical assumption that communities are maximally even, despite mounting evidence to the contrary (Wilsey and Polley, 2004; Huston, 1997; Grime, 1998; Smith and Knapp, 2003). This oversight also ignores the fact that damages due to human actions can affect the evenness of natural communities, often making them more vulnerable to invasion, stresses and disturbances (Wittebolle et al., 2009). While theoretical studies such as this one are beginning to increase in number, experimental studies to validate their conclusions are still lacking (Isbell et al., 2009b).

7

In silico substrate dependence increases community productivity but threatens biodiversity

7.1 Introduction

In Chapters 5 and 6 we formulated and analysed models incorporating non-transitive competition and variable initial evenness, two mechanisms known to strongly affect the biodiversity of a system. In this chapter, we go a step further and incorporate another mechanism shown to be key in mediating biodiversity, namely resource dependence. Existing microscopic models of communities with cyclic competition typically neglect the resource-dependent nature of demographic processes (Reichenbach et al., 2007; Kreft et al., 1998; Cheng et al., 2014). However, as discussed in Section 3.4, this is a key mechanism that can have significant effects on community composition and functioning.

As addressed in Chapters 5 and 6, community diversity can be promoted by non-transitive competition. In the particular case of competition in a microbial setting, nutritional resources are a particular focus (Hibbing et al., 2010). The resource ratio model of competition suggests that the prevalence of species in a community is mediated by the available nutrients, in particular their availability and rate of consumption (Tilman, 1977). Thus competing microbial species can coexist for certain ratios of nutrient concentrations. But if nutrient availability is limited, some species can die out as a result of being outcompeted, leading to a biodiversity loss. The resource ratio competition model has been shown to explain, for various ecosystems including microbial communities (Murray and Baird, 2008; Smith, 2002), some of the most typical dynamics between resource competitors, including for example that the species better able to survive at lower levels of a limiting resource will be the best competitor for that resource (Miller et al., 2005).

Competition for a limiting resource can be categorised as either scramble competition or contest competition (Hibbing et al., 2010). Scramble competition occurs when one species deprives its competitor(s) of the communal resource by depleting said resource, whereas contest competition occurs when one species actively harms its competitor(s), for example by producing harmful toxins or otherwise attacking other individuals (Nicholson, 1954). An example of scramble competition is the non-interference competition between different microbial strains for a common but limited carbon source needed to drive their growth (Smith, 2002). An example of contest competition is the secretion, by bacteria such as *Streptococcus thermophilus*, of toxic antimicrobial compounds that directly reduce the growth rates of its competitor species (Hibbing et al., 2010).

In microscopic models, the typical approach to resource limitation has been to represent it by imposing a constant limit on population size, rather than modelling the resource dynamics explicitly (Nowak, 2006; Riolo et al., 2001). More recently, this approach has been altered to consider resource fluxes and dynamic population sizes (Requejo and Camacho, 2013; Melbinger et al., 2010; Requejo and Camacho, 2011; Centler and Thullner, 2015). This echoes recent developments in the more specific case of modelling biofilm formation, where both the growth of cells and the diffusion of nutrients through the bulk liquid are taken into account (Lardon et al., 2011; Kragh et al., 2016; Ardré et al., 2015). This approach has in recent years been extended to IBMs (Centler and Thullner, 2015), where a typical example admits a limiting resource that constrains individuals' reproduction (Requejo and Camacho, 2012). However, some models of resource-limited reproduction assume that the population in question is well mixed, a typical yet significant simplifying assumption which by design does not permit any effects of spatial structure to emerge. This is despite the fact that spatially structured environments have been acknowledged to result in a significantly different population dynamics than well-mixed environments (Allison, 2005).

We therefore extend established models by incorporating these factors, as such

aligning them more closely with real-world microbial ecosystems, and to investigate how this more realistic approach affects community productivity and biodiversity, two key indicators of ecosystem functionality. In doing so, we address research questions 1, 4, and 5 (see Section 1.2):

- What effect does initial evenness have on maintaining community diversity?
- What effect does initial evenness have on maintaining community functionality?
- If interactions within a community are dependent on resource availability and use, how does this affect community diversity and functionality?

In Section 7.2 we describe the model developed for this purpose, and in Section 7.3 the set-up of the *in silico* experiments conducted to understand the model's behaviour. The results of these simulation studies are presented and discussed in Section 7.4. In the final section, some conclusions are drawn.

7.2 Model description

Our goal is to study the effects on the maintenance of biodiversity of both variable initial evenness and resource dependence in the demographic processes. We therefore formulate an IBM representing a community of three interacting bacterial species, and present a stochastic, spatial IBM simulating the system dynamics and allowing for the assessment of the effect of variable initial evenness and resource dependence. As a foundational framework, we use the model described in Chapter 6, which we extend by incorporating environmental substrate dynamics, individual substrate uptake and biomass growth, and linking the demographic processes to internal substrate level. We again describe our IBM using the ODD protocol introduced in Section 3.2.3.

7.2.1 Overview

7.2.1.1 Purpose

The aim of the model is to investigate how community evenness and resource dependence in demographic processes affect community biodiversity and productivity, which are key proxies of community functionality.

7.2.1.2 State variables and scales

The model is a two-dimensional representation of a closed experimental domain, mimicking the closed environment typically employed for *in vitro* studies, for example a Petri dish. The domain is divided into a regular grid of size $L \times L = N$. The model comprises two entities: bacterial cells and their local environment (grid sites). The grid is populated by a community of three species denoted by A, B and C. In terms of bacteria, each grid site is either occupied by a single individual, or is empty. The grid also contains a substrate, which is displaced by a diffusion process.

Individual microbes are characterized by three state variables: grid position (i, j), species identity $s \in \{A, B, C\}$, and internal substrate level $I(i, j)$. A grid site is characterized by two state variables, namely substrate concentration $s(i, j)$ and substrate uptake rate $r_s(i, j)$.

7.2.1.3 Process overview

We consider an *in silico* microbial community that is initially placed on the grid with a random spatial distribution. The community's initial species abundance distribution is constructed to obtain a desired level of evenness. There is an initial environmental substrate gradient so that all grid sites initially contain a small amount of substrate. The amount of substrate in the *in silico* environment is then maintained via a constant inflow from a source in the centre of the grid. Substrate is displaced via a diffusion process, resulting in a heterogeneous substrate gradient.

An individual can interact with its nearest neighbours, defined as those individuals in its von Neumann neighbourhood. Three possible interactions can occur: reproduction, competition and mobility. Interactions are dependent on an individual's biomass, which is fuelled by its uptake of the environmental substrate. The growth of an individual is modelled using Monod kinetics (Monod, 1948).

The basic mechanisms of these interactions are illustrated in Figure 7.1. An interaction can only occur if the individual's internal substrate (the amount of substrate at its disposal if its biomass is converted to substrate) is above a certain threshold.

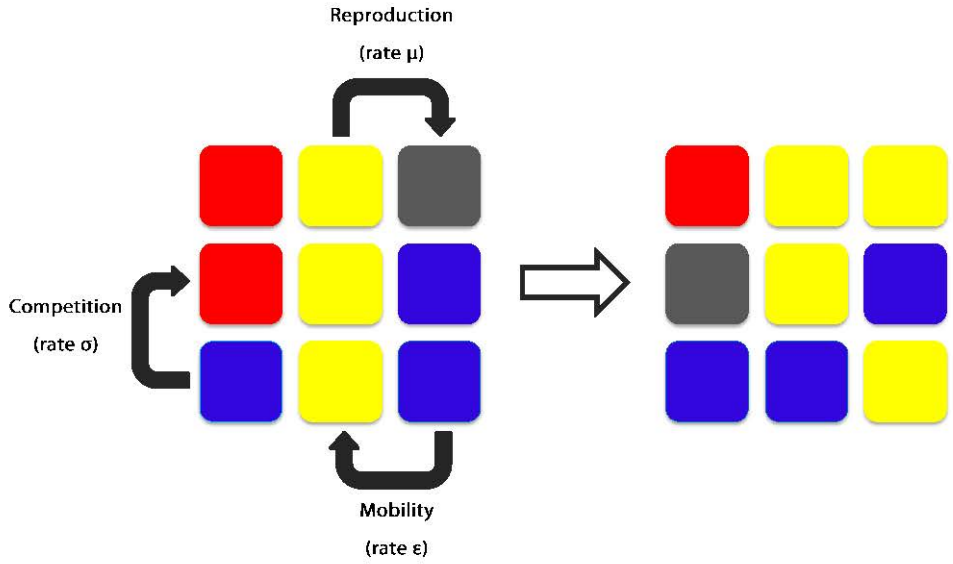


Figure 7.1: Mechanisms of demographic processes, following Reichenbach *et al.* (Reichenbach et al., 2007). Whether an interaction occurs or not depends on the individuals' internal substrate levels (details given in Section 7.2.3.3).

7.2.1.4 Scheduling

The IBM proceeds in discrete time steps. The following processes are performed sequentially at each time step:

- **Substrate inflow and diffusion:** substrate flows into the *in silico* environment from a source in the centre of the grid. It then diffuses around the grid. Local substrate concentrations are updated accordingly.
- **Uptake of substrate and conversion to biomass:** individuals uptake substrate (if they are not already saturated) and convert this to biomass according to Monod kinetics. After uptake, their biomass and the amount of substrate in their local environment are updated accordingly, so that the mass balance is respected.
- **Demographic interactions:** reproduction, competition and mobility.

The demographic interactions are simulated using a modified version of the Gillespie algorithm (Gillespie, 1976) described in Section 5.3.1.4. The only alteration is that now an additional check must be performed on the individuals' biomass to determine if the selected interaction can take place:

1. Set time to $t = 0$ and set the event rates (as in Chapter 6):

   (a) reproduction with rate μ
   (b) competition with rate σ
   (c) mobility with rate ϵ
2. Calculate the overall rate of events $r = \mu + \sigma + \epsilon$
3. Calculate inflow and diffusion of substrate; update each grid site with new substrate concentrations (where a change occurred)
4. Calculate each individual's substrate uptake and conversion to biomass (if this occurs); update each individual's biomass and the local environmental substrate concentration to respect the mass balance
5. Select an individual at random
6. Select one of the focal individual's nearest neighbours at random
7. Select an interaction event with the following probabilities, by drawing a random number from the interval $[0, r]$:
   (a) reproduction with a probability $\frac{\mu}{r}$
   (b) competition with a probability $\frac{\sigma}{r}$
   (c) mobility with a probability $\frac{\epsilon}{r}$
8. Execute the selected interaction event on the selected individual (if permitted by the governing rules and the individuals' internal substrate levels) and determine the outcome according to the governing rules
9. Update the grid according to the outcome of step 8
10. Update the time step to $t = t + 1$
11. Return to step 3 and continue until $t = t_{\text{end}}$

Details of the submodels mentioned above can be found in Section 7.2.3. We again evolve the model for 500 generations, where a generation is defined as the number of time steps required for each individual to be the subject of one interaction on average.

7.2.2 Design concepts

- **Emergence:** the population-level dynamics of the community, and the spatial patterns of the individuals and the environmental substrate emerge naturally from the local interactions.

- **Interactions:** individuals interact with each other and their environment by consuming substrate, reproducing if located next to an empty site, exchanging sites with their neighbours, or competing with their neighbours.
- **Stochasticity:** the stochasticity in the model arises from the initial spatial distribution of the grid; the interactions between individuals and the environment; and the interactions between individuals.
- **Sensing:** if selected for reproduction, individuals can sense whether their neighbouring site is empty; if so, they will reproduce. Individuals can also sense the substrate concentration within the grid site where they are located, which affects their uptake and growth rate,
- **Observation:** the data collected from the IBM includes: the spatial distribution of substrate, the population count of each species, the community evenness, a spatial aggregation measure, the biomass distribution of individuals, and their time to extinction. These are tracked for each time step.

7.2.3 Details

7.2.3.1 Initialization

The model is initialized with a random spatial distribution of individuals and empty sites. Initially, a certain proportion of grid sites are left empty; thus the system is initially below carrying capacity. The initial species abundance distribution is determined by a selected evenness value using the Gini index, which is used to generate a vector of population abundances as described in Section 5.3.3. Individuals are then randomly placed in the remaining grid sites according to these proportions. The domain is also initialized with a small amount of substrate in each grid site.

Aside from the input variables described in the next section, all parameters are fixed for all simulations, and are shown in Table 7.1.

7.2.3.2 Input

The input variables are: the initial community evenness, and substrate limitation scenario. The possibilities for the latter are described in detail in Section 7.3.

For each scenario, we check a range of initial evenness values, chosen so as to sample across the range of possible values of the Gini index, namely the unit interval.

Table 7.1: Parameters of the three-species IBM with substrate dynamics.

Parameter	Description	Value	Unit
L	Grid side length	100	-
$ø$	Initial proportion empty sites	0.1	-
μ	Reproduction rate	1	T^{-1}
σ	Competition rate	1	T^{-1}
ϵ	Mobility rate	4.25	T^{-1}
T	Generations evolved	500	-
I_0	Initial substrate concentration	10^{-17}	g
r_I	Substrate inflow rate	8.3×10^{-20}	$\mathrm{g\,\mu m^{-2}\,s^{-1}}$
D	Diffusion coefficient	1.7×10^{-3}	$\mathrm{\mu m^2\,s^{-1}}$
μ_{max}	Maximum growth rate	8.3×10^{-4}	$\mathrm{s^{-1}}$
K_s	Half-saturation constant	4.5×10^{-3}	$\mathrm{g\,L^{-1}}$
E_c	Substrate conversion efficiency	0.44	$\mathrm{g_{mass}/\,g_{sub}}$
m_{max}	Maximum biomass of individuals	2.5×10^{-15}	g
E_r	Reproductive efficiency	0.85	$\mathrm{g_{mass}/\,g_{mass}}$

7.2.3.3 Submodels

- **Substrate diffusion:** after the grid is initialized, substrate flows into the grid at each time step at a fixed rate r_I via a source located in the centre of the grid, and a fixed number of diffusion steps (with coefficient D) is then carried out. The time scales of the diffusion and individual interaction processes are separated so that diffusion occurrs at a faster time scale than individual interactions, since otherwise all substrate would be very quickly consumed.

  The diffusion process is implemented using the following scheme:

$$U^{t+1}(i,j) = U^t(i,j) + \frac{\Delta t}{\Delta x^2} D\big[U^t(i-1,j) + U^t(i+1,j) + U^t(i,j-1) + U^t(i,j+1) - 4U^t(i,j)\big], \tag{7.1}$$

  where Δt and Δx are respectively the time and space discretization step size. These parameters are chosen so that their ratio $\frac{\Delta t}{\Delta x}$ satisfies the Courant-Friedrichs-Lewy (CFL) condition which is necessary for stability of the time-explicit numerical scheme (Courant et al., 1928). An excess of environmental substrate (an atypical situation in natural systems) was avoided through the appropriate choice of the inflow and diffusion parameters.

- **Bacterial growth:** After the diffusion steps have taken place, the environmental substrate concentration is updated for each grid cell. Then each individual consumes substrate if it is not yet at maximum biomass m_{max}.

Substrate uptake is governed by the Monod equation (Monod, 1948):

$$r^t(i,j) = \mu_{max} \frac{I^{t-1}(i,j) U^{t-1}(i,j)}{K_s + U^{t-1}(i,j)}, \tag{7.2}$$

where $r^t(i,j)$ is the uptake rate of the individual at grid cell (i,j) at time step t, and the Monod parameters are as described in Table 7.1.

The individual's internal substrate concentration $I^t(i,j)$ and the environmental substrate concentration $S^t(i,j)$ are then updated for each grid cell where uptake occurred.

For the sake of completeness, it should be mentioned that we implemented several different bacterial growth models with the IBM, and tested these for significant differences in model output. The additional growth models tested were the Blackman (Koch, 2012), Tessier (Pinna et al., 2009) and Powell (Koch, 1982) models. These led to very similar model outputs, hence we opted for the Monod since it is the most well-established and commonly used growth model in the microbiological literature (Koch et al., 2012).

- **Reproduction/division:** if an individual is stochastically selected for reproduction, and located next to an empty grid site, its internal substrate level is checked. If it is higher than the threshold value τ_R, the parent with biomass m splits into two offspring with equal biomass $\frac{m-m_w}{2}$. The combined biomass of the offspring is less than that of the parent because reproduction is assumed to be less than fully efficient. The lost mass m_w is given by $(1-E_r)\,m$, where E_r is the reproductive efficiency reported in Table 7.1. The second offspring is placed in the neighbouring grid cell.

- **Competition:** if stochastically selected for competition and located next to an individual of a different species, competition occurs if at least one individual's internal substrate level is above the threshold τ_C. The outcome of the competition event is then determined as follows:

  - if one individual's internal substrate is above the threshold while the other individual's is below it, then the second individual is assumed to be too weak to compete and loses the competition regardless of its place in the cyclic competition structure;
  - if both individuals' internal substrate is above the threshold, the outcome of the competition is determined by the cyclic competition scheme (Figure 5.1(b)).

  When an individual is killed, its internal substrate is instantaneously released into its local environment.

- **Mobility:** if stochastically selected for mobility, individuals can exchange places if their internal substrate levels are above the threshold τ_M.

7.3 *In silico* scenarios

We can use our IBM to investigate the effects of resource dependence on community diversity and productivity in different ways. Firstly, by varying the amount of substrate required for a demographic process to occur (the substrate threshold τ), and secondly by specifying that the species in the community may have different substrate utilization profiles. Motivated by examples of microbial communities in nature (Allison, 2005; Ratledge, 1993; Velicer, 2003), we selected three substrate utilization scenarios. To facilitate analysis and comparison of the different scenarios, we also investigate a benchmark case.

In the benchmark case, demographic processes occur regardless of individuals' internal substrate level, at the rates μ, σ and ϵ given in Table 7.1. Individuals absorb environmental substrate and convert this to biomass, but their biomass has no impact on interaction events. This is analogous to the original Reichenbach model (Reichenbach et al., 2007), where there is no influence of biomass on interaction events, since in that model there is no substrate and individuals have no mass. Thus, in the benchmark case, a pair of neighbours is selected randomly, an interaction occurs and its outcome is calculated, the grid is updated and the process repeats.

In the following three substrate utilization scenarios, we impose a substrate threshold on a specific demographic process, for example reproduction. In that case, individuals, when stochastically selected for reproduction, may only carry out this process if their internal substrate level is above the substrate threshold.

In the first scenario, all three species are subject to the same substrate threshold and therefore have the same reproductive/competitive/mobile capacity: $\tau_A = \tau_B = \tau_C$. This symmetric limitation scenario represents the simplest case of an ecosystem with similar species that all depend on a common environmental resource in the same way (Smalla et al., 1998).

In the second scenario, one species is subject to a substrate threshold τ, while the other two species are not (asymmetric limitation). Then for example species A has substrate threshold $\tau_A > 0$ while species B and C have substrate thresholds equal to zero, i.e. $\tau_B = \tau_C = 0$. Thus only one species is constrained in its capacity for reproduction/competition/mobility. This scenario can represent, for example, a community with one species that must produce an extracellular enzyme to degrade the substrate into a usable form, whereas the other species do not and hence do not face the same cost in synthesizing the substrate to drive their growth (Ratledge, 1993).

In the third scenario, a hierarchy in terms of substrate limitation is imposed on the community, e.g. $\tau_A > \tau_B > \tau_C$. This scenario can represent, for example, a community where the common strategy of 'cheating' is present (Velicer, 2003).

Two species require an extracellular enzyme to degrade the substrate, but one of them intercepts the reaction products secreted by the other species, in this way avoiding the need to produce its own enzyme, and cheating to benefit more from the common resource (Allison, 2005). This phenomenon is also known as sequential cross-feeding, and has been recognized in many different microbial systems. A particular example is anaerobic methane oxidation, which involves two types of bacteria engaged in this relationship: methanogenic and sulfate-reducing bacteria (Hummert et al., 2014).

For each substrate limitation scenario (symmetric, asymmetric and hierarchical), we limit separately each of the three demographic processes (reproduction, competition and mobility) to avoid confounding effects. For example, reproduction may be limited asymmetrically among the species, while competition and mobility are not substrate-limited. With the addition of the benchmark scenario without substrate dependence, this results in ten scenarios. For each of these, we investigate three initial evenness settings using the Gini index: maximal ($E_0 = 1$), intermediate ($E_0 = 0.5$) and low ($E_0 = 0.2$).

Once the simulation set-up is specified (with specific substrate limitation and specific initial evenness), a set of simulations is carried out with the substrate threshold systematically varied from the lowest substrate threshold (10^{-16} g) to the highest (2×10^{-15} g), in twenty increments. Multiple outputs are tracked for each simulation, listed under “Observations” in Section 7.2.2.

7.4 Results and discussion

7.4.1 Impact of substrate limitations

Since productivity is a key indicator of ecosystem functionality (Isbell et al., 2009a), we compare in Figure 7.2 the mean final community biomass yield for each experiment as the substrate threshold is varied. Recall that the benchmark scenario by design involves no substrate limitation, and hence is insensitive to substrate threshold.

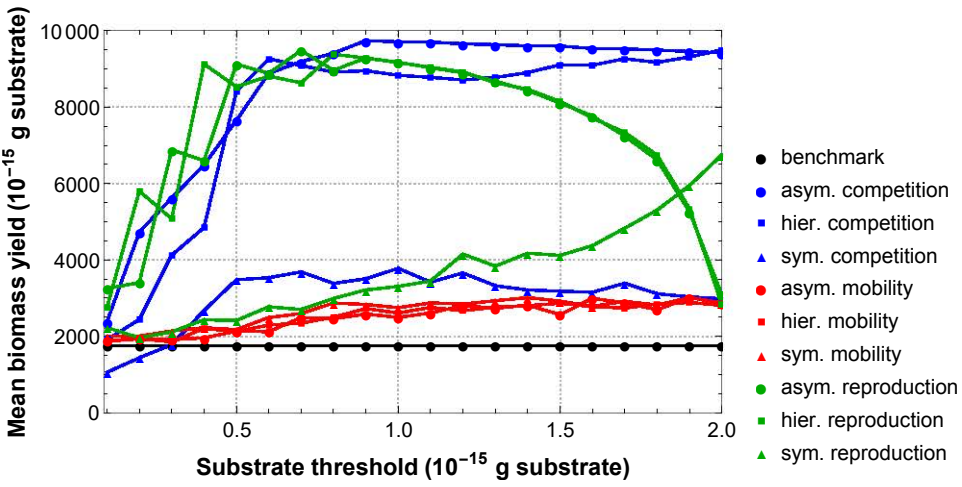


Figure 7.2: Comparison of average biomass yield for the different scenarios, as a function of substrate threshold. Each curve represents the mean of 200 simulations.

For nearly all substrate threshold values, all scenarios result in a higher biomass yield than the benchmark scenario. This can be explained by a "quality over quantity" phenomenon (Reznick et al., 2002). When the substrate threshold is increased, those individuals able to carry out the focal demographic process are fitter (in terms of biomass) than in the benchmark case. Since individuals may only reproduce once they gain sufficient biomass, fewer individuals can reproduce at any given time but these high-biomass individuals will produce high-biomass offspring, which more than compensates for the decrease in the number of reproduction events.

The demographic process most sensitive to this effect is reproduction: linking this process to internal substrate results in consistently higher biomass yields for all threshold values. However, this effect levels off as the substrate threshold increases further, before decreasing sharply to approach again the benchmark productivity. This occurs once the decrease in the number of reproduction events is no longer compensated by the increased biomass of the offspring.

Linking competition to substrate utilization also impacts productivity dramatically, even for high substrate thresholds. This is due to the fact that if individuals can only compete once they have sufficient biomass, they will release more substrate into the environment if killed, which can then be absorbed by neighbouring individuals. This effect saturates at a certain threshold value, indicating that the deaths of high-biomass individuals are releasing substrate into the environment at a rate higher than their neighbours can absorb it, thus individuals are now limited by their maximal uptake rate.

We also note that for both competition and reproduction, symmetric substrate limitation scenarios result in qualitatively different behaviour than their asymmetric

and hierarchical counterparts. These latter cases are more likely to result in monocultures, communities which have a higher biomass yield since no effort needs to be diverted to competition with other individuals. This effect can also be seen in Figure 7.3, where symmetric limitation scenarios result in significantly higher final community evenness than their asymmetric and hierarchical counterparts.

Comparing Figures 7.2 and 7.3, we note that those scenarios that give rise to the highest biomass yields also result in the least diverse communities. This difference in biomass output between monocultures and more diverse communities is caused by a negative dominance effect that originates from a trade-off between growth rate (due to a high substrate threshold) and final biomass level. In monocultures, species with a high substrate threshold can build up higher biomass levels than fast-growing ones (with a lower substrate threshold), albeit more slowly. However, in more diverse communities a fast-growing but low biomass productive species will monopolize most of the substrate and prevent competing species from producing the high biomass levels seen in monocultures. This effect has also been observed in experimental studies, which have noted an underyielding of diverse communities compared to their component monocultures due to the competitive suppression of highly productive species (Schmidtke et al., 2010; Hooper and Dukes, 2004; Loreau and Hector, 2001; Weis et al., 2007).

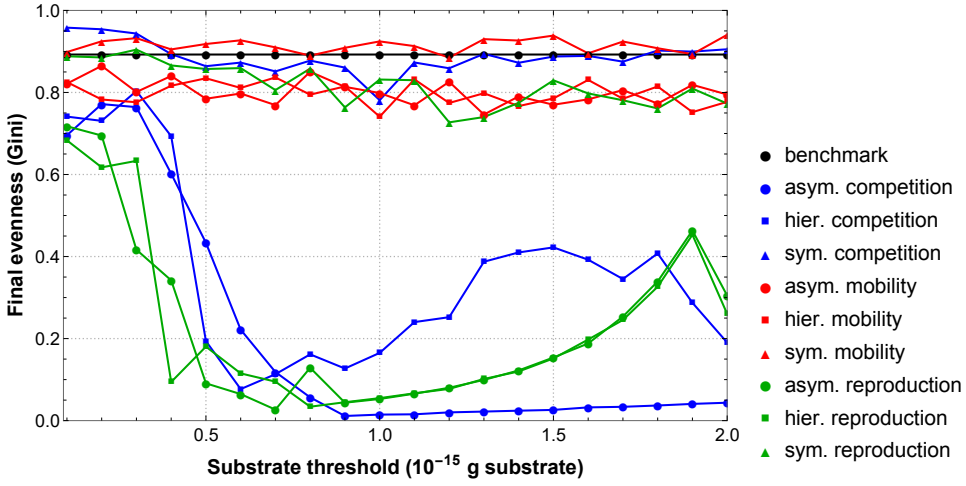


Figure 7.3: Comparison of average final evenness for the different scenarios, as a function of substrate threshold. Each curve represents the mean of 200 simulations.

Linking mobility to substrate utilization has a comparatively weak effect on community biomass yield. Increasing an individual's mobility results in a mild increase of its reproductive capacity, since it increases the chance it will encounter an empty site into which it can reproduce. It has been demonstrated that as long as the mobility rate remains below a critical value, biodiversity will be maintained

regardless of the details of the competition structure (Reichenbach et al., 2007). Linking mobility to substrate utilization does not push the mobility rate above this critical value, and hence does not greatly impact the system dynamics.

7.4.2 Impact of initial evenness

While substrate utilization has often been overlooked in modelling studies of microscopic communities, it is not the only factor that has been largely neglected. Community evenness has been shown to be a key factor in preserving the functional stability of ecosystems (Hillebrand et al., 2008), but has often been overlooked in experimental and modelling studies (Daly et al., 2015) where it is typically assumed to be maximal despite experimental evidence to the contrary (Huston, 1997; Grime, 1998; Smith and Knapp, 2003).

Varying the initial evenness of the community has a significant impact on biodiversity maintenance. This is shown in Figure 7.4 for the case of substrate-limited competition; other scenarios demonstrate similar behaviour. Lower initial evenness magnifies the impact of increasing the substrate threshold, negatively impacts the maintenance of biodiversity, and increases the probability of species extinctions and hence the tendency towards monoculture. Thus we find again the trade-off between biomass yield and biodiversity maintenance, due to the strong competition effects that are exacerbated by lower evenness, since species with low abundances relative to the rest of the community are more vulnerable to extinction.

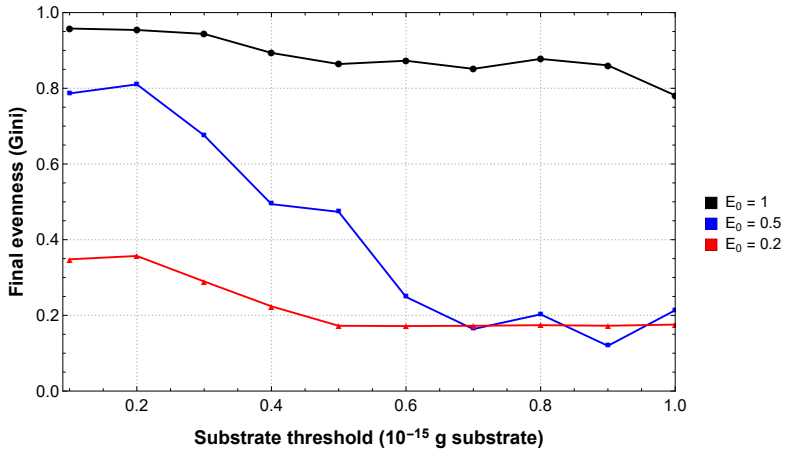


Figure 7.4: Comparison of average final evenness as a function of substrate threshold, for different initial evenness values E_0 under the symmetric competition limitation scenario. Each curve represents the mean of 200 simulations.

In the case of substrate-limited competition illustrated in Figure 7.4, this effect

saturates at higher substrate thresholds as individuals of different species become increasingly spatially disaggregated, acting as a further brake on the frequency of competition events. This effect will be discussed in more detail in Section 8.4.1.

In general, lower initial evenness unbalances community dynamics by magnifying destabilizing effects such as one species outpacing the others in terms of growth or competitive ability, and so on. When such effects occur in more even communities, the higher level of community diversity acts as a buffer against the destabilization, so that the community's functionality is more difficult to disturb. This phenomenon has also been observed in experimental studies of the behaviour of communities with different levels of evenness in the face of various stresses (Hillebrand et al., 2008; Wittebolle et al., 2009; De Roy et al., 2013).

Further study of this system using approximate analytic methods could be done on the basis of a system of PDEs mimicking both the species and substrate dynamics. The typical mean field approximation approach assumes a well-mixed environment, so spatial aspects are not accounted for, which would imply a homogeneous distribution of substrate due to the lack of diffusion. We consider this to be an excessive simplification. Essentially, an analytic treatment must be done using a system of four PDEs: one equation for each of the bacterial species, and one equation for the substrate. The rates at which the demographic processes occur would be functions of the substrate concentration. Such a system must be solved numerically since a steady-state analysis would result in the case of a homogeneous distribution of substrate.

7.5 Conclusions

By extending existing models to incorporate both the resource-dependent nature of demographic processes and the variability of community evenness, we provide a more realistic *in silico* representation of natural systems. *In silico* experiments reveal a trade-off between maintaining community diversity and increasing biomass yield. This result is consistent with experimental observations of a negative dominance effect. In addition, the important role that evenness plays in maintaining the functional stability of ecosystem is demonstrated, indicating the danger in overlooking this key feature in modelling or experimental studies.

8

The impact of resource dependence on spatial microbial population dynamics

8.1 Introduction

Most early models of population growth, ranging from the Malthusian or exponential growth model (Malthus, 1798) through to the logistic (density-dependent) growth model (Verhulst, 1845) and the Lotka-Volterra predator-prey model (Lotka, 1925; Volterra, 1926), were based on ODEs. Therefore these models did not take spatial considerations into account, and instead focused on determining the equilibria of the system in order to conduct stability analyses.

Models of this sort also typically use the mass-action law as the basis of the interactions they mimic. This law originates from the study of chemical reactions (Murray,

2002). The mass-action law dictates that if the different molecules must collide in order to start the reaction, and the experimental system is well mixed, then the collision and hence reaction rate is proportional to the product of the concentrations of the reactants (Song et al., 2014). In the ODE population models mentioned above, this law is used to describe the interactions between species as a function of their density. This description depends on the mean field assumption: that the environment is sufficiently well mixed to allow any individual to come in contact with any other. This assumption justifies the use of species or population-level averages (Zomorrodi and Segrè, 2016).

However, in many ecological settings in the real world, the mean field assumption of a well-mixed environment does not hold. Therefore, models should take into account heterogeneous space and local interactions in order to obtain a more realistic representation of reality (Hellweger et al., 2016a). This paradigm shift has been shown to result in representations and predictions significantly different, and more realistic, than those obtained using mean field models. More details can be found in Section 3.3, where we provided a more in-depth survey of spatial models and their characteristics.

The key conclusion is that spatial heterogeneity can promote species coexistence and thus help to maintain diversity. This heterogeneity can have two broad sources, namely environmental and population dynamical (Neuhauser, 2001).

If environmental factors (such as temperature, pH, salinity, substrate concentration, etc.) vary over sufficiently small scales, then species can coexist by specializing in particular conditions (Hibbing et al., 2010). This is a key biological factor underpinning coexistence. From a modelling perspective, this type of coexistence is less interesting, since the species in the community do not interact with each other in any meaningful way. Each population is adapted and specialized to its own local habitat, therefore if dispersal allows for immigration and emigration between habitats, then coexistence is trivial (Neuhauser, 2001).

Spatial heterogeneity due to population dynamics is mathematically more interesting, while of course still being biologically significant. It should be noted that spatial heterogeneity does not always support coexistence, hence the complex and interesting nature of the topic. For example, the fragmentation of habitats (due to ecological disturbances or human actions) has been identified as a significant driver of species extinctions (Neuhauser, 2001). On the other hand, theoretical models have shown that merely limiting the number of (possibly threatening) neighbours an individual comes into contact with, is not the key mechanism supporting coexistence. Somehow, the spatially explicit nature of the interactions are playing a further role in promoting diversity (Laird, 2014)

The explicit consideration of realistic spatial heterogeneities is therefore an additional factor that contributes to the variability between individuals, which can be amplified to differences in population-level characteristics, and ultimately the pop-

ulation dynamics, including species coexistence (Werner et al., 2001). This can be particularly important in models of microbial systems, since "practically everything [microbes] do is in response to their environment" (Hellweger et al., 2016a), and hence a more realistic depiction of this environment will result in a more realistic depiction of the microbial community's dynamics.

We have seen in Chapter 6 that the dynamics of our foundational modelling framework is largely mediated by mobility - if individuals' interactions are sufficiently localized, then stable spatial structures will form and enable coexistence of all species in the community, thereby maintaining community biodiversity. Importantly, this occurred in a spatially explicit but homogeneous environment. With the extended IBM developed in Chapter 7, our *in silico* landscape is now heterogeneous due to the resource gradient. What effect does this spatial heterogeneity have on the community's population dynamics, and most importantly its biodiversity and functionality?

To address this question, we extend the approach presented in Chapter 7, in order to present in this chapter a comprehensive study of how resource dependence impacts biodiversity maintenance for *in silico* communities in heterogeneous space. We thus focus on studying the emergence of spatial patterns and the population dynamics of the community, as well as their underlying mechanisms and the interplay between them. For this purpose, we employ the spatially explicit individual-based model described in Chapter 6 to conduct further *in silico* experiments. This allows us to address research questions 2, 4, 5, and 6:

- Which types of competitive interactions can help maintain community diversity, and which types can threaten it?
- What effect does initial evenness have on maintaining community functionality?
- If interactions within a community are dependent on resource availability and use, how does this affect community diversity and functionality?
- How does the spatial structure of a community affect its stability and functionality?

In Section 8.2, we give a brief description of the model and how it is deployed to address our research questions. Then in Section 8.3 we describe the set-up of the *in silico* experiments which aim to uncover the system's dynamics. Results of these experiments are presented and discussed in Section 8.4, before conclusions are drawn in Section 8.5.

8.2 Model description

Our goal is to study the effects of emergent spatial patterns, due to resource dependence in the demographic processes, on biodiversity maintenance. For this purpose, we use the model developed in Chapter 7 with a different goal. We describe briefly the IBM using the ODD protocol, highlighting the aspects which differ from the protocol described in Chapter 7.

8.2.1 Overview

8.2.1.1 Purpose

The aim of the model is to investigate how community evenness and resource dependence in demographic processes affect the community's spatial population dynamics, including the probability of extinction. We also wish to investigate the effects of multiple demographic processes being simultaneously substrate limited, in particular the effects on community diversity and productivity, as proxies of the community's functionality.

8.2.1.2 State variables and scales

The state variables and scales are the same as those described in Section 7.2.1.2.

8.2.1.3 Process overview

The processes are the same as those described in Section 7.2.1.3.

8.2.1.4 Scheduling

The scheduling of the IBM is the same as described in Section 7.2.1.4.

8.2.2 Design concepts

- **Emergence:** the spatial patterns of the individuals and the environmental substrate, and the population-level dynamics of the community emerge naturally from the interactions at the local scale.

- **Interactions:** individuals interact with each other and their environment by consuming substrate, reproducing if located next to an empty site, exchanging sites with their neighbours, or competing with their neighbours.
- **Stochasticity:** the stochastic processes in the model are: the initial spatial distribution of the grid; the interactions between individuals and the environment; and the interactions between individuals.
- **Sensing:** if selected for reproduction, individuals can sense whether their neighbouring site is empty; if so, they will reproduce. Individuals can also sense the substrate concentration within the grid site where they are located, which affects their uptake and growth rate,
- **Observation:** the data collected from the IBM includes: the spatial distribution of individuals, the spatial distribution of substrate, the population count of each species and their probability of extinction, the community evenness, a spatial aggregation measure per individuals, the biomass distribution of individuals, their time to extinction. These are tracked for each time step.

8.2.3 Details

8.2.3.1 Initialization

The model is initialized in the same way as in Section 7.2.3.1.

8.2.3.2 Input

The input variables are: the initial community evenness, and substrate limitation scenario. Aside from the scenarios described previously in Section 7.3, we investigate additional subtrate limitation scenarios which are described in more detail in Section 8.3.

For each scenario, we check a range of initial evenness values, chosen so as to sample across the range of possible values of the Gini index, namely the unit interval.

8.2.3.3 Submodels

The submodels are implemented in the same way as described in Section 7.2.3.3.

8.3 *In silico* scenarios

We investigate the three substrate utilization scenarios described in Section 7.3. To facilitate analysis and comparison of the different scenarios, we also investigate a benchmark case.

In the benchmark case, demographic processes occur regardless of individuals' internal substrate level. Individuals absorb environmental substrate and convert this to biomass, but their biomass has no impact on interaction events.

In the first substrate limitation scenario, all three species are subject to the same substrate threshold and therefore have the same reproductive/competitive/mobile capacity: $\tau_A = \tau_B = \tau_C$.

In the second scenario, one species is subject to a substrate threshold τ, while the other two species are not (asymmetric limitation). Then for example species A has substrate threshold $\tau_A > 0$ while species B and C have substrate thresholds equal to zero, i.e. $\tau_B = \tau_C = 0$.

In the third scenario, a hierarchy in terms of substrate limitation is imposed on the community, e.g. $\tau_A > \tau_B > \tau_C$.

This results in ten scenarios: the benchmark scenario, and three scenarios for each of the demographic processes (reproduction, competition and mobility).

Aside from studying the spatial population dynamics of these ten resource limitation scenarios in isolation to avoid confounding effects, we also investigate simultaneously substrate-limited processes – that is, scenarios where two demographic processes are both substrate-limited.

To investigate the interaction between these two phenomena, we assign the same substrate threshold in the same way (symmetrically, asymmetrically or hierarchically between species) to both processes. For example, in the simultaneous symmetric scenario, we assign the thresholds for reproduction as $\tau_A^r = \tau_B^r = \tau_C^r = 10^{-16}$ g, and the thresholds for competition as $\tau_A^c = \tau_B^c = \tau_C^c = 10^{-16}$ g. In an asymmetric scenario, we assign for example $\tau_A^r = 10^{-16}$ g and $\tau_A^c = 10^{-16}$ g to species A while the other species are not subject to substrate thresholds.

Once the simulation set-up is specified (with specific substrate limitation and specific initial evenness), a set of simulations is carried out with the substrate threshold systematically varied from the lowest substrate threshold (10^{-16} g) to the highest (2×10^{-15} g), in twenty increments. Multiple outputs are tracked for each simulation; these are listed under "Observations" in Section 8.2.2 above.

8.4 Results and discussion

In Chapter 7, we used this model to study the impact of substrate limitation on community productivity, as a proxy for ecosystem functioning. Here, we present a more comprehensive study of the model's behaviour, and the insights this can lead to with regard to the mechanisms under investigation. We present and discuss results related to pattern formation, population dynamics, the effects of more than one demographic process being simultaneously substrate-limited, and the effect of resource dependence on the system's critical mobility rate.

8.4.1 Pattern formation

To study the role of space in maintaining biodiversity in our system, we examined the spatial evolution of our modelled communities under the different resource limitation scenarios.

In the benchmark case, the different species arrange themselves into stable rotating spirals, as seen in the Reichenbach model (2007), and experimental results for cyclically competing *E. coli* strains (Kerr et al., 2002). Introducing resource dependence restricts the development of these spiral structures to those areas of the grid where there is firstly sufficient substrate (which we recall is diffusing outwards from the central source), and secondly individuals with the ability to exploit such substrate (Figure 8.1). These emergent spatial patterns broadly agree with those observed in experimental studies of three bacterial strains in cyclic competition (Weber et al., 2014).

We notice a significant difference in behaviour between resource-dependent reproduction and competition when compared to resource-dependent mobility. In the first two cases, a notable zone of activity emerges where interactions occur more frequently due to the relatively higher concentration of environmental substrate compared to the outer areas of the grid.

We observe little impact of resource dependence on mobility. As noted previously, strong effects are only observed when mobility exceeds a critical rate, allowing individuals to move over distances too great to permit local interactions. In this case of high mobility, we approach a setting where the population can be considered well mixed, and therefore the mean field approximation becomes relevant. This approximation predicts that the coexistence equilibrium is not asymptotically stable and therefore extinctions are frequent (Szabó and Fath, 2007). Hence as long as mobility in our model does not decline to zero nor exceed the critical rate, we notice little difference between low mobility and even lower mobility (as induced by increasing the substrate threshold).

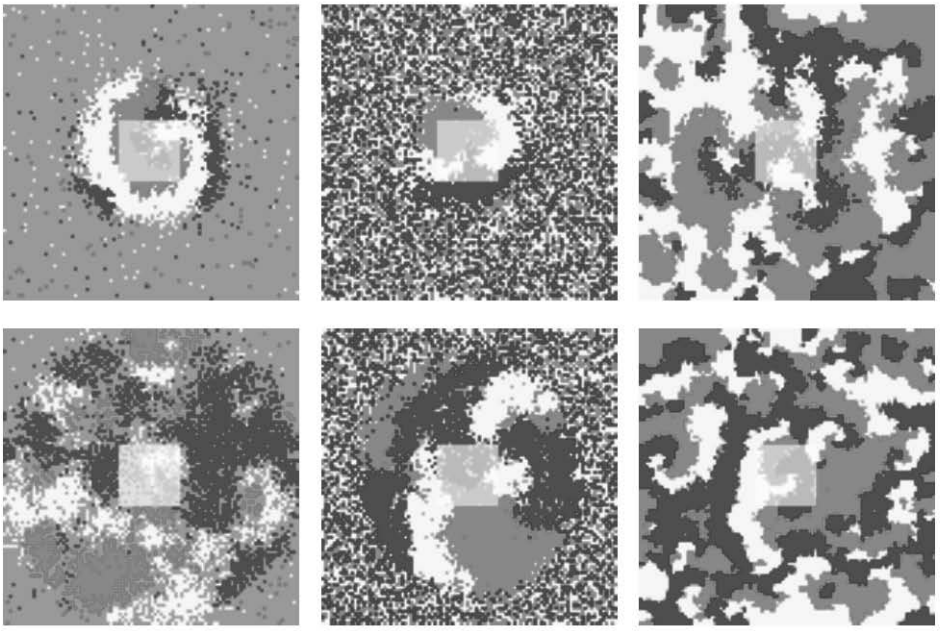


Figure 8.1: Comparison of pattern formation for resource-dependent demographic processes, after 500 generations. The box in the centre indicates the size of the substrate source. Top: substrate threshold $\tau = 10^{-16}$ g. Bottom: substrate threshold $\tau = 2 \times 10^{-15}$ g. Left to right: substrate-limited process — reproduction, competition, mobility.

In the case of substrate-limited reproduction, the substrate gradient implies that most reproductive events occur closer to the centre of the grid, where the high-biomass individuals are predominantly located. Outside this zone, reproductive events occur rarely since there is less environmental substrate to fuel growth. Thus individuals killed in competition events in this zone are not replaced, and the grid becomes depopulated. When the substrate threshold is increased (see bottom row of Figure 8.1) the zone of activity expands due to the increased substrate cost to reproduce.

The existence of the zone of activity is confirmed using the density plots in Figure 8.2, which illustrate the number of interaction events that occurred per cell over the course of a simulation. We observe similar behaviour in the cases of substrate-limited reproduction and competition (hence we show only one example of each in Figure 8.2): most of the interaction activity occurs in a central zone, which expands when the substrate threshold is increased. In the case of substrate-limited mobility, the interaction activity is not confined to any particular area, but occurs throughout the grid. It is possible in this case to distinguish the characteristic spiral formations, since there is much activity occurring at their edges, being the interfaces of different species aggregations.

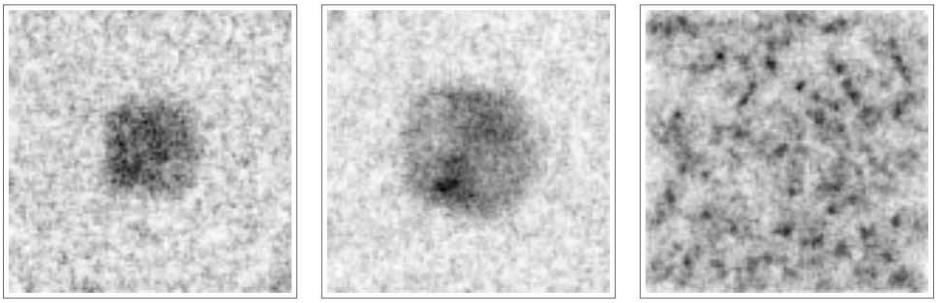


Figure 8.2: Activity for resource-dependent demographic processes, after 500 generations. Darker colour indicates that a higher number of interaction events were executed during the course of the simulation. Left to right, substrate-limited process: competition (threshold $\tau = 10^{-16}$ g), reproduction (threshold $\tau = 1.2 \times 10^{-15}$ g), mobility (threshold $\tau = 5 \times 10^{-16}$ g).

The mechanism driving the formation of the zone of activity is the following: individuals require more time to grow sufficient biomass to reproduce, and in this extended time the substrate has diffused farther away from the centre, expanding the zone of activity. However, within this expanded zone, comparatively fewer reproductive events are occurring. Hence individuals become more disaggregated, since empty grid cells are not filled as easily as for lower substrate thresholds. Thus increasing the substrate threshold does not greatly affect spatial aggregation per species, as measured by patchiness (Bez, 2000), since the expansion in the zone of activity is counterbalanced by the decrease in aggregation (van de Koppel et al., 2005).

An analogous phenomenon is seen for substrate-limited competition, where instead of depopulation in the outer areas of the grid, we observe a lack of spatial structure. The spiral structures characteristic of this type of model emerge as a result of the local cyclic competition scheme, and hence are confined to the zone of activity where individuals can find sufficient substrate to fuel competition. Outside this zone, few competition events occur and thus the community remains well mixed. Again the zone of activity expands when the substrate threshold is increased, while there is no disaggregation as seen for substrate-limited reproduction.

These results support the mechanistic explanation of spatial structures in microbial biofilms, which describes the formation of various structures in biofilms as a consequence of differences in local substrate availability (Wimpenny and Colasanti, 1997), more specifically as occurring under substrate-limited conditions such as those modelled here (Tolker-Nielsen and Molin, 2000).

In addition to the scenarios described above with a square source in the centre of the lattice, we also investigated different geometries for the substrate inflow region. If the substrate inflow was homogeneous across the lattice, the lattice became saturated homogeneously with substrate, and we did not observe any

spatial gradient in this case. In particular, the population aggregation patterns seen in Figure 8.1 do not occur. Instead, due to the lack of a significant spatial substrate gradient, the characteristic spiral structures may form over the entire grid in a similar way to the substrate-less case, but with a delay due to the time needed for individuals to uptake sufficient substrate. Thus the same population persistence typical of the original Reichenbach (2007) model can be observed. On the other hand, if the substrate flows into the lattice via the boundaries, we found the same qualitative effects as in the central source case, but the lattice became saturated with substrate more quickly since the diffusive front was larger. These results are omitted for reasons of length, and since we considered the case of a central source inflow more interesting due to the spatial gradients it produces, and its representation of an experimental set-up where substrate is provided via a deposit in the centre of a Petri dish, for example.

8.4.2 Population dynamics

The spatial structures described in Section 8.4.1 also have significant effects on the overall population dynamics. The mean field approximation of the rock-paper-scissors model predicts a single coexistence equilibrium (which is stable but not asymptotically so) and three stable homogeneous equilibria representing the three possible monocultures (Reichenbach et al., 2006). However, the mean field approximation relies on the assumption that the population is well mixed, which is not relevant in the spatially explicit case we consider. This explicit consideration of space therefore produces significantly different population dynamics, as does the resource dependence of the demographic processes.

An example of the evolution of the species proportions for hierarchically limited competition is shown in Figure 8.3. In this scenario, species A is subject to the highest substrate threshold for successful competition, with $\tau_A = 2 \times 10^{-16}$ g, while species B has a substrate threshold that is half that of species A, and species C has no substrate threshold and may therefore compete with no regard to its internal substrate level. The simulation is initialized with a completely even community. As the simulation evolves, species B and C benefit from lower substrate thresholds than species A, and prosper and persist in oscillating proportions. Meanwhile species A declines but does not collapse to extinction.

The mechanism underlying this behaviour has previously been noted in a voter-type model of cyclic competition between three populations, with the additional process of one population type being “externally supported”, in the sense of having stronger competitive and reproductive ability than the other two types (Tainaka, 1993). In our model, this “externally supported” species is species C, since it is not resource-limited in contrast to the other two species. Thus species C kills more of its prey (species A), whose population decreases. Thus species A kills fewer of

species B, which thereby achieves an advantage in the dynamical balance since it has more prey (species C) and less predators (species A). The advantage of species B is less dramatic in our model than in that of Tainaka (1993) and other models (Szabó and Fath, 2007), since in our model the externally supported species enjoys an advantage in only one process (competition) as opposed to two, as in the voter model studied for example by Tainaka (1993), where reproduction and competition are coupled, which is characteristic of such models and in contrast to rock-paper-scissors models.

Thus all three species may coexist in unequal proportions, behaviour not seen in the cases without a substrate limitation, or homogeneous substrate inflow. In these cases, the species can only coexist in roughly equal proportions, since any small differences in population proportions unbalance the rock-paper-scissors dynamics (Laird, 2014). These differences in proportions (due to stochasticity or otherwise) are quickly magnified and inevitably lead to the extinction of two species, with the third species persisting alone. A clue to the mechanism underlying this behaviour can be found in the mean field approximation of the three species model in the substrate-less case, where the only stable equilibria are those representing equal coexistence, or survival of a single species (see Section 5.2.2).

The existence of a stable unbalanced community also agrees qualitatively with the predicted outcome of siderophore-mediated bacterial competition between two species (Hibbing et al., 2010). In this system, both species require a compound, called a siderophore, to chelate iron. When one species does not produce a siderophore itself, but rather uses that produced by the other species, the predicted outcome of the competition for iron is that the 'cheating' species will dominate the 'honest' species, since it benefits from a lower cost by not producing its own siderophore (Hibbing et al., 2010). Our model produces analogous behaviour for a three-species equivalent, where the third species C also benefits from cheating.

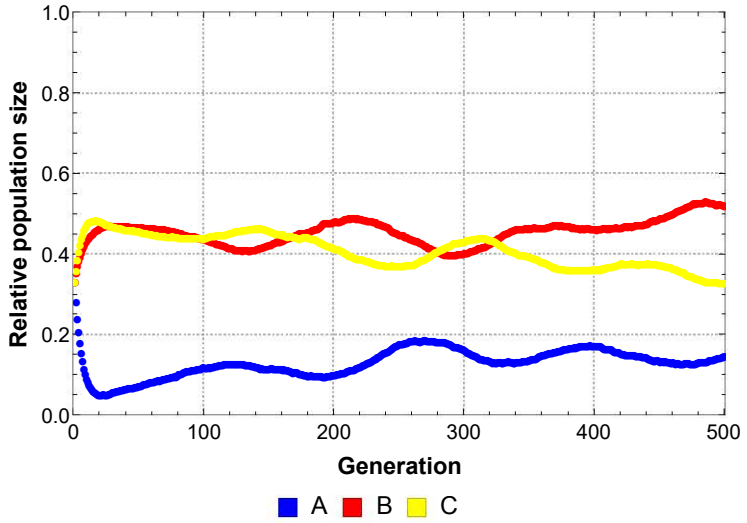


Figure 8.3: Evolution of species proportions for hierarchically limited competition ($\tau_A > \tau_B > \tau_C$) in a community which is initially completely even.

The existence of this type of stable unbalanced community in our model is facilitated by the emergence of the spatial structures discussed in Section 8.4.1, where most of the interaction events occur in a central 'zone of activity' determined by the substrate gradient, where all three species are present. Outside this zone, dynamics are significantly different (see Figure 8.4). This behaviour agrees qualitatively with patterns observed in juvenile mussel banks, where it was noted that "self-organization allows mussels to persist at algal concentrations that would not permit survival of mussels in a homogeneous bed" (van de Koppel et al., 2005). In our model, the spatial structures permit the persistence of species at lower levels than would be possible under homogeneous spatial conditions, as the spatial heterogeneities provide 'refuges' for these species, notably in the central zone of activity.

That these unbalanced communities are stable and persisting agrees with microbiological studies showing that the presence of spatial refuges can enhance community resistance to stress and disturbance, providing a buffer against adverse effects on community composition and function (Baho et al., 2012).

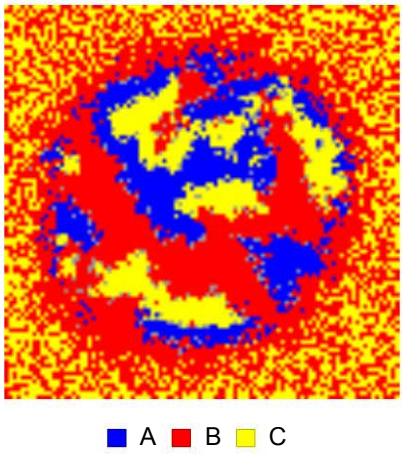


Figure 8.4: Grid configuration after 500 generations for hierarchically limited competition ($\tau_A > \tau_B > \tau_C$) with high initial evenness.

We note that the formation of spatial refuges observed in our model is qualitatively different from a phenomenon described as "coexistence by small numbers", which has been observed in other individual-based modelling studies of non-transitive competition (Abrudan et al., 2016). This term refers to the situation where a species collapses to near extinction due to competition effects, but does not disappear completely - a very small number of individuals remains present, completely disaggregated from other individuals of the same species, and entirely surrounded by non-competitors (e.g. an individual of species A survives in a neighbourhood consisting entirely of individuals of species B). Such a scenario is theoretically consistent with experimental studies that suggest that communities in nature are often dominated by only a few species, with many other species present in low quantities (McGill et al., 2007), but has also been suggested as an artefact of the neighbourhood structure that was used (Abrudan et al., 2016).

The stable unbalanced communities evolved using our model are much less unbalanced than those in the case of "coexistence by small numbers", where the least abundant species may number only a handful of individuals. In our case, the least abundant species is still present in a significant proportion relative to the rest of the community, and individuals remain fairly aggregated with their conspecifics (Figure 8.4).

A different type of behaviour is illustrated in Figure 8.5, where the evolution of the species proportions is shown for a simulation of the asymmetrically limited reproduction scenario with substrate thresholds $\tau_A = 2 \times 10^{-16}$ g and $\tau_B = \tau_C = 0$. In this case, a cheater species (B) again dominates the honest species (A), but here the second cheater species (C) does not prosper. This is due to the imposed cyclic competition structure, which in this case is not substrate limited. Hence species C is more vulnerable to its predator (species B) than in the case shown in Figure 8.3,

and is depressed by the increased competitive pressure despite its stronger reproductive capacity relative to the honest species (A). The stable persistence of all three species in unbalanced proportions is again permitted by the emergent spatial structure discussed in Section 8.4.1, which resembles the one shown in Figure 8.4, but now species C is also absent from the outer regions of the grid.

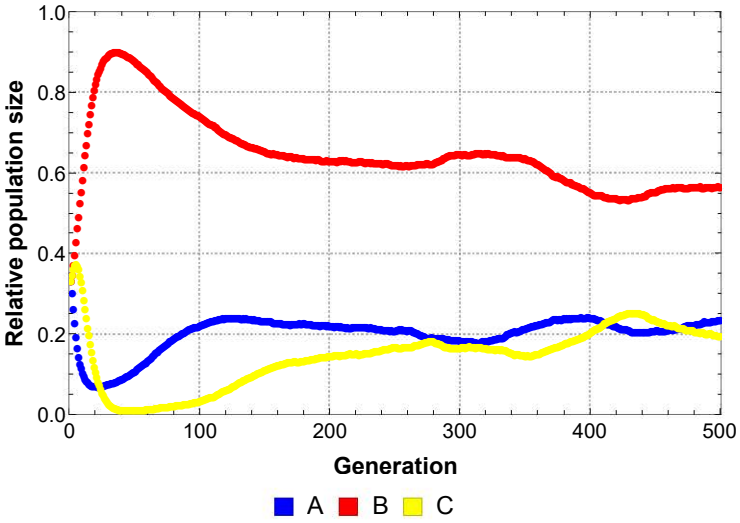


Figure 8.5: Evolution of species proportions for asymmetrically limited reproduction ($\tau_A > \tau_B = \tau_C = 0$) in a community which was initially completely even.

In a third and final example, we show in Figure 8.6 the evolution of the species proportions for the scenario of asymmetrically limited reproduction with substrate threshold $\tau_A = 2 \times 10^{-16}$ g and $\tau_B = \tau_C = 0$, with intermediate initial evenness. The behaviour again agrees with results observed in siderophore-mediated competition (Hibbing et al., 2010), where a species able to monopolize the available iron will force the other species into extinction. In our three species case, the species with the highest cost (A) quickly collapses to extinction as the others monopolize the available substrate. The subsequent resource competition between the remaining two species is then once more determined by the cyclic competition scheme, which explains why species B outcompetes species C until the latter species collapses to extinction. In this case, the unstable dynamics are magnified by the lower initial evenness, which we previously observed as increasing the probability of species extinctions by increasing the amplitudes of the population oscillations (see Section 7.4.2).

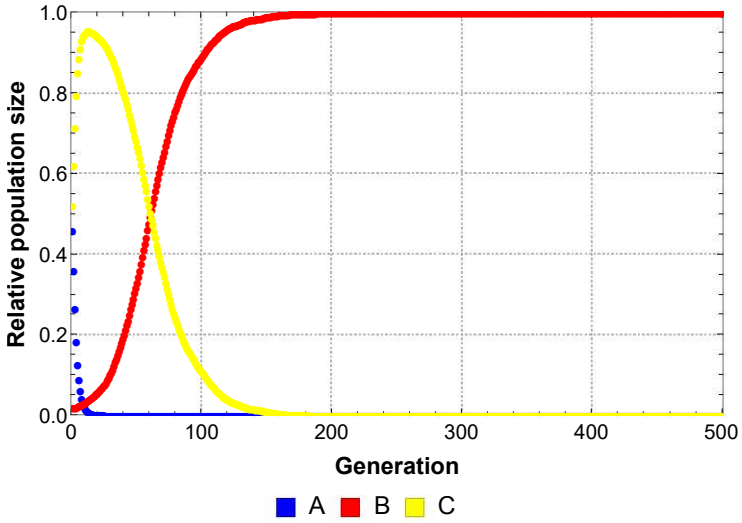


Figure 8.6: Evolution of species proportions for asymmetrically limited reproduction ($\tau_A > \tau_B = \tau_C = 0$) in a community with initially intermediate evenness.

8.4.3 Probability of extinction

In the benchmark case of no resource dependence (analogous to the model of Reichenbach *et al.* (2007)), the biodiversity of the system is moderated by the mobility rate; once the mobility exceeds a certain critical rate ϵ_c (which scales with system size), the interactions are no longer sufficiently localized to permit long-term coexistence, and the system will suffer extinctions and tend to monoculture.

When the demographic processes become resource-dependent, we have seen in Section 8.4.2 that extinction events become more common, and the tendency to monoculture increases, particularly for the asymmetric and hierarchical limitation scenarios. This raises the question of what effect the substrate threshold has on the extinction probability of the system, and whether strength of resource limitation or mobility rate takes precedence in moderating long-term coexistence.

Therefore, selecting a mobility rate significantly smaller than the critical value of the benchmark case ($\epsilon_c = 10.63$), which would in the benchmark case ensure long-term coexistence, we calculated the extinction probabilities obtained for simulations with varying substrate thresholds. In this way we observed the impact of the different resource limitation scenarios and substrate thresholds on the extinction probability of the system. The results are shown in Figure 8.7. For the symmetric limitation scenario, imposing a substrate threshold does not increase the probability of extinction past the benchmark level (approximately zero, implying long-term coexistence). This agrees with the results reported in Section 7.4.1, which show that imposing a substrate threshold reduces biodiversity levels slightly

below benchmark levels, but does not provoke extinction events to reduce biodiversity more dramatically. We observe a similar effect here.

For the asymmetric and hierarchical scenarios, the results of increasing the substrate threshold are more dramatic, again in agreement with results in Section 7.4.1 where it was observed that above very small substrate thresholds, for these two scenarios there was a significant decrease in biodiversity due to extinction events, in particular for the asymmetric limitation scenario (compare with Figure 7.4). There remains a significant range of the tested values of τ where extinction probabilities remain low or close to zero, permitting the type of coexistence seen for example in Figures 8.3, 8.4, and 8.5.

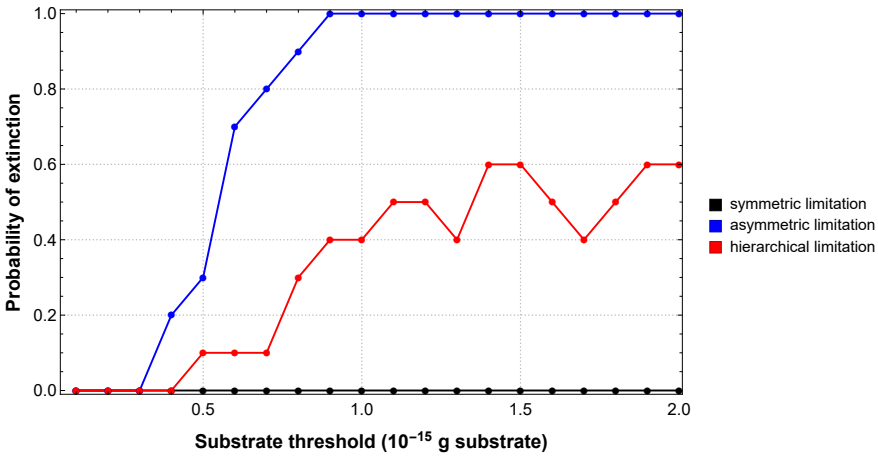


Figure 8.7: Probability of extinction as a function of substrate threshold. The symmetric, asymmetric and hierarchical limitation scenarios were tested, for the competition process. The mobility rate of the system was fixed at $\epsilon = 4.5$.

We therefore conclude that the main determinant of long-term coexistence becomes firstly the resource limitation scenario. If the scenario is asymmetric or hierarchical, this will imply significant extinctions despite a low mobility rate. Individuals may benefit from spatial refuges and persist in unbalanced but coexisting communities (cfr. Section 8.4.2), but these scenarios are vulnerable to stochastic effects which may provoke extinctions, and thus extinction probabilities for these cases can be significant. If the limitation scenario is symmetric (mirroring the more balanced dynamics of the benchmark case), increasing the substrate threshold within the same range does not induce extinctions, and therefore the mobility rate remains the principal determinant of long-term coexistence.

8.4.4 Simultaneous substrate limitations

Examples of the grid configurations produced by these simulations are shown in Figure 8.8, where a clear separation of regions can be observed. When reproduction and competition are both symmetrically limited, we observe two central zones: the characteristic spiral formations in the centre, enclosed firstly by empty space (as was seen for substrate-limited reproduction) and secondly by a region where individuals are present, but randomly mixed (as was seen for substrate-limited competition).

In the cases where reproduction and competition are both asymmetrically or hierarchically limited, we again observe unbalanced communities due to the spatial heterogeneities providing refuges for a vulnerable species. For asymmetric simultaneous limitations, the vulnerable species is species C; compare with Figures 8.5 and 8.6 where in the case of a single substrate-limited process, for high initial evenness species C persisted in similar proportion to species B and for lower initial evenness species C collapsed to extinction. In this simultaneous limitation case, the effect is somewhere in between: species C persists but at a lower proportion than the other two species.

When both reproduction and competition are hierarchically limited, the vulnerable species is species A, which in this scenario is subject to the highest substrate threshold. This behaviour is similar to the effects of a single hierarchically limited process (see Figure 8.4).

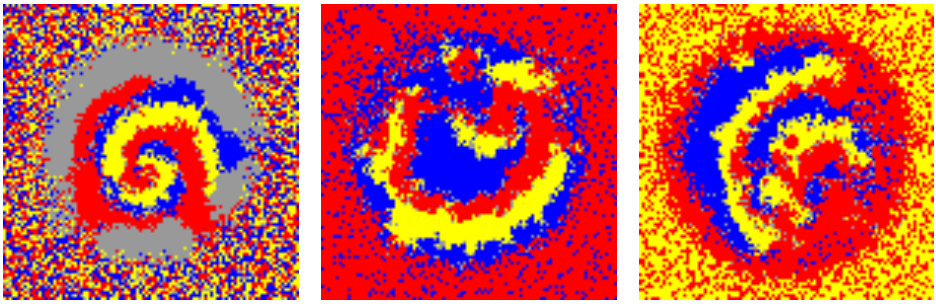


Figure 8.8: Grid configuration after 500 generations for a community where both competition and reproduction are (left to right) symmetrically, asymmetrically and hierarchically substrate-limited ($\tau = 10^{-16}$ g).

Making reproduction and competition simultaneously resource-dependent also has a significant effect on community productivity, which collapses compared to the cases of individual processes being resource-dependent. This is shown in Figures 8.9-8.11, where community productivity is compared for symmetric, asymmetric and hierarchical scenarios, respectively. In all cases, the productivity of

communities subject to simultaneous process limitations never exceeds the productivity of the benchmark case.

Thus while resource-limiting two processes simultaneously does not greatly alter the population dynamics by way of significantly increased extinctions compared to individual limitation scenarios, in the simultaneous case individuals must divide their resource allocation between reproducing and competing with their neighbours. This trade-off significantly depresses community productivity, although this effect is not greatly increased by increasing the substrate threshold, having reached a plateau at intermediate threshold values. This suggests that for high substrate thresholds, the resource allocation trade-off between reproduction and competition has constrained biomass growth to its minimum, and thus further increasing the substrate threshold has little effect on productivity.

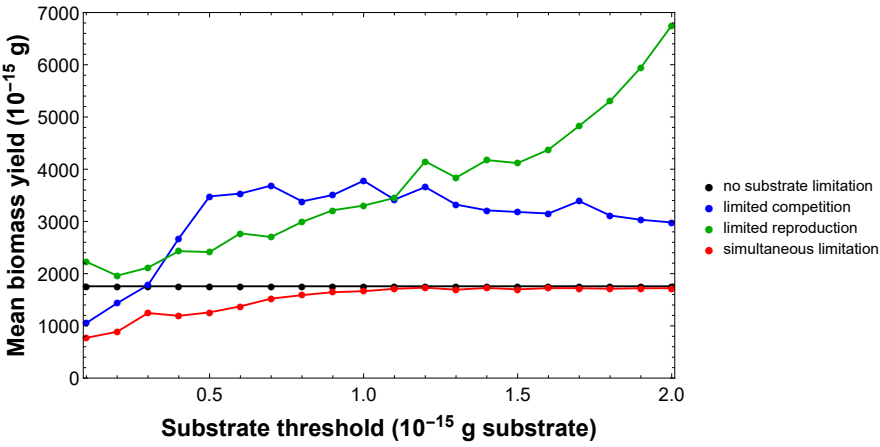


Figure 8.9: Comparison of average biomass yield for the symmetrically limited scenarios. Each curve represents the mean of 200 simulations.

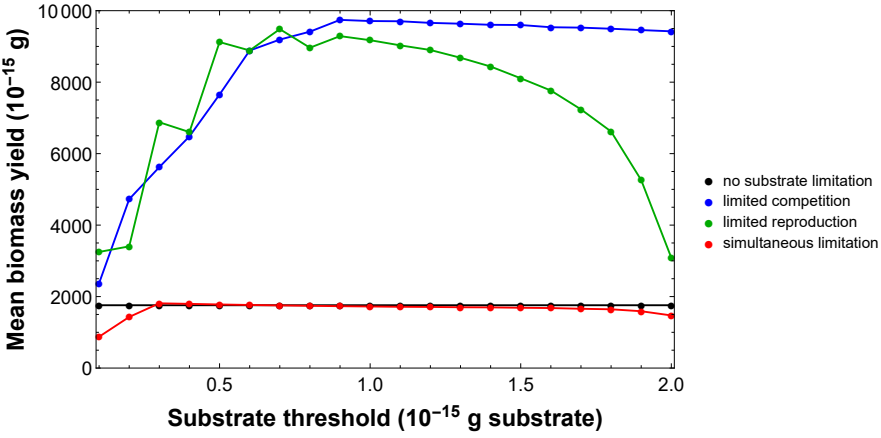


Figure 8.10: Comparison of average biomass yield for the asymmetrically limited scenarios. Each curve represents the mean of 200 simulations.

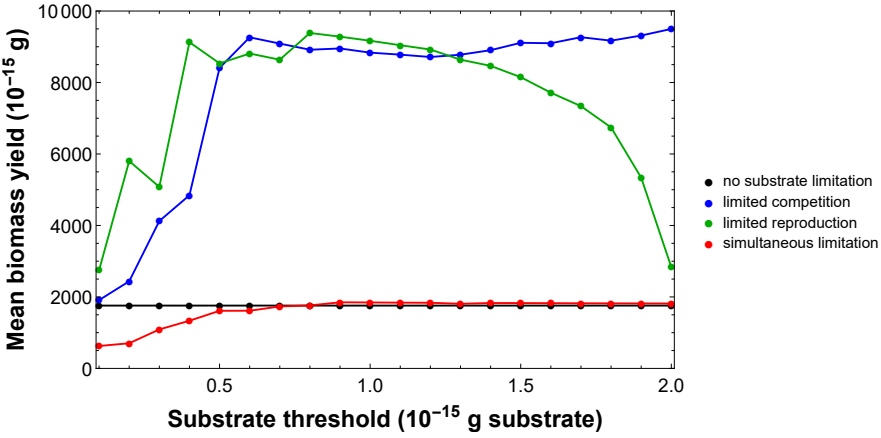


Figure 8.11: Comparison of average biomass yield for the hierarchically limited scenarios. Each curve represents the mean of 200 simulations.

8.5 Conclusions

We have extended existing microscopic models of three cyclically competing species by considering resource-dependent demographic processes in a spatially heterogeneous landscape, thereby providing a more realistic *in silico* representation of natural systems. The explicit treatment of space, which permits resource gradients, can induce dramatic effects in the system population dynamics. These effects, consistent with other modelling and experimental studies, are not seen in well-mixed models due to the absence of spatial heterogeneities in such models,

thereby neglecting this key facet of natural systems. Our findings have implications for the formation and maintenance of spatial patterns in microbial populations such as biofilms.

If validated with experimental data, such a model can be used to predict and visualize unobserved substrate gradients which can be experimentally impractical or infeasible to measure directly (Hellweger et al., 2016a). The validation would require the spatial distribution of the cells, acquired for example from image analysis of microscopy images, as well as the quantification of the substrate uptake kinetics.

PART III

INCORPORATING DATA FROM *IN VITRO* SYNTHETIC MICROBIAL COMMUNITIES

9

Towards a simulation framework for synthetic microbial communities

9.1 Introduction

In Chapters 7 and 8, we have studied *in silico* synthetic microbial communities under co-culture conditions, addressing research questions related to community diversity, functionality and productivity. In this chapter, we will draw on data from an *in vitro* synthetic community in order to investigate whether our *in silico* approach is capable of mimicking the dynamics of a similar *in vitro* counterpart.

We have discussed in Section 2.3 the use of co-culture experiments to study interactions between microbes, which we have applied *in silico* in Chapters 5–8. A more specific application of co-cultures is bioaugmentation, where the biomass in water treatment plants is altered by the addition of certain microbial strains that have been selected for their ability to degrade specific chemical compounds (Hairston

et al., 1997). For example, during the treatment of drinking water, the common groundwater pollutant 2,6-dichlorobenzamide (BAM) must be removed below a threshold concentration of 0.1 $\mu g\,L^{-1}$ to meet the EU Directive on Drinking Water (EU, 2006). However, the endogenous microbial communities in the sand filters (SFs) of such drinking water treatment plants are not capable of achieving BAM removal to concentrations below this threshold (Björklund et al., 2011). Therefore, bioaugmentation of SFs has been proposed as an alternative strategy, by the addition of a specialized BAM mineralizer such as *Aminobacter* sp. MSH1 (Sørensen et al., 2007). This has already been tested in laboratory-scale SFs containing different types of filter material (Albers et al., 2014) and in pilot scale rapid SFs (Albers et al., 2015). However, studies of this type of bioaugmentation of drinking water ecosystems rarely address how exactly the pesticide degrader interacts with the resident community, or other such fundamental ecological questions (Thompson et al., 2005).

From an ecological point of view, bioaugmentation represents a form of microbial invasion process (cfr. Section 3.4.2.3), where the strains introduced to augment resident community functionality are the invaders. The introduced strains must establish themselves in the resident community, by maintaining a metabolically active population for a significant period of time (Kinnunen et al., 2016).

In Vandermaesen *et al.* (2017, in prep), the authors hypothesize that the establishment of MSH1 and its subsequent BAM mineralization in SFs not only depend on exploitative competition effects, but also on other features such as interactions with resident community members. Therefore, the BAM mineralization activity of MSH1 was evaluated in sand microcosms in the presence of a selection of the 13 sand filter isolates (SFI) described in Vandermaesen *et al.* (2017, in prep). Synthetic microbial communities of MSH1 combined with SFI were subjected to an initial competition phase. Subsequently, BAM was added and the kinetics of BAM mineralization was evaluated as a measure of bioaugmentation success.

To characterize the interactions between resident community members, co-cultures of various combinations of SFI with MSH1 were inoculated, and their mineralization kinetics was followed. However, given the total number of strains in the community, it is practically impossible to experimentally study all possible co-culture combinations. In addition, for the limited number of combinations that is feasible to investigate, one runs into the issues with *in vitro* co-culture experiments described in Section 2.3.1. In such cases, predictive modelling is becoming more and more appreciated as a tool for identifying possible co-cultures of interest (Esser et al., 2015; Seshan et al., 2014; Poschet et al., 2005; Widder et al., 2016). If the mineralization kinetics can be predicted for all co-cultures, the subset of combinations that appear interesting for the particular application (in this case, bioaugmentation) can then be extensively studied *in vitro*, thereby reducing the experimental load.

To determine the feasibility of such an approach for the experimental system de-

scribed in more detail in Section 9.2, we will illustrate in Section 9.3 how to predict of the mineralization kinetics of co-cultures of two SFIs with MSH1, from the mineralization kinetics of monocultures of SFI with MSH1. In this case, we have a completely characterized dataset. This provides us with an ideal setting to investigate in Section 9.3 the possibilities of predictive modelling of co-culture growth in this well-defined system.

Then in Section 9.4, we use the modelling framework developed in previous chapters to construct the *in silico* counterpart of the *in vitro* synthetic community used in the experiments of Vandermaesen *et al.* (2017, in prep), with the goal of qualitatively reproducing the observed dynamics. This requires adjustments to our modelling framework to bring it closer to reality, particularly in modelling the competitive interactions between individual microbes, in order to take advantage of the knowledge and data gained from the Vandermaesen *et al.* experiment. This process is described in Section 9.4.2, and the resultant model is described in Section 9.4.3, as well as the set-up of the *in silico* experiments it is employed for. The results of these experiments are presented and discussed in Section 9.4.4. Finally, in Section 9.5 we summarize the conclusions of the modelling and simulation studies carried out in this chapter.

9.2 Materials and methods

In this section, we summarize the experimental set-up and procedure used by Vandermaesen *et al.* (2017, in prep) to obtain the dataset that we will use in the remainder of this chapter for our modelling and simulation studies.

9.2.1 Experimental set-up

The hypothesis of this *in vitro* study was that the establishment of MSH1 and its subsequent BAM mineralization in SFs depend on interactions with and between resident community members. Therefore, the BAM mineralization activity of MSH1 was evaluated in sand microcosm co-cultures in the presence of different combinations of 13 SFI. Synthetic microbial communities of MSH1 combined with SFI were co-cultured, then BAM was added and the kinetics of BAM mineralization was evaluated as a measure of bioaugmentation success.

9.2.1.1 Bacterial strains

The specific variant of the BAM mineralizing *Aminobacter* sp. MSH1 (Sørensen et al., 2007) used in this study, MSH1-GFP, was fluorescently tagged. The 13

SFI used were isolated from SF material from two drinking water treatment plants (Vandermaesen *et al.*, 2017, in prep): *Acidovorax* sp. S9, *Undibacterium* sp. S22, *Brachybacterium* sp. S51, *Mesorhizobium* sp. S158, *Acidovorax* sp. S164, *Rhodococcus* sp. K27, *Acidovorax* sp. K52, *Aeromonas* sp. K62, *Paucibacter* sp. K67, *Pelomonas* sp. K89, *Rhodoferax* sp. K112, *Rhodoferax* sp. K129, and *Piscinibacter* sp. K169. None of the selected SFI were capable of BAM mineralization, avoiding any confounding effects with the BAM mineralization performance of MSH1.

9.2.1.2 Microcosm set-up

Microcosms were created in deep 96-well plates, containing sterile sand in every well. MSH1 and SFI were cultured and prepared as described in Vandermaesen *et al.* (2017, in prep) and combined in synthetic communities in such a way that the number of cells of every strain was 10^7 cells/mL. Since each community included MSH1, the total richness of a community R_T is given by $R_T = R_{SFI} + 1$, where R_{SFI} is the number of SFI present. In addition to all combinations of individual SFI with MSH1 (13 combinations at $R_{SFI} = 1$), all 78 different pair combinations of two SFI with MSH1 ($R_{SFI} = 2$) were tested.

Sodium acetate was provided as the only carbon source at a concentration of 150 $\mu g\,L^{-1}$ in MMO (Minimal Medium ONPG) medium. Assuming that 50% of acetate-C is actually assimilated, this corresponds to an AOC concentration of 22 $\mu g\,C/L$, which is within the range of AOC values in drinking water ecosystems (20-100 $\mu g\,C/L$) (Lehtola et al., 2002). Of every synthetic community, 100 μL was inoculated in the sand microcosms. A reference microcosm inoculated with 100 μL MSH1 at 10^7 cells/mL ($R_{SFI} = 0$) was included in every deep well plate. In addition, to account for abiotic $^{14}CO_2$ production, one negative control ($R_T = 0$) was included, containing sand amended with 100 μL MMO+Ac. All synthetic communities and controls were replicated four times. No $^{14}CO_2$ production was observed in the abiotic control. The plates were sealed and incubated at 20°C for 7 days.

After this initial competition phase, all wells were spiked with ^{14}C-BAM, dissolved in 5 μL MMO, which corresponds to a final BAM concentration of 150 $\mu g\,L^{-1}$. BAM mineralization was then followed for $\pm$130 h by trapping the BAM-produced $^{14}CO_2$ with $Ca(OH)_2$. Trapped $^{14}CO_2$ radioactivity was quantified by digital autoradiography. The cumulative percentage $^{14}CO_2$ was plotted relative to the total amount of ^{14}C added as a function of the incubation time, and hence cumulative mineralization curves were obtained.

9.2.2 Modelling of mineralization kinetics

To describe the kinetics of BAM mineralization, the modified Gompertz model (Zwietering et al., 1990) was used. This model is one of the most commonly used

microbial growth models (Buchanan et al., 1997), and is given by

$$P = A \exp\left(-\exp\left(\frac{\mu e}{A}(\lambda - ct) + 1\right)\right), \tag{9.1}$$

where P (%) is the percentage mineralization at time t (h), A (%) is the total extent of mineralization after the exponential mineralization phase, λ (% h^{-1}) is the lag time, c (% h^{-1}) is the endogenous mineralization rate, and μ (% h^{-1}) is the maximum mineralization rate constant. The modified Gompertz model differs from the standard Gompertz model (Zwietering et al., 1990) in that its parameters each have a biological meaning, whereas the parameters of the standard model do not reflect any biological attribute but merely determine the function's shape.

The Gompertz parameters of the cumulative mineralization curves were determined by least squares curve fitting, using the Trust-Region-Reflective algorithm (Coleman and Li, 1994, 1996), at a termination tolerance of 10^{-14} and allowing at most 2×10^5 function evaluations and 3×10^5 iterations. Initial parameter estimates were set at 30, 5, 0.1, and 2 for A, μ, c, and λ, respectively (Vandermeeren et al., 2016). This was implemented using Matlab R2012b (Mathworks, USA). All values of c were zero or close to, and were hence excluded.

9.2.3 Description of the dataset

From the experimental set-up described in Section 9.2.1, we obtained a dataset representing 13 monocultures (the individual strains) and 78 co-cultures (the pair combinations). For each of these 91 conditions, we have two types of mineralization data.

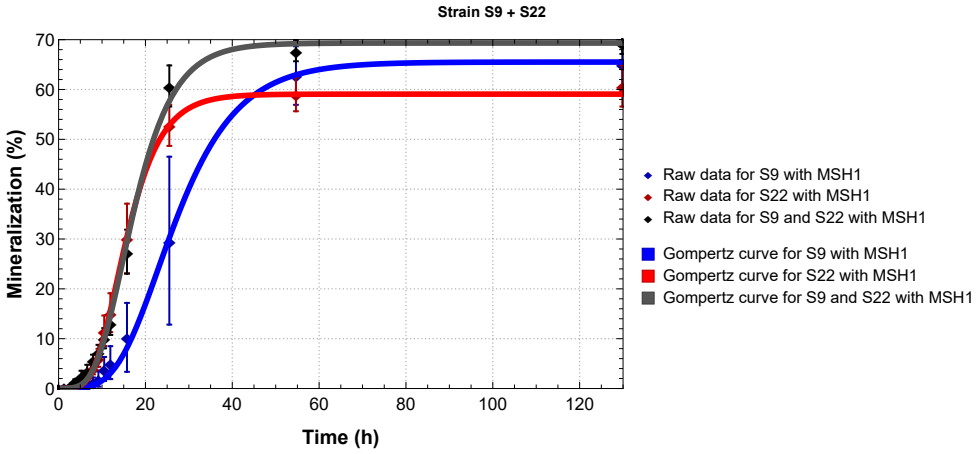
First, a cumulative BAM mineralization time series consisting of achieved mineralization values at 13 time points, from $t = 0$h to $t = 130$h. There are four biological replicates of each time series, except where some outliers were removed as indicated in Vandermaesen *et al.* (2017, under review). In total, 21 out of 364 time series were removed. After removal of these outliers, no condition had less than three replicates. The second data type consists of the fitted Gompertz parameters λ, μ and A describing the mineralization kinetics. Examples of both data types are shown in Figure 9.1.

9.3 Predictive modelling

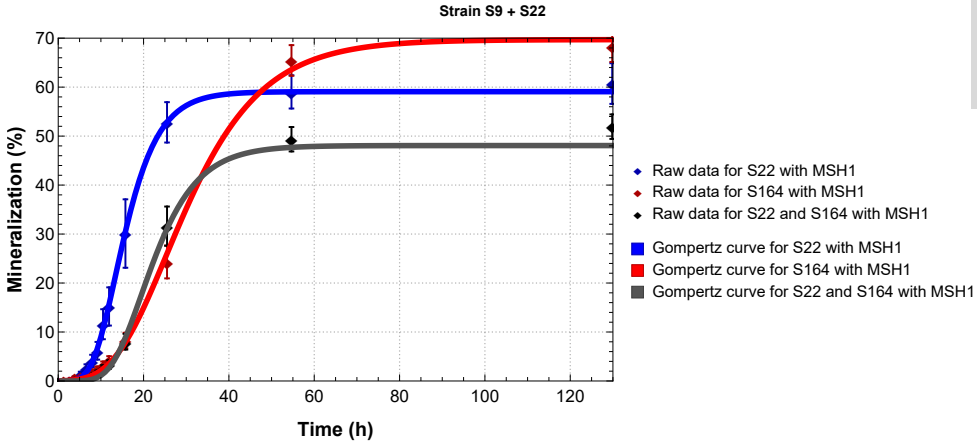
9.3.1 Modelling approach

An initial inspection of the experimental data reveals clear interaction and identity effects as hypothesized in Vandermaesen *et al.* (2017, in prep). Examples are shown in Figure 9.1. Figure 9.1(a) shows a case where the mineralization achieved by the co-culture has a lower lag time λ, a higher mineralization rate μ, and a higher mineralization extent A than the constituent monocultures (in this case, strains S9 and S22).

The most important factor in bioaugmentation success has been postulated (Ekelund et al., 2015) as the invader's ability to grow quickly and establish itself, reflected by the mineralization rate μ and the lag time λ, and not the total amount of accumulated mineralization A. Therefore, positive interactions between MSH1 and SFI in co-culture were defined as those increasing the mineralization rate and shortening the lag time, and vice versa for negative interactions. Therefore, the effect shown in Figure 9.1(a) is classified as positive. In contrast, some co-cultures result in poorer mineralization performance compared to the monoculture cases, i.e. a longer lag time λ and a lower mineralization rate μ (Figure 9.1(b)). These interaction effects are classified as negative. Finally, there are cases where the co-culture mineralization performance falls between the monoculture performances, and are therefore classified as neutral.



(a) Positive interaction



(b) Negative interaction

Figure 9.1: Observed mineralization curves (mean and standard deviation) for three pairs with different interactions, plotted with their constituent observed monoculture mineralization curves.

Such comparisons of the kinetic mineralization parameters of the pairwise combinations make it clear that strong interaction effects are occurring. This raises the question of whether we can predict these co-culture effects, based on monoculture mineralization performances, in order to pinpoint strain combinations that are interesting for BAM mineralization bioaugmentation in our experimental system.

To construct such a predictive model, we adopted a regression approach. This method is not complex, but given the limited dataset and the exploratory nature of this work, we opted for a straightforward approach, deciding to resort to more involved predictive techniques only if unsatisfactory results were obtained with the regression approach. Additionally, this method is well established in the literature of predictive models in microbiological settings (Seshan et al., 2014; Baty and Delignette-Muller, 2004; Song et al., 2014; Baranyi and Roberts, 1995; Gil et al., 2006), increasing the accessibility of this work for microbiologists.

Therefore, we performed a non-linear regression for each mineralization parameter. Each model took as input variables the key features of the monocultures, and their outputs were the respective mineralization parameters of the co-culture. The key features of the monocultures were: the identity of each strain (a numerical identifier from 1 to 13, i.e. $S_i \in \{1, \ldots, 13\}$, with the order assigned according to the list in Section 9.2.1.1), and its three mineralization parameters λ, μ and A.

Hence the predictive model for the lag time $\hat{\lambda}$ of the co-culture of S_1 and S_2 with MSH1 took the form:

$$
\begin{aligned}
&M_\lambda\left(S_1, A_1, \mu_1, \lambda_1, S_2, A_2, \mu_2, \lambda_2\right) \\
&= \beta_0 + \beta_1 S_1 + \beta_2 A_1 + \beta_3 \mu_1 + \beta_4 \lambda_1 \\
&\quad + \beta_5 S_2 + \beta_6 A_2 + \beta_7 \mu_2 + \beta_8 \lambda_2 \\
&\quad + \beta_9 \lambda_1 \lambda_2 \\
&= \hat{\lambda}
\end{aligned}
\tag{9.2}
$$

Likewise, the corresponding predictive models for $\hat{\mu}$ and $\hat{A}$ (the other co-culture mineralization parameters of S_1 and S_2 with MSH1) included instead an interaction term for the parameters μ and A: $\beta_9 \mu_1 \mu_2$ and $\beta_9 A_1 A_2$ respectively.

To determine the model coefficients $\beta_0, \ldots, \beta_9$, we performed a weighted least squares regression, which searches for the parameter values that minimize a weighted sum of squared residuals (where each weight is equal to the reciprocal of the variance) (Brown, 1978). This was done using the NonlinearModelFit function in Mathematica (version 11.0.1, Wolfram Research Inc., USA).

The models were validated using an exhaustive cross-validation (Friedman et al., 2008), where the models are trained and tested on every possible division of the original sample into a training and a test set of a certain size. Specifically, we used leave-one-out cross-validation (Friedman et al., 2008). In this procedure, each of the n observations is dropped in turn to form n training sets of $n-1$ samples. For each of the training sets, the model is fitted and then tested on the corresponding test set (the dropped observation). The performance of the models was evaluated using the cross-validated R^2 statistic, given by

$$
R^2_{\mathrm{CV}} = 1 - \frac{\sum\left(Y_{\mathrm{train}_i} - \hat{Y}_{\mathrm{train}_i}\right)^2}{\sum\left(Y_{\mathrm{train}_i} - \overline{Y}_{\mathrm{train}_i}\right)^2}
$$

where Y_{train_i} and $\hat{Y}_{\mathrm{train}_i}$ are, respectively, the observed and predicted values from the i-th training set obtained using leave-one-out cross-validation (Roy et al., 2015). We also computed the root-mean-square error (RMSE) of the models' predictions.

9.3.2 Results and discussion

We constructed a predictive model for each mineralization parameter as described in Section 9.3. For completeness, we report in Table 1 the best fit parameters of the regression models. The cross-validated R^2 values for the regression models were: 0.68 for the λ model, 0.55 for the μ model and 0.45 for the A model. Thus we found a range of predictive performances across the three parameters.

Table 9.1: Best fit regression parameters, to two significant figures, for λ, μ and A regression models.

Parameter	λ	μ	A
β_0	4.84	2.05	63.60
β_1	-0.011	0.0020	0.064
β_2	0.039	-0.029	0.027
β_3	-0.20	0.85	-2.31
β_4	1.42	-0.080	0.64
β_5	-0.0045	0.025	-0.10
β_6	-0.090	0.016	-0.77
β_7	0.26	-0.12	1.60
β_8	-0.20	0.076	-0.48
β_9	-0.035	-0.047	0.010

We also tested our predictive models' performance against the raw mineralization data, as well as the fitted parameters (since this may be a source of error, as will be discussed shortly). Thus for each pair, we used the model to predict the mineralization values for the 13 time points used in the lab experiment. By using these values, we reconstructed the cumulative mineralization curve, which was subsequently compared to the mean cumulative mineralization curve obtained from the experimental observations (Figure 9.3).

To further test our predictive models' performance, for each strain pair we used their predicted pairwise growth parameters to parameterize the modified Gompertz model (see Section 9.2.2). We then used this Gompertz model to compute the mineralization values for the 13 time points used in the experiment. With this time series, we obtained a predicted cumulative mineralization curve that could be compared with the observed cumulative mineralization curve.

The goodness-of-fit of the 78 predicted mineralization curves (one for each pair) was verified using the root-mean-square error and the R^2 statistic, given by $R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$ where SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares (Roy et al., 2015). The distribution of the R^2 values is shown in Figure 9.2. The mean R^2 value obtained for the predicted mineralization curves of the 78 pairs was 0.85 ± 0.2, while the average root-mean-square error of the mineralization percentage was 7.04 % ± 6.85 %.

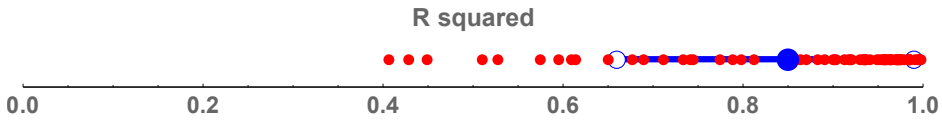


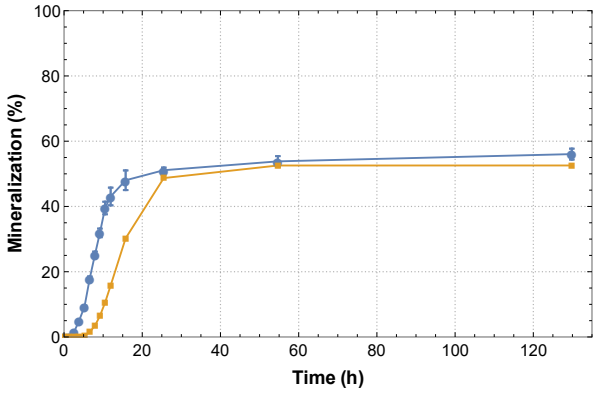
Figure 9.2: R^2_{CV} values for regression model predictions of mineralization curves, compared to experimental observations. Mean and standard deviation shown in blue.

In Figure 9.3, we show some examples of predicted mineralization curves compared to the observed mineralization curves. The first example, predicting the mineralization curve of strains S22 and K82 in co-culture with MSH1, shows a very good agreement with the corresponding observed mineralization curve, achieving a goodness-of-fit value of 0.98. Thus in this case the models led to a very good prediction of all three mineralization parameters, evidenced by the key features of the mineralization curve: the lag time λ, the mineralization rate μ and the mineralization extent A.

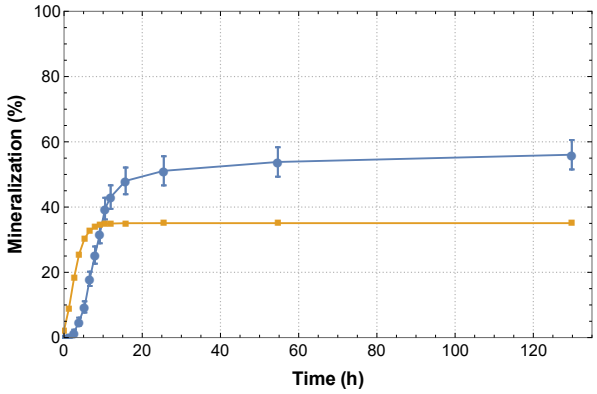
An example of a poorer goodness-of-fit is shown in Figure 9.3b, where the predicted mineralization curve of strains S158 and K169 in co-culture with MSH1 differs fairly significantly from the observed mineralization curve. However, as can be seen in Figure 9.2, most of the 78 pairs were predicted with a very good agreement to observations.

It is clear from Figure 9.2 that despite a relatively poor R^2_{CV} value for the predictive model for the mineralization extent A, reconstruction of the cumulative mineralization curves can still be achieved with a very good experimental agreement. This reinforces the relative importance of the mineralization parameters in the co-culture dynamics, as discussed in Section 9.3. It is also a positive point for this modelling approach that its best performances are for the most important parameters: if its best predictive performances were for parameter A, the model would be less useful for the bioaugmentation purpose for which it was conceived.

There are several possible explanations for the lower accuracy in the case of some pairs. Firstly, the natural variability between the experimental replicates, which led to a small number of outliers being removed from the dataset before the modelling stage. Although the removal of outliers led to a more consistent dataset, some natural variability between replicates undoubtedly remains, which leads to less accurate parameter estimates in the initial fitting procedure. Although regressing on the fitted parameters rather than the raw mineralization data therefore brings an additional source of error, we used this approach since it resulted in a reduced dataset that was easily comparable and biologically meaningful at a glance (in contrast to the mineralization time series).



(a) $R^2 = 0.985$



(b) $R^2 = 0.528$

Figure 9.3: Predicted vs observed mineralization curves for two pairs, showing examples of (a) excellent and (b) relatively poor fit.

Secondly, it is almost certainly true that some interactions are taking place in the co-cultures that are not captured by the models. Based on the specific strains involved, these interactions could include cooperation in the form of degradation of BAM metabolites (Little et al., 2008), interference competition between strains (Moons et al., 2009), and competition for additional resources such as metabolites or the scavenging of lysed cells (Kerr et al., 1999) (Vandermaesen *et al.*, 2017).

Despite these possible sources of error, the model still achieved a very good predictive performance even using the fitted parameters and without resorting to a detailed metabolic model, which would have involved a more intensive experimental and modelling characterization (Seshan et al., 2014; Poschet et al., 2005).

The same modelling approach could also be used for co-cultures involving higher numbers of species. This would require a larger but not necessarily exhaustive dataset. Although the possible number of co-culture combinations increases dramatically with the number of strains present (from our 13 strains, there are 286 possible co-cultures of three species, and 715 of four species), not all of these combinations would need to be tested. Instead, a combinatorial experimental design (Shasha et al., 2001) could be used to investigate certain subsets of combinations (based on interesting results at lower richness), and a predictive model could then be constructed on this basis. We believe that such an approach, if conducted using a careful selection of combinations, can still result in a good predictive performance without being experimentally impractical.

9.4 Emergent competition

9.4.1 Motivation and scope

Previous results with synthetic microbial communities with similar characteristics in terms of diversity and composition (Horemans *et al.*, 2017, in prep,) and also for this particular synthetic community of MSH1 and 13 SFI (Vandermaesen *et al.*, 2017, in prep), show that after the initial competition phase only some of the strains persist, forming a stable subcommunity of reduced richness. It is this subcommunity that is present at the moment of the BAM spike and during the subsequent mineralization period that determines the bioaugmentation success. In Section 9.3 we have demonstrated the possibilities of predictive modelling for highlighting SFI co-culture combinations that improve the BAM mineralization performance of MSH1, and thus support bioaugmention success. These results refer to the second phase, where the subcommunity and MSH1 interact as MSH1 mineralizes BAM.

Now we turn our attention to the first phase, before the addition of BAM, where all 13 SFI are inoculated in the co-culture, and competition between them results in extinctions that lead to a persisting subcommunity that would then be spiked with BAM. Can we retrieve this behaviour using the IBM framework developed throughout the previous chapters, supported with data and knowledge from the *in vitro* experiments?

For this purpose, we require information about the interactions between the SFI. Here it is important to note that the experiments described above were not purposely conceived and designed for use in a modelling study. Hence we do not have all information we would wish for in order to parameterize our model. However, we can still make use of the information that we do have.

9.4.2 Assessing strain interactions

Ideally, we would like to dispose of as much information as possible about the 13 SFI and, most importantly, the interactions between them. However, the experiments of Vandermaesen *et al.* (2017, in prep) focused on bioaugmentation success (the second experimental phase) and therefore collected data relating to BAM mineralization and MSH1 survival. The only data we have that directly relates to the SFI themselves are their monoculture growth curves (Figure 9.4), and their monoculture survival curves on acetate (Figure 9.5), relevant to the first experimental phase that we now seek to model. These data can give us an idea of how the SFI grow and persist in isolation, and on this basis Vandermaesen *et al.* (2017, in prep) classified the strains according to their "intrinsic competitiveness", a classification that we can use as an additional feature of the strains. However these data do not give us any information about how the SFI may interact, and in particular compete, when they are inoculated together in co-culture.

Thus if we were to attempt to use these data in a model based for example on exploitative competition between the SFI, we would somehow have to infer the interaction effects. One possibility is to look at the respective rankings of the different SFI in terms of their growth rates by the end of the initial growth period of seven days. However, this would result in straightforward hierarchical competition, where the SFI with the highest growth rate or survival rate would persist to the exclusion of all other SFI. This is not realistic given the *in vitro* experimental results which demonstrate the persistence of a coexisting sub-community of SFI.

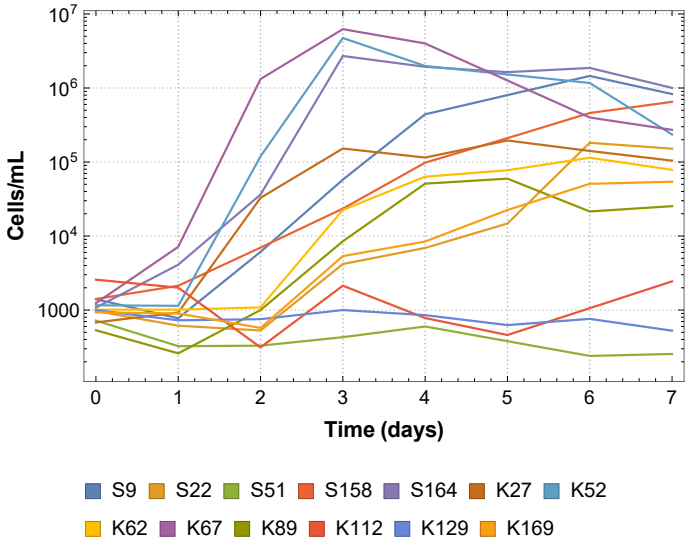


Figure 9.4: SFI monoculture growth curves on acetate (Vandermaesen *et al.*, 2017, in prep).

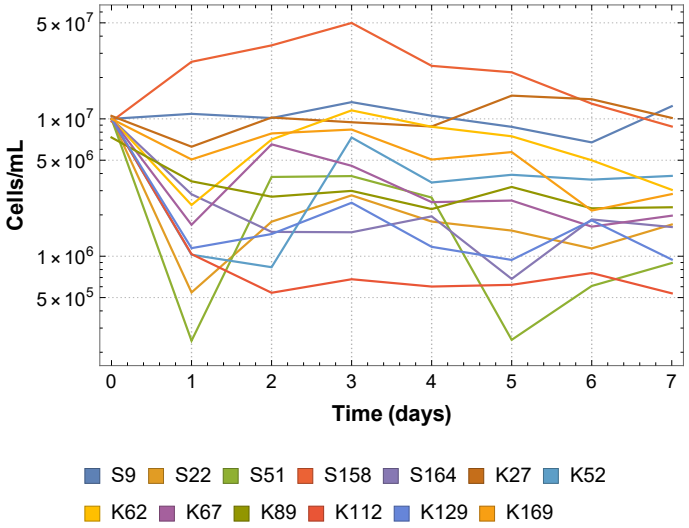


Figure 9.5: SFI monoculture survival curves on acetate (Vandermaesen *et al.*, 2017, in prep).

Another possibility is to use the classification of Vandermaesen *et al.* (2017, in prep) based on the intrinsic competitiveness of the SFI, which grouped the SFI in strong, intermediate or weak competitors. However, again without further information about how these classifications might change when the SFI interact with each other, we would merely find that the "strong" competitors persist to the exclusion of the other groups. Furthermore, we would not be able to infer any intra-group interactions and would therefore not be able to distinguish between different SFI with the same classification.

Thus we need further information regarding the interactions between SFI, which we do in fact dispose of —however, this information is indirect, namely the effect on MSH1's mineralization performance in the presence of different combinations of strains, in the second experimental phase. From the differences in mineralization parameters between these different co-culture combinations, we can infer when there are interaction effects occurring between strains, by comparing the mineralization performances of MSH1 alone, in co-culture with individual SFI, and in co-culture with both strains.

For these interaction effects, we focus on two of the three mineralization parameters: the lag time λ and the maximum mineralization rate μ. As discussed in Section 9.3, these two parameters have been highlighted as key to the success of bioaugmentation strategies and are more strongly linked with both positive and negative mineralization effects than the mineralization extent A (Ekelund et al., 2015).

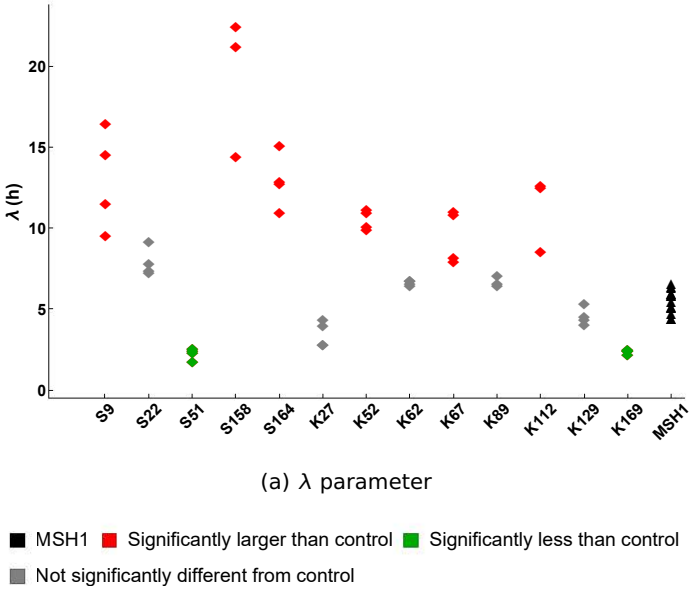
9.4.2.1 Identifying strain identity effects

To compare values for λ and μ across different R_{SFI} levels, the pairwise Tukey test (Tukey, 1949) was used. With this test it is possible to evaluate whether λ or μ values observed for a specific synthetic community were significantly different from the respective parameter values observed for a different community. The test statistic is $z = \frac{m_A - m_B}{SE}$, where m_A and m_B are the respective means of the observations of two populations being compared, and SE is the data's standard error (Haynes, 2013). The null hypothesis of the test is that the means are from the same population. The test statistic is then compared to a critical test statistic value z_{crit} which is obtained from the studentized range distribution (Keuls, 1952). If z is larger than z_{crit}, then the null hypothesis is rejected and it is concluded that the two populations are significantly different. Tests were performed at the 95% significance level, using Mathematica (version 11.0, Wolfram Research, Champaign, IL, USA).

Two types of tests were conducted. First, we compared λ or μ values for $R_{\text{SFI}} = 1$ communities against $R_{\text{SFI}} = 0$ (i.e. MSH1 alone) as a benchmark population. We have 16 replicates for this control population. These results are shown in Figures 9.6 and 9.7, where the λ and μ values, respectively, for each $R_{\text{SFI}} = 1$ community are plotted, and the points are coloured according to the Tukey test results. Recall that for λ, a decrease in this parameter is considered a positive effect while an increase is considered a negative effect. For μ, the opposite is true.

The second type of test required selecting one of the SFI as the focal strain. The test then compared λ or μ values for the $R_{\text{SFI}} = 2$ communities including this focal strain, against the λ or μ values for the corresponding $R_{\text{SFI}} = 1$ community for the non-focal strain. An example is shown in Figure 9.8, where S9 is the focal strain of the test and the parameter under consideration is λ. We therefore selected all $R_{\text{SFI}} = 2$ communities containing S9. One such community contained S9, S22 and MSH1. We then compared the λ values of this community against the λ values of the community containing S22 and MSH1. This allowed us to conclude if in this case there were significant differences in lag time due to the inclusion of S9. This was repeated for every strain other than the focal strain. The results of the equivalent test for μ with S9 as the focal strain are shown in Figure 9.9.

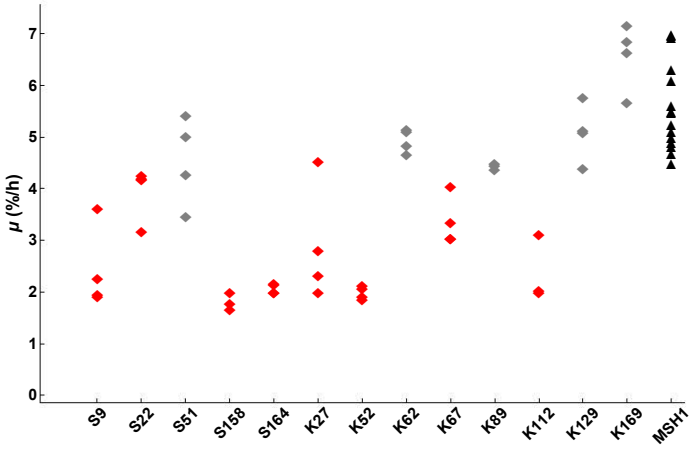
This test was done 13 times for each parameter, so that each of the strains was used once as the focal strain. The results of these tests are collected in the tables shown in Tables 9.10 and 9.11. In these tables, each row collects the results of Tukey tests with a particular focal strain, e.g. the first row shows the results of tests where S9 was the focal strain, and the columns indicate the other strains being tested for interaction effects with S9.



(a) λ parameter

MSH1 Significantly larger than control Significantly less than control
Not significantly different from control

Figure 9.6: Tukey test results when comparing λ parameters for $R_{SFI} = 1$ co-cultures against $R_{SFI} = 0$ (i.e. MSH1 alone).



(a) μ parameter

MSH1 Significantly less than control Significantly larger than control
Not significantly different from control

Figure 9.7: Tukey test comparing μ parameters for $R_{SFI} = 1$ co-cultures against $R_{SFI} = 0$ (i.e. MSH1 alone).

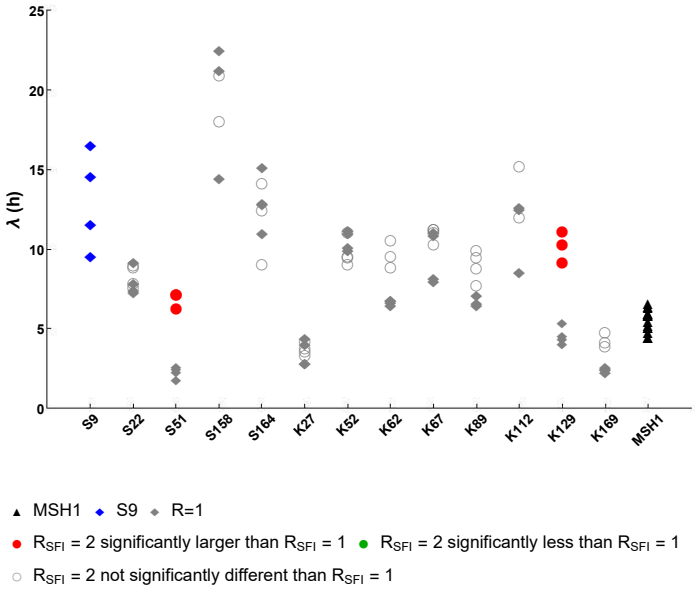


Figure 9.8: Tukey test with S9 as the focal strain, comparing λ parameters of $R_{SFI} = 2$ combinations including S9 against their corresponding $R_{SFI} = 1$ co-culture not including S9.

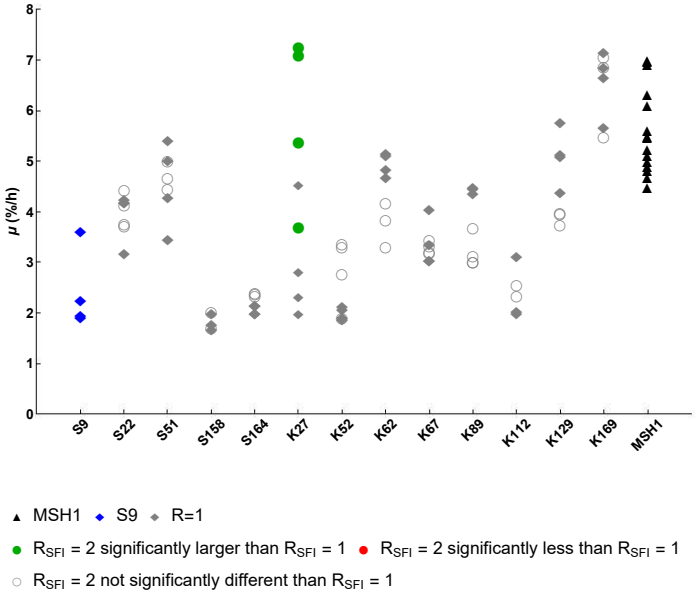


Figure 9.9: Tukey test with S9 as the focal strain, comparing μ parameters of $R_{SFI} = 2$ combinations including S9 against their corresponding $R_{SFI} = 1$ co-culture not including S9.

	S9	S22	S51	S158	S164	K27	K52	K62	K67	K89	K112	K129	K169
S9		+	0	+	+	0	+	+	+	+	+	+	–
S22	+		–	+	+	0	+	0	+	0	+	+	–
S51	0	–		0	0	–	0	–	0	–	0	–	–
S158	+	+	0		+	0	+	+	+	+	+	+	–
S164	+	+	0	+		0	+	+	+	+	+	+	–
K27	0	0	–	0	0		0	–	0	–	0	–	–
K52	+	+	0	+	+	0		+	+	+	+	+	–
K62	+	0	–	+	+	–	+		+	0	+	0	–
K67	+	+	0	+	+	0	+	+		+	+	+	–
K89	+	0	–	+	+	–	+	0	+		+	0	–
K112	+	+	0	+	+	0	+	+	+	+		+	0
K129	+	+	–	+	+	–	+	0	+	0	+		–
K169	–	–	–	–	–	–	–	–	–	–	0	–	

Figure 9.10: Table of Tukey test results for the λ parameter. Each row collects the results of Tukey tests with a particular strain as the focal strain, the columns then indicate the strains that were tested for interaction effects with the focal strain. The entry in cell (i, j) indicates the difference (if any) between the $R_{\text{SFI}} = 2$ community containing species i and species j, and the control $R_{\text{SFI}} = 0$ community: "+" indicates the $R_{\text{SFI}} = 2$ parameter values were significantly larger than the $R_{\text{SFI}} = 0$ values, "–" indicates they were significantly less, and "0" indicates no significant difference. The background colour of cell (i, j) indicates the difference (if any) between the $R_{\text{SFI}} = 2$ community containing species i and species j, and the $R_{\text{SFI}} = 1$ community containing species j: green indicates the $R_{\text{SFI}} = 2$ parameter values were significantly less than the $R_{\text{SFI}} = 0$ values, red indicates they were significantly larger, and no colour indicates no significant difference.

	S9	S22	S51	S158	S164	K27	K52	K62	K67	K89	K112	K129	K169
S9		–	0	–	–	0	–	–	–	–	–	–	0
S22	–		0	–	–	–	–	–	–	–	–	–	0
S51	0	0		–	–	0	–	–	–	–	–	0	0
S158	–	–	–		–	–	–	–	–	–	–	–	0
S164	–	–	–	–		–	–	–	–	–	–	–	–
K27	0	–	0	–	–		–	0	–	0	–	0	0
K52	–	–	–	–	–	–		–	–	–	–	–	0
K62	–	–	–	–	–	0	–		–	–	–	0	0
K67	–	–	–	–	–	–	–	–		–	–	–	0
K89	–	–	–	–	–	0	–	–	–		–	0	0
K112	–	–	–	–	–	–	–	–	–	–		–	0
K129	–	–	0	–	–	0	–	0	–	0	–		0
K169	0	0	0	0	–	0	0	0	0	0	0	0	

Figure 9.11: Table of Tukey test results for the μ parameter. Each row collects the results of Tukey tests with a particular strain as the focal strain, the columns then indicate the strains that were tested for interaction effects with the focal strain. The entry in cell (i, j) indicates the difference (if any) between the $R_{\text{SFI}} = 2$ community containing species i and species j, and the control $R_{\text{SFI}} = 0$ community: "+" indicates the $R_{\text{SFI}} = 2$ parameter values were significantly larger than the $R_{\text{SFI}} = 0$ values, "–" indicates they were significantly less, and "0" indicates no significant difference. The background colour of cell (i, j) indicates the difference (if any) between the $R_{\text{SFI}} = 2$ community containing species i and species j, and the $R_{\text{SFI}} = 1$ community containing species j: green indicates the $R_{\text{SFI}} = 2$ parameter values were significantly larger than the $R_{\text{SFI}} = 0$ values, red indicates they were significantly less, and no colour indicates no significant difference.

9.4.2.2 Building the competition structures

Using the information gathered in Tables 9.9 and 9.10, we will now consider how to model the competition occurring between the SFI. It is clear from the interaction effects noticeable in these tables that using the same approach of Chapters 4–8 will not be sufficiently realistic. Whereas before we imposed a competition structure on the community as a whole, to say for example that A always beats B which always beats C, we must now model competition in a different way. We dispose of data relating to (indirect) pairwise interaction effects, so it is more suitable to "build up" the competition structure in this way.

Using the information in Tables 9.9 and 9.10 will result in a so-called tournament matrix. Such a matrix M for s species has dimensions $s \times s$. If the species represented by row i outcompetes the species represented by column j, then $M_{ij} = 1$. On the other hand, if the species represented by row i is outcompeted by the species

represented by column j, we have $M_{ij} = -1$. If $i = j$, then $M_{ij} = 0$. Using the information in Tables 9.9 and 9.10, we can compile such a tournament or competition matrix. The question remains how precisely to do so.

We have two possibilities: to merge the information about λ and μ interaction effects, or to treat the parameters separately. The latter option is justified by considering that the parameters represent different biological attributes and different underlying processes (Ekelund et al., 2015). This is most noticeable in their opposing effects on mineralization performance in particular; an increased λ parameter is considered a negative effect while an increased μ parameter is considered a positive effect.

We will however consider both cases, which results in three different competition matrices. The first is based on λ and μ interaction effects, the second on λ interaction effects only, and the third on μ interaction effects only. We look in Table 9.9 (λ interaction effects) or Table 9.10 (μ interaction effects), or both (for λ and μ interaction effects) for pairs of SFI which appear to interact with each other, and check what kind of interaction appears to be taking place: is it positive or negative with respect to each of the SFI?

This corresponds in Tables 9.9 and 9.10 to both the cell entries and the cell background colours. The cell entries indicate which kind of difference (if any) exists between the control community and the $R_{\text{SFI}} = 2$ community containing the particular species corresponding to the cell row and column. These relationships can be positive, negative, or not significant. The cell background colours indicates the difference (if any) between the $R_{\text{SFI}} = 1$ community containing the species corresponding to the cell column, and the $R_{\text{SFI}} = 2$ community containing the particular species corresponding to the cell row and column. These relationships can also be positive, negative, or not significant. Recall that positive or negative relationships are defined in the biological sense: a positive relationship increases μ or decreases λ, while a negative relationship decreases μ or increases λ.

This approach results in matrices that are noticeably sparser than the competition structures used in previous chapters, but this attribute is also more realistic, since the SFI do not necessarily interact and compete with every other strain they meet. In fact, it is clear from Tables 9.9 and 9.10 that some strains do not interact in any significant way, and this must be reflected in the competition structures.

We then obtain the following matrices representing competition between the SFI. When considering interactions based on λ or μ effects, the matrix has the form:

$$M_{\lambda\mu} = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ -1 & -1 & 0 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & -1 & 1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ -1 & -1 & 0 & -1 & -1 & 0 & -1 & -1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{9.3}$$

When considering interactions based only on λ effects, the matrix reads:

$$M_{\lambda} = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & -1 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{9.4}$$

When considering interactions based only on μ effects, the matrix has the form:

$$M_\mu = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{9.5}$$

An additional extension of our modelling approach that will bring it closer to reality is to also consider non-deterministic competition. Thus far we have considered deterministic competition, where if the competition structure specifies that A beats B, this will always occur: it will never be possible for B to beat A. But this is not always realistic (Planque et al., 2014). Variation between individuals can result in an individual of species A that is a particularly weak competitor, and an individual of species B that is a particularly strong competitor. If these two specific individuals meet, the outcome of the competition can be in doubt. It may be more realistic to specify a so-called winning probability: the probability that A beats B. Including a winning probability allows for different competition outcomes to occur, and the value of the winning probability allows us to account for the relative strengths of the individuals.

Therefore we will also consider non-deterministic competition between the SFI, not only in terms of its effects on the diversity and stability of the community (and possible subcommunity), but in comparison with the same effects due to deterministic competition. Our immediate question is then how to assign the winning probabilities to the different pairwise competitions.

Using data relating to the SFI's monoculture growth and survival curves, Vandermaesen *et al.* (2017, in prep) classified the "intrinsic competitiveness" of the SFI and on this basis grouped them into strong, intermediate and weak competitors (2017, in prep). Using this information, we can assign winning probabilities to each pairwise competition based on the differences in intrinsic competitiveness between the two strains. For example, competition between a weak intrinsic competitor and a strong intrinsic competitor will most likely result in the success of the latter. It should also be clear that this winning probability should be higher than the

winning probability assigned to an intermediate intrinsic competitor when faced with a weak intrinsic competitor. Using this approach, we replace the 1's and -1's populating our matrices $M_{\lambda\mu}$, M_λ and M_μ with rational numbers of absolute value less than 1, representing the appropriate winning probability.

Using this approach, we obtain the following matrices representing non-deterministic competition. When considering interactions based on λ or μ effects, the matrix has the form:

$$M^*_{\lambda\mu} = \begin{pmatrix} 0 & 0 & .9 & .9 & 0 & .9 & 0 & 0 & 0 & 0 & 0 & .9 & 0 \\ 0 & 0 & .7 & -.6 & .7 & 0 & 0 & 0 & 0 & 0 & .7 & .7 & 0 \\ -.9 & -.7 & 0 & 0 & -.9 & 0 & -.9 & 0 & -.7 & 0 & 0 & 0 & 0 \\ -.9 & .6 & 0 & 0 & .9 & -.9 & .9 & .9 & .9 & .9 & .9 & .9 & 0 \\ 0 & -.7 & .9 & -.9 & 0 & 0 & 0 & -.7 & 0 & 0 & 0 & .9 & -.7 \\ -.9 & 0 & 0 & .9 & 0 & 0 & .9 & 0 & 0 & -.6 & 0 & 0 & .9 \\ 0 & 0 & .9 & -.9 & 0 & -.9 & 0 & .9 & 0 & 0 & 0 & .9 & 0 \\ 0 & 0 & 0 & -.9 & .7 & 0 & -.9 & 0 & 0 & 0 & 0 & .7 & 0 \\ 0 & 0 & .7 & -.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7 & -.7 \\ 0 & 0 & 0 & -.9 & 0 & .6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -.7 & 0 & -.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -.6 & 0 \\ -.9 & -.7 & 0 & -.9 & -.9 & 0 & -.9 & -.7 & -.7 & 0 & .6 & 0 & 0 \\ 0 & 0 & 0 & 0 & .7 & -.9 & 0 & 0 & .7 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{9.6}$$

When considering interactions based only on λ effects, the matrix has the form:

$$M_{\lambda}^{*} = \begin{pmatrix} 0 & 0 & .9 & .9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .9 & 0 \\ 0 & 0 & 0 & -.9 & .7 & 0 & 0 & 0 & 0 & 0 & .7 & .7 & 0 \\ -.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -.9 & .9 & 0 & 0 & .9 & 0 & 0 & .9 & 0 & 0 & .9 & .9 & 0 \\ 0 & -.7 & 0 & -.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .9 & 0 \\ 0 & 0 & 0 & -.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -.7 & 0 & -.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -.6 & 0 \\ -.9 & -.7 & 0 & -.9 & -.9 & 0 & -.9 & 0 & 0 & 0 & .6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{9.7}$$

When considering interactions based only on μ effects, the matrix has the form:

$$M_{\mu}^{*} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -.9 & .7 & 0 & 0 & 0 & 0 & 0 & .7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -.9 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .9 & 0 & 0 & 0 & 0 & 0 & .9 & 0 & .9 & 0 & .9 & 0 \\ 0 & -.7 & 0 & 0 & 0 & 0 & 0 & -.7 & 0 & -.6 & 0 & .9 & -.7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .9 & 0 & 0 & 0 & 0 & .9 & 0 & .9 & 0 & .9 & 0 \\ 0 & 0 & 0 & -.9 & .7 & 0 & -.9 & 0 & .7 & 0 & .7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -.7 & 0 & 0 & 0 & 0 & -.7 \\ 0 & 0 & 0 & -.9 & .6 & 0 & -.9 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -.7 & 0 & 0 & 0 & 0 & 0 & -.7 & 0 & 0 & 0 & -.6 & 0 \\ 0 & 0 & 0 & -.9 & -.9 & 0 & -.9 & 0 & 0 & 0 & .6 & 0 & 0 \\ 0 & 0 & 0 & 0 & .7 & 0 & 0 & 0 & .7 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{9.8}$$

9.4.3 Model description

To understand how the different competition structures affect the dynamics of the system, we consider the *in silico* counterpart of the synthetic community of 13 SFI. We model this community using an individual-based approach similar to previous chapters, which we again describe using the ODD protocol.

9.4.3.1 Overview

Purpose

The aim of the model is to study how more realistic competition structures affect the *in silico* dynamics, particularly in terms of community diversity and stability, and investigate whether this approach can qualitatively reproduce the dynamics observed in similar *in vitro* studies, namely a persisting subcommunity.

State variables and scales

The model is a two-dimensional representation of an experimental domain divided into a regular grid of size $L \times L = N$, and populated by a community of 13 SFI. We use the same labels as in Section 9.3. Namely, we assign to each strain a numerical label between one and 13, in the order given in Section 9.2.1.1: S9, S22, S51, S158, S164, K27, K52, K62, K67, K89, K112, K129, K169.

Each grid site is either occupied by a single individual, or is empty. Individuals are characterized by two state variables: grid position (x, y) and species identity $s \in \{1, \ldots, 13\}$.

Process overview

We consider an *in silico* microbial community that is initially placed on the grid with a random spatial distribution. The community's initial species abundance distribution is completely even, to mimic the *in vitro* experimental set-up.

An individual can interact with its nearest neighbours, defined as those individuals in its von Neumann neighbourhood (the four grid cells with which it shares an edge). Three possible interactions can occur, representing the three key demographic processes discussed in Section 3.4: reproduction, competition and mobility.

The mechanisms of these interactions are the same as in Chapters 5–8. Reproduction can occur when an individual is located adjacent to an empty grid site, which is then filled with a new individual of the same species. In order to provide a form of mobility, all individuals can exchange their position with a nearest neighbour or move to a neighbouring empty site. Competition can occur between two neighbouring individuals that do not represent the same species. The outcome of the competition event is determined by the governing competition matrix; the defeated individual is removed from the grid and the grid site becomes empty.

Scheduling

The IBM proceeds in a similar way as the three- and four-species models in Chapters 6–8, using the modified version of the Gillespie algorithm described in Section 5.3.1.4, to determine which interaction occurs at each time step and calculate the interaction outcome. The algorithm iterates over the following steps:

1. Set time to $t = 0$ and set the event rates:
   (a) reproduction with rate μ
   (b) competition with rate σ
   (c) mobility with rate ϵ
2. Calculate the overall rate of events $r = \mu + \sigma + \epsilon$
3. Select an individual at random
4. Select one of the focal individual's nearest neighbours at random
5. Select an interaction event with the following probabilities, by drawing a random number from the interval $[0, r]$:
   (a) reproduction with a probability $\frac{\mu}{r}$
   (b) competition with a probability $\frac{\sigma}{r}$
   (c) mobility with a probability $\frac{\epsilon}{r}$
6. Execute the selected interaction event on the selected individual (if permitted) and determine the outcome according to the governing rules:
   (a) reproduction occurs deterministically (it is always carried out if possible)
   (b) mobility occurs deterministically
   (c) competition can occur:
      i. deterministically: the winner is determined by the appropriate entry (being 1 or -1) in the competition matrix $M_{\lambda\mu}, M_{\lambda}$ or M_{μ}
      ii. non-deterministically: a random number r_c is drawn from the unit interval and compared to the appropriate winning probability M_{ij} in the competition matrix $M^*_{\lambda\mu}, M^*_{\lambda}$ or M^*_{μ}, where species i and species j are competing.
      If $M_{ij} > 0$:
      - species i wins the competitive event if $r_c < M_{ij}$
      - species j wins the competitive event $r_c > M_{ij}$

      If $M_{ij} < 0$:
      - species i wins the competitive event if $r_c > |M_{ij}|$

- species j wins the competitive event $r_c < |M_{ij}|$

7. Update the grid according to the outcome of step 6
8. Update the time step to $t = t + 1$
9. Return to step 3 and continue until $t = t_{\text{end}}$

9.4.3.2 Design concepts

- **Emergence:** the spatial patterns and population-level dynamics of the community emerge naturally from the interactions occurring between individuals.
- **Competition based on pairwise interaction effects:** we no longer impose a competition scheme on the community as a whole, as in Chapters 5–8. Instead, the competition scheme is constructed based on pairwise interaction effects, encoded in a competition matrix.
- **Non-deterministic competition:** In addition to deterministic competition, we also investigate the effects of non-deterministic competition, where the victor of any competition event is not predetermined but is instead probability-based.
- **Interactions:** individuals interact with each other and their environment by reproducing if located next to an empty site, exchanging sites with their neighbours, or competing with their neighbours.
- **Stochasticity:** the stochasticity in the model arises from the initial spatial distribution of the grid; the interactions between individuals and the environment (reproduction); the interactions between individuals (mobility, competition); and from the non-deterministic competition.
- **Sensing:** if selected for reproduction, individuals can sense whether their neighbouring site is empty; if so, they will reproduce. If the site is occupied by an individual, no reproduction will occur.
- **Observation:** the data collected from the IBM includes the population count of each species, the community evenness, the spatial distribution of individuals, and their time to extinction. These are tracked for each time step.

9.4.3.3 Details

Initialization

The model is initialized with a random spatial distribution of individuals and empty sites. Initially, a certain proportion of grid sites are left empty; thus the system

is initially below carrying capacity. The initial species abundance distribution is completely even ($E_0 = 1$). Aside from the input variables, all other parameters used to initialize the model are fixed for all simulations, and are shown in Table 9.2.

9.4.3.4 Input

The input variable is the competition matrix. There are six different matrices:

- $M_{\lambda\mu}$: deterministic competition based on λ and μ interaction effects (matrix (9.3))
- M_{λ}: deterministic competition based on λ interaction effects (matrix (9.4))
- M_{μ}: deterministic competition based on μ interaction effects (matrix (9.5))
- $M^*_{\lambda\mu}$: non-deterministic competition based on λ and μ interaction effects (matrix (9.6))
- M^*_{λ}: non-deterministic competition based on λ interaction effects (matrix (9.7))
- M^*_{μ}: non-deterministic competition based on μ interaction effects (matrix (9.8))

For each of these initial conditions, we run 200 replicates.

9.4.4 Results and discussion

9.4.4.1 Richness

To study the effects of the different types of competition on the diversity of the *in silico* synthetic community, we first examine the richness effects, by determining the number of surviving species after 1000 generations to see what levels of richness are maintained under the different competition structures.

Table 9.2: Parameters of the individual-based model of 13 SFI.

Parameter	Description	Value
L	Grid side length	200
$\varnothing$	Initial proportion empty sites	0.1
μ	Reproduction rate	1
σ	Competition rate	1
ϵ	Mobility rate	4.25
T	Number of generations evolved	1000

In Figure 9.12, we show the probability of observing a certain species richness after 1000 generations, for deterministic and non-deterministic competition based on combined λ and μ interaction effects. The majority of the simulations result in monoculture. Indeed, more than 80% for deterministic competition and more than 75% for non-deterministic competition. In this sense, non-deterministic competition has a minor stabilizing effect on the community dynamics, relative to deterministic competition, by slightly reducing the probability of monoculture (and thus maximal extinctions). For both types of competition, a small percentage of the simulations results in communities of more than one species. This happens more frequently for non-deterministic competition ($\sim$ 25%) than for deterministic competition ($\sim$ 15%), however in neither case do more than four species persist. Thus competition based on λ and μ interaction effects results in very low richness levels in both the deterministic and non-deterministic cases.

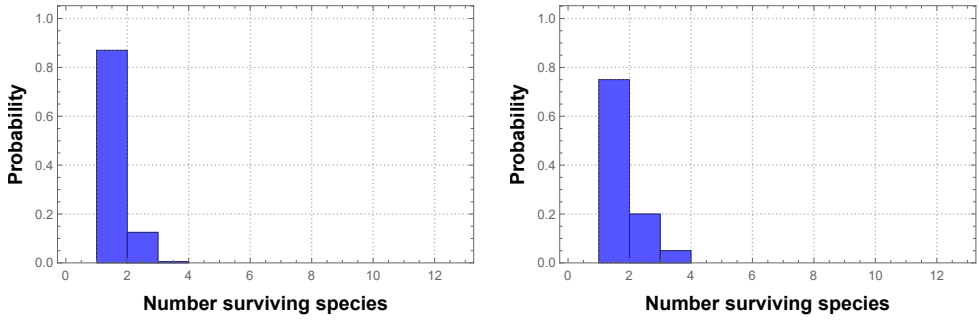


Figure 9.12: Probability of observing a particular species richness after 1000 generations for (a) deterministic and (b) non-deterministic emergent competition based on λ and μ effects. Probabilities calculated from 200 replicates.

In Figure 9.13, we show the corresponding results for deterministic and non-deterministic competition based on λ interaction effects only. We immediately notice higher richness levels. With this competition structure, we observe monocultures very rarely in the deterministic case, and never in the non-deterministic case. We find final richness levels as high as eight (deterministic case) or nine species (non-deterministic case). In the deterministic case, approximately 70% of simulations result in communities of five or six species, and the same for the non-deterministic case. The distribution of final richness is more skewed towards higher richness for the non-deterministic case, again indicating a stabilizing effect on the dynamics in terms of fewer extinctions and thus higher richness. This effect is not surprising, since non-deterministic competition results in fewer prey extinctions and more predator extinctions compared to deterministic competition, and thus decreasing extinction probabilities of the most vulnerable species.

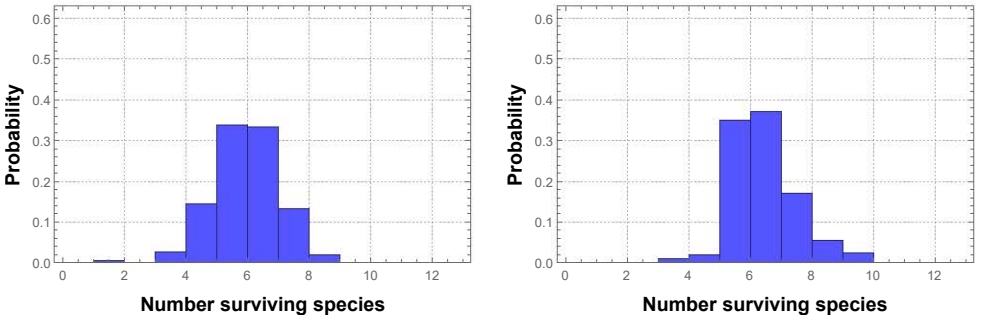


Figure 9.13: Probability of observing a particular species richness after 1000 generations for (a) deterministic and (b) non-deterministic emergent competition based on λ effects. Probabilities calculated from 200 replicates.

Even higher richness levels are observed for deterministic and non-deterministic competition based on μ interaction effects only (Figure 9.14). No monocultures are ever observed, and in fact richness never drops below four (deterministic case) or five species (non-deterministic case). In the deterministic case, approximately 95% of simulations result in communities of five or six species, in the non-deterministic case approximately 95% of simulations result in communities of five, six or seven species. The distribution of final richness is again more skewed towards higher richness for the non-deterministic case, indicating a stabilizing effect on the dynamics in terms of fewer extinctions and thus higher richness.

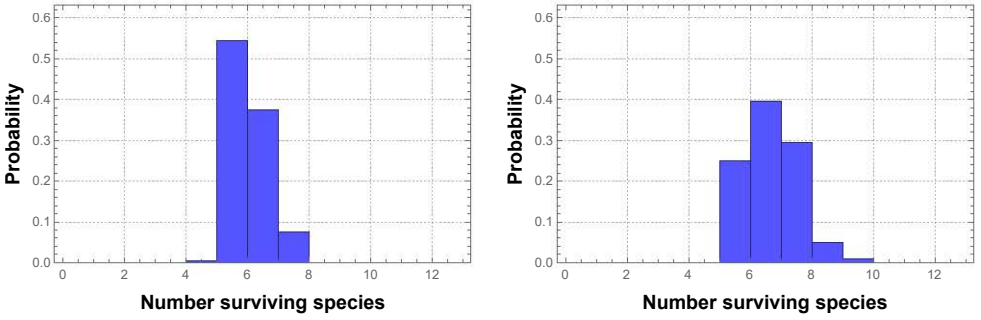


Figure 9.14: Probability of observing a particular species richness after 1000 generations for (a) deterministic and (b) non-deterministic emergent competition based on μ effects. Probabilities calculated from 200 replicates.

9.4.4.2 Diversity

After observing the richness effects due to the different forms of competition, we now consider community diversity. We do so using the Leinster-Cobbold diversity index (see Eq. (4.11) in Section 4.2.6.2). Recall that this is an effective number

index which is also multivariate, since it includes a sensitivity parameter q that determines how much weight is assigned to rare or common species. For $q < 1$, more weight is given to rare species ($q = 0$ is exactly species richness), while for $q > 1$ more weight is given to common species. All species are weighed equally for $q = 1$.

For each of the six competition matrices, we calculate the Leinster-Cobbold diversity index over time, for different values of q so that we may gather information about the composition and balance of the communities, as well as their changes in diversity as the different simulations evolve.

In Figure 9.15 we show the average Leinster-Cobbold diversity over time for deterministic and non-deterministic competition based on combined λ and μ interaction effects, for $q \in \{0, 1, 20\}$. With these different orders of diversity, we can infer changes in species richness ($q = 0$), evenness ($q = 20$) and diversity ($q = 1$). We firstly note that there is little variation between simulations, as evidenced by the low levels of variability in diversity values for all values of q.

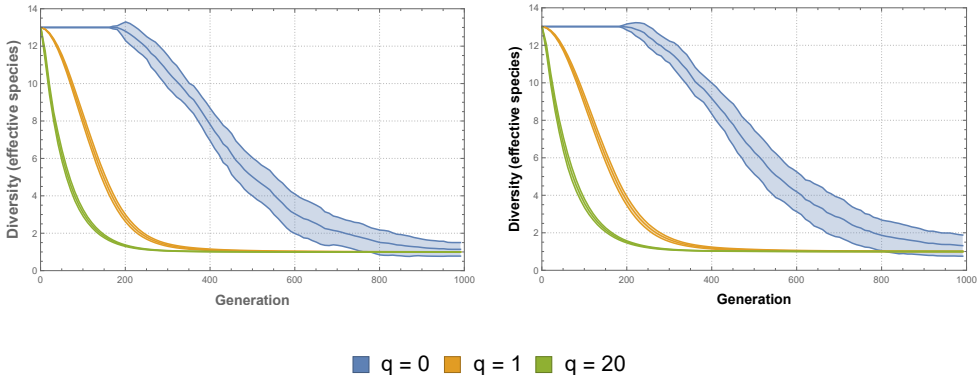


Figure 9.15: Mean diversity profiles (Leinster-Cobbold index) for $q \in \{0, 1, 20\}$, for (a) deterministic and (b) non-deterministic emergent competition based on λ and μ effects. Mean and standard deviation calculated from 200 replicates.

Initially, the community undergoes a sharp drop in evenness (seen in differences between the two curves for $q > 0$ relative to the $q = 0$ curve) while richness is maintained at its initial level. The time to the first species extinction is roughly similar for all replicates, namely around 200 generations. This period represents the time required for the spirals to begin to form, and the first species to be entirely surrounded by its predator(s) and killed off. Following the first extinction, others rapidly follow, also enabled by the spatial structures as the species have aggregated sufficiently to begin to chase each other around the grid. Extinctions occur frequently and continuously, resulting in a final diversity of approximately one effective species, as in Figure 9.12. The higher order diversities ($q = 1$ and

$q = 20$) reach their minimum much earlier than the diversity of order zero (i.e. richness), indicating that while more than one species can persist in late simulation time, this occurs in extremely uneven communities. This can also be inferred by the similarity between the $q = 1$ and $q = 20$ curves. Increasing q has the effect of giving more weight to common species and thus less to rare species, so when this has little impact on diversity values it implies that there are very few rare species and thus ignoring them has little effect (since they have already collapsed to extinction).

There is little difference between the cases of deterministic and non-deterministic competition aside from a slightly longer time to first extinction and a slightly higher final diversity for the non-deterministic case. Thus non-deterministic competition has a minor stabilizing effects on the community dynamics, but only in the sense of delaying the onset of monoculture.

In Figure 9.16 we show the equivalent results for deterministic and non-deterministic competition based on λ interaction effects only. As observed in Figure 9.13, richness is higher than for the cases of competition based on λ interaction effects. There is again an initial phase where initial richness is sustained before extinctions begin. This period is longer compared to Figure 9.15, and final community diversity is higher for all values of q. Additionally, the decrease in similarity between the $q = 1$ and $q = 20$ curves compared to Figure 9.15 implies that communities are more even when competition is based on λ interaction effects. In Figure 9.16 these curves approach each other significantly later, indicating that higher evenness is maintained for longer than in Figure 9.15. However, the higher order diversities are significantly less than the zero order diversity (richness), indicating that while multiple species continue to coexist, these communities are quite uneven, though not so uneven as the communities subject to competition based on both λ and μ interaction effects. Finally, we again observe a stabilizing effect when considering non-deterministic rather than deterministic competition, in terms of time to first extinction and final community diversity.

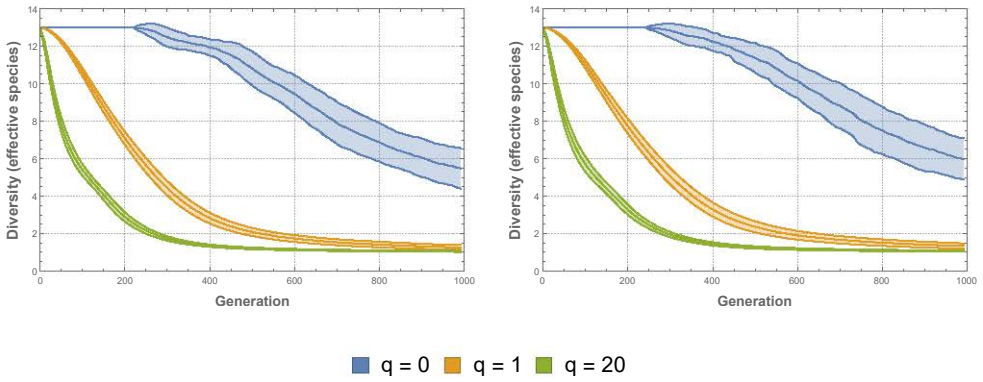


Figure 9.16: Mean diversity profiles (Leinster-Cobbold index) for $q \in \{0, 1, 20\}$, for (a) deterministic and (b) non-deterministic emergent competition based on λ effects. Mean and standard deviation calculated from 200 replicates.

In Figure 9.17, we compare the changes in diversity for communities subject to deterministic and non-deterministic competition based on μ interaction effects. Diversity is higher here than for the four previous competition matrices, for all values of q. Additionally, the communities are more even. Notably, in Figure 9.17 the $q = 1$ and $q = 20$ curves never overlap, indicating higher levels of evenness compared to the previous competition matrices which resulted in curves that converged (see Figures 9.15 and 9.16). This can also be inferred by the smaller distance between the $q = 0$ curve and the $q > 0$ curves in Figure 9.17, which indicates relatively more species coexisting in relatively more even communities. The minor stabilizing effect of non-deterministic competition compared to deterministic competition can also be observed in terms of diversity maintenance and time to first extinction.

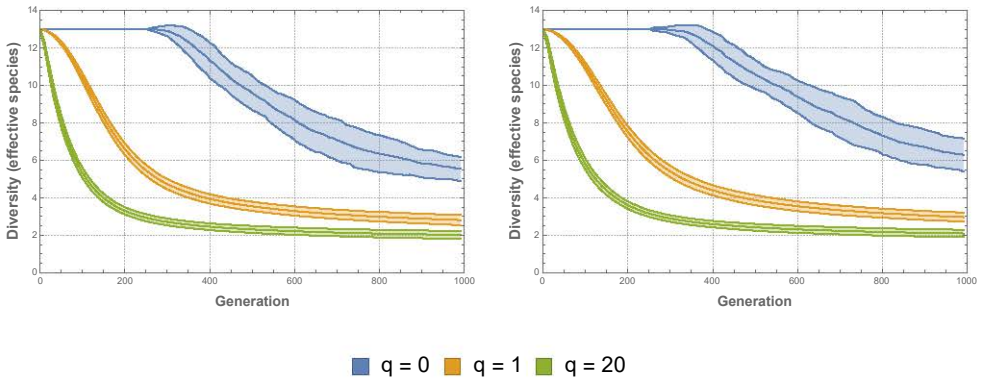


Figure 9.17: Mean diversity profiles (Leinster-Cobbold index) for $q \in \{0, 1, 20\}$, for (a) deterministic and (b) non-deterministic emergent competition based on μ effects. Mean and standard deviation calculated from 200 replicates.

9.4.4.3 Spatial structures

These diversity effects, and the spatial dynamics underlying them, can also be observed in Figure 9.18, where we show two representative examples of the grid configuration at $T = 1000$ generations for non-deterministic competition based on λ (Fig. 9.18(a)) or μ (Fig. 9.18(b)) interaction effects only. As was observed in Figures 9.16 and 9.17, competition based on the former results in more uneven communities than competition based on the latter. In the former case, sufficient species are present in sufficient numbers to form the spiral patterns characteristic of the models studied in Chapters 4–8, which were shown to help maintain coexistence. This occurs even though we no longer impose a competition structure on the community as a whole. These patterns also qualitatively resemble those observed in *in vitro* experiments where a similar synthetic community of SFI was co-cultured with MSH1 in the presence of BAM (Horemans *et al.*, 2017, in prep). The spiral formations also enable spatial refuges, which we previously observed in Chapter 8 as supporting species coexistence by allowing vulnerable species to persist at low but still significant levels. Such refuges can be observed for example in Figure 9.18(b) for multiple species.

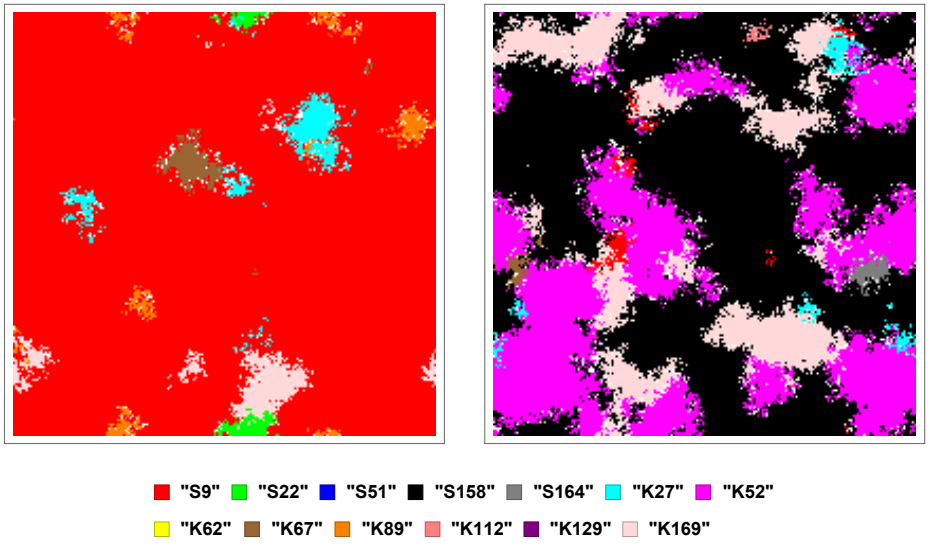


Figure 9.18: Examples of the grid configurations at $T = 1000$ generations with emergent non-deterministic competition based on (a) λ effects, and (b) μ effects.

The link between the spatial dynamics and diversity maintenance can be further explored by examining patchiness, which is the average number of conspecifics in individuals' nearest neighbourhoods. The evolution of patchiness over time is shown in Figure 9.19 for deterministic and non-deterministic competition based on

λ or μ interaction effects only. We use relative patchiness, which is the patchiness normalized to the unit interval through division by four (the maximum number of nearest neighbours in a von Neumann neighbourhood, and thus the maximum value that patchiness can take). We also plot the average time to first extinction for these four competition structures (see Figures 9.16 and 9.17). For all simulations, patchiness is initially approximately $0.28 = \frac{4}{14}$, which represents the initially random and well-mixed spatial distribution of the SFI around the grid. An individual's four nearest neighbours have the same probability of being occupied by a conspecific, namely $\frac{1}{14}$ (accounting for the fact they might also be empty). Patchiness then increases sharply during the initial phase, as the spirals begin to form. By the time the first extinction occurs, patchiness is typically around 0.4, representing the value at which the spirals coalesce and the weakest species begin to die off. As extinctions begin to occur, the variability between the simulations increases due to the increased stochasticity, since the order of the species extinctions can change between different replicates.

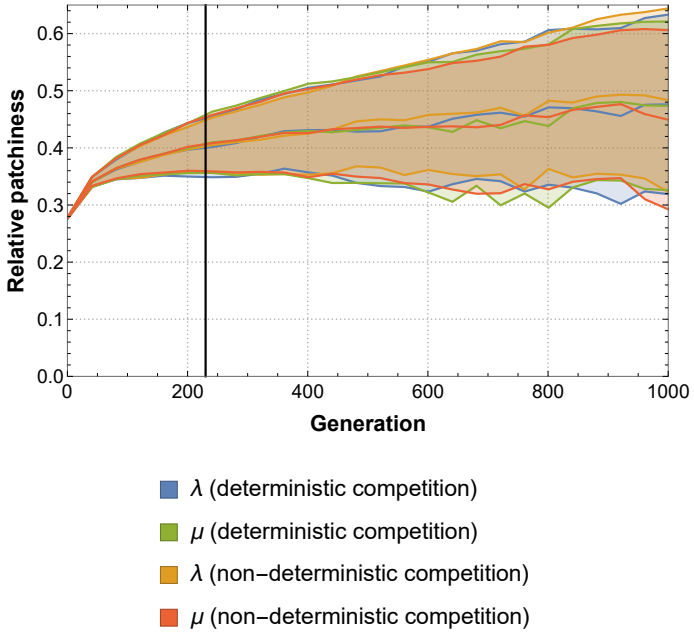


Figure 9.19: Mean relative patchiness over time for deterministic and non-deterministic competition based on λ or μ effects. The black line indicates the mean time of first extinction for these conditions. Means and standard deviations calculated for 200 replicates.

9.4.4.4 Community composition

Having studied community diversity effects, we can now turn our attention to the composition of these persisting communities. In Figure 9.20, we show the persis-

tence probability for each SFI, for deterministic and non-deterministic competition based on combined λ and μ interaction effects. The results reflect the dynamics illustrated in Figure 9.18(a): S9 is the dominant strain. It is always present, while the next most frequently observed strain, K89, only persists in 40% of the simulations. This is unsurprising, since S9 was the strongest competitor in the two competition structures based on combined λ and μ interaction effects ($M_{\lambda\mu}$ and $M^*_{\lambda\mu}$). In the deterministic case, S9 is able to defeat more individuals than any other strain, while in the non-deterministic case it also has the highest possible winning probability in all its pairwise competitions.

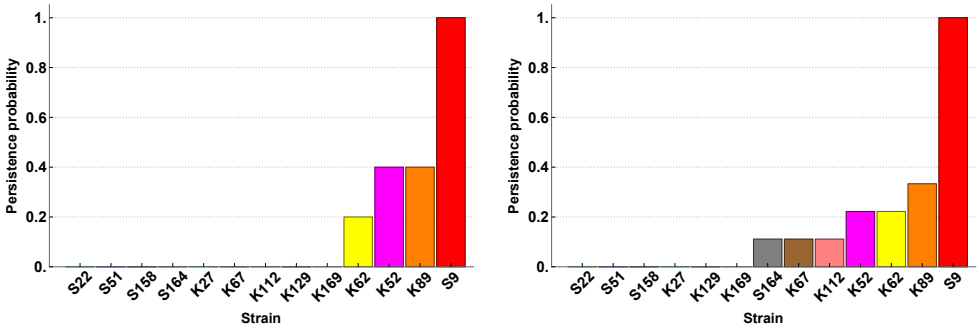


Figure 9.20: Probability of finding each strain in the community after 1000 generations for (a) deterministic and (b) non-deterministic emergent competition based on λ and μ effects.

In Figure 9.21, we show the corresponding results for deterministic and non-deterministic competition based on λ interaction effects only. We find that again S9 is always present in the final community, but now it is part of a subgroup of SFI that are present in the majority of the simulations. In more than 80% of the simulations, we observe the same SFI persisting together: S9, K67, K169, K27 and K89. This is true for both the deterministic and non-deterministic competition cases. Thus our model is able to qualitatively reproduce the previously observed *in vitro* dynamics of a persisting smaller subcommunity (Vandermaesen *et al.*, 2017, in prep; Horemans *et al.*, 2017, in prep). S9 is again the strongest competitor and thus it is again the dominant SFI in the persisting subcommunity, which we recall is quite uneven (see e.g. Figures 9.18 and 9.16).

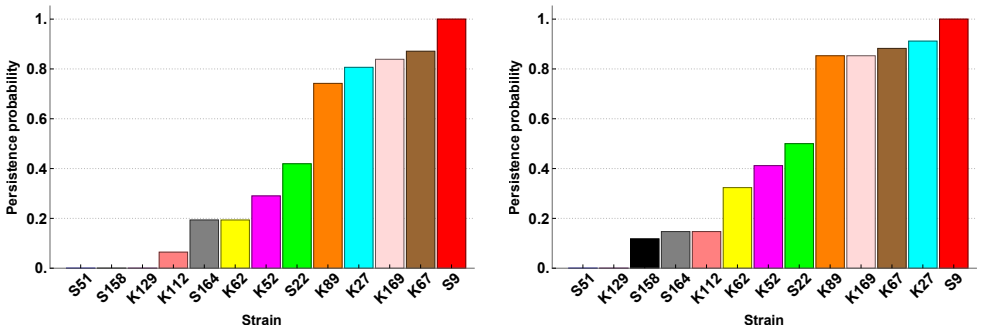


Figure 9.21: Probability of finding each strain in the community after 1000 generations for (a) deterministic and (b) non-deterministic emergent competition based on λ effects.

Another subgroup of persisting SFI is found for deterministic and non-deterministic competition based on μ interaction effects only (Figure 9.22), once again matching qualitatively the dynamics observed in *in vitro* synthetic communities (Vandermaesen *et al.*, 2017, in prep; Horemans *et al.*, 2017, in prep). The members of this subgroup are not the same as for Figure 9.21. Instead we find K169, K52, S158, K27 and S9 coexisting in more than 80% of the simulations. The strains in the persisting subcommunity are also more equal in terms of their persistence probabilities (and hence their extinction probabilities) than was the case for competition based on λ interaction effects only (Figure 9.21). These SFI are also more equally matched in terms of their competitive strengths (see M_μ and M_μ^*). These factors result in these subcommunities being able to maintain significantly higher evenness levels than the other competition structures, as we noted when studying the diversity of these communities (Figure 9.17).

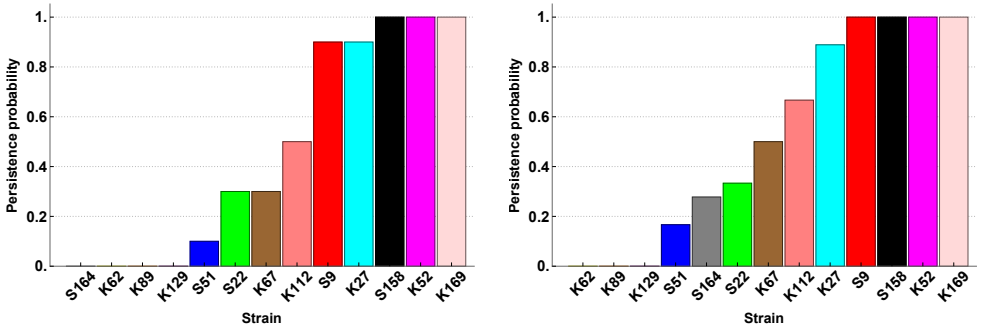
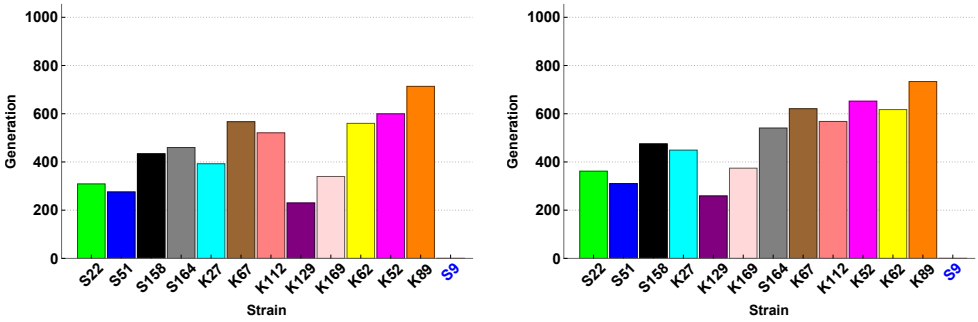


Figure 9.22: Probability of finding each strain in the community after 1000 generations for (a) deterministic and (b) non-deterministic emergent competition based on μ effects.

Finally, we examine extinctions in our different communities. We have seen that extinctions are frequent, but generally limited to the same set of SFI. In Figure 9.23,

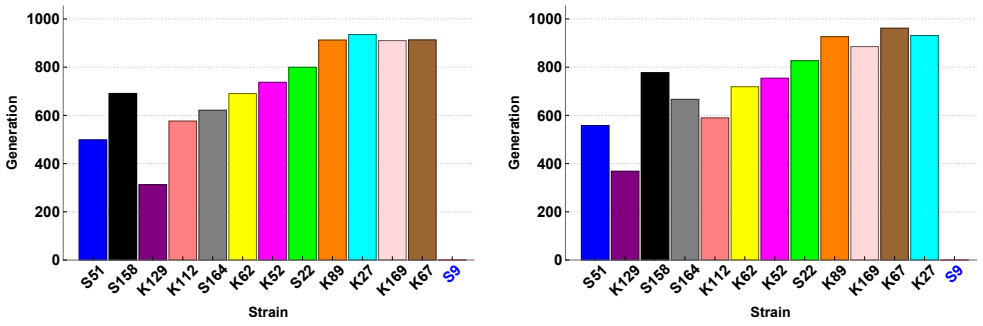
we show the average time to extinction for each SFI, for deterministic and non-deterministic competition based on combined λ and μ interaction effects. We note again that these are slightly longer for non-deterministic competition compared to deterministic competition, and always occur after an initial period of spiral formation ($\sim$ 200 generations), after which extinctions occur rapidly and continuously until the community is reduced to S9 monoculture before $T = 800$ generations.



(a) Deterministic competition (b) Non-deterministic competition

Figure 9.23: Mean time to extinction per strain for (a) deterministic and (b) non-deterministic emergent competition based on μ and λ effects. Blue labels indicate strains for which no extinctions occurred. Means calculated from 200 replicates.

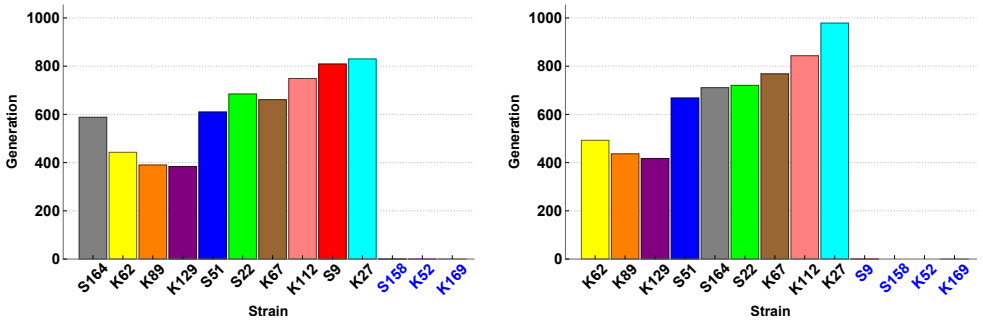
In comparison, the average time to extinction increases for every SFI in the cases of deterministic and non-deterministic competition based on λ interaction effects (Figure 9.24). One strain, K129, collapses to extinction not long after spiral formation has commenced ($\sim$ 300 generations); this strain is the weakest in both these two competition structures. After it disappears, there is another lapse before extinctions recommence and thereafter proceed fairly continuously until the community is reduced to the persisting uneven subcommunity dominated by S9 (which never suffers any extinctions) and the other strains in small proportions.



(a) Deterministic competition (b) Non-deterministic competition

Figure 9.24: Mean time to extinction per strain for (a) deterministic and (b) non-deterministic emergent competition based on λ effects. Blue labels indicate strains for which no extinctions occurred. Means calculated from 200 replicates.

For deterministic and non-deterministic competition based on μ interaction effects (Figure 9.25), we notice a reduction in extinction times compared to competition based on λ interaction effects. This may seem counterintuitive given that we have already observed these communities to be more stable, however the key point is that fewer species go extinct. Those that do collapse to extinction do so more quickly, but this does not affect the stability of the persisting subcommunity. Now S9 is not the only SFI to never suffer extinctions, but it is joined by the other members of the persisting subcommunity, again indicating that this subcommunity is more even and thus more stable than in the cases of competition based on λ interaction effects.



(a) Deterministic competition (b) Non-deterministic competition

Figure 9.25: Mean time to extinction per strain for (a) deterministic and (b) non-deterministic emergent competition based on μ effects. Blue labels indicate strains for which no extinctions occurred. Means calculated from 200 replicates.

9.5 Conclusions

In this chapter, we have studied the *in silico* counterpart of an *in vitro* synthetic community of 13 SFI in co-cultures of varying richness with MSH1. These SFI had been selected based on their potential for improving the BAM mineralization performance of MSH1 for bioaugmentation applications. The *in vitro* experimental approach consisted of two phases. First, synthetic microbial communities of MSH1 combined with different numbers of SFI were subjected to an initial competition phase. Subsequently, BAM was added and the kinetics of BAM mineralization was evaluated as a measure of bioaugmentation success.

In Section 9.3, we have addressed the second phase. We used straightforward regression modelling to illustrate the potential of predictive models for highlighting SFI co-culture combinations that improve the BAM mineralization performance of MSH1, and thus support bioaugmentation success.

Although this modelling approach is straightforward and makes minimal assumptions about the kinetics of the strains and their interactions, the use of Gompertz mineralization parameters and non-linear regression is able to capture the essentials of the interactions taking place, allowing for a very good predictive performance on this dataset.

This approach has promise for reducing experimental load, since good prediction of the mineralization kinetics of all co-cultures can allow the subset of combinations that appear interesting for particular bioaugmentation strategies to be identified and then extensively studied *in vitro*. This approach can be extended to co-cultures of higher richness by the careful choice of co-culture combinations to test, by using for example combinatorial design principles.

In Section 9.4, we have addressed the second experimental phase, where all 13 SFI were inoculated in co-culture during an initial competition phase. We therefore developed an IBM representing the *in silico* counterpart of this synthetic community, where a competition structure was no longer imposed on the community as a whole, but rather pairwise competition outcomes were determined based on data relating to λ and μ interactions effects in terms of mineralization performance. In this way, we built up different competition matrices. This approach is more realistic than the top-down imposition of a competition structure. We also considered non-deterministic competition, where winning probabilities were assigned based on the relative intrinsic competitiveness of each SFI.

Our model was able to recover the qualitative dynamics observed in *in vitro* experiments with similar synthetic sand filter communities: the majority of the community collapsing to extinction and a subcommunity persisting (Vandermaesen *et al.*, 2017, in prep; Horemans *et al.*, 2017, in prep). It should be emphasized that the results in this chapter do not represent a validation of our model, since we have

imposed the competition structures based on the experimental data. Our results instead demonstrate the consistency of our modelling approach in that they are able, based on some experimental input, to reproduce the qualitative behaviour patterns observed *in vitro*. The memberships of the *in silico* subcommunities were consistent, and their presence could be explained based on their attributes as represented in the competition matrix. The simulation outcomes were explained based on the underlying competition structures and the resulting spatial dynamics. Our results again highlighted the importance of diversity and in particular evenness in stabilizing the community dynamics, as shown previously in Chapters 5 and 7.

This work therefore serves as a proof-of-concept for using IBMs as *in silico* counterparts of *in vitro* synthetic communities, as we were able to find a qualitative agreement between the *in silico* and *in vitro* dynamics. It is important to note that the data used in this chapter came from experiments not purposely designed for modelling use, and thus we did not dispose of all the data we would wish for. For example, it is important to have more information about the SFI themselves, and their interactions with each other, not just MSH1. The *in vitro* experimental focus on MSH1 was understandable given the application to bioaugmentation strategies, but it would also be informative for this purpose to examine in more detail the interactions between the SFI, not just in terms of effects on BAM mineralization. This could be done for example by tracking the growth and survival of SFI in pairwise co-cultures. Despite this, our model was able to retrieve the observed qualitative dynamics, allowing us to interrogate their development, and thereby illustrating the potential of this modelling approach for addressing ecological theories relating to synthetic communities.

PART IV

CONCLUSIONS AND PERSPECTIVES

10

General conclusions and perspectives

10.1 General conclusions

In recent years, synthetic microbial communities have gained increasing interest from microbiologists due to their reduced complexity and increased controllability, which favours them over more complex natural systems for examining ecological theories. In this thesis, the *in silico* counterpart of this approach was considered for the purpose of testing ecological theories relating to biodiversity and functionality. Individual-based models of synthetic microbial communities were developed and used in simulation studies to answer research questions relating to community diversity, stability, productivity and functionality.

10.1.1 Part I: background and literature reviews

In Part I, we gathered and synthesized the knowledge necessary to motivate and properly underpin this modelling endeavour. This began in Chapter 2, where we discussed the importance of microbial communities in various vital domains of life, as well as their significant practical applications. The importance of microbial communities, and in particular synthetic microbial communities, was highlighted for the purpose of theory development. This field, known as synthetic microbial ecology, has undoubted potential but is currently in its infancy and therefore presently subject to different limitations in terms of its complexity and applicability. One approach that has been proposed to help synthetic microbial ecology make the jump to further progress is the use of complementary *in silico* studies.

Therefore, in Chapter 3 we surveyed the modelling literature related to this topic, structured in terms of the different possible basic modelling unit: communities, populations, or individuals. We surveyed the modelling approaches that have been developed to study functionality effects due to community spatial structure and dynamics, the interactions taking place within the community, and the interactions occurring with the environment. The gaps remaining in our knowledge of the fundamental mechanisms and processes underlying these features were highlighted, as well as the suitability and promise of IBMs for addressing these open questions.

In Chapter 4, we studied the use of diversity and evenness indices. These measures are key when addressing theories about the relationship between ecosystem stability and biodiversity, which have been tested in various ways using natural and synthetic microbial ecosystems (discussed in more detail in Section 2.5). It is not straightforward to define and select diversity and evenness indices, for reasons specific to microbial ecology, but also for similar reasons as in classical ecology, starting with the most basic question of all, namely how to define diversity itself.

Key differences between microbial and macro-scale communities lead to difficulties when microbial ecologists attempt to use classical macro-ecological methods such as diversity indices to analyze their data. First, microbial communities often contain organisms of wildly different types, involving different domains of life in a variability not typically seen in macro-scale ecological communities. Second, the notion of “species” can be difficult to apply to microbial organisms. Classical measures of diversity typically require a clear differentiation between species, which can often be difficult to achieve in microbial communities due to their particular features such as nonhomologous recombination and a lack of sexual reproduction. Third, characterizing and classifying microbes is complicated by the difficulties in directly observing microbes and their distinguishing characteristics.

Researchers in the field have begun to grapple with these issues in recent years, and have started to turn away from classical diversity indices due to their shortcomings when used with microbiological data. Two notable advances relate to

the use of effective number and similarity-sensitive diversity indices. Effective number indices permit easy interpretation and comparison of the diversities of different communities, while similarity-sensitive indices account for the realistic varying degrees of similarity between microbial species or strains. In cases where microbial strains cannot easily be separated into distinct classes, a similarity measure enables researchers to study the community's diversity without confronting the sometimes tricky issue of species. Numerous such similarity measures can be found in the literature, based on notions ranging from functionality to phylogenetics. Ultimately, the choice of similarity measure can be based on the type of data being generated.

10.1.2 Part II: modelling studies

Armed with the knowledge gathered in Part I, we then embarked in Part II on our modelling study, guided by the research questions highlighted in Section 1.2 which are related to community diversity, stability, productivity and functionality.

In Chapter 5, we considered research questions 1 and 2:

- What effect does initial evenness have on maintaining community diversity?
- Which types of competitive interactions can help maintain community diversity, and which types can threaten it?

To address these questions, we implemented an IBM of a community of three microbial species, which included the three key demographic processes of reproduction, competition and mobility. Our model also allowed the initial evenness of the community to be varied, in order to investigate the consequent effects on community diversity.

Two competition schemes, hierarchical competition and cyclic competition, were investigated using simulation studies which modelled various possible communities, resulting in qualitatively different coexistence and extinction scenarios. System behaviour was strongly dependent on both initial community evenness and the particular competition scheme to which the community was subject.

Coexistence of all species was not permitted by the hierarchical competition scheme, due to frequent and rapid extinction events. Varying the community's initial evenness could not counteract the competitive dynamics which necessarily resulted in monoculture of the dominant species. Very low initial evenness could merely extend the initial transient period before the system settled to its steady state. For the cyclic competition scheme, low initial evenness could counteract the stabilizing dynamics of the competition scheme and provoke extinctions. In

contrast, higher initial evenness could stabilize the dynamics by significantly extending the time until the first species extinction. By extending the region of biodiversity in this way, there was sufficient time for system behaviour to be affected by other factors such as competition scheme, rates of competition and mobility. These results support experimental observations that biodiversity is promoted by increasing evenness.

In Chapter 6, we extended our IBM to include a fourth species. This admitted more competition schemes, which induced more complex behaviour than the three-species case. In addition to studying the effects of varying initial community evenness on the community's stability and diversity, we also made use of two different *in silico* experimental set-ups, motivated by results related to their *in vitro* counterparts. First, we used a co-culture set-up where all species were inoculated at the same time and allowed to evolve together. This set-up permitted the study of four different competition schemes. We also considered an invasion set-up, where three species were inoculated and allowed to evolve to their coexisting steady state, at which point a fourth species was added as an invader. This framework allowed us to address research questions 1 and 2 in a more complicated setting than in Chapter 5, while also considering question 3:

- What effect does initial evenness have on maintaining community diversity?
- Which types of competitive interactions can help maintain community diversity, and which types can threaten it?
- What effect does initial evenness have when a community is faced with invasion?

The four-species system was generally unstable for all competition schemes under both experimental set-ups. There were frequent extinction events, which typically occurred very rapidly. The dynamics induced by the competition schemes worked against the coexistence of all species, and these effects could not be mitigated by varying the evenness of the community.

System behaviour was strongly dependent on initial evenness and competition scheme. The importance of initial evenness was confirmed by means of a sensitivity analysis. Low initial evenness could counteract the dynamics of the competition scheme in the sense that the identity of the first species to collapse to extinction could change. But generally, low initial evenness could only extend the initial transient period before the system settled to its steady state. If initial evenness was excessively low, system biodiversity was lost before other emergent behaviours could be noticed.

In contrast, higher initial evenness could have a small stabilizing effect, in the sense that the time until the first species extinction was slightly extended as initial evenness increased. The time until the first extinction was generally quite

short for all competition schemes except Scheme 3, which was notable as being the most intransitive competition scheme. In this case, the time until the first species extinction could vary significantly. By extending the parameter range permitting biodiversity in this way, there was sufficient time for system behaviour to be affected by other factors such as competition scheme, rates of competition and mobility. These results support experimental observations that biodiversity is promoted by increasing evenness (Isbell et al., 2009b).

When considering an invasion experiment, we found similar evenness effects. Higher resident community evenness before invasion led to a less successful invasion, in terms of invader proportion at the end of the simulation, probability of extinction and time until the first extinction of a resident community member. These results agree with empirical studies from different natural and synthetic ecosystems (Wilsey and Polley, 2004; Hillebrand et al., 2008; Hodgson et al., 2002).

The results of Chapters 4 and 5 demonstrate the danger in overlooking variable community evenness and making the typical assumption that communities are maximally even, despite mounting evidence to the contrary (Huston, 1997; Grime, 1998; Smith and Knapp, 2003). This oversight also ignores the fact that damages due to human actions can affect the evenness of natural communities, often making them more vulnerable to invasion, stresses and disturbances (Wittebolle et al., 2009). While theoretical studies such as this one are beginning to increase in number, experimental studies to validate their conclusions are still lacking (Isbell et al., 2009b).

After studying non-transitive competition and variable initial evenness, two mechanisms known to strongly affect the biodiversity of a system, we moved a step further in Chapter 7 by extending our IBM to include another mechanism shown to be key in mediating biodiversity, namely resource dependence. The resource-dependent nature of demographic processes in real world microbial systems is typically neglected by microscopic models of communities with cyclic competition. However, resource dependence is a key mechanism that can have significant effects on community composition and functioning.

We therefore extended established models by incorporating these three factors, as such aligning them more closely with real-world microbial ecosystems, and permitting us to investigate how this more realistic approach affected community productivity and biodiversity, two key indicators of ecosystem functionality. This allowed us to address research questions 1, 4, and 5:

- What effect does initial evenness have on maintaining community diversity?
- What effect does initial evenness have on maintaining community functionality?
- If interactions within a community are dependent on resource availability and use, how does this affect community diversity and functionality?

In Chapter 7, *in silico* experiments revealed a trade-off between maintaining community diversity and increasing biomass yield. This result is consistent with experimental observations of a negative dominance effect. In addition, the important role that evenness plays in maintaining the functional stability of ecosystem was again demonstrated, indicating the danger in overlooking this key feature in modelling or experimental studies.

Our *in silico* experiments revealed that the dynamics of our IBM framework is mediated in a key way by individuals' mobility. If their interactions were sufficiently localized, this permitted the formation of stable spatial structures, which facilitated the coexistence of all species in the community. In Chapters 5 and 6, this occurred in a spatially explicit but homogeneous environment. The extension of our IBM to include environmental resource dynamics resulted in a heterogeneous *in silico* landscape. In Chapter 8, we studied the effect of this spatial heterogeneity on the community's stability, biodiversity and functionality. We therefore focused on the emergence of spatial patterns and the population dynamics of the community, as well as their underlying mechanisms and the interplay between them. This addressed research questions 2, 4, 5, and 6:

- Which types of competitive interactions can help maintain community diversity, and which types can threaten it?
- What effect does initial evenness have on maintaining community functionality?
- If interactions within a community are dependent on resource availability and use, how does this affect community diversity and functionality?
- How does the spatial structure of a community affect its stability and functionality?

The explicit treatment of space in our model permitted the formation of resource gradients, which induced dramatic effects in the community population dynamics. These effects, consistent with other modelling and experimental studies, are not seen in well-mixed models due to the absence of spatial heterogeneities in such models, thereby neglecting this key facet of natural systems. Our findings have implications for the formation and maintenance of spatial patterns in microbial populations such as biofilms.

If validated with experimental data, our model could be used to predict and visualize unobserved substrate gradients, which can be experimentally impractical or infeasible to measure directly (Hellweger et al., 2016a). The validation would require the spatial distribution of the cells, acquired for example from image analysis of microscopy images, as well as the quantification of the substrate uptake kinetics.

10.1.3 Part III: making use of data from *in vitro* synthetic communities

In Part III, we expanded our scope to work with experimental data obtained from *in vitro* studies of bioaugmentation strategies, building on the insights gained in Chapters 7 and 8. In Chapter 9, we made use of experimental data relating to an *in vitro* synthetic community in order to investigate whether our *in silico* approach was capable of mimicking the dynamics of a similar *in vitro* counterpart.

We therefore formulated an *in silico* counterpart of an *in vitro* synthetic community of 13 SFI in co-cultures of varying richness with MSH1. These SFI had been selected based on their potential for improving the BAM mineralization performance of MSH1 for bioaugmentation applications. The *in vitro* experimental approach consisted of two phases. First, synthetic microbial communities of MSH1 combined with different numbers of SFI were subjected to an initial competition phase. Subsequently, BAM was added and the kinetics of BAM mineralization was evaluated as a measure of bioaugmentation success.

We addressed the second experimental phase through the use of straightforward regression modelling to illustrate the potential of predictive models for highlighting SFI co-culture combinations that improve the BAM mineralization performance of MSH1, and thus support bioaugmentation success. Although this modelling approach was straightforward and made minimal assumptions about the kinetics of the strains and their interactions, through the use of Gompertz mineralization parameters and non-linear regression our approach was able to capture the essentials of the interactions taking place, allowing for a very good predictive performance on this dataset.

This approach has promise for reducing experimental load, since a good prediction of the mineralization kinetics of all co-cultures can allow the subset of combinations that appear interesting for particular bioaugmentation strategies to be identified and then extensively studied *in vitro*. This approach can be extended to co-cultures of higher richness by the careful choice of the co-culture combinations to test, by using for example combinatorial design principles.

We then addressed the first experimental phase, where all 13 SFI were inoculated in co-culture during an initial competition phase. We therefore developed an IBM representing the *in silico* counterpart of this synthetic community, where a competition structure was no longer imposed on the community as a whole. Instead, pairwise competition outcomes were determined based on data relating to lag time λ and mineralization rate μ interactions effects in terms of mineralization performance. This approach is more realistic than the top-down imposition of a competition structure. We also considered non-deterministic competition, where winning probabilities were assigned based on the relative intrinsic competitiveness of each SFI.

Our model was able to recover the qualitative dynamics observed in *in vitro* experiments with similar synthetic sand filter communities: the majority of the community collapsing to extinction and a subcommunity persisting (Vandermaesen *et al.*, 2017, in prep; Horemans *et al.*, 2017, in prep). The compositions of these subcommunities were consistent and could be explained based on the attributes of the community member as represented in the competition matrix. The simulation outcomes were explained based on the underlying competition structures and the resulting spatial dynamics. Our results again highlighted the importance of diversity and in particular evenness in stabilizing the community dynamics, as shown previously in Chapters 5 and 7.

Our work therefore serves as a proof-of-concept for using IBMs as *in silico* counterparts of *in vitro* synthetic communities, as we were able to find qualitative agreement between the *in silico* and *in vitro* dynamics, despite the use of data from experiments not expressly designed for modelling use, and thus not entirely optimal for the modelling endeavour. For example, significant information was lacking regarding the features of the SFI, as well as their interactions with each other. Despite this, our model was able to retrieve the observed qualitative dynamics, namely the development of a persisting sub-community with consistent composition. This allowed us to interrogate the development of these communities, and thereby illustrate the potential of this modelling approach for addressing microbial ecological theories relating to synthetic communities.

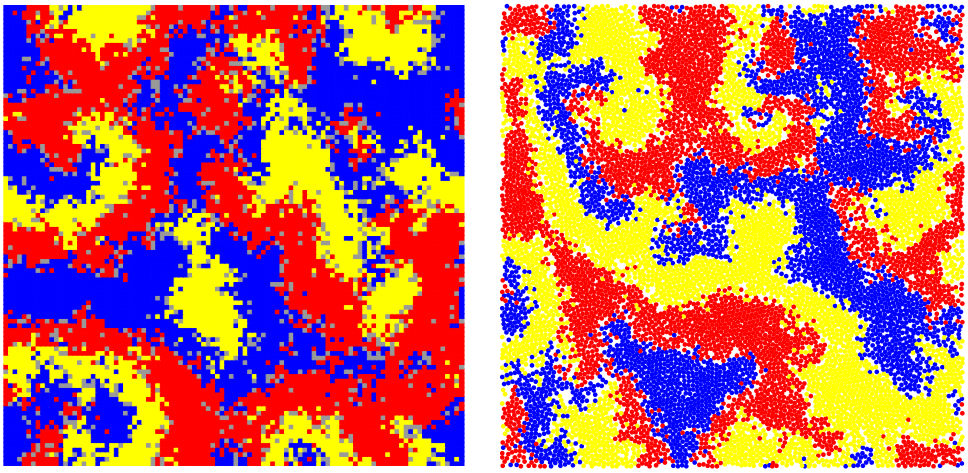
10.2 Perspectives

The work in this thesis demonstrates the potential of *in silico* synthetic microbial ecology studies for the purpose of theory testing and development. This field is still in its infancy, and there is still much more progress to be achieved. The modelling framework we have developed, and the individual-based techniques it incorporates, are very flexible. Hence various extensions are possible that would certainly open new avenues of research, building upon the insights and techniques developed in this thesis. We outline in the subsequent sections several extensions that have promise for further progress in this field.

10.2.1 More realistic movement mechanisms

A first extension would be to remove the lattice structure that characterizes the *in silico* space in our modelling framework. When considering a lattice-free *in silico* space, individuals' locations are described by coordinates in continuous space. Individuals are therefore permitted more spatial degrees of freedom. To enable

coexistence, interactions should still be localized to an individual's nearest neighbourhood, but a new definition of neighbourhood would need to be developed. This could be, for example, those individuals with which one is in direct contact, or could be those individuals located within a certain distance, depending on how far an individual is able to sense and interact with other microbes.



(a) (b)

Figure 10.1: Comparison of spatial structures obtained using (a) lattice-based and (b) lattice-free models.

Preliminary studies of a lattice-free counterpart of the model developed in Chapter 5 show that for sufficiently low mobility, the same characteristic spatial structures are obtained (Figure 10.1), which are lost when the mobility rate exceeds a certain critical value that is a function of system size (Quaghebeur, 2017). Hence the lattice-free model produces identical behaviour to the lattice-based model in this case.

The advantage of this approach is an increased realism in the spatial characteristics and dynamics which the model can produce. Then, for example, theories relating to biofilm formation could be studied. The formation of these structures is key to several different ecological and industrial processes, and has gained much attention in the modelling of waste water treatment strategies in particular (Esser et al., 2015). Thus far, these modelling studies have generally been context-specific, and are not often used for the purpose of theory development.

The disadvantage of an extension to continuous space is the increase in computational cost. Depending on the particular application, this increase in cost may not be worth the added realism. For example, removing the lattice structure from the model studied in Chapter 5 would increase the necessary computation time without adding much to our understanding of the dynamics under consideration,

since these are well represented by a lattice-based model, and finer spatial characterization is not required to address the research hypotheses under consideration. Thus researchers would need to consider if added realism in movement modelling is necessary for their particular purpose and application. It may be that for theory development objectives, the more nimble lattice-based models may suffice, but in cases where the spatial dynamics require finer and more realistic treatment, shifting to continuous *in silico* space can represent an important extension.

One area where a more realistic movement mechanism would be important is the study of directed movement. The movement mechanism in our modelling framework is random, since *in silico* individuals do not orient themselves in any particular direction, but rather wherever there is space available for them. Including mechanisms such as chemotaxis, where microbes follow substrate gradients with a preference for regions of high concentration (Shklarsh et al., 2011), would increase the realism of the *in silico* mobility process. This would also permit the study of other behavioural dynamics, and increase the realism of the interactions with the *in silico* environment. It could also be used to study theories relating to community assembly, where these processes are hypothesized to play key roles, and for which theory development studies are increasing in number (Mensens et al., 2015).

10.2.2 Extension to three dimensions

Related to increasing the realism of the model's spatial characteristics is its extension to three dimensions. An additional dimension is important if this approach is to be applied to study biofilms (Van Loosdrecht et al., 2002) or sand filter columns, where the spatial matrix can play a key role in the assembly and functioning of the resident microbial community (Liu et al., 2012). As noted in Chapter 3, such biofilm IBMs have been developed with extensive microbiological detail and complexity, for example by Lardon *et al.* 2011, Bucci *et al.* 2011, and Picioreanu *et al.* 2004 among others.

For example, Momeni *et al.* (2013) used a three-dimensional IBM of three engineered yeast strains to study heterotypic cooperation, meaning cooperation between two populations that are exchanging different benefits that each incur a production cost. This cooperation can be threatened by cheaters who take advantage of the public benefits without incurring any production cost. The authors observed self-organization of the population, driven by the asymmetric fitness effects of cooperators and cheaters on their partners. This self-organization meant that cooperators interacted more with other cooperators, while cheaters interacted more with other cheaters. This phenomenon of "positive assortment" is the mechanism that stabilized and maintained the heterotypic cooperation, and its emergence from the model relied strongly on the three-dimensional spatial treatment (Momeni et al., 2013).

Furthermore, modelling microbes in three dimensions would permit more realistic representations of their morphologies. In two-dimensional modelling studies, microbes are typically represented by circles or squares. In reality, microbes demonstrate a wide range of different shapes, which are generally classified as: coccus in case of a spherical shape, bacillus in case of a rod shape, spirillum in case of a spiral shape, and pleomorphic in case of no defined shape (Cabeen and Jacobs-Wagner, 2005). These different morphologies may have importance in some microbial ecological theories and in these cases should be accounted for in any corresponding modelling study, for example when studying aggregation phenomena (Peruani et al., 2006).

Implementing this extension for the modelling framework presented in this thesis would have several important implications. First, the possibility of admitting a wider range of microbial body shapes and sizes would increase the realism of the obtained spatial patterns, as well as having important effects on the spatial dynamics. For example, restricting to circular (coccus) morphologies implies an important symmetry in the spatial patterns, whereas a rod shape would allow for different kinds of packing or space-filling patterns, which could have particular significance in cases where the mobility of the *in silico* microbes is limited, and hence the topology of the neighbourhood strongly affects the interactions. In general, asymmetric space-filling patterns can increase the system's spatial heterogeneity, which may have strong positive or negative effects on coexistence (see Section 8.1).

Second, increasing the spatial degrees of freedom to three dimensions would permit the study of more realistic and more complex scenarios, by permitting simulation of the system dynamics over multiple scales of interest. This would require further extensions to account for the increased complexity of the relevant microbial ecology processes, but would be an important step towards the future development of integrated models. However, as with the lattice-free extension discussed in Section 10.2.1, this extension would represent a significantly increased cost in terms of computation time, and hence it should be determined which level of trade-off between computing time and model realism is appropriate for the particular research question.

10.2.3 Increasing community richness

As demonstrated in Chapter 9, our modelling framework can easily be extended to represent communities of high richness. Synthetic microbial ecology studies are characterized by their reduced complexity compared to natural systems, and while this reduced complexity permits their increased manipulability and controllability, it also limits the applicability of the insights and theories gained from their experimental exploitation. Indeed, synthetic microbial ecology studies typically consider

communities consisting of four species or fewer (De Roy et al., 2014), as we have considered in Chapters 5 to 8. Hence our IBM of 13 microbial strains in Chapter 9 is in this sense an innovative outlier.

Increasing the richness of the *in silico* community would allow for the study of more complex interactions and dynamics, but would incur a computational cost. This may be justifiable and indeed necessary for *in silico* studies for theory development purposes, but developing testable models with higher richness for predictive purposes is currently limited not by our modelling or computational tools, but by the difficulties in obtaining appropriate *in vitro* data (Hellweger et al., 2016a) for their calibration and validation, necessary steps before models can be used for generating predictions.

Hence refining the techniques for constructing these models of richer communities is a worthy endeavour, but their deployment is not yet feasible except in cases of very simplified synthetic communities.

10.2.4 Cooperative interactions

In this thesis, we have focused on the effects of competitive interactions in community stability, diversity and functionality. Our modelling framework is equally able to mimic cooperative interactions. As discussed in Section 2.3.3 cooperation is one of the key types of interaction that drives community functionality.

Cooperative interactions can take the form of the shared production of public goods (Tanouchi et al., 2012), or the formation of alliances against common enemies (Frey and Reichenbach, 2011). Such dynamics can, like competitive dynamics, emerge from an individual-based modelling of the underlying interactions (Wintermute and Silver, 2010). Spatial heterogeneity, the modelling of which IBMs are particularly suited for, also plays a key role in mediating cooperation dynamics, since localization and spatial structure affect the interactions between partners (Nadell et al., 2010).

Incorporating cooperative interactions would also allow the study of social strategies and how these are established and maintained in microbial communities. A particularly interesting scenario is cheater/cooperator dynamics (Hibbing et al., 2010), which is at the interface of cooperative and competitive interactions. Such “selfish” social behaviour has been found in a variety of microbial communities (Velicer, 2003). Modelling studies of these communities can help to understand the fundamental dynamics underlying cooperation, as well as their limits.

10.2.5 Adaptive dynamics

Evolution is typically ignored in ecological models, since it is assumed to occur over excessively long time scales compared to the ecological processes being modelled, such as metabolism or reproduction (Jessup et al., 2004). However, this assumption does not always hold true for microbial communities, where ecological and evolutionary time scales can be similar due to short generation times, large population sizes and high rates of horizontal gene transfer (Widder et al., 2016). Hence an important extension to current models is to account for both ecology and evolution. Individual-based models are very well suited for this purpose, as discussed in Section 3.2.3.

Evolutionary or adaptive dynamics can be incorporated into our modelling framework by having dynamic interaction rules, which can adapt in response to changes in the environmental conditions or changes in the behaviour of other strains. This would allow for example the study of the role of mutants in developing new community functionalities (Wintermute and Silver, 2010), or selection due to environmental pressures (Hellweger et al., 2016b). This extension could be tested for example by considering evolutionary rescue, a phenomenon whereby an advantageous genetic change can permit a population to recover from a disturbance (Schiffers and Travis, 2014). Testing would involve tracking the effect of a local adaptation in response to an environmental change on the population's potential for evolutionary rescue. It should however be noted that such an experiment would be difficult to replicate *in vitro*, due to confounding effects which make it difficult to disentangle evolutionary rescue from demographic rescue (due to immigration) or genetic rescue (due to genetic mixing and immigration) except in highly controlled experiments (Kliman, 2016).

10.2.6 Integration with individual-level *in vitro* data

Incorporating additional and more realistic features of microbes requires appropriate data, which relates to the developing field of microbial individual-based ecology (μIBE) (Kreft et al., 2013), involving the combination of individual-based modelling and experimental approaches.

This approach is facilitated by the recent technological advances in single-cell analyses, such as microfluidic devices (Hellweger et al., 2016a). These fabricated environments allow researchers to position individual microbes, as well as to control their environmental conditions and their interactions (Kim et al., 2008). Through the use of genetically engineered reporter cells, metabolic and functional activities can be tracked and their outputs measured (Hol et al., 2014).

Further work with the modelling framework we have developed in this thesis would strongly benefit from such complementary *in vitro* studies. These could be used to

assess the impact of the initial experimental conditions, to give only one example. The experiments need not only be focused on the individual level, since population or even community-level experiments could be used to study the emergent properties of the model community, which we recall as occurring at the population level.

Once such data is obtained and used to parameterize a model, then calibration and validation can be carried out. An important tool in this respect for IBMs (and indeed other types of stochastic models) is Approximate Bayesian Computation (Beaument, 2010). This is a Monte Carlo-based procedure, whose advantage relative to other inverse modelling methods is that it permits the computation of the parameters' posterior probability distributions, which can then directly be used as input for probabilistic assessment methods (Franssen et al., 2009). However, like all Monte Carlo-based method, it requires more computing time.

The procedure for estimating model parameters using Approximate Bayesian Computation can be summarized as follows (van der Vaart et al., 2015):

1. Obtain an empirical dataset relevant for the study (i.e. the data to which the model will be fitted), and construct the IBM
2. Define prior distributions of the model's parameters (e.g. within which ranges are they likely to lie)
3. Using random samples from the prior distributions, run the IBM repeatedly
4. Select the subset of simulation outcomes that best agree with the dataset
5. Using these selected parameters, construct the posterior distributions of the parameters
6. Using these selected parameters, check how well the IBM fits the data
7. Check the accuracy of the estimation, for example using cross-validation

Models may only be used for quantitative predictive purposes — the ultimate goal of synthetic microbial ecology — after proper calibration and validation. These procedures are more complex for IBMs, but still very necessary. Hence the recent and growing appreciation for Approximate Bayesian Computation in this context represents an important advance for mathematical microbial ecology studies that are individual-based, the number of which is increasing every year as their merits are also increasingly appreciated (Hellweger et al., 2016a).

10.3 Concluding summary

In this thesis, we have used *in silico* synthetic microbial communities to test ecological theories. Using an individual-based framework, we assembled communities with different levels of microbial interactions and complexity. This mathematical approach to ecological theory development has allowed us to test microbial ecological theories relating to community stability, diversity, functionality and productivity.

Our work has implications for the management of natural communities, and the engineering of synthetic communities for various applications. The modelling framework we have developed is flexible, extendable to other avenues of research, and furthermore the modelling techniques described in this thesis are not limited in their applicability to microbial ecology, but can be used in other disciplines and fields.

Mathematical approaches like those contained in this thesis will play an important role in the development of Microbial Resource Management theory and tools, which are necessary to permit the management and protection of natural communities, as well as the rational design of engineered communities for industrial applications. Advancing the MRM field will allow researchers to develop new processes and products as well as to manage and improve the natural environment, and to achieve this progress in a suitably sustainable way.

Bibliography

Abrudan, M., You, L., Staňková, K., Thuijsman, F., 2016. A game theoretical approach to microbial coexistence. In: Advances in Dynamic and Evolutionary Games. Springer, Switzerland, pp. 267–282.

Acosta, F., Zamor, R., Najar, F., Roe, B., Hambright, K., 2015. Dynamics of an experimental microbial invasion. Proceedings of the National Academy of Sciences 112 (37), 11594–11599.

Adamson, M., Morozov, A., 2012. Revising the role of species mobility in maintaining biodiversity in communities with cyclic competition. Bulletin of Mathematical Biology 74 (9), 2004–2031.

Adler, P., HilleRisLambers, J., Levine, J., 2007. A niche for neutrality. Ecology Letters 10 (2), 95–104.

Alatalo, R., 1981. Problems in the measurement of evenness in ecology. Oikos 37, 199–204.

Albers, C., Feld, L., Ellegaard-Jensen, L., Aamand, J., 2015. Degradation of trace concentrations of the persistent groundwater pollutant 2, 6-dichlorobenzamide (BAM) in bioaugmented rapid sand filters. Water Research 83, 61–70.

Albers, C., Jacobsen, O., Aamand, J., 2014. Using 2, 6-dichlorobenzamide (BAM) degrading Aminobacter sp. MSH1 in flow through biofilters - initial adhesion and BAM degradation potentials. Applied Microbiology and Biotechnology 98 (2), 957–967.

Allison, S., 2005. Cheaters, diffusion and nutrients constrain decomposition by microbial enzymes in spatially structured environments. Ecology Letters 8 (6), 626–635.

An, D., Danhorn, T., Fuqua, C., Parsek, M., 2006. Quorum sensing and motility mediate interactions between Pseudomonas aeruginosa and Agrobacterium tumefaciens in biofilm cocultures. Proceedings of the National Academy of Sciences of the United States of America 103 (10), 3828–3833.

Ardré, M., Henry, H., Douarche, C., Plapp, M., 2015. An individual–based model for biofilm formation at liquid surfaces. Physical Biology 12 (6), 066015.

Armitage, D., Gallagher, K., Youngblut, N., Buckley, D., Zinder, S., 2012. Millimeter-scale patterns of phylogenetic and trait diversity in a salt marsh microbial mat. Frontiers in Microbiology 3, 293.

Attard, E., Recous, S., Chabbi, A., De Berranger, C., Guillaumaud, N., Labreuche, J., Philippot, L., Schmid, B., Le Roux, X., 2011. Soil environmental conditions rather than denitrifier abundance and diversity drive potential denitrification after changes in land uses. Global Change Biology 17 (5), 1975–1989.

Augusiak, J., Van den Brink, P., Grimm, V., 2014. Merging validation and evaluation of ecological models to "evaludation": a review of terminology and a practical approach. Ecological Modelling 280, 117–128.

Bader, J., Mast-Gerlach, E., Popović, M., Bajpai, R., Stahl, U., 2010. Relevance of microbial coculture fermentations in biotechnology. Journal of Applied Microbiology 109 (2), 371–387.

Baho, D., Peter, H., Tranvik, L., 2012. Resistance and resilience of microbial communities–temporal and spatial insurance against perturbations. Environmental Microbiology 14 (9), 2283–2292.

Bakken, S., Borody, T., Brandt, L., Brill, J., Demarco, D., Franzos, M., Kelly, C., Khoruts, A., Louie, T., Martinelli, L., Moore, T., Russell, G., Surawicz, C., 2011. Treating Clostridium difficile infection with fecal microbiota transplantation. Clinical Gastroenterology and Hepatology 9 (12), 1044–1049.

Balagaddé, F., Song, H., Ozaki, J., Collins, C., Barnet, M., Arnold, F., Quake, S., You, L., 2008. A synthetic Escherichia coli predator–prey ecosystem. Molecular Systems Biology 4 (1), 187.

Baranyi, J., Roberts, T., Jul 1995. Mathematics of predictive food microbiology. International Journal of Food Microbiology 26 (2), 199–218.

Baty, F., Delignette-Muller, M.-L., 2004. Estimating the bacterial lag time: which model, which precision? International Journal of Food Microbiology 91 (3), 261–277.

Beaument, M., 2010. Approximate Bayesian computation in evolution and ecology. Annual Review of Ecology, Evolution, and Systematics 41, 379–406.

Beisel, J.-N., Usseglio-Polatera, P., Bachmann, V., Moreteau, J.-C., 2003. A comparative analysis of evenness index sensitivity. International Review of Hydrobiology 88 (1), 3–15.

Bell, T., Newman, J., Silverman, B., Turner, S., Lilley, A., 2005. The contribution of species richness and composition to bacterial services. Nature 436 (7054), 1157–1160.

Benner, J., Helbling, D., Kohler, H.-P., Wittebol, J., Kaiser, E., Prasse, C., Ternes, T., Albers, C., Aamand, J., Horemans, B., Springael, D., Walravens, E., Boon, N., 2013. Is biological treatment a viable alternative for micropollutant removal in drinking water treatment processes? Water Research 47 (16), 5955–5976.

Benner, S., Sismour, A., 2005. Synthetic biology. Nature Reviews Genetics 6, 533–543.

Berger, W., Parker, F., 1970. Diversity of planktonic foraminifera in deep-sea sediments. Science 168, 1345–1347.

Bernet, M.-F., Brassart, D., Neeser, J.-R., Servin, A., 1994. Lactobacillus acidophilus LA 1 binds to cultured human intestinal cell lines and inhibits cell attachment and cell invasion by enterovirulent bacteria. Gut 35 (4), 483–489.

Berr, M., Reichenbach, T., Schottenloher, M., Frey, E., 2009. Zero–one survival behavior of cyclically competing species. Physical Review Letters 102 (4), 048102.

Bertrand, J.-C., Caumette, P., Lebaron, P., Matheron, R., Normand, P., Sime-Ngando, T., 2015. Environmental microbiology: Fundamentals and applications. Springer, Dordrecht, Netherlands.

Bewick, S., Staniczenko, P., Li, B., Karig, D., Fagan, W., 2017. Invasion speeds in microbial systems with toxin production and quorum sensing. Journal of Theoretical Biology 420, 290–303.

Bez, N., 2000. On the use of Lloyd's index of patchiness. Fisheries Oceanography 9 (4), 372–376.

Billiard, S., Smadi, C., 2015. The interplay of two mutations in a population of varying size: a stochastic eco-evolutionary model for clonal interference. arXiv preprint (arXiv:1505.04550).

Björklund, E., Anskær, G., Hansen, M., Styrishave, B., Halling-Sørensen, B., 2011. Analysis and environmental concentrations of the herbicide dichlobenil and its main metabolite 2, 6-dichlorobenzamide (BAM): a review. Science of the Total Environment 409 (12), 2343–2356.

Bonanomi, G., Capodilupo, M., Incerti, G., Gaglione, S., Scala, F., 2014. Fungal diversity increases soil fungistasis and resistance to microbial invasion by a non resident species. Biological Control 72, 38–45.

Borenstein, D., Meir, Y., Shaevitz, J., Wingreen, N., 2013. Non-local interaction via diffusible resource prevents coexistence of cooperators and cheaters in a lattice model. PLOS ONE 8 (5), e63304.

Borenstein, E., 2012. Computational systems biology and in silico modeling of the human microbiome. Briefings in Bioinformatics 13 (6), 769–780.

Borlee, B., Geske, G., Robinson, C., Blackwell, H., Handelsman, J., 2008. Quorum-sensing signals in the microbial community of the cabbage white butterfly larval midgut. The ISME Journal 2 (11), 1101–1111.

Botta-Dukát, Z., 2005. Rao's quadratic entropy as a measure of functional diversity based on multiple traits. Journal of Vegetation Science 16 (5), 533–540.

Brenner, K., You, L., Arnold, F., 2008. Engineering microbial consortia: a new frontier in synthetic biology. Trends in Biotechnology 26 (9), 483–489.

Brown, K., 1978. Estimation of variance components using residuals. Journal of the American Statistical Association 73 (361), 141–146.

Bruno, J., Cardinale, B., 2008. Cascading effects of predator richness. Frontiers in Ecology and the Environment 6 (10), 539–546.

Bucci, V., Nadell, C., Xavier, J., 2011. The evolution of bacteriocin production in bacterial biofilms. The American Naturalist 178 (6), E162–E173.

Buchanan, R., Whiting, R., Damert, W., 1997. When is simple good enough: a comparison of the Gompertz, Baranyi, and three–phase linear models for fitting bacterial growth curves. Food Microbiology 14 (4), 313–326.

Bulla, L., 1994. An index of evenness and its associated diversity measure. Oikos 70 (1), 167–171.

Cabeen, M., Jacobs-Wagner, C., 2005. Bacterial cell shape. Nature Reviews Microbiology 3, 601–610.

Case, S. O., Durney, C. H., Pleimling, M., Zia, R., 2010. Cyclic competition of four species: Mean–field theory and stochastic evolution. EPL (Europhysics Letters) 92 (5), 58003.

Centler, F., Thullner, M., 2015. Chemotactic preferences govern competition and pattern formation in simulated two-strain microbial communities. Frontiers in Microbiology 6, 40.

Chao, A., Chiu, C.-H., Hsieh, T., 2012. Proposing a resolution to debates on diversity partitioning. Ecology 93 (9), 2037–2051.

Chao, A., Chiu, C.-H., Jost, L., 2010. Phylogenetic diversity measures based on Hill numbers. Philosophical Transactions of the Royal Society B: Biological Sciences 365 (1558), 3599–3609.

Cheng, H., Yao, N., Huang, Z.-G., Park, J., Do, Y., Lai, Y.-C., 2014. Mesoscopic interactions and species coexistence in evolutionary game dynamics of cyclic competitions. Scientific Reports 4 (7486), 1–7.

Chesson, P., 2000. Mechanisms of maintenance of species diversity. Annual Review of Ecology and Systematics 31 (1), 343–366.

Chu, C., Adler, P., 2015. Large niche differences emerge at the recruitment stage to stabilize grassland coexistence. Ecological Monographs 85 (3), 373–392.

Coleman, T., Li, Y., 1994. On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds. Mathematical Programming 67 (1-3), 189–224.

Coleman, T., Li, Y., 1996. An interior trust region approach for nonlinear minimization subject to bounds. SIAM Journal on Optimization 6 (2), 418–445.

Cornu, M., Billoir, E., Bergis, H., Beaufort, A., Zuliani, V., 2011. Modeling microbial competition in food: Application to the behavior of Listeria monocytogenes and lactic acid flora in pork meat products. Food Microbiology 28 (4), 639–647.

Courant, R., Friedrichs, K., Lewy, H., 1928. Über die partiellen Differenzengleichungen der mathematischen Physik. Mathematische Annalen 100, 32–74.

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

Crowder, D., Northfield, T., Strand, M., Snyder, W., 2010. Organic agriculture promotes evenness and natural pest control. Nature 466 (7302), 109–112.

Czachor, M., Naudts, J., 2002. Thermostatistics based on Kolmogorov–Nagumo averages: unifying framework for extensive and nonextensive generalizations. Physics Letters A 298 (5), 369–374.

D'Acunto, B., Frunzo, L., Klapper, I., Mattei, M., 2015. Modeling multispecies biofilms including new bacterial species invasion. Mathematical Biosciences 259, 20–26.

Dalton, H., 1920. The measurement of the inequality of incomes. Econ. J. 31, 124–126.

Daly, A., Baetens, J., De Baets, B., 2015. The impact of initial evenness on biodiversity maintenance for a four–species in silico bacterial community. Journal of Theoretical Biology 387, 189–205.

Daly, A., Baetens, J., De Baets, B., 2016. The impact of resource dependence of the mechanisms of life on the spatial population dynamics of an in silico microbial community. Chaos 26 (12), 123121.

Damodaran, S., Eberhard, S., Boitard, L., Rodriguez, J., Wang, Y., Bremond, N., Baudry, J., Bibette, J., Wollman, F.-A., 2015. A millifluidic study of cell-to-cell heterogeneity in growth-rate and cell-division capability in populations of isogenic cells of Chlamydomonas reinhardtii. PLOS ONE 10 (3), e0118987.

Davidson, R., 2009. Reliable inference for the Gini index. Journal of Econometrics 150 (1), 30–40.

De Bello, F., Lavergne, S., Meynard, C., Lepš, J., Thuiller, W., 2010. The partitioning of diversity: showing Theseus a way out of the labyrinth. Journal of Vegetation Science 21 (5), 992–1000.

De Deyn, G., Raaijmakers, C., Zoomer, H., Berg, M., de Ruiter, P., Verhoef, H., Bezemer, T., van der Putten, W., 2003. Soil invertebrate fauna enhances grassland succession and diversity. Nature 422 (6933), 711–713.

De Meester, N., Gingold, R., Rigaux, A., Derycke, S., Moens, T., 2016. Cryptic diversity and ecosystem functioning: a complex tale of differential effects on decomposition. Oecologia 182 (2), 559–571.

De Roy, K., 2014. Microbial resource management: introducing new tools and ecological theories. Ph.D. thesis, Universiteit Gent.

De Roy, K., Marzorati, M., Negroni, A., Thas, O., Balloi, A., Fava, F., Verstraete, W., Daffonchio, D., Boon, N., 2013. Environmental conditions and community evenness determine the outcome of biological invasion. Nature Communications 4, 1383.

De Roy, K., Marzorati, M., Van den Abbeele, P., Van de Wiele, T., Boon, N., 2014. Synthetic microbial ecosystems: an exciting tool to understand and apply microbial communities. Environmental Microbiology, 1472–1481.

DeAngelis, D., Yurek, S., 2017. Spatially explicit modeling in ecology: a review. Ecosystems 20 (2), 284–300.

Derry, A., Staddon, W., Trevors, J., 1998. Functional diversity and community structure of microorganisms in uncontaminated and creosote–contaminated soils as determined by sole–carbon–source–utilization. World Journal of Microbiology and Biotechnology 14 (4), 571–578.

Dethlefsen, L., Relman, D., Mar 2011. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proceedings of the National Academy of Sciences of the United States of America 108 Suppl 1, 4554–61.

Dieckmann, U., Law, R., Metz, J., 2000. The geometry of ecological interactions.

Dobrinevski, A., Frey, E., 2012. Extinction in neutrally stable stochastic Lotka–Volterra models. Physical Review E 85 (5), 051903.

Doering, C., Sargsyan, K., Sander, L., 2005. Extinction times for birth-death processes: exact results, continuum asymptotics, and the failure of the Fokker–Planck approximation. Multiscale Modeling & Simulation 3 (2), 283–299.

Doll, H., Armitage, D., Daly, R., Emerson, J., Goltsman, D., Yelton, A., Kerekes, J., Firestone, M., Potts, M., 2013. Utilizing novel diversity estimators to quantify multiple dimensions of microbial biodiversity across domains. BMC Microbiology 13 (1), 259.

Durney, C., Case, S., Pleimling, M., Zia, R., 2011. Saddles, arrows, and spirals: deterministic trajectories in cyclic competition of four species. Physical Review E 83 (5), 051108.

Durney, C., Case, S., Pleimling, M., Zia, R., 2012. Stochastic evolution of four species in cyclic competition. Journal of Statistical Mechanics: Theory and Experiment 2012 (06), P06014.

Dutton, R., Brigham, R., 1983. A characterization of competition graphs. Discrete Applied Mathematics 6 (3), 315–317.

Ebenman, B., Jonsson, T., 2005. Using community viability analysis to identify fragile systems and keystone species. Trends in Ecology & Evolution 20 (10), 568–575.

Editorial, 2011. Microbiology by numbers. Nature 9, 628.

Eisenhauer, N., Scheu, S., Jousset, A., 2012. Bacterial diversity stabilizes community productivity. PLOS ONE 7 (3), e34517.

Eisenhauer, N., Schulz, W., Scheu, S., Jousset, A., 2013. Niche dimensionality links biodiversity and invasibility of microbial communities. Functional Ecology 27 (1), 282–288.

Ekelund, F., Harder, C., Knudsen, B., Aamand, J., 2015. Aminobacter MSH1-mineralisation of BAM in sand-filters depends on biological diversity. PLOS ONE 10 (6), e0128838.

Eliazar, I., 2011. Randomness, evenness, and Rényi's index. Physica A: Statistical Mechanics and its Applications 390 (11), 1982–1990.

Eliazar, I., Sokolov, I., 2010. Maximization of statistical heterogeneity: From Shannon's entropy to Gini's index. Physica A: Statistical Mechanics and its Applications 389 (16), 3023–3038.

Eliazar, I., Sokolov, I., 2012. Measuring statistical evenness: A panoramic overview. Physica A: Statistical Mechanics and its Applications 391 (4), 1323–1353.

Esser, D., Leveau, J., Meyer, K., 2015. Modelling microbial growth and dynamics. Applied Microbiology and Biotechnology 99 (21), 8831–8846.

EU, 2006. Directive 2006/118/EC of the European Parliament and of the council of 12 December on the protection of groundwater against pollution and deterioration. Official Journal of the European Union, 19–31.

Faith, D., 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation 61 (1), 1–10.

Falkowski, P., Fenchel, T., Delong, E., 2008. The microbial engines that drive earth's biogeochemical cycles. Science 320 (5879), 1034–1039.

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

Ferrer, J., Prats, C., López, D., 2008. Individual–based modelling: an essential tool for microbiology. Journal of Biological Physics 34 (1-2), 19–37.

Frachebourg, L., Krapivsky, P. L., Ben-Naim, E., 1996. Spatial organization in cyclic Lotka–Volterra systems. Physical Review E 54 (6), 6186.

Franssen, H., Alcolea, A., Riva, M., Bakr, M., Van der Wiel, N., Stauffer, F., Guadagnini, A., 2009. A comparison of seven methods for the inverse modelling of groundwater flow. application to the characterisation of well catchments. Advances in Water Resources 32 (6), 851–872.

Frey, E., Reichenbach, T., 2011. Principles of Evolution. Springer.

Friedman, J., Hastie, T., Tibshirani, R., 2008. The Elements of Statistical Learning, 2nd Edition. Vol. 1. Springer.

Friedman, J., Higgins, L., Gore, J., 2017. Community structure follows simple assembly rules in microbial microcosms. Nature Ecology & Evolution 1, 0109.

Fulthorpe, R., Roesch, L., Riva, A., Triplett, E., Sep 2008. Distantly sampled soils carry few species in common. The ISME Journal 2 (9), 901–10.

Germerodt, S., Bohl, K., Lück, A., Pande, S., Schröter, A., Kaleta, C., Schuster, S., Kost, C., 2016. Pervasive selection for cooperative cross-feeding in bacterial communities. PLoS Computational Biology 12 (6), e1004986.

Gibson, D., 1988. The relationship of sheep grazing and soil heterogeneity to plant spatial patterns in dune grassland. Journal of Ecology 76, 233–252.

Gil, M., Brandao, T., Silva, C., 2006. A modified Gompertz model to predict microbial inactivation under time-varying temperature conditions. Journal of Food Engineering 76 (1), 89–94.

Gilbert, E., Walker, A., Keasling, J., 2003. A constructed microbial consortium for biodegradation of the organophosphorus insecticide parathion. Applied Microbiology and Biotechnology 61 (1), 77–81.

Gilg, O., Hanski, I., Sittler, B., 2003. Cyclic dynamics in a simple vertebrate predator-prey community. Science 302 (5646), 866–868.

Gillespie, D., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics 22 (4), 403–434.

Ginebra, J., Puig, X., 2010. On the measure and the estimation of evenness and diversity. Computational Statistics & Data Analysis 54 (9), 2187–2201.

Ginovart, M., López, D., Gras, A., 2005. Individual–based modelling of microbial activity to study mineralization of c and n and nitrification process in soil. Nonlinear Analysis: Real World Applications 6 (4), 773–795.

Ginovart, M., Xifré, J., López, D., Silbert, M., 2007. INDISIM-YEAST, an individual–based model to study yeast population in batch cultures. Communicating Current Research and Educational Topics and Trends in Applied Microbiology (1), 401–409.

Glendinning, P., 1994. Stability, instability and chaos: an introduction to the theory of nonlinear differential equations. Cambridge University Press, Cambridge, United Kingdom.

Goers, L., Freemont, P., Polizzi, K., 2014. Co-culture systems and technologies: taking synthetic biology to the next level. Journal of The Royal Society Interface 11 (96), 20140065.

Gosselin, F., 2001. Lorenz partial order: the best known logical framework to define evenness indices. Community Ecology 2, 197–207.

Gras Moreu, A., Ginovart Gisbert, M., Valls Ribas, J., Baveye, P., et al., 2011. Individual–based modelling of carbon and nitrogen dynamics in soils: parameterization and sensitivity analysis of microbial components. Ecological Modelling 222, 1998–2010.

Greenblum, S., Turnbaugh, P., Borenstein, E., 2012. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity

and inflammatory bowel disease. Proceedings of the National Academy of Sciences 109, 594–599.

Gregorius, H.-R., Kosman, E., 2017. On the notion of dispersion: from dispersion to diversity. Methods in Ecology and Evolution 8 (3), 278–287.

Grime, J., 1998. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. Journal of Ecology 86 (6), 902–910.

Grimm, V., Augusiak, J., Focks, A., Frank, B., Gabsi, F., Johnston, A., Liu, C., Martin, B., Meli, M., Radchuk, V., Thorbek, P., Railsback, S., 2014. Towards better modelling and decision support: documenting model development, testing, and analysis using trace. Ecological Modelling 280, 129–139.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S., Huse, G., et al., 2006. A standard protocol for describing individual-based and agent-based models. Ecological Modelling 198 (1), 115–126.

Grimm, V., Berger, U., DeAngelis, D., Polhill, J., Giske, J., Railsback, S., 2010. The ODD protocol: A review and first update. Ecological Modelling 221 (23), 2760–2768.

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W., Railsback, S., Thulke, H.-H., Weiner, J., Wiegand, T., DeAngelis, D., 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. Science 310 (5750), 987–991.

Guill, C., Drossel, B., Just, W., Carmack, E., May 2011. A three–species model explaining cyclic dominance of Pacific salmon. Journal of Theoretical Biology 276 (1), 16–21.

Hairston, D., Huban, C., Plowman, R., Chemicals, S., 1997. Bioaugmentation: Put microbes to work. Chemical Engineering 104 (3), 74.

Hallam, A., Wignall, P., 1997. Mass Extinctions and Their Aftermath. Oxford University Press.

Han, X., Chen, B., Hui, C., 2016. Symmetry breaking in cyclic competition by niche construction. Applied Mathematics and Computation 284, 66–78.

Hanemaaijer, M., Röling, W., Olivier, B., Khandelwal, R., Teusink, B., Bruggeman, F., 2015. Systems modeling approaches for microbial community studies: from metagenomics to inference of the community structure. Frontiers in Microbiology 6, 213.

Harch, B., Correll, R., Meech, W., Kirkby, C., Pankhurst, C., 1997. Using the Gini coefficient with BIOLOG substrate utilisation data to provide an alternative quantitative measure for comparing bacterial soil communities. Journal of Microbiological Methods 30 (1), 91–101.

Hardy, O., Senterre, B., 2007. Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. Journal of Ecology 95 (3), 493–506.

Harpole, W., Suding, K., Dec 2007. Frequency–dependence stabilizes competitive interactions among four annual plants. Ecology Letters 10 (12), 1164–9.

Haynes, W., 2013. Tukey's test. In: Dubitzky, W., Wolkenhauer, O., Cho, K.-H., Yokota, H. (Eds.), Encyclopedia of Systems Biology. Springer New York, pp. 2303–2304.

Heip, C., 1974. A new index measuring evenness. Journal of the Marine Biological Association of the United Kingdom 54 (03), 555–557.

Heip, C., Warwick, R., Carr, M., Herman, P., Huys, R., Smol, N., Van Holsbeke, K., 1988. Analysis of community attributes of the benthic meiofauna of Frierfjord/Langesundfjord. Marine Ecology Progress Series 46, 171–180.

Hellweger, F., 2008. Spatially explicit individual-based modeling using a fixed super-individual density. Computers & Geosciences 34 (2), 144–152.

Hellweger, F., Bucci, V., 2009. A bunch of tiny individuals — individual-based modelling for microbes. Ecological Modelling 220 (1), 8–22.

Hellweger, F., Clegg, R., Clark, J., Plugge, C., Kreft, J.-U., 2016a. Advancing microbial sciences by individual-based modelling. Nature Reviews Microbiology 14, 461–471.

Hellweger, F., van Sebille, E., Calfee, B., Chandler, J., Zinser, E., Swan, B., Fredrick, N., 2016b. The role of ocean currents in the temperature selection of plankton: Insights from an individual-based model. PLOS ONE 11 (12), e0167010.

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

Hill, M., 1973. Diversity and evenness: a unifying notation and its consequences. Ecology 54 (2), 427–432.

Hill, T., Walsh, K., Harris, J., Moffett, B., Feb 2003. Using ecological diversity measures with bacterial communities. FEMS Microbiology Ecology 43 (1), 1–11.

Hillebrand, H., Bennett, D., Cadotte, M., 2008. Consequences of dominance: a review of evenness effects on local and regional ecosystem processes. Ecology 89 (6), 1510–1520.

Ho, A., De Roy, K., Thas, O., De Neve, J., Hoefman, S., Vandamme, P., Heylen, K., Boon, N., 2014. The more, the merrier: heterotroph richness stimulates methanotrophic activity. The ISME Journal 8 (9), 1945.

Hodgson, D., Rainey, P., Buckling, A., Nov 2002. Mechanisms linking diversity, productivity and invasibility in experimental bacterial communities. Proceedings of the Royal Society of London B: Biological Sciences 269 (1506), 2277–83.

Hoffmann, S., 2008. Generalized distribution based diversity measurement: Survey and unification. Tech. rep., Otto-von-Guericke University Magdeburg, Faculty of Economics and Management.

Hol, F. J., Voges, M. J., Dekker, C., Keymer, J. E., 2014. Nutrient-responsive regulation determines biodiversity in a colicin-mediated bacterial community. BMC Biology 12 (1), 1.

Hooper, D., Chapin, F., Ewel, J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J., Lodge, D., Loreau, M., Naeem, S., Schmid, B., Setala, H., Symstad, A., Vandermeer, J., Wardle, D., 2005. Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. Ecological Monographs 75 (1), 3–35.

Hooper, D., Dukes, J., 2004. Overyielding among plant functional groups in a long-term experiment. Ecology Letters 7 (2), 95–105.

Horňák, K., Corno, G., 2012. Every coin has a back side: invasion by Limnohabitans planktonicus promotes the maintenance of species diversity in bacterial communities. PLOS ONE 7 (12), e51576.

Hughes, A., Inouye, B., Johnson, M., Underwood, N., Vellend, M., 2008. Ecological consequences of genetic diversity. Ecology Letters 11 (6), 609–623.

Hummert, S., Bohl, K., Basanta, D., Deutsch, A., Werner, S., Thei, G., Schroeter, A., Schuster, S., 2014. Evolutionary game theory: cells as players. Molecular BioSystems 10 (12), 3044–3065.

Hurlbert, S., 1971. The nonconcept of species diversity: a critique and alternative parameters. Ecology 52 (4), 577–586.

Huston, M., 1997. Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. Oecologia 110 (4), 449–460.

Hutkins, R., 2006. Microbiology and Technology of Fermented Foods. Wiley-Blackwell, New Jersey, USA.

Iman, R., 2008. Encyclopedia of quantitative risk analysis and assessment.

Intoy, B., Pleimling, M., 2013. Extinction in four species cyclic competition. Journal of Statistical Mechanics: Theory and Experiment 2013 (08), P08011.

Isbell, F., Polley, H., Wilsey, B., May 2009a. Biodiversity, productivity and the temporal stability of productivity: patterns and processes. Ecology Letters 12 (5), 443–51.

Isbell, F., Polley, H., Wilsey, B., 2009b. Species interaction mechanisms maintain grassland plant species diversity. Ecology 90 (7), 1821–1830.

Jackson, J., Buss, L., Dec 1975. Alleopathy and spatial competition among coral reef invertebrates. Proceedings of the National Academy of Sciences of the United States of America 72 (12), 5160–3.

Jenkinson, H., Lamont, R., 2005. Oral microbial communities in sickness and in health. Trends in Microbiology 13 (12), 589–595.

Jensen, J., 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta Mathematica 30 (1), 175–193.

Jessup, C., Kassen, R., Forde, S., Kerr, B., Buckling, A., Rainey, P., Bohannan, B., 2004. Big questions, small worlds: microbial model systems in ecology. Trends in Ecology & Evolution 19 (4), 189–197.

Johnson, J., 1973. Use of nucleic-acid homologies in the taxonomy of anaerobic bacteria. International Journal of Systematic Bacteriology 23 (4), 308–315.

Jost, L., 2006. Entropy and diversity. Oikos 113 (2), 363–375.

Jost, L., 2007. Partitioning diversity into independent alpha and beta components. Ecology 88 (10), 2427–2439.

Jost, L., 2009. Mismeasuring biological diversity: response to Hoffmann and Hoffmann (2008). Ecological Economics 68 (4), 925–928.

Jost, L., 2010. The relation between evenness and diversity. Diversity 2 (2), 207–232.

Jost, L., Chao, A., 2008. Diversity Analysis. Taylor & Francis, United Kingdom.

Kerr, B., Riley, M., Feldman, M., Bohannan, B., 2002. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. Nature 418 (6894), 171–174.

Kerr, M., Fitzgerald, M., Sheridan, J., McDowell, D., Blair, I., Dec 1999. Survival of Escherichia coli O157:H7 in bottled natural mineral water. Journal of Applied Microbiology 87 (6), 833–41.

Keuls, M., 1952. The use of the "studentized range" in connection with an analysis of variance. Euphytica 1, 112–122.

Keylock, C., 2005. Simpson diversity and the Shannon–Wiener index as special cases of a generalized entropy. Oikos 109 (1), 203–207.

Kim, H., Boedicker, J., Choi, J., Ismagilov, R., Nov 2008. Defined spatial structure stabilizes a synthetic multispecies bacterial community. Proceedings of the National Academy of Sciences of the United States of America 105 (47), 18188–18193.

Kinnunen, M., Dechesne, A., Proctor, C., Hammes, F., Johnson, D., Quintela-Baluja, M., Graham, D., Daffonchio, D., Fodelianakis, S., Hahn, N., Boon, N., Smets, B., 2016. A conceptual framework for invasion in microbial communities. The ISME Journal 10 (12), 2773–2779.

Kirkpatrick, S., Gelatt, C., Vecchi, M., 1983. Optimization using simulated annealing. Science 220 (4598), 671–680.

Kliman, R. (Ed.), 2016. Encyclopedia of Evolutionary Biology, Volume 1. Elsevier Science Technology Books.

Klimenkoa, A., Mustafina, Z., Chekantseva, A., Zudina, R., Matushkina, Y., Lashina, S., 2016. A review of simulation and modeling approaches in microbiology. Russian Journal of Genetics: Applied Research 6 (8), 845–853.

Klitgord, N., Segre, D., 2010. Environments that induce synthetic microbial ecosystems. PLoS Computational Biology 6 (11), e1001002.

Klitgord, N., Segre, D., 2011. Ecosystems biology of microbial metabolism. Current Opinion in Biotechnology 22 (4), 541–546.

Koch, A., 1982. Multistep kinetics: choice of models for the growth of bacteria. Journal of Theoretical Biology 98 (3), 401–417.

Koch, A., 2012. The Monod Model and Its Alternatives. Springer Science & Business Media.

Koch, A., Robinson, J., Milliken, G., 2012. Mathematical modeling in microbial ecology. Springer Science & Business Media.

Koenigstein, S., Mark, F., Gö-Reisemann, S., Reuter, H., Poertner, H.-O., 2016. Modelling climate change impacts on marine fish populations: process-based integration of ocean warming, acidification and other environmental drivers. Fish and Fisheries 17 (4), 972–1004.

König, S., Worrich, A., Centler, F., Wick, L., Miltner, A., Kästner, M., Thullner, M., Frank, K., Banitz, T., 2017. Modelling functional resilience of microbial ecosystems: Analysis of governing processes. Environmental Modelling & Software 89, 31–39.

Kraft, B., Tegetmeyer, H., Meier, D., Geelhoed, J., Strous, M., 2014. Rapid succession of uncultured marine bacterial and archaeal populations in a denitrifying continuous culture. Environmental Microbiology 16 (10), 3275–3286.

Kragh, K., Hutchison, J., Melaugh, G., Rodesney, C., Roberts, A., Irie, Y., Jensen, P., Diggle, S., Allen, R., Gordon, V., Bjarnsholt, T., 2016. Role of multicellular aggregates in biofilm formation. mBio 7 (2), 1–11.

Kreft, J.-U., Booth, G., Wimpenny, J., 1998. BacSim, a simulator for individual–based modelling of bacterial colony growth. Microbiology 144 (12), 3275–3287.

Kreft, J.-U., Plugge, C., Grimm, V., Prats, C., Leveau, J., Banitz, T., Baines, S., Clark, J., Ros, A., Klapper, I., Topping, C., Field, A., Schuler, A., Litchman, E., Hellweger, F., 2013. Mighty small: Observing and modeling individual microbes becomes big science. Proceedings of the National Academy of Sciences 110 (45), 18027–18028.

Kunin, V., He, S., Warnecke, F., Peterson, S. B., Martin, H., Haynes, M., Ivanova, N., Blackall, L., Breitbart, M., Rohwer, F., McMahon, K., Hugenholtz, P., 2008. A bacterial metapopulation adapts locally to phage predation despite global dispersal. Genome Research 18 (2), 293–297.

Kvalseth, T., 2015. Evenness indices once again: critical analysis of properties. SpringerPlus 4 (1), 232.

Laird, R., 2014. Population interaction structure and the coexistence of bacterial strains playing “rock–paper–scissors”. Oikos 123 (4), 472–480.

Laird, R., Schamp, B., 2006. Competitive intransitivity promotes species coexistence. The American Naturalist 168 (2), 182–193.

Laird, R., Schamp, B., 2008. Does local competition increase the coexistence of species in intransitive networks. Ecology 89 (1), 237–247.

Laird, R., Schamp, B., 2015. Competitive intransitivity, population interaction structure, and strategy coexistence. Journal of Theoretical Biology 365, 149–158.

Lam, K., Cheng, J., Engel, K., Neufeld, J., Charles, T., 2015. Current and future resources for functional metagenomics. Frontiers in Microbiology 6, 196.

Lambert, P., Aronson, J., 1993. Inequality decomposition analysis and the Gini coefficient revisited. The Economic Journal, 1221–1227.

Lande, R., 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos 76, 5–13.

Lardon, L., Merkey, B., Martins, S., Dötsch, A., Picioreanu, C., Kreft, J.-U., Smets, B., 2011. iDynoMiCS: next-generation individual-based modelling of biofilms. Environmental Microbiology 13 (9), 2416–2434.

Larsen, P., Hamada, Y., Gilbert, J., 2012. Modelling microbial communities: Current, developing, and future technologies for predicting microbial community interaction. Journal of Biotechnology 160 (1), 17–24.

Laspidou, C., Kungolos, A., Samaras, P., 2010. Cellular–automata and individual–based approaches for the modeling of biofilm structures: Pros and cons. Desalination 250 (1), 390–394.

Lehtola, M., Miettinen, I., Vartiainen, T., Martikainen, P., 2002. Changes in content of microbially available phosphorus, assimilable organic carbon and microbial growth potential during drinking water treatment processes. Water Research 36 (15), 3681–3690.

Leinster, T., Cobbold, C., 2012. Measuring diversity: the importance of species similarity. Ecology 93 (3), 477–489.

Lemieux, J., Cusson, M., 2014. Effects of habitat-forming species richness, evenness, identity, and abundance on benthic intertidal community establishment and productivity. PLOS ONE 9 (10), e109261.

Leonard, E., Nielsen, D., Solomon, K., Prather, K., 2008. Engineering microbes with synthetic biology frameworks. Trends in Biotechnology 26 (12), 674–681.

Levine, J., Bascompte, J., Adler, P., Allesina, S., 2017. Beyond pairwise mechanisms of species coexistence in complex communities. Nature 546 (7656), 56–64.

Levy, R., Borenstein, E., 2012. Reverse ecology: from systems to environments and back. Advances in Experimental Medicine and Biology 751, 329–45.

Lewis, M., Petrovskii, S., Potts, J., 2016. The Mathematics Behind Biological Invasions. Springer International Publishing.

Lilburne, L., Tarantola, S., 2009. Sensitivity analysis of spatial models. International Journal of Geographical Information Science 23 (2), 151–168.

Lin, J., 1991. Divergence measures based on the shannon entropy. IEEE Transactions on Information Theory 37 (1), 145–151.

Little, A., Robinson, C., Peterson, S., Raffa, K., Handelsman, J., 2008. Rules of engagement: Interspecies interactions that regulate microbial communities. Annual Review of Microbiology 62, 375–401.

Liu, M., Bjørnlund, L., Rønn, R., Christensen, S., Ekelund, F., 2012. Disturbance promotes non-indigenous bacterial invasion in soil microcosms: analysis of the roles of resource availability and community structure. PLOS ONE 7 (10).

Lloyd, D., Allen, R., 2015. Competition for space during bacterial colonization of a surface. Journal of The Royal Society Interface 12 (110), 20150608.

Loreau, M., Hector, A., 2001. Partitioning selection and complementarity in biodiversity experiments. Nature 412 (6842), 72–76.

Lotka, A., 1925. Elements of Physical Biology. Williams and Wilkins.

Lütz, A., Risau-Gusman, S., Arenzon, J., 2013. Intransitivity and coexistence in four species cyclic games. Journal of Theoretical Biology 317, 286–292.

Mabrouk, N., 2010. Analyzing individual–based models of microbial systems. Ph.D. thesis, Université Blaise Pascal-Clermont-Ferrand II.

MacArthur, R., 1965. Patterns of species diversity. Biological Reviews 40 (4), 510–533.

MacPherson, B., Gras, R., 2016. Individual-based ecological models: Adjunctive tools or experimental systems? Ecological Modelling 323, 106–114.

Madigan, M., Martinko, J., Stahl, D., Clark, D., 2008. Brock Biology of Microorganisms 13th edn. International Microbiology 11, 65–73.

Magurran, A., 2004. Measuring Biological Diversity. Blackwell Science Ltd, Oxford, UK.

Maignan, C., Ottaviano, G., Pinelli, D., Rullani, F., 2003. Bio-ecological diversity vs. socio-economic diversity: A comparison of existing measures.

Mallon, C., van Elsas, J., Salles, J., 2015. Microbial invasions: the process, patterns, and mechanisms. Trends in Microbiology 23 (11), 719–729.

Malthus, T., 1798. An Essay On The Principle Of Population. Penguin Classics.

Manukyan, L., Montandon, S., Fofonjka, A., Smirnov, S., Milinkovitch, M., 2017. A living mesoscopic cellular automaton made of skin scales. Nature 544 (7649), 173–179.

Martin, M., 2002. Predatory prokaryotes: an emerging research opportunity. Journal of Molecular Microbiology and Biotechnology 4 (5), 467–477.

Matos, A., Kerkhof, L., Garland, J., 2005. Effects of microbial community diversity on the survival of Pseudomonas aeruginosa in the wheat rhizosphere. Microbial Ecology 49 (2), 257–264.

Matthiessen, B., Ptacnik, R., Hillebrand, H., 2010. Diversity and community biomass depend on dispersal and disturbance in microalgal communities. Hydrobiologia 653, 65–78.

McGill, B., Etienne, R., Gray, J., Alonso, D., Anderson, M., Benecha, H., Dornelas, M., Enquist, B., Green, J., He, F., Hurlbert, A., Magurran, A., Marquet, P., Maurer, B., Ostling, A., Soykan, C., K.I., U., White, E., 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecology Letters 10 (10), 995–1015.

Melbinger, A., Cremer, J., Frey, E., 2010. Evolutionary game theory in growing populations. Physical Review Letters 105 (17), 178101.

Mendes, R., Evangelista, L., Thomaz, S., Agostinho, A., Gomes, L., 2008. A unified index to measure ecological diversity and species rarity. Ecography 31 (4), 450–456.

Mensens, C., De Laender, F., Janssen, C., Sabbe, K., De Troch, M., 2015. Stressor-induced biodiversity gradients: revisiting biodiversity–ecosystem functioning relationships. Oikos 124 (6), 677–684.

Merkey, B., Lardon, L., Seoane, J., Kreft, J.-U., Smets, B., 2011. Growth dependence of conjugation explains limited plasmid invasion in biofilms: an individual-based modelling study. Environmental Microbiology 13 (9), 2435–2452.

Miller, T., Burns, J., Munguia, P., Walters, E., Kneitel, J., Richards, P., Mouquet, N., Buckley, H., 2005. A critical review of twenty years' use of the resource–ratio theory. The American Naturalist 165 (4), 439–448.

Mills, A., Wassel, R., 1980. Aspects of diversity measurement for microbial communities. Applied and Environmental Microbiology 40 (3), 578–586.

Moayyedi, P., Ford, A., Talley, N., Cremonini, F., Foxx-Orenstein, A., Brandt, L., Quigley, E., 2008. The efficacy of probiotics in the therapy of irritable bowel syndrome: a systematic review. Gut 53 (3), 325–332.

Momeni, B., Waite, A., Shou, W., 2013. Spatial self-organization favors heterotypic cooperation over cheating. eLIFE 2, e00960.

Monod, J., 1948. The growth of bacterial cultures. Annual Review of Microbiology 3, 371–394.

Moons, P., Michiels, C., Aertsen, A., 2009. Bacterial interactions in biofilms. Critical Reviews in Microbiology 35, 157–168.

Mora Villarrubia, R., Ruiz-Castillo, J., 2010. Entropy-based segregation indices. Tech. rep., Universidad Carlos III, Madrid, Spain.

Mouritsen, K., Poulin, R., 2005. Parasites boosts biodiversity and changes animal community structure by trait-mediated indirect effects. Oikos 108 (2), 344–350.

Murphy, J., Walshe, R., Devocelle, M., 2008. A computational model of antibiotic-resistance mechanisms in Methicillin-Resistant Staphylococcus aureus (MRSA). Journal of Theoretical Biology 254 (2), 284–293.

Murray, J., 2002. Mathematical Biology. Springer-Verlag, New York, USA.

Murray, M., Baird, D., 2008. Resource-ratio theory applied to large herbivores. Ecology 89 (5), 1445–1456.

Nadell, C., Bassler, B., 2011. A fitness trade-off between local competition and dispersal in Vibrio cholerae biofilms. Proceedings of the National Academy of Sciences 108 (34), 14181–14185.

Nadell, C., Foster, K., Xavier, J., 2010. Emergence of spatial structure in cell groups and the evolution of cooperation. PLOS Computational Biology 6 (3), e1000716.

Nei, M., 1972. Genetic distance between populations. American Naturalist 106 (949), 283–292.

Neuhauser, C., 2001. Mathematical challenges in spatial ecology. Notices of the AMS 48 (11), 1304–1314.

Nicholson, A. J., 1954. An outline of the dynamics of animal populations. Australian Journal of Zoology 2 (1), 9–65.

Nowak, M., 2006. Five rules for the evolution of cooperation. Science 314, 1560–1563.

Ogunseitan, O., 2005. Microbial diversity. Blackwell Science Ltd.

Orth, J., Thiele, I., Palsson, B., 2010. What is flux balance analysis? Nature Biotechnology 28, 245–248.

Pallmann, P., Schaarschmidt, F., Hothorn, L., Fischer, C., Nacke, H., Priesnitz, K., Schork, N., 2012. Assessing group differences in biodiversity by simultaneously testing a user-defined selection of diversity indices. Molecular Ecology Resources 12 (6), 1068–1078.

Parker, M., Kamenev, A., 2010. Mean extinction time in predator–prey model. Journal of Statistical Physics 141 (2), 201–216.

Patil, G., 2002. Diversity profiles. Encyclopedia of Environmetrics.

Patil, G., 2013. Encyclopedia of Econometrics.

Patra, A., Abbadie, L., Clays-Josserand, A., Degrange, V., Grayston, S., Guillaumaud, N., Loiseau, P., Louault, F., Mahmood, S., Nazaret, S., Phillipot, L., Poly, F., Prosser, J., Le Roux, X., 2006. Effects of management regime and plant species on the enzyme activity and genetic structure of N-fixing, denitrifying and nitrifying bacterial communities in grassland soils. Environmental Microbiology 8 (6), 1005–1016.

Pavoine, S., Ollier, S., Pontier, D., 2005. Measuring diversity from dissimilarities with Rao's quadratic entropy: are any dissimilarities suitable? Theoretical Population Biology 67 (4), 231–239.

Peet, R., 1974. The measurement of species diversity. Annual Review of Ecology and Systematics, 285–307.

Pérez-Velázquez, J., Gölgeli, M., García-Contreras, R., 2016. Mathematical modelling of bacterial quorum sensing: A review. Bulletin of Mathematical Biology 78 (8), 1585–1639.

Permogorskiy, M., 2015. Competitive intransitivity among species in biotic communities. Biology Bulletin Reviews 5 (3), 213–219.

Peruani, F., Deutsch, A., Bär, M., 2006. Nonequilibrium clustering of self-propelled rods. Physical Review E 74 (3), 030904.

Petchey, O., Gaston, K., 2006. Functional diversity: back to basics and looking forward. Ecology Letters 9 (6), 741–758.

Pham, L., van Spanning, R., Röling, W., Prosperi, A., Terefework, Z., Jacob, M., Crielaard, W., Zaura, E., 2009. Effects of probiotic Lactobacillus salivarius W24 on the compositional stability of oral microbial communities. Archives of Oral Biology 54 (2), 132–137.

Picioreanu, C., Kreft, J.-U., Van Loosdrecht, M., May 2004. Particle–based multidimensional multispecies biofilm model. Applied and Environmental Microbiology 70 (5), 3024–40.

Pielou, E., 1966. The measurement of diversity in different types of biological collections. Journal of Theoretical Biology 13, 131–144.

Pinna, A., Lallai, A., Mura, G., Grosso, M., 2009. Comparison across different models for the description of batch biodegradation processes. Chemical Engineering Transactions 149, 1227–1232.

Piper, C., Siciliano, S., Winsley, T., Lamb, E., 2015. Smooth brome invasion increases rare soil bacterial species prevalence, bacterial species richness and evenness. Journal of Ecology 103 (2), 368–396.

Planque, B., Lindstrom, U., Subbey, S., 2014. Non-deterministic modelling of food-web dynamics. PLOS ONE 9 (10), e108243.

Poggiale, J.-C., Dantigny, P., De Wit, R., Steinberg, C., 2015. Environmental microbiology: Fundamentals and applications.

Polley, H., Wilsey, B., Derner, J., 2003. Do species evenness and plant density influence the magnitude of selection and complementarity effects in annual plant species mixtures? Ecology Letters 6 (3), 248–256.

Pollmächer, J., Figge, M., 2014. Agent-based model of human alveoli predicts chemotactic signaling by epithelial cells during early Aspergillus fumigatus infection. PLOS ONE 9 (10), e111630.

Ponomarova, O., Patil, K. R., 2015. Metabolic interactions in microbial communities: untangling the Gordian knot. Current Opinion in Microbiology 27, 37–44.

Poschet, F., Vereecken, K., Geeraerd, A., Nicolaï, B., Van Impe, J., 2005. Analysis of a novel class of predictive microbial growth models and application to coculture growth. International Journal of Food Microbiology 100 (1), 107–124.

Presley, S., Scheiner, S., Willig, M., 2014. Evaluation of an integrated framework for biodiversity with a new metric for functional dispersion. PLOS ONE 9, e105,818.

Prosser, J., Bohannan, B., Curtis, T., Ellis, R., Firestone, M., Freckleton, R., Green, J., Green, L., Killham, K., Lennon, J., et al., 2007. The role of ecological theory in microbial ecology. Nature Reviews Microbiology 5 (5), 384–392.

Quaghebeur, W., 2017. Individual-based modelling of biodiversity in microbial communities. Master's thesis, Ghent University.

Railsback, S., Grimm, V., 2012. Agent-based and Individual-based Modelling: a Practical Introduction. Princeton University Press.

Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., Medzhitov, R., 2004. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. Cell 118 (2), 229–241.

Rastall, R., Gibson, G., Gill, H., Guarner, F., Klaenhammer, T., Pot, B., Reid, G., Rowland, I., Sanders, M., 2005. Modulation of the microbial ecology of the human colon by probiotics, prebiotics and synbiotics to enhance human health: an overview of enabling science and potential applications. FEMS Microbiology Ecology 52 (2), 145–152.

Ratledge, C. (Ed.), 1993. Biochemistry of Microbial Degradation. Kluwer Academic Publishers.

Reed, J., Palsson, B., May 2003. Thirteen years of building constraint-based in silico models of Escherichia coli. Journal of Bacteriology 185 (9), 2692–9.

Reichenbach, T., Mobilia, M., Frey, E., 2006. Coexistence versus extinction in the stochastic cyclic Lotka–Volterra model. Physical Review E 74 (5), 051907.

Reichenbach, T., Mobilia, M., Frey, E., 2007. Mobility promotes and jeopardizes biodiversity in rock–paper–scissors games. Nature 448 (7157), 1046–1049.

Reichenbach, T., Mobilia, M., Frey, E., 2008. Self-organization of mobile populations in cyclic competition. Journal of Theoretical Biology 254 (2), 368–383.

Rényi, A., 1961. On measures of entropy and information. In: Fourth Berkeley Symposium on Mathematical Statistics and Probability. pp. 547–561.

Requejo, R., Camacho, J., Mar 2011. Evolution of cooperation mediated by limiting resources: connecting resource based models and evolutionary game theory. Journal of Theoretical Biology 272 (1), 35–41.

Requejo, R., Camacho, J., 2012. Analytical models for well-mixed populations of cooperators and defectors under limiting resources. Physical Review E 85 (6), 066112.

Requejo, R., Camacho, J., 2013. Scarcity may promote cooperation in populations of simple agents. Physical Review E 87 (2), 022819.

Reznick, D., Bryant, M., Bashey, F., 2002. r-and K-selection revisited: the role of population regulation in life-history evolution. Ecology 83 (6), 1509–1520.

Ricotta, C., 2003. On parametric evenness measures. Journal of Theoretical Biology 222 (2), 189–197.

Ricotta, C., 2004. A recipe for unconventional evenness measures. Acta Biotheoretica 52 (2), 95–104.

Ricotta, C., Szeidl, L., 2006. Towards a unifying approach to diversity measures: bridging the gap between the Shannon entropy and Rao's quadratic index. Theoretical Population Biology 70 (3), 237–243.

Ricotta, C., Szeidl, L., 2009. Diversity partitioning of Rao's quadratic entropy. Theoretical Population Biology 76 (4), 299–302.

Riolo, R., Cohen, M., Axelrod, R., 2001. Evolution of cooperation without reciprocity. Nature 414 (6862), 441–443.

Röling, W., Ferrer, M., Golyshin, P., 2010. Systems approaches to microbial communities and their functioning. Current Opinion in Biotechnology 21 (4), 532–538.

Rollié, S., Mangold, M., Sundmacher, K., 2012. Designing biological systems: Systems engineering meets synthetic biology. Chemical Engineering Science 69, 1–29.

Rousseau, R., 2011. Lorenz curves determine partial orders for comparing network structures. DESIDOC Journal of Library & Information Technology 31 (5).

Rousseau, R., Van Hecke, P., 1999. Measuring biodiversity. Acta Biotheoretica 47 (1), 1–5.

Roy, K., Kar, S., Das, R., 2015. Statistical methods in QSAR/QSPR. In: A primer on QSAR/QSPR modeling. Springer, pp. 37–59.

Sabra, W., Dietz, D., Tjahjasari, D., Zeng, A.-P., 2010. Biosystems analysis and engineering of microbial consortia for industrial biotechnology. Engineering in Life Sciences 10 (5), 407–421.

Saleem, M., Fetzer, I., Dormann, C., Harms, H., Chatzinotas, A., 2012. Supplmentary information for Predator richness increases the effect of prey diversity on prey yield. Nature Communications 3, 1305.

Sanderson, R., 1992. Diversity and evenness of Hemiptera communities on naturally vegetated derelict land in NW England. Ecography 15, 154–160.

Scheiner, S., 2012. A metric of biodiversity that integrates abundance, phylogeny, and function. Oikos 121, 1191–1202.

Schiffers, K., Travis, J., 2014. ALADYN — a spatially explicit, allelic model for simulating adaptive dynamics. Ecography 37 (12), 1288–1291.

Schloss, P., Handelsman, J., 2004. Status of the microbial census. Microbiology and Molecular Biology Reviews 68 (4), 686–691.

Schmidtke, A., Gaedke, U., Weithoff, G., 2010. A mechanistic basis for underyielding in phytoplankton communities. Ecology 91, 212–221.

Schreiber, S., Killingback, T., 2013. Spatial heterogeneity promotes coexistence of rock–paper–scissors metacommunities. Theoretical Population Biology 86, 1–11.

Seal, J., Alverdy, J., Zaborina, O., An, G., 2011. Agent-based dynamic knowledge representation of Pseudomonas aeruginosa virulence activation in the stressed gut: towards characterizing host-pathogen interactions in gut-derived sepsis. Theoretical Biology and Medical Modelling 8 (1), 33.

Sekhar, A., Horemans, B., Aamand, J., Sörensen, S., Vanhaecke, L., Vanden Bussche, J., Hofkens, J., Springael, D., 2016. Surface colonization and activity of the 2, 6-dichlorobenzamide (BAM) degrading aminobacter sp. strain MSH1 at macro- and micropollutant BAM concentrations. Environmental Science & Technology 50 (18), 10123–10133.

Seshan, H., Goyal, M., Falk, M., Wuertz, S., 2014. Support vector regression model of wastewater bioreactor performance using microbial community diversity indices: Effect of stress and bioaugmentation. Water Research 53, 282–296.

Shannon, C., 1948. A mathematical theory of communication. The Bell System Technical Journal 27, 379–423.

Shasha, D., Kouranov, A., Lejay, L., Chou, M., Coruzzi, G., 2001. Using combinatorial design to study regulation by multiple input signals. A tool for parsimony in the post-genomics era. Plant Physiology 127 (4), 1590–1594.

Shimatani, K., 2001. On the measurement of species diversity incorporating species differences. Oikos 93 (1), 135–147.

Shklarsh, A., Ariel, G., Schneidman, E., Ben-Jacob, E., 2011. Smart swarms of bacteria-inspired agents with performance adaptable interactions. PLOS Computational Biology 7 (9), e1002177.

Shurin, J., Allen, E., 2001. Effects of competition, predation, and dispersal on species richness at local and regional scales. The American Naturalist 158 (6), 624–637.

Sinclair, L., Osman, O., Bertilsson, S., Eiler, A., 2015. Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the Illumina platform. PLOS ONE 10 (2), e0116955.

Sinervo, B., Lively, C., 1996. The rock–paper–scissors game and the evolution of alternative male strategies. Nature 380 (6571), 240–243.

Smalla, K., Wachtendorf, U., Heuer, H., Liu, W., Forney, L., Apr 1998. Analysis of BIOLOG GN substrate utilization patterns by microbial communities. Applied and Environmental Microbiology 64 (4), 1220–1225.

Smith, B., Wilson, J., 1996. A consumer's guide to evenness indices. Oikos 76, 70–82.

Smith, M., Knapp, A., 2003. Dominant species maintain ecosystem function with non-random species loss. Ecology Letters 6 (6), 509–517.

Smith, V., Aug 2002. Effects of resource supplies on the structure and function of microbial communities. Antonie van Leeuwenhoek 81, 99–106.

Sobol, I., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation 55 (1), 271–280.

Song, H.-S., Cannon, W., Beliaev, A., Konopka, A., 2014. Mathematical modeling of microbial community dynamics: A methodological review. Processes 2 (4), 711–752.

Sørensen, S., Holtze, M., Simonsen, A., Aamand, J., 2007. Degradation and mineralization of nanomolar concentrations of the herbicide dichlobenil and its persistent metabolite 2, 6-dichlorobenzamide by Aminobacter spp. isolated from dichlobenil-treated soils. Applied and Environmental Microbiology 73 (2), 399–406.

Stefani, F., Bell, T., Marchand, C., Ivan, E., El Yassimi, A., St-Arnaud, M., Hijri, M., 2015. Culture-dependent and-independent methods capture different microbial community fractions in hydrocarbon-contaminated soils. PLOS ONE 10 (6), e0128272.

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

Stenuit, B., Agathos, S., 2015. Deciphering microbial community robustness through synthetic ecology and molecular systems synecology. Current Opinion in Biotechnology 33, 305–317.

Stewart, E., 2012. Growing unculturable bacteria. Journal of Bacteriology 194 (16), 4151–4160.

Stewart, J., Allison, P., Johnson, R., 2001. Putting a price on biotechnology. Nature Biotechnology 19 (9), 813.

Stirling, G., Wilsey, B., 2001. Empirical relationships between species richness, evenness, and proportional diversity. The American Naturalist 158 (3), 286–299.

Stone, L., Roberts, A., 1991. Conditions for a species to gain advantage from the presence of competitors. Ecology 72 (6), 1964–1972.

Streit, W., Daniel, R., Jaeger, K.-E., 2004. Prospecting for biocatalysts and drugs in the genomes of non-cultured microorganisms. Current Opinion in Biotechnology 15 (4), 285–290.

Szabó, G., Fath, G., 2007. Evolutionary games on graphs. Physics Reports 446 (4), 97–216.

Szabó, G., Sznaider, G., 2004. Phase transition and selection in a four-species cyclic predator–prey model. Physical Review E 69 (3), 031911.

Tainaka, K., 1993. Paradoxical effect in a three-candidate voter model. Physics Letters A 176 (5), 303–306.

Tan, J., Zuniga, C., Zengler, K., 2015. Unraveling interactions in microbial communities-from co-cultures to microbiomes. Journal of Microbiology 53 (5), 295–305.

Tanouchi, Y., Smith, R., You, L., Oct 2012. Engineering microbial systems to explore ecological and evolutionary dynamics. Current Opinion in Biotechnology 23 (5), 791–7.

Terry, J., Morris, R., Bonsall, M., 2017. Trophic interaction modifications: an empirical and theoretical framework. Ecology Letters 20 (10), 1219–1230.

Thompson, I., van der Gast, C., Ciric, L., Singer, A., 2005. Bioaugmentation for bioremediation: the challenge of strain selection. Environmental Microbiology 7 (7), 909–15.

Tilman, D., 1977. Resource competition between plankton algae: an experimental and theoretical approach. Ecology 58 (2), 338–348.

Tilman, D., Isbell, F., Cowles, J., 2014. Biodiversity and ecosystem functioning. Annual Review of Ecology, Evolution, and Systematics 45, 471–493.

Tolker-Nielsen, T., Molin, S., 2000. Spatial organization of microbial biofilm communities. Microbial Ecology 40 (2), 75–84.

Tukey, J., Jun 1949. Comparing individual means in the analysis of variance. Biometrics 5 (2), 99–114.

Tuomisto, H., 2011. Commentary: do we have a consistent terminology for species diversity? Yes, if we choose to use it. Oecologia 167 (4), 903–911.

Tuomisto, H., 2012. An updated consumer's guide to evenness and related indices. Oikos 121 (8), 1203–1218.

Turnbaugh, P., Hamady, M., Yatsunenko, T., Cantarel, B., Duncan, A., Ley, R., Sogin, M., Jones, W., Roe, B., Affourtit, J., et al., 2009. A core gut microbiome in obese and lean twins. Nature 457 (7228), 480–484.

Valm, A., Welch, J., Rieken, C., Hasegawa, Y., Sogin, M., Oldenbourg, R., Dewhirst, F., Borisy, G., 2011. Systems-level analysis of microbial community organization through combinatorial labeling and spectral imaging. Proceedings of the National Academy of Sciences 108 (10), 4152–4157.

van de Koppel, J., Rietkerk, M., Dankers, N., Herman, P., 2005. Scale-dependent feedback and regular spatial patterns in young mussel beds. The American Naturalist 165 (3), E66–E77.

van der Vaart, E., Beaumont, M., Johnston, A., Sibly, R., 2015. Calibration and evaluation of individual–based models using Approximate Bayesian Computation. Ecological Modelling 312, 182–190.

van Elsas, J., Chiurazzi, M., Mallon, C., Elhottovā, D., Krištůfek, V., Salles, J., 2012. Microbial diversity determines the invasion of soil by a bacterial pathogen. Proceedings of the National Academy of Sciences 109 (4), 1159–1164.

Van Elsas, J., Hill, P., Chroňáková, A., Grekova, M., Topalova, Y., Elhottová, D., Krištůfek, V., 2007. Survival of genetically marked Escherichia coli O157: H7 in soil as affected by soil microbial community shifts. The ISME Journal 1 (3), 204–214.

Van Loosdrecht, M., Heijnen, J., Eberl, H., Kreft, J., Picioreanu, C., 2002. Mathematical modelling of biofilm structures. Antonie van Leeuwenhoek 81 (1-4), 245–256.

Vandermeeren, P., Baken, S., Vanderstukken, R., Diels, J., Springael, D., 2016. Impact of dry-wet and freeze-thaw events on pesticide mineralizing populations and their activity in wetland ecosystems: A microcosm study. Chemosphere 146, 85–93.

Vane-Wright, R., Humphries, C., Williams, P., 1991. What to protect? – Systematics and the agony of choice. Biological Conservation 55 (3), 235–254.

Vannecke, T., 2015. Modelling of microbial populations in biofilm reactors for nitrogen removal from wastewater. Ph.D. thesis, Universiteit Gent.

Velicer, G., 2003. Social strife in the microbial world. Trends in Microbiology 11 (7), 330–337.

Veresoglou, S., Powell, J., Davison, J., Lekberg, Y., Rillig, M., 2014. The Leinster and Cobbold indices improve inferences about microbial diversity. Fungal Ecology 11, 1–7.

Verhulst, P.-F., 1845. Recherches mathématiques sur la loi d'accroissement de la population. Nouv. mém. de l'Academie Royale des Sci. et Belles-Lettres de Bruxelles 18, 1–41.

Villenave, C., Saj, S., Attard, E., Klumpp, K., Le Roux, X., 2011. Grassland management history affects the response of the nematode community to changes in above-ground grazing regime. Nematology 13 (8), 995–1008.

Vincke, S., Van de Vijver, B., Gremmen, N., Beyens, L., 2006. The moss dwelling testacean fauna of the Stromness Bay (South Georgia). Acta Protozoologica 45 (1), 65.

Vivant, A.-L., Garmyn, D., Maron, P.-A., Nowak, V., Piveteau, P., 2013. Microbial diversity and structure are drivers of the biological barrier effect against Listeria monocytogenes in soil. PLOS ONE 8 (10), e76991.

Volij, O., 2014. Segregation: theoretical approaches. Tech. rep., Beersheba, Israel.

Volterra, V., 1926. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. Mem. Acad. Lincei Roma 2, 31–113.

Wade, M., Harmand, J., Benyahia, B., Bouchez, T., Chaillou, S., Cloez, B., Godon, J.-J., Boudjemaa, B., Rapaport, A., Sari, T., Arditi, R., Lobry, C., 2016. Perspectives in mathematical modelling for microbial ecology. Ecological Modelling 321, 64–74.

Walker, A., Ince, J., Duncan, S., Webster, L., Holtrop, G., Ze, X., Brown, D., Stares, M., Scott, P., Bergerat, A., Louis, P., McIntosh, F., Johnstone, A., Lobley, G., Parkhill, J., Flint, H., 2011. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. The ISME Journal 5 (2), 220–230.

Wang, Q., Zhang, T., 2010. Review of mathematical models for biofilms. Solid State Communications 150 (21), 1009–1022.

Wanner, O., Gujer, W., 1986. A multispecies biofilm model. Biotechnology and Bioengineering 28, 314–328.

Warwick, R., Clarke, K., et al., 1995. New “biodiversity” measures reveal a decrease in taxonomic distinctness with increasing stress. Marine Ecology Progress Series 129 (1), 301–305.

Watve, M., Gangal, R., 1996. Problems in measuring bacterial diversity and a possible solution. Applied and Environmental Microbiology 62 (11), 4299–4301.

Weber, M., Poxleitner, G., Hebisch, E., Frey, E., Opitz, M., 2014. Chemical warfare and survival strategies in bacterial range expansions. Journal of The Royal Society Interface 11 (96), 20140172.

Weis, J., Cardinale, B., Forshay, K., Ives, A., 2007. Effects of species diversity on community biomass production change over the course of succession. Ecology 88 (4), 929–939.

Werner, F., Quinlan, J., Lough, R., Lynch, D., 2001. Spatially–explicit individual based modeling of marine populations: a review of the advances in the 1990’s. Sarsia 86 (6), 411–421.

Wessel, A., Hmelo, L., Parsek, M., Whiteley, M., May 2013. Going local: technologies for exploring bacterial microenvironments. Nature Reviews Microbiology 11 (5), 337–48.

Widder, S., Allen, R., Pfeiffer, T., Curtis, T., Wiuf, C., Sloan, W., Cordero, O., Brown, S., Momeni, B., Shou, W., Kettle, H., Flint, H., Haas, A., Laroche, B., Kreft, J.-U., Rainey, P., Freilich, S., Schuster, S., Milferstedt, K., van der Meer, J., Grosskopf, T., Huisman, J., Free, A., Picioreanu, C., Quince, C., Klapper, I., Labarthe, S., Smets, B., Wang, H., Soyer, O., 2016. Challenges in microbial ecology: building predictive understanding of community function and dynamics. The ISME Journal 10 (11), 1–12.

Wilcove, D., Rothstein, D., Dubow, J., Phillips, A., Losos, E., 1998. Quantifying threats to imperiled species in the United States. BioScience 48 (8), 607–615.

Wills, C., Harms, K., Condit, R., King, D., Thompson, J., He, F., Muller-Landau, H., Ashton, P., Losos, E., Comita, L., Hubbell, S., LaFrankie, J., Bunyavejchewin, S., Dattaraja, H., Davies, S., Esufali, S., Foster, R., Gunatilleke, N., Gunatilleke, S., Hall, P., Itoh, A., John, R., Kiratiprayoon, S., de Lao, S., Massa, M., Nath, C., Noor, N., Kassim, A., Sukumar, R., Suresh, H., Sun, I.-F., Tan, S., Yamakura, T., Zimmerman, J., 2006. Nonrandom processes maintain diversity in tropical forests. Science 311 (5760), 527–531.

Wilsey, B., Polley, H., 2004. Realistically low species evenness does not alter grassland species–richness–productivity relationships. Ecology 85 (10), 2693–2700.

Wilsey, B., Potvin, C., 2000. Biodiversity and ecosystem functioning: importance of species evenness in an old field. Ecology 81 (4), 887–892.

Wilsey, B., Teaschner, T., Daneshgar, P., Isbell, F., Polley, H., May 2009. Biodiversity maintenance mechanisms differ between native and novel exotic–dominated communities. Ecology Letters 12 (5), 432–42.

Wimpenny, J., 1999. Modelling in microbiology. In: Microbial Biosystems: New Frontiers (Proceedings of the 8th International Symposium on Microbial Ecology), Halifax, Canada. Atlantic Canada Society for Microbial Ecology.

Wimpenny, J., Colasanti, R., 1997. A unifying hypothesis for the structure of microbial biofilms based on cellular automaton models. FEMS Microbiology Ecology 22 (1), 1–16.

Wintermute, E., Silver, P., 2010. Emergent cooperation in microbial metabolism. Molecular Systems Biology 6 (1), 407.

Wittebolle, L., Marzorati, M., Clement, L., Balloi, A., Daffonchio, D., Heylen, K., De Vos, P., Verstraete, W., Boon, N., 2009. Initial community evenness favours functionality under selective stress. Nature 458 (7238), 623–626.

Wolfe, B., Dutton, R., 2015. Fermented foods as experimentally tractable microbial ecosystems. Cell 161 (1), 49–55.

Xavier, J., Picioreanu, C., Van Loosdrecht, M., 2005. A framework for multidimensional modelling of activity and structure of multispecies biofilms. Environmental Microbiology 7 (8), 1085–1103.

Yachi, S., Loreau, M., 1999. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. Proceedings of the National Academy of Sciences of the United States of America 96 (4), 1463–1468.

Yoshida, S., Ogawa, N., Fujii, T., Tsushima, S., 2009. Enhanced biofilm formation and 3-chlorobenzoate degrading activity by the bacterial consortium of Burkholderia sp. NK8 and Pseudomonas aeruginosa PAO1. Journal of Applied Microbiology 106 (3), 790–800.

Yu, Z., Krause, S., Beck, D., Chistoserdova, L., 2016. A synthetic ecology perspective: How well does behavior of model organisms in the laboratory predict microbial activities in natural habitats? Frontiers in Microbiology 7, 946.

Zhang, R., Weinbauer, M., Qian, P.-Y., 2007. Viruses and flagellates sustain apparent richness and reduce biomass accumulation of bacterioplankton in coastal marine waters. Environmental Microbiology 9 (12), 3008–3018.

Zhang, Y., Chen, H., Reich, P., 2012. Forest productivity increases with evenness, species richness and trait variation: a global meta-analysis. Journal of Ecology 100 (3), 742–749.

Zomorrodi, A., Segrè, D., 2016. Synthetic ecology of microbes: mathematical models and applications. Journal of Molecular Biology 428 (5), 837–861.

Zwietering, M., Jongenburger, I., Rombouts, F., Van't Riet, K., 1990. Modelling of the bacterial growth curve. Applied and Environmental Microbiology 56 (6), 1875–1881.

Appendix 1

Representative comparison of results using Gini, Simpson, and Shannon indices

To illustrate the similarities between results obtained with the Gini, Simpson, and Shannon indices, and motivate the inclusion in the main text of only those corresponding to the Gini index, we provide here a representative example of the comparison of these three indices. These figures relate to the model developed in Chapter 5 representing three species in cyclic competition, and show mean final community evenness as a function of initial community evenness.

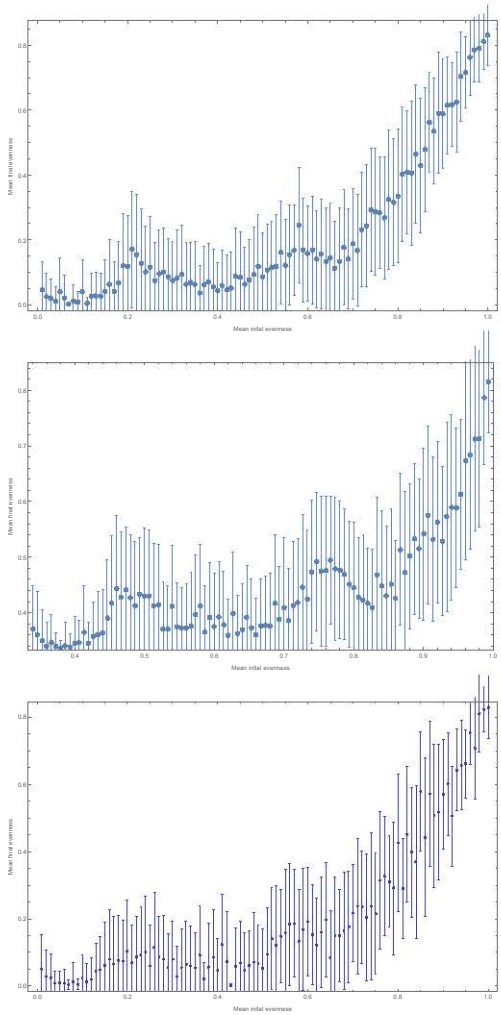


Figure 10.2: Comparison of results obtained using three evenness indices (from top to bottom: Gini, Simpson, Shannon). The results show mean final community evenness as a function of initial community evenness, after 500 generations. The model used for simulation is the three species cyclic competition model developed in Chapter 5.

Curriculum vitae

Personalia

Name	Aisling Jane Daly
Date of birth	05/06/1990
Place of birth	Tokyo, Japan
Nationality	Irish & Canadian
E-mail	Aisling.Daly@UGent.be

Education

University

2011 - 2012

M.Sc. Mathematical Sciences, University of Bath, Bath, United Kingdom

2008 - 2011

B.Sc. Mathematics, University of Warwick, Coventry, United Kingdom

Primary and secondary education

2002 - 2008

Secondary school, International School of Brussels, Brussels, Belgium

2000 - 2002

Primary school, International School of Brussels, Brussels, Belgium

1996 - 2000

Primary school, Rathmichael Parish School, Dublin, Ireland

Employment

June 2013 - present

Full-time researcher at the Research Unit Knowledge-Based Systems, Department of Mathematical Modelling, Statistics and Bioinformatics, Faculty of Bioscience Engineering, Ghent University

Tutorship of master and bachelor theses

Individual-based modelling of biodiversity in microbial communities. Master thesis, academic year 2016-2017 (Ward Quaghebeur).

Predicting coexistence in transitive graphs. Master thesis, academic year 2016-2017 (Zeno Van Moerkerke).

Transient coexistence in non-deterministic cyclic species competition. Master thesis, academic year 2014-2015 (Tim Depraetere).

Scientific output

Publications in international journals (ISI-papers)

A.J. Daly, J.M. Baetens, B. De Baets (2015). The impact of initial evenness on biodiversity maintenance for a four-species in silico bacterial community. Journal of Theoretical Biology 387, 189-205.

A.J. Daly, J.M. Baetens, B. De Baets (2016). *In silico* substrate dependence increases community productivity but threatens biodiversity. Physical Review E 93, 042414.

A.J. Daly, J.M. Baetens, B. De Baets (2016). The impact of resource dependence of the mechanisms of life on the spatial population dynamics of an *in silico* microbial community. Chaos 26, 123121.

Submitted manuscripts

A.J. Daly, J.M. Baetens, J. Vandermaesen, N. Boon, D. Springael and B. De Baets. Individual-based modelling of invasion in bioaugmented sand filter communities (submitted 2017). Processes (in revision)

Conference proceedings

J. Vandermaesen, A.J. Daly, J.M. Baetens, B. De Baets, N. Boon, D. Springael (2015). Unravelling the relationship between indigenous community diversity and success of bioaugmentation using synthetic microbial ecosystems. 13th Symposium on Bacterial Genetics and Ecology, June 2015, Milan, Italy.

Conference Abstracts

A.J. Daly, J.M. Baetens, B. De Baets (2017). The impact of resource dependence on the spatial population dynamics of an in silico microbial community. Mathematical Models in Ecology and Evolution, July 2017, London, United Kingdom.

A.J. Daly, J.M. Baetens, B. De Baets (2016). In silico substrate dependence increases community productivity, threatens biodiversity. 10th European Conference on Mathematical and Theoretical Biology, July 2016, Nottingham, United Kingdom.

A.J. Daly, J.M. Baetens, B. De Baets (2016). The impact of initial evenness on competitive dynamics: an individual-based approach. Mathematical Models in Ecology and Evolution, July 2015, Paris, France.

A.J. Daly, J.M. Baetens, B. De Baets (2015). Individual-based modelling of the impact of initial evenness on biodiversity preservation for in silico bacterial communities. 6th Congress of European Microbiologists, June 2015, Maastricht, The Netherlands.

A.J. Daly, J.M. Baetens, B. De Baets (2015). The impact of initial evenness on the preservation of biodiversity in bacterial communities: a stochastic, spatial individual-based model. 20th National Symposium on Applied Biological Sciences, January 2015, Louvain-la-Neuve, Belgium.

A.J. Daly, J.M. Baetens, B. De Baets (2014). A stochastic spatial individual-based model for three competitively interacting microbial populations. 19th National Symposium on Applied Biological Sciences, February 2014, Gembloux, Belgium.

Active participation at international scientific events

2nd International Symposium on Microbial Resource Management, September 2017, Ghent University, Ghent, Belgium.

Mathematical Models in Ecology and Evolution, July 2017, City, University of London, London, United Kingdom.

10th European Conference on Mathematical and Theoretical Biology, July 2016, University of Nottingham, Nottingham, United Kingdom.

Mathematical Models in Ecology and Evolution, July 2015, Collège de France, Paris, France.

6th Congress of European Microbiologists, June 2015, Maastricht, the Netherlands.

Sustainability and Complex Systems workshop, July 2013, Mathematical Biosciences Institute, Ohio State University.

Memberships of international conference organizing committees

2nd International Symposium on Microbial Resource Management, September 2017, Ghent University, Ghent, Belgium.