TOWARDS BIOENGINEERED ECOSYSTEMS: 
THREE STUDIES IN INVASION BIOLOGY

BY
NICOLAS HAWLEY-WELD

THESIS
Submitted in partial fulfillment of the requirements 
for the degree of Master of Science in Agricultural and Biological Engineering 
in the Graduate College of the 
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Master’s Committee:

Associate Professor Kaustubh Bhalerao, Chair
Research Assistant Professor Seppe Kuehn
Associate Professor Michael Miller
Abstract

An important feature of ecosystems is their invasibility, i.e. their resistance to invasion by nonnative organisms. Invasion is a driver of change at scales as small as a bioreactor and as large as the scale of biogeography. Invasion can be either a desired outcome, for example in the use of probiotics to replace healthful gut microflora after a disturbance to the gut microbial ecosystem, or an undesired outcome, for example in the contamination of open culture systems such as algal raceway ponds. Biological invasion theory attempts to predict what makes a community vulnerable to invasion, what characteristics make an invader successful, and what consequences following invasion might arise. In this thesis, a review of recent literature on microbial invasion biology was conducted, and three separate studies in invasion biology were pursued.

In Chapter 1, two overarching yet poorly understood factors determining microbial community invasibility were identified: environmental modulation, and community structure. The first factor, community structure, describes the set of species in an ecosystem and how they are trophically and otherwise related. While the majority of work in invasion biology has focused on the relationship between diversity and invasibility, we argue that it is imperative to move beyond simple diversity metrics towards more information-rich descriptors of community structure, by casting traditional ecological concepts like plasticity, redundancy, dormancy, and diversity into the language of networks. The second factor, environmental modulation, describes how communities can alter their own environment, deterring potential invaders by making the environment less consumable or habitable. We argue that this process has received too little attention in invasion biology and that insights can be drawn from the field of biopreservation, which utilizes beneficial microorganisms to preserve food environments from unwanted invaders.

Chapter 2 introduces wild sourdough fermentations as a model system for the study of invasion biology through the lens of biopreservation. An experimental investigation of the robustness of traditional sourdough cultures to invasion was performed, exploring the role of environmental modulation of pH in deterring invaders. Two invader organisms were tested for their ability to invade, and a mathematical model was developed describing the bacterial fraction of a wheat flour dough. First, a simple experiment with a laboratory strain
of *Escherichia coli* introduced into a traditionally propagated sourdough culture indicated that the production of organic acids is the primary mechanism of invasion resistance of sourdoughs against *E. coli*, which is ubiquitous in human environments. Second, experiments with the acidophile *Alicyclobacillus acidoterrestris* showed that while acid-tolerant invaders exist, the presence of environmental hurdles such as low temperature and high osmolarity are major factors in preventing their establishment in sourdough cultures. Finally, a mathematical model for the growth, metabolic output, and self-inhibition of the dominant sourdough organism *Lactobacillus sanfranciscensis* in wheat flour dough was developed. This model was able to predict population density, pH, lactate concentration, and maltose concentration in a manner consistent with experimental data.

Chapter 3 considers the general problem of invasion in a three-species ecosystem, which was simulated mathematically assuming that the community structure of the ecosystem is completely defined. The aim of this chapter was to determine any relationships between the invasibility of a particular ecosystem and its community structure. We considered the scenario in which a rare species in an ecosystem suddenly becomes abundant as a result of a discrete mutation in one or more ecological parameters describing the system, representing a sudden change in ecological strategy, an actual mutation, or a sudden change in the environment. The concepts of *invasion distance* and *invasion direction* were formulated, the former referring to the magnitude of the change in parameters required for a rare species to become abundant, and the latter referring to the direction of that parameter change. By performing simulations for 44 out of the 138 possible community structures in a simple 3-species ecosystem, we found that the invasion direction varied significantly with community structure but the invasion distance did not. This result set up future work investigating analytically how invasion direction varies with community structure.

Chapter 4 introduces a mathematical model for a phage-bacterium ecosystem. From an invasion biology standpoint, either the phage or the bacterium can be considered as an invader, depending on the ecosystem they inhabit and also on one’s point of view. A model was developed to describe an ecosystem containing three populations: phage-infected bacteria, uninfected bacteria, and an environmental reservoir of free bacteriophages. Five system parameters were used as inputs to the model: intrinsic host death rate, virus degradation rate, environmental transmission rate, lysis rate, and burst size. The model was set up to capture logarithmic growth of bacteria, as well as both lytic and lysogenic life stages of the phage. All pairwise relationships between the five system parameters were investigated, by examining two payoff functions plotted as contour surfaces over each pair of parameters. We found that a high burst size, low virus degradation rate, and low host intrinsic death rate always resulted in the highest payoff to the phage. In contrast, payoffs always reached a maximum when lysis rate and environmental transmission rate took
on intermediate values. Moreover, a third payoff, the basic epidemiological reproduction number $R_0$, was shown to be limited in expressivity as it could only capture monotonic payoff relationships. These results were relevant to the classic "trade-off" hypothesis in virulence theory, which states that lysis rate (virulence) and transmission rate are coupled. While no coupling relationship was assumed in our model, the relationship between lysis rate and transmission rate was markedly the most complex out of all pairwise relationships considered, thus validating the substantial amount of interest historically in this relationship as opposed to other relationships between system parameters. Finally, conditions were found for a stable steady-state solution representing a desirable outcome in a phage therapy context, which attempts to control unwanted bacterial populations by using bacteriophages as therapeutic agents. These conditions were overlayed upon the phage payoff surfaces, and it was found that the phage therapy stability region was always located at higher lysis rate and host intrinsic death rate values than the payoff optima.

Taken together, the work here contributes to the fields of biological and ecological engineering in the following ways. First, the role of ecosystem structure and environmental modulation were elucidated, and gaps in the literature in both of these areas were identified. Principles were then abstracted from both physical (e.g. sourdough fermentation) and simulated systems (e.g. phage-bacterium interaction) towards a greater understanding of robustness towards invasion and evolutionary trade-off relationships, respectively. Finally, several future directions for research were identified, including the generation of mathematical models to describe yeast-bacterium interactions in sourdoughs, further work on the effect of community structure on invasion direction, and additional experimental and theoretical work towards the realization of phage therapy.
Dedicated to Thomas K. Hawley

"If such be the case, the wonderful noonday silence of a tropical forest is, after all, due only to the
dullness of our hearing; and could our ears catch the murmur of these tiny maelstroms, as they whirl
in the innumerable myriads of living cells which constitute each tree, we should be stunned, as with
the roar of a great city." —T.H. Huxley
Acknowledgments

Before embarking upon my journey at Illinois, a mentor wrote to me that my time here would be a great one of both scientific and self-discovery. Two years later, I couldn’t put it in better words myself. Barely knowing anyone in the state of Illinois, I arrived on campus and immediately had a home in the Bhalerao lab. I want to thank, firstly, Dr. Bhalerao for creating a space that merged three things: a deep respect for biology, a testbed for creating electronics, hardware, and algorithms, and an ensemble of amazing people. Few research groups are able to do work that is both cutting edge and has potential for immediate commercial impact; this is especially true in the life sciences. But both of these were the bread and butter of my time here. While my own thesis is abstract in nature, it was the exposure to the people around me and their pursuits of truth, beauty, and impact that defined my experience.

At the granular level, I have to thank the people who helped teach me how to stay afloat in a microbiology lab. Sadia Bekal, Safyre Anderson, Derek Jacoby, Christine Xiaoji Liu, and Rekha Balachandran: without you all, I’d still be banging my head on the door. SeokChan Yoo, you were my partner for two years and everything I learned was as a student alongside you. Abhishek Dhoble, Pratik Lahiri, Chinmay Soman, and Farhan Abbas, you provided friendship and support and most importantly were sparring partners when I needed it.

I am indebted to the College of Aces, the Department of Agricultural and Biological Engineering, and the Department of Mathematics for supporting me financially, and to Dr. Yanhui Zhang to making sure I came to Illinois. My research would not have been possible without the support of the Jin Lab and their people and facilities.

To my committee members Dr. Seppe Kuehn and Dr. Mike Miller, you provided much-needed guidance, know-how, and inspiration during the home stretch. Every conversation with each of you forced me to grow in new ways, the best I could have hoped for.

My friends at Illinois, in particular Macky, Zenobia, Kaneeka, Bruce, and Sid, you taught me what matters in life, and I am proud to have shared the realest moments with you. Martina and Tim, there’s no such thing as better people to come home to. Sophie, mom, and dad, you guys have been there every step of the way.

Lastly, Dr. Bruce Vojak, thank you for giving me a roadmap for my next chapter, and
thank you to everyone else who’s inspired me and corrected my course along the way. Most of all, Dr. Bhalerao, none of this would have been possible without the amazing space for innovation and discovery that you create.
# Table of Contents

List of Tables .............................................................. x
List of Figures .............................................................. xi
List of Abbreviations ....................................................... xiii

Chapter 1  Microbial Invasion Biology: A Review .......................... 1
  1.1 Abstract ........................................................................ 1
  1.2 Introduction .................................................................... 2
  1.3 Defining Robustness ..................................................... 2
  1.4 Invisible Invaders: A New Frontier in Invasion Biology ........ 3
  1.5 Properties of Successful Invaders ................................. 4
  1.6 Community and Environmental Aspects of Invasion Resistance 5
  1.7 Fridley’s Synthesis ....................................................... 10
  1.8 Implications for Microbiology ....................................... 11
  1.9 Aims of This Thesis .................................................... 13
  1.10 References ............................................................... 14

Chapter 2  Sourdough: A Model System for Invasion Biology ......... 20
  2.1 Abstract ........................................................................ 20
  2.2 Background .................................................................... 20
  2.3 Objectives ...................................................................... 22
  2.4 Sourdough Biochemistry and Microbiology .................... 23
    2.4.1 Types of Sourdoughs .............................................. 23
    2.4.2 Composition of Flour ............................................ 23
    2.4.3 Stages of Ecological Succession ............................. 24
    2.4.4 Sourdough Microbiota ........................................... 25
  2.5 Investigation of Sourdough Invasion Resistance ............... 27
    2.5.1 Methods ............................................................... 27
    2.5.2 Results ................................................................. 30
    2.5.3 Discussion ............................................................ 39
  2.6 Modeling Growth of *L. sanfranciscensis* in Wheat Flour Sourdough 39
    2.6.1 Methods ............................................................... 39
    2.6.2 Results and Discussion ......................................... 44
  2.7 Future Work ............................................................... 45
  2.8 References ............................................................... 54
# Chapter 3  Invasion By a Rare Species in a Three-Species Ecosystem

- 3.1 Abstract
- 3.2 Background and Problem Statement
- 3.3 System overview
- 3.4 Classification Task
- 3.5 Method and Results
- 3.6 Discussion and Future Work
- 3.7 References

# Chapter 4  The Emergence of Trade-Offs in a Phage-Bacterium System

- 4.1 Abstract
- 4.2 Background
- 4.3 Objectives
- 4.4 Methods
  - 4.4.1 Development of Mathematical Model
  - 4.4.2 Implementation of Payoff Functions
- 4.5 Results
- 4.6 Discussion and Further Work
- 4.7 References

# Appendix A  Additional Sourdough Results

- Growth of *L. sanfranciscensis* on Different Sugars
- Volumetric Rise of *S. cerevisiae*

# Appendix B  Additional Phage Results

# Appendix C  Next Generation Matrix Methods

- Derivation of $R_0$ and $T$
- Derivation for a Phage-Bacterium Model
- References
List of Tables

2.1 Chemical Composition of Flour .................................................. 23
2.2 Formulations of LB and YSG Media ................................. 27
2.3 Composition of Invasion Cultures ............................................. 32
2.4 Growth Observed in Cultures C1–C8 and D1–D6 ............. 36
2.5 sMRS Media Ingredients ......................................................... 40
2.6 Dough Recipes ................................................................. 42
2.7 L. sanfranciscensis Model Parameters and Variables ....... 44

4.1 Parameters and State Variables in the Phage-Bacterium Model .... 79
4.2 Parameter Ranges in the Phage-Bacterium Model ............. 82
4.3 Relationship between Payoff Functions and System Parameters .... 93

B.1 Relationship between Payoffs (P_{Inf1}, P_{Inf2}) and System Parameters .... 102
B.2 Relationship between Payoffs (P_{Susc1}, P_{Susc2}) and System Parameters .... 103

C.1 Placement of Terms into Matrices $\mathbf{F}$ and $\mathbf{V}$ ............ 106
List of Figures

2.1 Growth of E. Coli in Sourdough ............................................. 34
2.2 Growth of A. acidoterrestris in YSG with Added Sourdough or Flour .... 35
2.3 Optical Density of YSG Broth With Varying Amounts of Flour Added .... 35
2.4 Growth of A. acidoterrestris in YSG With Flour Added During Logarithmic Phase ................................................................. 37
2.5 Histogram Surface of Optimum pH and Temperature Range for 28 Acidophile Bacteria Species .................................................. 38
2.6 Sourdough Starter Microbes Photographed ........................................ 46
2.7 Volumetric Rise and pH Change in Three Sourdoughs ......................... 47
2.8 HPLC Readings of Three Sourdoughs ........................................... 48
2.9 HPLC Readings of Pure Doughs, Without Sourdough Starter Added .......... 49
2.10 OD–CFU Standard Curve for L. sanfranciscensis .................................... 49
2.11 pH Profile of Doughs with L. sanfranciscensis Added ........................ 50
2.12 HPLC Readings of L. sanfranciscensis in Dough ................................. 51
2.13 Data on L. sanfranciscensis for Parameter Estimation .......................... 52
2.14 Model of Population Growth, pH, Lactate Production, and Maltose Consumption ................................................................. 53
3.1 First 15 Ecosystem Topologies ...................................................... 63
3.2 2-Nodes in Topology 8 .................................................................... 66
3.3 Variance in Invasion Distance Within and Between Topologies ............. 67
3.4 Relationship between Invasion Distance and r₁ for Topology 6 .............. 68
3.5 Angular Histogram of Invasion Direction for Topologies 6 and 8 .......... 68
3.6 Correspondence of Gradient Field With Invasion Direction .................. 71
4.1 Invariance of Payoff Surfaces ....................................................... 84
4.2 Payoffs R₀ ............................................................................... 85
4.3 Payoffs P₁ and P₂ .............................................................. 88
A.1 Growth of *L. sanfranciscensis* on Different Sugars .......................... 99
A.2 Comparison of Leavening Activity of Baker’s Yeast to Wild Sourdough Yeasts100
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFU</td>
<td>Colony Forming Unit</td>
</tr>
<tr>
<td>DFE</td>
<td>Disease-Free Equilibrium</td>
</tr>
<tr>
<td>FMT</td>
<td>Fecal Microbiota Transplantation</td>
</tr>
<tr>
<td>HPLC</td>
<td>High Performance Liquid Chromatography</td>
</tr>
<tr>
<td>LAB</td>
<td>Lactic Acid Bacteria</td>
</tr>
<tr>
<td>LA</td>
<td>Lysogeny Agar Media (also referred to as LB agar)</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny Broth Media</td>
</tr>
<tr>
<td>LV</td>
<td>Lotka-Volterra System</td>
</tr>
<tr>
<td>MIC</td>
<td>Minimum Inhibitory Concentration</td>
</tr>
<tr>
<td>OD&lt;sub&gt;600nm&lt;/sub&gt;</td>
<td>Measurement of Absorbance at 600 nm</td>
</tr>
<tr>
<td>ODE</td>
<td>Ordinary Differential Equation</td>
</tr>
<tr>
<td>sMRS</td>
<td>Simplified De Man, Rogosa and Sharpe Media</td>
</tr>
<tr>
<td>YSG</td>
<td>Yeast Starch Glucose Media</td>
</tr>
</tbody>
</table>
Chapter 1

Microbial Invasion Biology: A Review

1.1 Abstract

Microbial ecosystems are ubiquitous in the natural environment and provide important functionality to both engineered and unengineered systems, from digestion to wastewater treatment to maintenance of soil health. An important feature of ecosystems is their invasibility, i.e. their resistance to invasion by nonnative organisms. Invasion is a driver of change at scales as small as a bioreactor and as large as the scale of biogeography. Invasion can be either a desired outcome, for example in the use of probiotics to replace healthful gut microflora after a disturbance to the gut microbial ecosystem, or an undesired outcome, for example in the contamination of open culture systems such as algal raceway ponds. Biological invasion theory attempts to predict what makes a community vulnerable to invasion, what characteristics make an invader successful, and what consequences following invasion might arise. Although there exists a long tradition of research on invasion by macroorganisms, the study of invasion by microorganisms, especially free-living nonpathogenic microorganisms, has until very recently received comparatively little attention. This literature review gives an overview of recent research on invasion in microbial communities and identifies that microbial community structure and environmental modulation are two overarching yet poorly understood factors determining microbial community invasibility, both of which can confer either an increase or decrease in community invasibility depending on the conditions. This review suggests that casting traditional ecological concepts like diversity, dormancy, redundancy, and plasticity into the language of graph theory will be a fruitful approach to understanding community invasibility and that the influence of environmental modulation in particular can be better understood by drawing insight from the field of biopreservation.
1.2 Introduction

The invasion of microbial ecosystems by non-native microbiota is of wide interest in ecology and engineering. Microbial invasions pervade much of epidemiology, and the ability of an invader to establish itself in an environment—or not—has repercussions at scales as small as a bioreactor and as large as biogeography [1]. From an engineering standpoint, successful invasion can be either a desired outcome, as with gut probiotics or the introduction of soil bacteria and fungi for suppressing plant disease [2], or it can be an undesired outcome, e.g. in the contamination of algal raceway ponds or other open culture systems [3]. Thus, the invasibility of engineered microbial communities is an important class of biological robustness. Numerous microbial ecosystems are naturally robust towards invasion, such as those in the digestive and vaginal tracts [2] as well as those in lactic acid fermented foods [4]. Yet the mechanisms behind this robustness are in general poorly understood, and, although it has been suggested that microbial consortia should be more robust than microbial monocultures [3, 5, 6], the prospect of designing and synthesizing consortia for robustness has remained an open frontier in synthetic biology [5–7].

1.3 Defining Robustness

Robustness, in general, is what allows a system to maintain relevant aspects of its structure and function despite a perturbation either to the system’s components or to the environment [8, 9]. In the context of a microbial ecosystem, component-wise perturbation can include, for example, transcriptional or translational errors or genetic change as a result of mutation or horizontal gene transfer, whereas environmental perturbation includes changes in temperature, pH, nutrient or metabolite levels, and physical features, not to mention the introduction or knockout of particular organisms, viruses and other entities in the environment. Besides its spatial structure, the major structural feature of a microbial ecosystem is its community composition, which describes the network of species and/or genomes in the community and their interactions. In this review, "function" is defined is biochemical function, i.e. which inputs are converted metabolically to which outputs by the entire system [10, 11]. A microbial ecosystem, then, can be robust both functionally and structurally, and it has been shown experimentally that these two types of robustness are not necessarily correlated [11]. Finally, to quantify these ideas, it is useful to use the concepts of resistance (insensitivity to perturbation) and resilience (rate of recovery after perturbation) as commonly used in ecology [12, 13].

There are four possible outcomes following a microbial invasion [14], which are here presented in adapted form. The first, community augmentation, occurs when the invader
is able to establish itself beyond a certain population threshold but has a negligible effect on overall ecosystem structure and function. Second, community alteration, occurs when the invader both exceeds a certain population threshold and has an appreciable effect on overall ecosystem structure or function. Third, rejection failure, occurs when the invader neither exceeds a certain population threshold (i.e. is rejected) nor has an appreciable effect on ecosystem structure or function. Finally, indirect failure, occurs when the invader fails to exceed a certain population threshold but nevertheless exerts appreciable change (e.g. through production of metabolites) on ecosystem structure or function. These four possible outcomes were originally presented in the context of a computer-simulated invasion experiment [14], in which the only effect an invader can have on a community is the extinction of one of the original community residents. Here, the more general case is considered, in which the invader can affect community structure and function in any reasonably feasible way, e.g. by production of antibiotics and growth factors, depletion of co-substrates, horizontal gene transfer, or influence on viral load. Therefore to characterize a microbial ecosystem’s invasibility, one must consider not only its invasion resistance (defined simply as its ability to resist the establishment of invader organisms beyond a certain population threshold [14–16]), but also its ability to maintain ecosystem structure and function despite the occurrence of invasion.

1.4 Invisible Invaders: A New Frontier in Invasion Biology

Biological invasion theory attempts to predict, across many scales and environments, what properties make a successful invader and conversely what properties make a community easy or difficult to invade. Until recently the majority of experimental, observational, and theoretical work on biological invasions has focused on invasions by macroorganisms such as plants and animals, which are important drivers, for example, of ecological succession and species distributions in island and other geographies [1]. Yet invasions by microorganisms are far more numerous by their sheer number and have wide repercussions in medicine, agriculture, and biosecurity. Furthermore, it is recognized that entirely different ecological principles may apply to microscopic environments and communities, compared to their macroscopic counterparts [10,17]. For example, the spatial structure of an environment can have greatly different effects on a community depending on its spatial scale [17–20]; moreover, there has been difficulty establishing concepts such as "nativeness" versus "non-nativeness" in the microbial world, not to mention the basic difficulty of applying the species concept to certain microbial taxa [1]. Existing work has treated the subject of invasion by pathogenic microorganisms quite extensively, but almost by definition this usually concerns the relation between an invasive pathogen and some multicellular host,
rather than the relationship between microbe and microbial community. Nonpathogenic microbes are in general harder to detect, but as suggested by Lichtman [1], an understanding of their invasiveness could contribute greatly to both to microbial population ecology and to an integrated understanding of biological invasion at all scales.

1.5 Properties of Successful Invaders

There are certain traits characteristic of successful invaders which apply equally to micro- and macroscopic organisms. Lichtman [1] provides a summary of these traits, as well as a shorter list of traits hypothesized to be operant for microorganisms in particular. From the literature on macroorganisms, traits that make an invader invasive include high growth rate and high resource use efficiency, high dispersal capability (both vector-assisted and unassisted), ability to enter into dormancy (for example to survive strongly nutrient-limited environments), high phenotypic plasticity, high toxicity, and low specificity for host environment [1]. Several additions to this list have been proposed for microorganisms, including large genome size and high frequency of horizontal gene transfer (for adapting to new environments different from the invader’s previous environment), as well as the existence of highly specific metabolic pathways (to utilize parts of the environment that are unconsumable by native species) [1]. Moreover, microbial invaders can be incredibly diverse across life domains, including archaea, bacteria, protists, and fungi (viruses are also powerful invaders and can have considerable impact on their hosts [21], but despite their potency are not considered in this review). Finally, in the microbial world it is rare for one single species of microbe to arrive in a new environment completely alone; much more likely is that an invader arrives as part of a small invading community that reaches its new environment as a single inoculum [22]. In light of this it is worth exploring whether a small community of invaders can in general be more successful than an invasion by a single species.

The competitive exclusion principle [23] is a well-established tenet in ecology that gives additional insight into what properties might make a successful invader. The principle states that two species, occupying the same geographical area and competing for the same resources using identically similar resource utilization strategies, cannot coexist. The "weak form" of the principle postulates that if the ecological strategies of two species are identical, then whichever species has a greater reproduction and/or resource utilization rate will run the other species out of existence. However in practice it is always possible to use the "strong form" of the principle, which states that in the real world no two reproduction rates are ever precisely identical; therefore, one of the two identically competing species will always either evolve to adopt a different ecological strategy ("niche partitioning"), or become
extinct [23]. In the context of invasion biology, it has been popularly suggested [3, 24] that, as a corollary, invaders with a similar resource utilization strategy as native community members should be expected to be less successful, in general, than invaders with a different ecological strategy. An example of this effect in the literature is the competitive exclusion of toxigenic, invasive strains of *Escherichia coli* by non-toxigenic *E. coli* strains already present in the gut [25]. A similar effect has been proposed to be operant in intestinal infections by *Clostridium difficile* [26]. A recent meta-analysis on macroorganisms, however, has shown that the competitive exclusion effect, instead of preventing invasion entirely, in practice simply limits the population levels of an invader once it has initially established itself in an environment [27]. Furthermore, while the competitive exclusion effect can indeed be strong between certain individuals in the native community and certain invaders [24], greater insight is needed to fully understand whether there are additional factors that enable invasion resistance in the community as a whole.

One other concept that has been developed to address the question of invasibility is the "invasional meltdown hypothesis" [28]. Simply put, this states that the establishment of an initial invader in an environment can have a positive feedback effect with secondary (i.e. later-arriving) invaders, hence facilitating a secondary invasion [28]. Green [28] points out that although the idea is oft cited, there has been relatively little supporting empirical evidence; furthermore, it is likely that invader-invader interactions are highly specific to the environment and species in question. Nevertheless Green [28] gives an excellent empirical example showing how the principle can work in practice. The yellow crazy ant *Anoplolepis gracilipes*, an established invader on several islands in the Indian Ocean, has dramatically reconfigured the islands’ native ecology by strongly suppressing population levels of red land crab *Gecarcinidae natalis*, killing them (for food) by spraying formic acid in their eyes and mouthparts. This in turn has enabled the invasion and growth of the giant African land snail *Achatina fulica*, due to reduced numbers of *G. natalis*, a keystone species and a natural predator of *A. fulica* [28, 29]. Although one cannot directly extrapolate to make conclusions about dynamics of other systems, the invasional meltdown hypothesis would support the idea that a community of microbial invaders may in general be more successful than invasion by a single microbe species.

### 1.6 Community and Environmental Aspects of Invasion Resistance

The question of what characteristics make a community resistant to invasion, specifically in the context of microbial communities, is substantially more complex the question of what makes a successful invader. Based on studies on macroorganisms, a number of charac-
characteristics have been proposed as contributors to community invasion resistance, including high diversity (functional, genetic, and taxonomic), high isolation (e.g. island ecosystems), high or fluctuating resource supply, absence of predators and pathogens, and high level of disturbance or perturbation [1]. Note that these can be classified either as characteristics of the community or characteristics of the environment, which in ecology are termed as biotic and abiotic factors, respectively. For microorganisms in particular, three major ideas have been proposed: that microbial consortia are more robust than single-species cultures [3,5,6], that disturbed (e.g. exposed to high temperatures, antibiotics, or host diet change) cultures are more easily invasible, and that functional diversity confers invasion resistance. These three ideas are related and will be explored in the following sections.

The notion that microbial consortia are more robust than monocultures has received wide support by proponents of synthetic biology [3,5,6], a field which began with the design of simple programmable genetic circuits but is now beginning to integrate these modules into higher order functional systems such as synthetic tissues and ecosystems [30,31]. A major issue with engineering "designer organisms" is that no organism is insulated from the processes of ecology and evolution; therefore it is highly possible, in practice, for the engineered entity to evolve away from its intended design or underperform due to interactions with co-inhabitant species that find their way into the system [32,33]. As a result it has been proposed, first, that a deeper understanding of both evolution and ecology will be critical to synthetic biology’s future success [33], and second, that research on microbial consortia will be crucial for the discovery and application of evolutionary and ecological principles to biologically engineered systems [3,5,6,34]. Although the majority of microbial production systems used in the biotechnology industry today are pure single-species cultures, the design of mixed culture systems is a recognized and promising new direction for biotechnology as well as a direct analog of how microbes are organized in the natural world [35,36].

Consortia, first of all, allow for greater functional complexity than single-species systems [6,18]. The classic example of this is breakdown and conversion of lignocellulose to value-added products, which is a two-stage process for which no single species has yet been developed to perform both stages at high yields [37–39]. The first stage, breakdown of lignocellulose to soluble saccharides, is performed far more effectively by fungi such as *Trichoderma reesei*, a natural degrader of woody biomass [38], whereas the second stage, conversion of soluble saccharides to ethanol, isobutanol, and other fuels or bioproducts, is performed more effectively by easily programmable bacteria such as *E. coli*. As a result, several recent bioprocessing efforts have focused on the creation of synthetic co-cultures which functionally compartmentalize the hydrolysis and fermentation steps into two or more separate species [37–39].
The second positive engineering aspect of consortia is their robustness to environmental fluctuations [6, 18]. It has been shown experimentally that, for example, in the bio-production of polyhydroxyalkanoates from crude glycerol and other inputs, mixed bacterial cultures can use lower quality substrates more efficiently and can cope better with fluctuations in substrate composition, compared to pure single-species cultures [40]. Furthermore there is evidence in other systems that by demonstrating greater functional plasticity than single-species cultures, consortia can more fully utilize the resources available in a fluctuating or seasonally varying environment [17].

There is evidence, additionally, that consortia are more robust against invasion. The mechanisms behind this can be non-obvious, as in the case of biofilms, which may be able to generate a more viscous, impermeable extracellular matrix by utilizing a variety of matrix polymers from multiple species rather than from a single species alone [41]. Yet in general the competitive exclusion principle provides a catch-all argument for the invasion resistance of consortia, namely that consortia are statistically more likely to demonstrate a high diversity of resource-utilization strategies, thereby enabling them to compete via competitive exclusion with a wider array of potential invaders. This is the classic form of the "biotic resistance" argument (which in general can refer to any mechanism through which a community deters invaders through biological, not environmental means) as proposed by Elton in 1958 [17, 42].

Another important ecological process, described in the theory of ecological succession, shows intuitively why naturally occurring microbial consortia should in general be expected to be more robust to invasion than pure single-species cultures [7]. Put simply, ecological succession theory predicts that regardless of how a habitable environment starts out, it will be colonized in successive waves by members of new speciation and invasion events as the physical environment changes over time [7, 43]. Ecological succession in general is understood to be directional, that is, it drives community structure towards an evolutionarily stable state, regardless of whether this stability is expressed using the somewhat outdated concept of a "climax community" [7, 44] or in newer terms describing the balance between stochastic speciation, invasion, and extinction events. Succession theory in particular predicts the outcomes of changes to the physical environment that are caused the community itself, such as changes in shade, impact on aquifers, or buildup of detritus. These "autogenic" environmental changes make the environment either more or less habitable to the native community [43], thereby acting either as positive feedback that makes the community harder to displace, or opening the door for alteration of community structure via new speciation and invasion events. Over evolutionary time, then, this would suggest (as has been proposed by Kazamia [7] in the context of algal cultures) that microbial communities become less and less invasible as the most successful invaders become "native"
members of the community and the least successful community members become locally extinct.

This process is illustrated theoretically by Case [14], using a competitive Lotka-Volterra simulated system in which a resident community is constantly bombarded with invaders. Like all Lotka-Volterra systems, the precise mechanism of interaction between residents and invaders is not delineated, but rather described by a real number denoting the interaction’s strength. Case shows that in the simulated system, the community size at equilibrium is that which causes the probability of resident species extinction and the probability of foreign species invasion to be equal. More often than not, the equilibrium community size is almost always greater than unity (i.e. more than one species) [14]. The logical extension of this is that microbial consortia in the natural environment (e.g. in soils, manures, and fermented foods), unlike the quickly assembled monocultures utilized elsewhere across biotechnology, have had the chance to benefit from the stabilizing processes of ecological succession of over evolutionary time and therefore on balance should be more robust to invasion. There is strong support for this idea in the fact that even synthetic consortia, created by arbitrarily combining multiple species, are often unstable in their community composition and susceptible both to extinction and to domination by certain component species after the consortium is assembled [39].

Nowhere is this distinction between naturally evolved and engineered microbial ecosystems more pronounced than in the microbial jungle of the mammalian gut. First of all, there is the difficulty of reestablishing the natural gut microflora after a disturbance event such as disease or significant diet change; the fact that fecal microbiota transplantation (FMT) is in many cases the most cutting edge tool for community reestablishment [45] shows how difficult it is to synthesize a robust and functional consortium from scratch. Second, even while the gut microbiome is functioning normally, there is the related challenge of successfully adding new members to the intestinal community (e.g. probiotics), as evidenced by the rapid disappearance of many orally administered bacterial strains shortly after entrance into the gastrointestinal tract in both adult pigs and humans [13]. Shade [13] points out that there has been a long history of experiments suggesting that the adult human gut microbiome is in general resistant to colonization, beginning with Ilya Metchnikoff finding in 1908 that lactic acid bacteria orally introduced into his own gut were unable to establish themselves in the community and needed constant re-inoculation [13]. As above, the potential mechanisms behind this resistance are numerous, ranging from direct effects such as competitive exclusion by native gut microbes, to indirect effects such as modulation of the host immune system or host epithelial cells, production of antimicrobials, and breakdown of metabolites that promote growth of potential invaders [46, 47]. Note, however, that these robustness properties attain their full strength only after the gut microflora has
matured—which occurs by succession from its sterile state at birth, via exposure first to vaginal and fecal inocula from the mother, then breast milk, and finally solid foods [48, 49].

Experimentally there are in general two classes of experiments that have been performed to investigate the invasibility of consortia. The first type of experiment begins with a handful of species, testing for the invasibility of each species on its own as well as of consortia built by combining the individual species combinatorically [41]. Other similar experiments have used functional groups and isolates from soil and marine environments (as opposed to single species) as the basic unit of combination [16, 50, 51]. Results have shown that greater species-richness in general leads to increased invasion resistance, but the identities of the individual species and functional groups as well as the nature of the physical environment (e.g. soil type, presence of sunlight) are big determining factors [16, 41, 50, 51].

A second class of experiments perform the reverse combinatoric operation, partitioning an initially diverse microbial community into a subset of its original size. This subsetting can be performed either by the method of "dilution to extinction", whereby an initial culture or inoculum is serially diluted thereby removing its least abundant community members [51, 52], or by subjecting the community to a disturbance event such as heat stress, antibiotic exposure, or other modalities of environmental change [2, 53]. Results have shown strong support for the hypothesis that disturbed microbial communities are in general more susceptible to invasion—an idea which originated shortly after the introduction of the first antibiotics, when it was noticed that the mammalian gut microbiome was more easily colonized by pathogenic bacteria after exposure to streptomycin [10]. Recent work by Liu [53] showed that soil microcosms are more susceptible to invasion by Pseudomonas fluorescens after being disturbed by heating to above 75 °C [53], and work by Robinson showed that the cabbage white butterfly midgut is more easily invaded by Pantoea stewartii after exposure to antibiotics [2]. A common theme from experiments such as these is that disturbances alter native community structure, reduce diversity, and free up resources by reducing the native community’s competitiveness as a consumer [53]. Nevertheless, experimental work on microbial disturbances has so far been performed on a case-by-case basis, and the richness of effects that disturbances may have on native and invasive community members has yet to be brought into a broadly applicable predictive and diagnostic framework of understanding.

To illustrate how difficult it can be to reconcile conflicting truths from different branches of ecology, consider the "intermediate disturbance hypothesis" [54], which suggests that diversity is maximized in a community subjected to moderate levels (i.e. frequency and severity) of disturbance, as opposed to an absence of disturbances entirely. This phenomenon is derived principally from work in terrestrial, macroscopic ecosystems in which it is well known that disturbances such as fires, floods, and storms contribute to the maintenance
of ecosystem functionality and diversity [54]. The intermediate disturbance hypothesis, as originally proposed in 1978 [55], conjectures that at extremely low disturbance levels, community composition is shaped by inter-species competitive exclusion, whereby competitors in a temporally homogeneous environment drive one another to extinction. At high disturbance levels, conversely, the theory suggests that only those species capable of rapidly recolonizing their newly disturbed environment will survive [55]. The applicability of these ideas to the microbial world is in the proposition that, while a single disturbance event may increase a community’s invasibility in the immediate short term, these disturbance-mediated invasions, if moderate in degree, are one potential factor that could allow communities to become robust and less invasible over evolutionary time.

1.7 Fridley’s Synthesis

One of the most beautiful syntheses in invasion biology was made by Fridley in 2007 [17], in a conclusive review on the relationship between community diversity and the invasibility of macroorganisms. This review will outline Fridley’s work in detail and also explore its applicability as a springboard for the development of similar syntheses for microbial invasions.

Fridley begins with the observation that the relationship between the diversity and invasibility of communities, and indeed between the diversity and stability of communities in general, has been the cornerstone of all experimental, observational, and theoretical work on biological invasions. In the context of microbial invasions this is no less true, since the ideas developed above on the robustness of consortia and the invasibility of disturbed communities are essentially about differences between monocultures and consortia and disturbed versus undisturbed ecosystems. For invasions by macroorganisms, Fridley explains that there have been significant differences in the results obtained from broad-scale studies investigating invasions in geographic areas 1 km$^2$ or larger, and results obtained from fine-scale studies investigating invasions in environments 10 m$^2$ or smaller. On aggregate, broad-scale studies have shown a positive correlation between diversity and invasibility, whereas fine-scale studies have shown a negative correlation. A similar trend can be found by comparing the results of experimental versus observational studies on invasions, but as Fridley points out, this is largely because the majority of experimental studies are by nature limited to a geographically fine scale, whereas the majority of observational studies by nature seek patterns across a geographically broad scale. It was a significant breakthrough simply to point this out [17].

Fridley reconciles this “invasion paradox” of conflicting trends at small and large scales, by making the case that as spatial scale increases, there is a shift from biotic to environmental
drivers as the primary influencer of the diversity-invasibility relationship. At small scales, Fridley argues, Eltonian biotic resistance through competitive exclusion is at its strongest, because competitive exclusion is fundamentally a local effect between species occupying exactly the same environment—a condition which is strongest at small (defined relative to organism size or motility) spatial scales. As spatial scale increases, Fridley reports a number of abiotic processes emerge that can facilitate invasion. First, there is evidence that the environment itself can accommodate more unique resources as its scale increases [17], thereby making it easier for an invader to be able to access resources that are unconsumed by the native community. Second, the larger a geographic environment, the better it can support the creation of physically heterogeneous microenvironments (e.g. canopy versus undergrowth levels) as a result of the ecological strategies and activities of native community members. A more diverse native community can create a greater diversity of invasible microenvironments, but only in a spatial area large enough for this to be possible. Finally, a sampling effect exists, such that a more diverse community is more statistically likely to include species that by themselves are strongly invasion resistance, thus contributing to the overall invasibility of the community. Although there are some processes which can operate at all spatial scales, the synthesis offered by Fridley provides tremendously useful insight into those processes which do scale spatially.

1.8 Implications for Microbiology

Despite the wide applicability of Fridley’s work, several additional considerations must be made in order to make similar advances in the study of microorganisms. The first of these nuances relates to the concept of scale. Obviously the concepts of "spatially small" and "spatially large" must be defined relative to the size and dispersal ability of the organisms in question. However, many aspects the physical world scale absolutely, such as why surface tension matters to water striders but is irrelevant for the locomotion of larger organisms, why insects can bite but not punch, and why falling has different consequences based on an organism’s absolute size [56]. As a result, the spatial aspects of the physical world that are of relevance to microorganisms are likely different than those that matter to macroorganisms. Just as a mammal cannot "see," for example, the biofilms on every surface and diffusion-limited processes happening all around its body, so the microorganism cannot "see" processes such as fires, rainfall, or geological change. Spatial dynamics certainly should be expected to influence microbial community invasibility, and indeed have been shown to matter. Kim [19], for example, showed that spatial heterogeneity achieved by microfluidics was both necessary and sufficient to stabilize a synthetic multispecies bacterial community, whereas Brenner [18] showed that the layered spatial structure of a
synthetic biofilm community allowed for increased ability to colonize downstream surface environments. However, despite the importance of both biotic and environmental factors on microbial community dynamics, there is no reason to believe that either the "invasion paradox" or its resolution can be usefully brought to bear upon the study of microbial invasions.

A second proposition is that, in light of recent advances in culture-independent methods for enumerating the abundance of microbial population levels [57], it is imperative at this point in time to move beyond simple diversity metrics as a proxy for microbial community structure. To complement current efforts investigating the relationship between community diversity and invasibility, the tools are available today to begin asking the broader question of how community structure and invasibility are related in general. If ecosystems are described as networks, then a profusion of graph theoretical properties become available as potential metrics of community structure, such as concepts like connectedness, average diameter, and existence of cycles [58, 59], which, taken together, offer a much more information-rich picture of community structure than simple metrics of diversity alone [60, 61]. These concepts have already been adopted advantageously in metabolic engineering [58] and in studies on the stability of ecosystems [62, 63]. Microbial communities, due to their inherent structure comprising of usually single-celled organisms (nodes) and metabolite-mediated interactions between organisms (edges), are by nature best described as networks. The robustness properties and interactions between native community members and invaders, in the opinion of this review, are ripe for study in graph theoretical terms. Putting established ecological concepts like redundancy, plasticity, dormancy, and even diversity into the language of networks will allow for the development of incisive, insightful new theory and falsifiable experimental hypotheses that will benefit the study of microbial invasions.

A final proposition is that future research on microbial invasions should look to the history of food preservation for inspiration—a field which, since the beginning of civilization, has been successfully deterring microbial invaders from highly consumable environments, principally by setting up extreme environmental conditions that make foods temporarily less consumable. Besides techniques that actually kill microorganisms (e.g. pasteurization and irradiation) or physically prevent them from entering an environment (e.g. canning), food preservationists have utilized every possible weapon in the environmental arsenal to prevent invaders from increasing their numbers after initial exposure to a food environment. These include control of moisture (curing, drying), temperature (freezing, chilling), salinity (pickling), and alkalinity (fermentation); control of atmosphere (e.g. reducing oxygen or increasing nitrogen and CO2); control of physical state (gelatinization, crystallization); and production of antimicrobials and antioxidants. These individual effects are standardly
applied in combination as "hurdles" which together are, by design, difficult for unwanted microorganisms to overcome [64].

The most interesting type of preservation, at least in the context of invasion biology, is biopreservation. Biopreservation refers dually to the use of beneficial microorganisms for preservation purposes and to the use of antimicrobials in foods [64]. By utilizing microorganisms that are either native to or introduced into foods, biopreservation achieves a full suite of environmental control measures mediated by the release of metabolites such as organic acids and bacteriocins into the food environment [4]. The classic, ubiquitous example of this is fermentation, in which lactic acid bacteria (LAB) and other food microorganisms not only protect the food environment from invasion but also are responsible for producing the finished food product (e.g. yogurt, sourdough) itself [4,64] and in many cases are themselves recognized for their continued beneficial properties once ingested [65,66].

Nevertheless, the usefulness of environmental control is almost entirely absent from the literature on microbial invasions—principally because invasion biology is concerned with invasions into existing biological communities that can withstand only a limited range of environmental parameters, which is not the case in largely abiotic environments such as preserved foods. Moreover, while the idea that extreme environmental conditions make a community less invasible has certainly been cited in the literature, it is usually cited as a drawback of a particular system rather than a feature that can be productively and systematically exploited. For example, Kazamia [7] points out that of all known algal monocultures attempted for cultivation in the biotechnology industry, the only ones that have been able to survive in outdoor environments are those which are maintained under extreme environmental conditions, such as high salinity in the case of Dunaliella production or high alkalinity for production of Spirulina. This review posits, however, that environmental control can be any system’s first line of defense against microbial invaders, whether in an animal feedlot, a surgical bed, or a bioreactor. Modulation of the environment in some cases is the only possible engineering strategy and in all cases is a potentially viable option.

1.9 Aims of This Thesis

The promising field of biopreservation offers a bridge between the success of traditional food preservation methods and the challenge of understanding and manipulating the robustness of microbial ecosystems against invasion. This review has identified that community structure, rather than simply diversity, is likely a driver of invasion resistance and that a community’s ability to alter its environment can result in both increased (as in biopreservation) or decreased (as in autogenic ecological succession) resistance to invasion.
This review proposes that both community structure and environmental modulation, on their own, can lead either to success or to failure at deterring invaders, and that it is the interaction between community structure and environment that ultimately determines a community’s robustness to invasion. Understanding both of these factors, and their interaction, will be fundamental both to the deployment of successful probiotic invaders and to the development of synthetic microbial communities for robustness.

Three separate studies were pursued in this thesis. First, an experimental investigation of the robustness of traditional sourdough cultures to invasion was performed, exploring the role of the environmental modulation of pH in deterring invaders. Two invader organisms were tested for their ability to invade, and a mathematical model was developed for the bacterial fraction of a wheat flour dough. Second, a computational investigation of the effect of community structure on invasibility was conducted. A simple three-species ecosystem was modeled, and different community structures were probed for their ability to allow a rare species to become abundant as a result of perturbations to the system’s parameters. Finally, a phage-bacterium system was modeled, describing the interaction between a population of bacteriophages and a population of susceptible and infected bacteria. Here, the relationships between system parameters were explored visually and various trade-offs and coupling relationships were elucidated.

1.10 References


Chapter 2

Sourdough: A Model System for Invasion Biology

2.1 Abstract

This chapter introduces wild sourdough culture as a model system for the study of invasion biology and the design of novel microbial consortia. In the first part of this study, a simple experiment with a laboratory strain of *Escherichia coli* indicated that the production of organic acids is the primary mechanism of invasion resistance of a traditionally propagated sourdough starter against invasion by *E. coli*. Further experiments with the acidophile *Alicyclobacillus acidoterrestris* showed that while acid-tolerant invaders exist, the presence of environmental hurdles such as low temperature and high osmolarity are major factors in preventing their establishment. In the second part of this study, a mathematical model for the growth, metabolic output, and self-inhibition of *Lactobacillus sanfranciscensis* in wheat flour dough was developed, as a precursor for further work in sourdough process control and synthetic community design. This model was able to predict population density, pH, lactate concentration, and maltose concentration in a manner consistent with experimental data.

2.2 Background

Traditional sourdoughs are one of the most ancient and culturally significant biotechnologies in the human repertoire, spanning their initiation in ancient Egypt, adoption in ancient Greece and Rome, refinement in early modern continental Europe, and use by the first pioneers in the American West [1,2]. While today’s bakeries routinely produce many other types of breads, sourdoughs were for most of human history—until 150 years ago—the only naturally leavened breads that were possible to make [3]. Factors such as flavor, texture, shelf life, nutritional value, and artisanal appeal have contributed to a recent resurgence of interest in and use of sourdough production worldwide [2,4–6].

The chief reason for the prevalence of sourdoughs historically is the profound robustness
of culture establishment. Mix together flour and water, let sit at ambient temperature, refresh periodically with fresh flour and water, and over a period of several days to several weeks a mature sourdough culture, called a starter, will emerge [2, 4, 5, 7]. The starter is then used as an ingredient in many types of doughs, providing both leavening and acidification. Combined with the fact that dry cereal grains are undigestable unless ground and mixed with water [5], the repeated emergence and discovery of these cultures was a biological inevitability.

Perhaps the most interesting feature of a starter is its community composition, which is a distinctive combination of wild yeasts and wild bacteria that are selected for from the air, flour, and surfaces in the environment [1, 8]. Leavening is provided primarily by the wild yeasts, which are slower growing and more acid-tolerant than commercial yeasts, and acidification by the bacteria, primarily lactic acid bacteria (LAB) ubiquitous in the soil, water and plant environments associated with cereal grains [9–11]. While it was once believed that microbial diversity in sourdoughs had primarily a geographic origin, it is now understood that controllable process parameters such as temperature, hydration, fermentation time, and frequency of refreshments, as well as uncontrollable or seasonally varying parameters such as flour type and house microbiota (i.e. organisms native to the specific facility), are the primary drivers of community composition in traditionally propagated starters [1, 7, 8, 12]. This is due to a substantial drift in community composition from the initial flour microbial ecology to the final composition of the starter, as the sourdough maturation process involves several stages of ecological succession and selection into the sourdough environmental niche before reaching a stable state [1, 12–14].

Once established, sourdough starters are known to exhibit legendary functional and structural robustness over time. Individual mother-cultures are routinely maintained continuously for decades, with very little drift in both function (flavor production, acidification and leavening) and in community composition (dominance of certain species) despite continual exposure to open air, aseptic conditions, potential contaminants, and varying environments [8, 12]. Additionally, while experimental studies in other microbial ecosystems (e.g. methanogenic reactors) have shown that functional and structural robustness need not be coupled [15], stable sourdoughs repeatedly exhibit both [12].

Thus, at a very high level, the sourdough ecosystem provides an opportunity to evaluate one of the major open questions identified in Chapter 1—namely, which confers greater invasion resistance to a community, the (functional) ability to modulate its environment, or (structural) "consortium effects" such as diversity, redundancy, dormancy, plasticity, and connectedness? Traditional sourdough starters demonstrate both strong environmental modulation via the release of organic acids and antimicrobials, and potentially strong consortium effects arising from community size and, frequently, mutualism between the
dominant LAB and the dominant yeasts [1,3,4,12,16–18]. As a result, sourdoughs represent an important data point for understanding robustness against, in particular, free-living non-pathogenic microbial invaders.

2.3 Objectives

In this study, two aspects of sourdough were investigated. First, a short series of experiments was conducted to investigate the robustness of mature starter cultures against invasion by exogenous microorganisms. Two organisms were chosen as potential invaders of a traditional sourdough starter. The first, *Escherichia coli*, was selected for its ubiquitousness in human environments, and its known sensitivity to low pH. In testing *E. coli* as an invader, the objective was to determine whether the production of organic acids could serve as the primary mechanism by which a sourdough starter resists invasion. A second organism, the acidophile and orange juice contaminant *Alicyclobacillus acidoterrestris*, was also tested as an invader organism, in order to evaluate other environmental hurdles that could play a role in the sourdough environment. Finally, a survey of bacteria and archaea known to be acidophiles was conducted, to evaluate the degree of overlap between the temperature range in traditional sourdoughs and the temperature range preferred by known acidophiles. Through these experiments, aspects of the role of environmental modulation and environmental hurdles in sourdough starters were elucidated.

In the second part of this study, a mathematical model for the bacterial fraction of a traditional sourdough starter was developed, as a first step towards describing the full sourdough community structure analytically. While both the yeast and bacterial fractions are important, in traditional sourdoughs the bacterial fraction is metabolically upstream. Moreover, while several yeast species are known to dominate in mature starters [1, 6], two LAB species, *Lactobacillus sanfranciscensis* and *Lactobacillus plantarum*, are ubiquitous in sourdoughs around the world [1, 17]. Of these, *L. sanfranciscensis* is celebrated for its role in the unique flavor profile of San Francisco sourdoughs. Thus, a simple, extensible model characterizing the growth, metabolic output, and self-inhibition of *L. sanfranciscensis* in wheat flour dough was developed. Future work aims to integrate this with additional modules describing interactions with yeasts and also the physico-chemical processes of baking [19].
2.4 Sourdough Biochemistry and Microbiology

2.4.1 Types of Sourdoughs

Modern sourdoughs are actually of three types. Type I starters are synonymous with traditional sourdoughs, which are initiated and then periodically refreshed with fresh flour and water, a process known as backslopping, every 8 to 48 hours, and kept at 20–30 °C [1, 2, 7, 12]. Type II sourdoughs are semi-fluid, kept at 30 °C or above, and are continuously fed and dispatched via pumping. Type III are in dried powder form used as a flavor additive [1, 8, 12]. Type II and III sourdoughs deviate from traditional practices and will not be considered here. Unless otherwise stated, all sourdoughs and starters referred to in this study are Type I.

2.4.2 Composition of Flour

Sourdough biochemistry can be seen as a metabolic cascade beginning with the raw materials in flour and ending with the production of biomass and waste products. The composition of wheat flour, by weight, is given in Table 2.1:

<table>
<thead>
<tr>
<th>Component</th>
<th>Percent Composition (by weight)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starch</td>
<td>60–72</td>
</tr>
<tr>
<td>Moisture</td>
<td>14</td>
</tr>
<tr>
<td>Gluten</td>
<td>6–13</td>
</tr>
<tr>
<td>Other Proteins</td>
<td>2</td>
</tr>
<tr>
<td>Lipids</td>
<td>1</td>
</tr>
<tr>
<td>Nonstarch Polysaccharides</td>
<td></td>
</tr>
<tr>
<td>Pentosans</td>
<td>1.5–2.5</td>
</tr>
<tr>
<td>Glucofructosans</td>
<td>1</td>
</tr>
<tr>
<td>β-glucans</td>
<td>0.5–2</td>
</tr>
<tr>
<td>Mono- and Disaccharides</td>
<td></td>
</tr>
<tr>
<td>Sucrose</td>
<td>0.19–0.26</td>
</tr>
<tr>
<td>Maltose</td>
<td>0.07–0.10</td>
</tr>
<tr>
<td>Glucose</td>
<td>0.01–0.09</td>
</tr>
<tr>
<td>Fructose</td>
<td>0.02–0.08</td>
</tr>
</tbody>
</table>

Immediately upon addition of water, the chemical composition of flour rapidly changes. The major carbon sources in flour—starch and nonstarch polysaccharides—are rapidly
hydrolyzed to single sugar units by enzymes endogenous to cereal grains and the sourdough microflora. Upon hydration, β-amylases present in the flour are activated, converting starch to maltose by cleaving every second α-1,4 bond in unbranched starch polymer chains [20]. The initial maltose concentration increases by a factor of 10–15 almost immediately due to rapid hydrolysis of mechanically damaged starch in the flour [22], and then five-fold again as unbranched starch (amylose) units are broken down completely to maltose and branched starch (amylopectin) units are broken down partially, in their outer chains only [20,23]. Pentosans are broken down to arabinose and xylose by pentosanases endogenous to sourdough microflora [24,25], and glucofructans are broken down primarily to fructose monomers by secreted and cellular yeast invertase [12]. Out of all mono- and disaccharides released, maltose is by far the most prevalent, followed by fructose, and then in much smaller quantities the other sugars mentioned above [12,20]. Maltose is released close to completion over a period of 1-2 hours depending on dough conditions [23], but timescale data on the release via hydrolysis of other sugars is not as readily available in the literature. Nevertheless, sourdough fermentations often are conducted with as little as 8 hours between refreshments, so the relevant timescale for saccharide release is of a similar order or less [12,26].

2.4.3 Stages of Ecological Succession

Ecological succession in a traditionally propagated starter occurs in approximately three phases [7,8,13,14], beginning with the microbes already present in the flour. Typically there no more than 10^5 CFU/g of any given species in the flour, including aerobic bacteria, LAB, yeasts, and molds [8]. As soon as water is added, the redox potential of the medium begins to decrease, as a result of the consumption of oxygen via aerobic respiration by aerobes [8,27]. Once the environment turns anaerobic, which occurs quickly, before the first backslopping, anaerobes are favored, including facultative anaerobes such as Enterobacteriaceae and yeasts, and LAB primarily of the genera Enterococcus, Lactococcus, and Leuconostoc (phase I) [7,8,14]. After the first backslopping, for the next several days, the buildup of fermentation products such as lactic and acetic acid causes the pH to drop, reducing the numbers of acid-sensitive Enterobacteriaceae and Leuconostoc species and favoring the growth of more acid-tolerant genera such as Lactobacillus, Pediococcus and Weissella (phase II) [8,14]. Finally, as environmental conditions are maintained through an increasing number of backslopings, selection pressure leads to the emergence of highly adapted LAB and yeasts suited both to the acidity and to the carbohydrate makeup of the sourdough environment (phase III) [7,8,13,14]. This phase, characteristic of mature Type I starters, is stable over time as long as process conditions are maintained, and most commonly includes the LAB Lactobacillus fermentum, L. sanfranciscensis, and L. plantarum, as well as wild sourdough
yeasts [1,7,8,13,14]. In a mature sourdough starter the ratio of LAB to yeasts is also stable over time, on the order of 100:1 or greater, with LAB populations in the range $10^6$–$10^9$ CFU/g and yeast populations one or two orders of magnitude fewer [1,8,12].

### 2.4.4 Sourdough Microbiota

Upwards of 60 LAB and 30 yeast species have been identified from mature sourdough starter isolates around the globe [1,7]. Half of these are considered common to the sourdough environment, rather than transient or fortuitous [1,5]. Traditionally propagated starters have, on average, 2–6 identifiable LAB and 1–3 identifiable yeast species, with one or two of each typically found to be dominant [1,5,8,27].

**Sourdough LAB**

Sourdough LAB can either be heterofermentative, i.e. producing both lactic and acetic acids from the carbon sources available, or homofermentative, i.e. producing just lactic acid. The most prevalent LAB in type I sourdoughs, *L. sanfranciscensis* and *L. plantarum*, are obligately and facultatively heterofermentative, respectively. A 2014 analysis of 125 separate studies showed that *L. sanfranciscensis* and *L. plantarum* are present in upwards of 43% and 67% of sourdough isolates sampled from around the globe [7]. While *L. plantarum* is more prevalent, *L. sanfranciscensis*, which was originally isolated from several San Francisco bakeries in the 1970s, is uniquely adapted to the sourdough environmental conditions, is very rarely found elsewhere in the wild [1,28], and demonstrates well-characterized synergistic interactions with wild sourdough yeasts [1,12].

Heterofermentative fermentation by *L. sanfranciscensis* occurs as follows. Maltose is transported across the cell membrane via an H$^+$ symport, and is broken down into one unit of glucose and one unit of glucose-6-phosphate [17]. In stress conditions or if maltose is abundant, the unphosphorylated glucose is released back into the environment via facultative diffusion [12,17,29], which enables *L. sanfranciscensis* to exclusively metabolize the phosphorylated glucose, thereby gaining one unit of ATP [17]. This release of glucose into the environment serves both as a food source for maltose-negative sourdough yeasts, and as an antagonistic mechanism causing maltose-positive competitors to switch to glucose metabolism even when maltose levels are high [4]. As glucose-6-phosphate units are further reduced, one CO$_2$ unit is produced, and the substrate is cleaved into a two-carbon molecule and a three-carbon molecule. What happens next is dependent on the presence or absence of co-substrates, such as fructose, malate, and citrate, which serve as electron acceptors for *L. sanfranciscensis* and help restore the NAD+/NADH redox balance. In the absence of an appropriate electron acceptor as a co-substrate, one unit of lactate and one unit of
ethanol are produced, generating a net surplus of 2 ATP units. In the presence of an additional electron acceptor, acetate is produced instead of ethanol, and three ATP units are generated in total. The switch from ethanol to acetate is possible because ethanol production regenerates two necessary units of NAD+ from NADH, but this same effect is achieved by the co-substrate, making ethanol production unnecessary [17]. Finally, the reduction of pH and the accumulation of organic acids start to have an inhibitory effect on *L. sanfranciscensis* as metabolic products build up [6], such that growth stops before the levels of maltose and fructose are depleted [12]. While other sugars such as arabinose or xylose are available in doughs and starters, most strains of *L. sanfranciscensis* are unable to metabolize either [27].

**Sourdough Yeasts**

Three yeast species are worth noting. The closely related *Kazachstania exigua* (currently synonymous with *Saccharomyces exiguus*) and *Candida humilis* (currently synonymous with *Candida milleri*) were isolated from the same 1970s San Francisco bakeries as *L. sanfranciscensis* [30, 31], and are, besides baker’s yeast, the most prevalent yeasts found in sourdough isolates around the globe [1,7]. Baker’s yeast, or *Saccharomyces cerevisiae*, is frequently found in sourdough isolates [1,7], but is at a disadvantage metabolically due to its high sensitivity to acetic acid and the potential repression of genes necessary for maltose metabolism [5], and thus does not survive more than a few backblings in a type I sourdough environment [17]. While it is frequently added as a leavening agent to type II and type III sourdoughs, its presence in type I isolates has been attributed to background contamination levels in the bakery environment [7].

*K. exigua* and *C. humilis*, on the other hand, form stable associations with the LAB *L. sanfranciscensis* [26]. While *K. exigua* is more prevalent in San Francisco breads and *C. humilis* is more prevent in Panettone, both are maltose-negative and thus thrive when free glucose molecules are secreted into the sourdough environment by *L. sanfranciscensis* [12,26]. Additional associations with *L. sanfranciscensis* include amino acid production by these yeasts, which stimulates *L. sanfranciscensis* growth [5], and a high tolerance to acetic acid produced by heterofermentative LAB [17]. Experiments with the *L. sanfranciscensis–C. milleri* relationship have shown that associations between these two species are stable against invasion by other LAB [32]. However, while *S. cerevisiae* is easily packagable for commercial use, both *K. exigua* and *C. humilis* easily attenuate when subjected to processes such as freeze drying or flash freezing [33,34], thus making them difficult to use commercially.
2.5 Investigation of Sourdough Invasion Resistance

2.5.1 Methods

Organisms and Media

A laboratory strain of *Escherichia coli*, a fluorescent, kanamycin-resistant DH5α, was obtained from the C.V. Rao Research Group at the University of Illinois [49]. An acidophile, the orange and apple juice contaminant *Alicyclobacillus acidoterrestris*, was obtained from the Japan Collection of Microorganisms, strain 21546 (JCM, Japan), originally isolated from apple juice. Both were kept on 50% glycerol stocks at -80 °C, in Lysogeny Broth (LB) and Yeast Starch Glucose (YSG) media, respectively. Formulations for LB and YSG growth media are presented in Table 2.2.

Table 2.2: Formulations of LB and YSG Media. From [35, 36].

<table>
<thead>
<tr>
<th>(a) LB Media</th>
<th>(b) YSG Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ingredient</td>
<td>Quantity (per L)</td>
</tr>
<tr>
<td>Tryptone</td>
<td>10 g</td>
</tr>
<tr>
<td>Yeast Extract</td>
<td>5 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>10 g</td>
</tr>
<tr>
<td>Agar</td>
<td>20 g (for plates)</td>
</tr>
<tr>
<td>Final pH</td>
<td>7.0 ± 0.1 (at 25 °C)</td>
</tr>
</tbody>
</table>

*Autoclave for 20 minutes. Adjust pH by adding 5N NaOH.*

*Escherichia coli as a Sourdough Invader*

The *E. coli* strain listed above was tested for its ability to invade a traditional sourdough starter. The strain was grown in LB broth at 37 °C to logarithmic phase and inoculated into three cultures: (a) pure flour and water, (b) flour, water, and sourdough starter, and (c) a flour and water mixture subject to the addition of organic acids at regular intervals to mimic the pH drop exhibited by the starter. The inoculation size was 1:100 ratio by weight, meaning 1 mL logarithmic phase culture was added per 99 g sourdough. (Sourdough cultures were always measured by weight, not volume, due to their characteristic volume change during fermentation). The three cultures were labeled A1, A2, and A3, respectively, and kept at 28.5 °C. The population size of *E. coli* was monitored at 2, 4, 8, and 12 hours.
after inoculation, using aerobic plate counts on LB agar prepared with 50 mg/L kanamycin. Every two hours, organic acids were added to culture A3 by introducing the appropriate amount of a mixture consisting of 1 mL anhydrous acetic acid, 4 mL filtered deionized water, and 5 g flour. The flour used in all cases was an all-purpose, bleached and enriched H&R wheat flour (Gordon Food Supply, Champaign IL). All cultures were maintained in 300 g batches kept in polypropylene tubs with loosely fitted lids, at a 1.0 hydration ratio (mass of water to mass of solids). The starter had been initiated de novo in Urbana-Champaign, IL and maintained for several months at room temperature with refreshments every 12 hours, adding two parts flour and two parts water to one part starter, by weight, and mixing vigorously with a stirring rod once after each refreshment.

Alicyclobacillus acidoterrestris as a Sourdough Invader

Next, an acidophile, the orange juice contaminant Alicyclobacillus acidoterrestris, was prepared for a similar experiment, designed to determine the effects of factors, not including the production of organic acids, that confer invasion resistance to sourdough starters. The species was chosen for its tolerance both to low pH and to the low temperatures characteristic of Type I sourdough environments. Following preliminary results indicating the fastidiousness of A. acidoterrestris in a flour environment (it grew neither in sourdough nor in a flour and water control), several experiments were performed to investigate whether the presence of flour itself might be an inhibitor of growth. In the first experiment, A. acidoterrestris was grown in YSG broth at 45 °C to logarithmic phase and inoculated at a 1:15 ratio by volume into seven cultures, B1 through B7, kept at 45 °C. Each culture contained 15–30 mL of YSG broth and varying amounts of either sourdough starter or an acidified mixture of flour and water. The flour and water mixture was prepared by mixing equal weights flour and water, letting sit for 30 minutes (to let the flour amylases work), then adding 0.5% glucose by weight (to mimic the glucose levels in the sourdough starter), and finally adjusting the pH with acetic acid to pH 3.7. The population size of A. acidoterrestris was monitored at 1 and 9 hours after inoculation, using aerobic plate counts on YSG agar.

A second experiment was conducted to evaluate whether the antagonistic ingredient in the flour might either be bleach, an enrichment ingredient (iron and B vitamins), a component of the milled cereal grain, or a generalized osmolarity effect. To test this, eight liquid cultures were prepared, cultures C1 through C8. Six of the cultures contained 25 g YSG broth and an added 5 g of a different flour type (bleached enriched all-purpose flour; unbleached all-purpose flour; cake flour; whole wheat flour; organic whole wheat flour). A first control contained 25 g of YSG broth and an added 5 g of soluble starch. A second control contained 25 g of YSG broth an an added 5 g of dry YSG powder. A third control contained 30 g of autoclaved YSG broth only. A. acidoterrestris was grown in YSG at 45 °C.
to logarithmic phase (OD$_{600\text{nm}}$ 0.24) and inoculated at a 1:20 ratio by volume into all eight cultures. Cultures were kept at 45 °C and aerobic plate counts on YSG agar were taken 2 and 12 hours after inoculation.

This experiment was repeated, this time separating the components of dry YSG media (glucose, yeast extract, and soluble starch) into three separate cultures. An all-purpose flour control, a cake flour control, and a blank control were used as well. Cultures were labeled D1 through D6. Since the addition of flour and yeast extract significantly altered the pH, each culture was prepared first without any acids added, and then adjusted to a standard pH with 10N acetic acid. Before running this experiment, a quick sanity check was performed to ascertain the pH range over which strain JCM21546 can grow in YSG broth. The midpoint value of this range was then used as the setpoint to calibrate the pH of cultures D1 through D6. Logarithmic phase A. acidoterrestris (OD$_{600\text{nm}}$ 0.39) was grown in YSG at 45 °C and inoculated at a 1:20 ratio by volume into all cultures. Cultures were incubated at 45 °C and aerobic plate counts on YSG agar were taken 1 and 7 hours after inoculation.

One final experiment with A. acidoterrestris was performed, this time adding different amounts of flour to a logarithmically growing culture with already high population density, to determine the extent to which the addition of flour can halt or slow the growth of cultures that are already growing. Ten cultures were prepared, E1 through E10, each consisting of YSG broth inoculated with an overnight culture of A. adicoterrestris. All cultures were allowed to grow for 4.5 hours in order to reach logarithmic phase, at which point varying amounts of flour were added. Culture E1 received no flour, whereas E2 through E10 received a range between 0.2 mg and 50 mg flour per liter of YSG broth. The population density of all cultures were monitored every hour turbidimetrically, and cultures were kept at 45 °C. Because the addition of flour to YSG broth caused a significant increase in the absorbance of the broth, baseline data on this effect was obtained, by measuring the absorbance at 600 nm (OD$_{600\text{nm}}$) of 11 different YSG broth cultures, each with between 0 and 20 g/L of all-purpose flour added. A best-fit curve was fitted to this baseline data using a generalized additive model computed with the mgcv package in R [37].

**Measurement Methods**

Measurements of pH were taken using a benchtop pH meter calibrated to pH 4.0, 7.0, and 10.0 at room temperature. To measure the pH of flour mixtures, a 5 g sample was diluted 2x with filtered deionized water, vortexed, and three pH readings of the resultant fluid were taken. Then, the actual pH was back-calculated using equation 2.1, which solves for the pH by assuming that the amount of H$^+$ ions in the water added for dilution is negligible.
compared to the amount of $H^+$ ions in the dough:

$$pH_{\text{actual}} = \log_{10} \left( \frac{1.85}{0.85} \times 10^x \right)$$  \hspace{1cm} (2.1)

Here, $x$ denotes the measured pH and the ratio $1.85/0.85$ represents the mass of the 2x diluted fluid divided by that of the original undiluted sample, given that the specific volume of flour mixtures was 0.85 mL/g. The pH meter used was the Orion 3-Star benchtop pH meter with the Orion Ross Ultra 8102BNUWP pH electrode (Fisher Scientific, Beverly MA).

Aerobic plate counts of flour media were performed by diluting the media 2x by volume with filtered deionized water, vortexing, serially diluting, and spreading 20 µL of the diluted solution on 100mm diameter agar plates. Plate counts of *E. coli* were performed on LB agar with 50 mg/L kanamycin incubated at 37 °C. Counts of cultures A1–A3 were performed in duplicate at a dilution factor of 2000x. Plate counts of *A. acidoterrestris* were performed on YSG agar incubated at 45 °C. Cultures B1–B7 were diluted to 20x, 200x, 2,000x, and 20,000x, and counts were made only on plates with 25 Colony Forming Units (CFUs) or greater. Cultures C1–C8 and D1–D6 were diluted to 20x, 200x, 2,000x, 20,000x, and 200,000x.

Turbidimetric measurements of absorbance at 600 nanometers, also known as OD$_{600\text{nm}}$, were made in transparent 96 well plates, with each well containing 200 µL of liquid media, and 16 readings made per well with 12 flashes per reading. The equipment used was the Infinite®200 PRO NanoQuant (Tecan, Switzerland).

### 2.5.2 Results

As shown in Figure 2.1, the invasion experiment with *E. coli* showed that the flour, water, and acetic acid mixture was most effective at limiting growth and reducing the population size of *E. coli*. The sourdough starter also reduced *E. coli* population size, but somewhat less. *E. coli* did grow in the pure flour and water culture.

The experiments with *A. acidoterrestris* led to an interesting set of results. While *A. acidoterrestris* does grow at sourdough temperature and pH levels (growth at 20–55 °C and pH 2.2–5.8 [38]), and while soluble starch is a major ingredient in YSG media, the strain did not grow in a simple flour and water mixture as *E. coli* did. When *A. acidoterrestris* was grown in YSG media with varying amounts of sourdough starter and a flour/water mixture added (cultures B1–B7), aerobic plate counts after 1 h showed that the greater the amount of starter or flour added, the smaller the population size, as presented in Figure 2.2. When *A. acidoterrestris* was added to YSG broth with several different flour types, growth was observed in none of the cultures (C1–C5), nor in the culture with dry YSG ingredients added (C7). Growth only was observed in the control with nothing added (C8) and in the control with soluble starch added (C6). When this was repeated with cultures containing all
three ingredients in YSG separately—soluble starch, glucose, and yeast extract—growth occurred in the cultures with glucose and soluble starch added, but not in the cultures with yeast extract or flour. These findings are presented in Table 2.4.

Results were less obvious when when flour was added to logarithmically growing cultures of *A. acidoterrestris*. The addition of flour caused a marked increase in turbidity in all cultures, due to the intrinsic effect of the flour itself. For cultures with 50 g/L, 25 g/L, and 12.5 g/L flour added, the absorbance (OD$_{600nm}$) first increased upon immediate addition of the flour, then increased slightly again after 30 minutes, and finally dropped back to intermediate levels after several hours. For all other cultures (0.39 g/L to 6.25 g/L flour added), the absorbance increased after flour was added, and continued to increase over several hours representing continued growth. These findings suggest that the flour did indeed have an inhibitory effect causing the population size to drop, but only at a minimum inhibitory concentration (MIC) of 12.50 g/L or greater. However this effect had a delayed onset of at least a half hour. These findings are presented in Figure 2.4.

Taking a step back, a quick analysis was performed to visualize the temperature and pH ranges that are optimal for 28 other species of acidophilic bacteria as listed in [38]. The frequency of acidophiles that grow at a given temperature and pH optimum was plotted as a probability density surface over two axes representing temperature and pH. This was compared to the temperature and pH range characteristic of type I sourdough starters, namely 20–30 °C and pH 3.8–5.5 [6]. The acidophilic bacteria listed in [38] demonstrated a possible bias towards mammalian temperature ranges (37 °C), as shown by the ridge near that location on the probability density surface. Nevertheless, the sourdough environmental niche overlapped substantially with the region most inhabited by the acidophiles. Thus, temperature and pH alone would not be sufficient hurdles for preventing growth of these acidophiles in sourdoughs. Instead, other hurdles such as high osmolarity and the presence of inhibitory compounds would have to operant in order to prevent growth.
Table 2.3: Composition of Invasion Cultures

(a) Composition of Cultures A1–A3

<table>
<thead>
<tr>
<th>Culture ID</th>
<th>Flour</th>
<th>Water</th>
<th>Sourdough Starter</th>
<th>Organic Acids</th>
<th>E. coli Inoculum</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>150 g</td>
<td>147 mL</td>
<td>0 g</td>
<td>None</td>
<td>3 mL</td>
</tr>
<tr>
<td>A2</td>
<td>100 g</td>
<td>97 mL</td>
<td>100 g</td>
<td>None</td>
<td>3 mL</td>
</tr>
<tr>
<td>A1</td>
<td>150 g</td>
<td>147 mL</td>
<td>0 g</td>
<td>Acetic acid added every 2 h</td>
<td>3 mL</td>
</tr>
</tbody>
</table>

(b) Composition of Cultures B1–B7

<table>
<thead>
<tr>
<th>Culture Name</th>
<th>YSG Broth</th>
<th>Sourdough Starter</th>
<th>Flour/Water Mixture</th>
<th>A. acidoterrestris Inoculum</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure YSG (B1)</td>
<td>30 mL</td>
<td>0 g</td>
<td>0 g</td>
<td>2 mL</td>
<td>3.74</td>
</tr>
<tr>
<td>16% Sourdough (B2)</td>
<td>25 mL</td>
<td>5 g</td>
<td>0 g</td>
<td>2 mL</td>
<td>3.72</td>
</tr>
<tr>
<td>31% Sourdough (B3)</td>
<td>20 mL</td>
<td>10 g</td>
<td>0 g</td>
<td>2 mL</td>
<td>3.71</td>
</tr>
<tr>
<td>47% Sourdough (B4)</td>
<td>15 mL</td>
<td>15 g</td>
<td>0 g</td>
<td>2 mL</td>
<td>3.71</td>
</tr>
<tr>
<td>16% Flour Mix (B5)</td>
<td>25 mL</td>
<td>0 g</td>
<td>5 g</td>
<td>2 mL</td>
<td>3.74</td>
</tr>
<tr>
<td>31% Flour Mix (B6)</td>
<td>20 mL</td>
<td>0 g</td>
<td>10 g</td>
<td>2 mL</td>
<td>3.72</td>
</tr>
<tr>
<td>47% Flour Mix (B7)</td>
<td>15 mL</td>
<td>0 g</td>
<td>15 g</td>
<td>2 mL</td>
<td>3.71</td>
</tr>
</tbody>
</table>

(c) Composition of Cultures C1–C8

<table>
<thead>
<tr>
<th>Culture ID</th>
<th>YSG Broth</th>
<th>Other Ingredients</th>
<th>A. acidoterrestris Inoculum</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>25 mL</td>
<td>5 g Bleached Enriched All-Purpose Flour</td>
<td>1.56 mL</td>
</tr>
<tr>
<td>C2</td>
<td>25 mL</td>
<td>5 g Unbleached All-Purpose Flour</td>
<td>1.56 mL</td>
</tr>
<tr>
<td>C3</td>
<td>25 mL</td>
<td>5 g Cake Flour</td>
<td>1.56 mL</td>
</tr>
<tr>
<td>C4</td>
<td>25 mL</td>
<td>5 g Whole Wheat Flour</td>
<td>1.56 mL</td>
</tr>
<tr>
<td>C5</td>
<td>25 mL</td>
<td>5 g Organic Whole Wheat Flour</td>
<td>1.56 mL</td>
</tr>
<tr>
<td>C6</td>
<td>25 mL</td>
<td>5 g Soluble Starch</td>
<td>1.56 mL</td>
</tr>
<tr>
<td>C7</td>
<td>25 mL</td>
<td>5 g Dry YSG</td>
<td>1.56 mL</td>
</tr>
<tr>
<td>C8</td>
<td>30 mL</td>
<td>None</td>
<td>1.56 mL</td>
</tr>
</tbody>
</table>
### Table 2.3: (Continued.)

#### (d) Composition of Cultures D1–D6

<table>
<thead>
<tr>
<th>Culture ID</th>
<th>YSG Broth</th>
<th>Other Ingredients</th>
<th>( A.\ acidoterrestris ) Inoculum</th>
<th>Hydrochloric Acid</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>25 mL</td>
<td>5 g All-Purpose Flour</td>
<td>1.56 mL</td>
<td>18 µL</td>
<td>4.62</td>
</tr>
<tr>
<td>D2</td>
<td>25 mL</td>
<td>5 g Cake Flour</td>
<td>1.56 mL</td>
<td>16.5 µL</td>
<td>4.60</td>
</tr>
<tr>
<td>D3</td>
<td>25 mL</td>
<td>5 g Soluble Starch</td>
<td>1.56 mL</td>
<td>8.5 µL</td>
<td>4.65</td>
</tr>
<tr>
<td>D4</td>
<td>25 mL</td>
<td>5 g Glucose</td>
<td>1.56 mL</td>
<td>4 µL</td>
<td>4.64</td>
</tr>
<tr>
<td>D5</td>
<td>25 mL</td>
<td>5 g Yeast Extract</td>
<td>1.56 mL</td>
<td>400 µL</td>
<td>4.67</td>
</tr>
<tr>
<td>D6</td>
<td>30 mL</td>
<td>None</td>
<td>1.56 mL</td>
<td>4 µL</td>
<td>4.66</td>
</tr>
</tbody>
</table>

#### (e) Composition of Cultures E1–E10

<table>
<thead>
<tr>
<th>Culture ID</th>
<th>Flour Added (per L YSG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>0 g</td>
</tr>
<tr>
<td>E2</td>
<td>0.2 g</td>
</tr>
<tr>
<td>E3</td>
<td>0.39 g</td>
</tr>
<tr>
<td>E4</td>
<td>0.78 g</td>
</tr>
<tr>
<td>E5</td>
<td>1.56 g</td>
</tr>
<tr>
<td>E6</td>
<td>3.13 g</td>
</tr>
<tr>
<td>E7</td>
<td>6.25 g</td>
</tr>
<tr>
<td>E8</td>
<td>12.50 g</td>
</tr>
<tr>
<td>E9</td>
<td>25.00 g</td>
</tr>
<tr>
<td>E10</td>
<td>50.00 g</td>
</tr>
</tbody>
</table>
Figure 2.1: Growth of *E. coli* in Sourdough. Top: Comparison of pH of sourdough starter and flour, water, and acetic acid mixture over time. Bottom: Population density of *E. coli* after 2 h and 12 h of incubation in flour mixtures. Error bars represent sample standard deviations of population counts. Culture A1 contained flour and water only; culture A2 contained flour, water, and starter; culture A3 contained flour, water, and added acetic acid. *E. coli* grows well in A1, but not in A2 or A3, indicating that both sourdough starter and acetic acid inhibit growth.
Figure 2.2: Growth of *A. acidoterrestris* in YSG with Added Sourdough or Flour. Either sourdough starter or a flour-water mixture were added to logarithmically growing *A. acidoterrestris* in YSG broth. Population counts were made 1 hour after inoculation. Error bars represent sample standard deviations of population counts. Results indicate that both sourdough starter and the flour-water mixture inhibit growth.

Figure 2.3: Optical Density of YSG Broth With Varying Amounts of All-Purpose Flour. Best fit curve obtained using a generalized additive model with the *mgcv* package in R® [37].
Table 2.4: Growth Observed in Cultures C1–C8 and D1–D6

(a) Growth Observed in Cultures C1–C8

<table>
<thead>
<tr>
<th>Culture ID</th>
<th>Ingredient Added to YSG</th>
<th>Growth Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Bleached Enriched All-Purpose Flour</td>
<td>No</td>
</tr>
<tr>
<td>C2</td>
<td>Unbleached All-Purpose Flour</td>
<td>No</td>
</tr>
<tr>
<td>C3</td>
<td>Cake Flour</td>
<td>No</td>
</tr>
<tr>
<td>C4</td>
<td>Whole Wheat Flour</td>
<td>No</td>
</tr>
<tr>
<td>C5</td>
<td>Organic Whole Wheat Flour</td>
<td>No</td>
</tr>
<tr>
<td>C6</td>
<td>Soluble Starch</td>
<td>Yes</td>
</tr>
<tr>
<td>C7</td>
<td>Dry YSG</td>
<td>No</td>
</tr>
<tr>
<td>C8</td>
<td>None</td>
<td>Yes</td>
</tr>
</tbody>
</table>

(b) Growth Observed in Cultures D1–D6

<table>
<thead>
<tr>
<th>Culture ID</th>
<th>Ingredient Added to YSG</th>
<th>Growth Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>All-Purpose Flour</td>
<td>No</td>
</tr>
<tr>
<td>D2</td>
<td>Cake Flour</td>
<td>No</td>
</tr>
<tr>
<td>D3</td>
<td>Soluble Starch</td>
<td>Yes</td>
</tr>
<tr>
<td>D4</td>
<td>Glucose</td>
<td>Yes</td>
</tr>
<tr>
<td>D5</td>
<td>Yeast Extract</td>
<td>No</td>
</tr>
<tr>
<td>D6</td>
<td>None</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Figure 2.4: Growth of *A. acidoterrestris* in YSG with flour added during logarithmic phase. Black circles represent OD$_{600\text{nm}}$ readings before flour was added. Triangles represent OD$_{600\text{nm}}$ readings after flour was added.
Figure 2.4: (Continued.) Cultures with 6.25 g/L flour or less demonstrated continued growth after the flour was added, whereas cultures with 12.5 g/L or greater demonstrated first a dramatic rise in turbidity, due to the addition of flour itself, and then a dropoff in turbidity soon after, possibly due to an inhibitory effect.

Figure 2.5: Histogram Surface of Optimum pH and Temperature Range for 28 Acidophile Bacteria Species (data from [38]). The blue region represents type I sourdough environmental conditions.
2.5.3 Discussion

Our Champaign, IL sourdough starter was shown to resist invasion by *E. coli*, and this effect was shown to be likely due primarily to the production of organic acids. While growth of *A. acidoterrestris* was inhibited when flour was present at concentrations greater than 12.5 g/L, growth was also inhibited in YSG broth with extra dry yeast extract added. Growth did occur, however, when extra dry glucose or soluble starch was added at similar amounts. We identified several possible explanations for this behavior. Firstly, there may be surface effects by which the flour and yeast extract are sequestering *A. acidoterrestris* cells by adsorbing them onto the surface of flour/yeast extract particles, thus keeping cells out of the liquid phase and thus unaccounted for in aerobic plate counts. Second, there may be compounds such as iron present in both flour and yeast extract that may inhibit growth of *A. acidoterrestris*; however, the mechanism must be explored further. If the inhibition effect is robust, there may be a commercial possibility of using this effect to purposefully control *A. acidoterrestris* population sizes in juice or other environments where contamination can be an issue.

2.6 Modeling Growth of *L. sanfranciscensis* in Wheat Flour Sourdough

2.6.1 Methods

Organisms and Media

The LAB *L. sanfranciscensis* was obtained from the American Type Culture Collection, type strain 27651 (ATCC, Manassas VA). This strain was grown on a simplified de Man, Rogosa, and Sharpe (sMRS) media as shown in Table 2.5. The identity of the strain was verified by 16s rDNA sequencing and then kept on 50% glycerol freezer stocks.

Sequencing was performed as follows. *L. sanfranciscensis* in sMRS culture was grown to logarithmic phase and approximately 2 mL was centrifuged at 13,000 rpm for 1 minute. The cell pellet was resuspended in 0.5 mL ddH$_2$O, heated at 90 °C for 10 minutes, and centrifuged again at 13,000 rpm for 1 minute. Then 3 µL of supernatant was used as the template for PCR amplification of the 16S rDNA, using primers 8F (5’–AGAGTTTGATCCTGGCTCAG–3’) and 1391R (5’–GACGGGCGGTGWGTRCA–3’) for amplification. The initial denaturing step was performed for 5 minutes at 95 °C, followed by 35 cycles of 30 seconds at 95 °C, 60 s at 54 °C, 2 minutes at 72 °C and a final extension step at 72 °C for 7 minutes. The PCR product was separated by electrophoresis at 110 V on a 1% agarose gel stained with ethidium bromide in Tris-acetate-EDTA (TAE) buffer. The PCR product was then purified...
using a QIAquick PCR Purification Kit (Qiagen, CA) and the concentration of DNA was determined by NanoDrop (Thermo Scientific, Wilmington DE). The purified PCR product was submitted to the University of Illinois at Urbana-Champaign Core Sequencing Facility for sequencing (Roy J. Carver Biotechnology Center, Champaign IL).

Table 2.5: sMRS Media Ingredients. Yeast Extract and Beef Extract levels were kept low in order to be able to use this as a minimal media with semi-defined carbon source. When higher growth rates were needed, 5 g of each was used instead. A total of 20 g of sugars was always used. For maximum growth rate, 13.33 g of maltose and 6.66 g of fructose was used.

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Quantity (per L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeast Extract</td>
<td>1.25 g</td>
</tr>
<tr>
<td>Beef Extract</td>
<td>1.25 g</td>
</tr>
<tr>
<td>Peptone</td>
<td>10 g</td>
</tr>
<tr>
<td>Total Sugars</td>
<td>20 g</td>
</tr>
<tr>
<td>Fructose</td>
<td></td>
</tr>
<tr>
<td>Glucose</td>
<td></td>
</tr>
<tr>
<td>Maltose</td>
<td></td>
</tr>
<tr>
<td>MgSO₄ · 7H₂O</td>
<td>200 mg</td>
</tr>
<tr>
<td>MnSO₄ · 4H₂O</td>
<td>50 mg</td>
</tr>
<tr>
<td>Tween 80</td>
<td>1 g</td>
</tr>
<tr>
<td>K₂PO₄</td>
<td>2 g</td>
</tr>
<tr>
<td>Agar</td>
<td>20 g (for plates)</td>
</tr>
<tr>
<td>Final pH</td>
<td>5.8 ± 0.1 (at 25 °C)</td>
</tr>
</tbody>
</table>

*Autoclave for 20 minutes. Adjust pH by adding 10N HCl.*

**Baseline Sourdough Data**

Three sourdough starters were obtained in order to establish baseline behavior for pH change, rise, and metabolite production and consumption in doughs. The first sourdough starter had been initiated *de novo* in Urbana-Champaign, IL, whereas the other two were obtained from commercial vendors. The commercially obtained starters were King Arthur Flour Classic Fresh Sourdough Starter (King Arthur Flour, Norwich VT) and BreadTopia Live Sourdough Starter (BreadTopia, Fairfield IA), respectively. All three starters were maintained at room temperature with refreshments every 12 hours, adding two parts
flour and two parts water to one part starter, by weight, and mixing vigorously with a stirring rod once after each refreshment. These starters were labeled S1, S2, and S3 (referring to Champaign, King Arthur, and BreadTopia, respectively). Based on personal observation, doughs made from S1 were found to be vinegary and metallic; doughs made from S2 were tangy and sweet; and doughs made from S3 were tart and fruity and had the greatest "authentic San Francisco" taste. Pictures of the microbiota in all three starters are presented in Figure 2.6, obtained by diluting the starters 10x with 0.85% saline solution, vortexing, letting settle, and pipetting the supernatant onto slides photographed under a light microscope.

Baseline data on the performance of these sourdoughs was collected by producing sourdough doughs and measuring pH, rise, and metabolite concentrations over a 8–10 hour period. Doughs were produced from the starters by mixing ingredients according to Table 2.6 in a bowl for 3 minutes, and then hand kneading for 12 minutes. Doughs were partitioned into 150 g pieces and allowed to rise in narrow glass jars at 28.5 °C. pH was measured with an Orion AquaPro electrode 9104APWP calibrated at pH 4.0, 7.0, and 10.0, which could be used to measure the doughs directly without the need for dilution. Volumetric rise was calculated by letting the doughs rise the glass jars and measuring the height of the top of the dough as well as of the height of where the dough met the sides of the jars, making 5 measurements of each, and calibrating these measurements to a volume-to-height standard made with filtered water. Metabolite measurements of maltose, glucose, fructose, ethanol, lactate, and acetate were made using High Performance Liquid Chromatography (HPLC). Samples were prepared for HPLC by freezing at -80 °C until ready for analysis. Then, samples were diluted 3x by weight with 7% perchloric acid and mixed thoroughly using a glass stir rod as a pestle. Samples were then centrifuged for 10 minutes at 15,000 rpm. The supernatant was diluted another 5x with ddH2O and filtered through a 22 µm syringe filter unit and loaded into HPLC vials for analysis. HPLC analysis was performed on equipment at the Jin Research Group at the University of Illinois at Urbana-Champaign [39].

For each dough, pH and rise were measured every hour in triplicate, and HPLC measurements were made every 1.5 hours. HPLC measurements were made in triplicate for the dough made from the Champaign starter, and once for the doughs made from the other two starters. These results are presented in Figures 2.7–2.8. To ascertain the baseline release of free sugars into the dough by amylases in the flour, an additional set of HPLC measurements for maltose, glucose, and fructose were made on doughs with no sourdough starter added. These results are presented in Figure 2.9.
Table 2.6: Dough Recipes. Both doughs have 1% salt content, and are formulated to have a hydration ratio (weight of liquids / weight of solids) of 0.60, as is standard for classic French breads. Bleached and enriched all-purpose flour was used in all cases.

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Quantity (With Starter)</th>
<th>Quantity (Without Starter)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flour</td>
<td>281 g</td>
<td>309 g</td>
</tr>
<tr>
<td>Filtered Deionized Water</td>
<td>157 g</td>
<td>186 g</td>
</tr>
<tr>
<td>Sourdough Starter</td>
<td>56 g</td>
<td>0 g</td>
</tr>
<tr>
<td>Salt</td>
<td>5 g</td>
<td>5 g</td>
</tr>
</tbody>
</table>

*L. sanfranciscensis* Baseline Data

A standard OD–CFU curve was made by growing *L. sanfranciscensis* in sMRS media in triplicate, and taking turbidimetric readings as well as aerobic plate counts every hour. The curve was generated from the dataset using a generalized additive model computed using the mgcv package in R® [37], using only counts greater than 10 CFUs per plate. Plate counts were prepared by adding approximately 2 g of dough into a 50 mL conical tube, then adding 5 mL of 0.85% saline solution and mixing with a glass rod as a pestle. Then, additional 0.85% saline solution was added until the sample was diluted 10x by weight. This was vortexed, and allowed to settle for 2 minutes. Finally, the supernatant was pipetted into an Eppendorf tube and serially diluted a further $10^1$–$10^7$x. The resulting OD–CFU standard curve is presented in Figure 2.10.

Next, *L. sanfranciscensis* was grown to logarithmic phase and inoculated into 500 g of fresh (0.60 hydration ratio) dough, at 3, 6, 12, 25, and 50 mL inoculation size. The pH was monitored every 1.5 hours. To determine which inoculation amount gave rise to a pH profile most similar to that of the three sourdoughs, the pH values were compared to those in the Champaign, King Arthur, and BreadTopia sourdoughs. The pH profile of the doughs with *L. sanfranciscensis* added was plotted on the same axes as the best-fit curve of the pH profile of the three wild sourdoughs, obtained using a generalized additive model computed using the mgcv package in R® [37]. These plots are presented in Figure 2.11.

Once the inoculation size of *L. sanfranciscensis* that gave rise to a pH profile most closely matching that of the sourdoughs was determined, another set of doughs was prepared in order to obtain estimates of growth parameters of *L. sanfranciscensis* in dough. Levels of lactate, acetate, ethanol, maltose, glucose, and fructose were monitored using HPLC, while population density and pH were measured using aerobic plate counts and direct pH readings, respectively. Doughs with *L. sanfranciscensis* were made in triplicate. The HPLC results are presented in Figure 2.12, and were used in the mathematical model developed in the next section.
Development of Model

A simple model was developed to model the growth of *L. sanfranciscensis* in a wheat flour dough. A Monod model describing growth limited by a single substrate was used. The release of maltose into the dough via amylases was modeled according to the best fit curve in Figure 2.9. HPLC results showed that lactic acid was produced in far greater quantities than acetic acid. Thus, lactic acid was the only organic acid included in the model. Based on previous experimental work by [6], the inhibition due to pH on the growth of *L. sanfranciscensis* was included in the model using the function $\gamma(pH)$, which was obtained experimentally by [6]. Finally, as lactic acid is produced, the pH was adjusted using a function relating lactate concentration in the dough to pH.

\[
\frac{dL(t)}{dt} = \mu L(t) \gamma(pH(t)) \frac{M(t)}{k_s + M(t)} \quad (2.2)
\]
\[
\frac{dM(t)}{dt} = f_1(t) - \frac{1}{Y_m} \frac{dL(t)}{dt} \quad (2.3)
\]
\[
\frac{dA(t)}{dt} = \frac{1}{Y_a} \frac{dL(t)}{dt} \quad (2.4)
\]
\[
pH(t) = f_2(A(t)) \quad (2.5)
\]

In Equations 2.2–2.5, $L$ is the population density of *L. sanfranciscensis*, $A$ is the lactic acid concentration, $M$ is the maltose concentration concentration, and pH represents the pH. The function $f_1$ represents the release of maltose into the dough, corresponding to the HPLC data on dough with no starter or *L. sanfranciscensis* added. The function $f_2$ relates the acetate concentration to the pH, and was obtained experimentally. The parameters $\mu$, $Y_a$, and $Y_m$ represent the specific growth rate and biomass conversion rates (CFU per mmol maltose and CFU per mmol acetate). The parameter $\mu$ was obtained by plotting the population growth on log scale, fitting a best fit curve using a generalized additive model in R® [37], finding the point of maximum growth rate, and taking the slope of the best fit curve at this point. The parameter $Y$ was obtained by fitting a generalized additive model to the lactate concentration curve, picking two endpoints (3 h and 9 h), and associating these endpoints with the same endpoints in the population growth curve. The initial value for lactic acid concentration was set at zero. Since the Monod constant $k_s$ is difficult to estimate accurately, several values were tried and the resulting system dynamics were visualized and compared against the experimental data.
2.6.2 Results and Discussion

Photographs of each sourdough starter showed the dominant population to be some species of LAB, but the morphology was different in each of the starters. The pH change and volumetric rise in each of the starters were similar, however. HPLC results showed that maltose, glucose, and fructose levels changed in a nontrivial manner, due to being released into the dough and consumed simultaneously. Glucose showed the most clear pattern, first increasing in concentration and then being depleted. However, the levels of ethanol, lactate, and acetate all increased mostly monotonically.

When HPLC readings were taken on doughs with no sourdough starter added, the release of maltose followed a clear saturating relationship, whereas the glucose and fructose levels did not vary significantly. As predicted by the literature, Figure 2.8 indicates that there is a correlation between high levels of fructose in the dough, and high levels of acetate production. This is because fructose is often used as a co-substrate by *L. sanfranciscensis* and other sourdough LAB, allowing for acetate to be produced rather than ethanol by using fructose as an electron acceptor restoring the balance of NADH and NAD$^+$, and also generating one additional molecule of ATP [17].

Once the *L. sanfranciscensis* OD–CFU curve was generated (Figure 2.10), the inoculation amounts of 3 mL, 6 mL, 12 mL, 25 mL, and 50 mL of OD 1.1 *L. sanfranciscensis* into a 500 g dough were shown to correspond to $8.7 \cdot 10^6$, $1.7 \cdot 10^7$, $3.5 \cdot 10^7$, $7.3 \cdot 10^7$, and $1.5 \cdot 10^8$ CFU/g, respectively. Of these, the 50 mL inoculation size gave rise to a pH profile that most closely

---

Table 2.7: *L. sanfranciscensis* Model Parameters and Variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
<th>Unit</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.4627</td>
<td>1/h</td>
<td>Specific Growth Rate of <em>L. sanfranciscensis</em></td>
</tr>
<tr>
<td>$Y_m$</td>
<td>$4.6 \cdot 10^{10}$</td>
<td>CFU/mmol</td>
<td>Maltose to Biomass Conversion Rate</td>
</tr>
<tr>
<td>$Y_a$</td>
<td>$2.3 \cdot 10^{10}$</td>
<td>CFU/mmol</td>
<td>Lactate to Biomass Conversion Rate</td>
</tr>
<tr>
<td>$k_s$</td>
<td>0.04 mmol/g</td>
<td>Monod Constant</td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>—</td>
<td>Dimensionsless</td>
<td>Inhibition of Growth as Function of pH</td>
</tr>
<tr>
<td>$L$</td>
<td>—</td>
<td>CFU/g</td>
<td><em>L. sanfranciscensis</em> Population Density</td>
</tr>
<tr>
<td>$A$</td>
<td>—</td>
<td>mmol/g</td>
<td>Lactic Acid Population Density</td>
</tr>
<tr>
<td>$M$</td>
<td>—</td>
<td>mmol/g</td>
<td>Maltose Concentration</td>
</tr>
<tr>
<td>pH</td>
<td>—</td>
<td>—</td>
<td>pH</td>
</tr>
<tr>
<td>$L_0$</td>
<td>$1.77 \cdot 10^7$</td>
<td>CFU/g</td>
<td>Initial Value</td>
</tr>
<tr>
<td>$A_0$</td>
<td>0</td>
<td>mmol/g</td>
<td>Initial Value</td>
</tr>
<tr>
<td>$M_0$</td>
<td>0.02478</td>
<td>mmol/g</td>
<td>Initial Value</td>
</tr>
</tbody>
</table>

---
matched those of the Champaign, King Arthur, and BreadTopia sourdoughs.

This inoculation size was used to inoculate a new set of doughs that were used to obtain parameter values for the Monod model. HPLC results for the doughs inoculated with a pure culture of \textit{L. sanfranciscensis} are shown in Figure 2.12. Interestingly, both acetate and ethanol levels were undetectable throughout the fermentation period. This is a significant inconsistency because metabolically, \textit{L. sanfranciscensis} always either produces lactate and acetate, or lactate and ethanol, as fermentation products. The HPLC results, performed in triplicate, indicate that neither acetate or ethanol were produced, which is biochemically inconsistent with what the known fermentation pathway of \textit{L. sanfranciscensis} is. Thus, for the purposes of the mathematical model, only lactate buildup was incorporated. The model would not be significantly altered if acetate buildup were incorporated, since the inhibitory effect on \textit{L. sanfranciscensis} is a computed as a function of pH rather than directly as function of organic acid concentration. Nevertheless, for future work, these HPLC results should be revisited and the data obtained again.

The model described the relevant features of the system accurately, as it was able to predict population density, pH, lactate concentration, and maltose concentration in a manner consistent with the experimental data. Figure 2.13 shows the experimental data used to estimate parameter values for the model, and Figure 2.14 shows the model dynamics with \( k_s \) values set to 0, 0.01, 0.02, 0.03, 0.04, and 0.05. The best fit was obtained when \( k_s = 0.04 \). Two important components were necessary in the model: (1) the inhibition of growth caused by reduction of pH, and (2) Monod kinetics overlayed on the release of maltose into the dough. While the Monod constant \( k_s \) is in general difficult to obtain experimentally, we showed that a value between 0.01 and 0.05 produced the most accurate results.

### 2.7 Future Work

Both of the sections above—the invasion experiments using \textit{E. coli} and \textit{A. acidoterrestris}, and the modeling of \textit{L. sanfranciscensis}—contribute to a larger aim, which is to investigate whether the functional features (leavening, acidification, and robustness) of traditional sourdoughs are attainable via a defined culture of two or three species alone, or whether they can only be obtained using a consortium consisting of many organisms. In the first section, we showed that the mechanisms of invasion resistance against \textit{E. coli} and \textit{A. acidoterrestris} are likely to be pH-based and potentially an inhibitory aspect of a component in the flour itself. Both of these are environmental effects. In the second section, we showed that a simple Monod model of the growth of \textit{L. sanfranciscensis} in a wheat flour dough, combined with a function describing inhibition by acid buildup as obtained from the literature in [6], was able to accurately describe the population growth and metabolite consumption and
production in the dough. This model is extensible and future work could add interactions with yeasts, as well as aspects of the physico-chemical process of baking.

Figure 2.6: Sourdough Starter Microbes Photographed. Sourdough samples from doughs S1, S2, and S3 were diluted and stained using the Gram Stain (top). Starters were also photographed in jars used for measuring volumetric rise (bottom).
Figure 2.7: Volumetric Rise and pH Change in Three Sourdoughs. Best fit curves obtained using a generalized additive model computed with the mgcv package in R® [37].
Figure 2.8: HPLC Readings of Three Sourdoughs. Notice the correlation between high fructose levels and higher production of acetate, as indicated by the hollow triangle data points. This corresponds to the fact that fructose is used as a co-substrate for acetate production in \textit{L. sanfranciscensis} and other heterofermentative LAB [17].
Figure 2.9: HPLC Readings of Pure Doughs, Without Sourdough Starter Added. A saturating curve was fit to the maltose data points. For fructose data, no significant change in concentration was observed over the 14 hour measuring period.

Figure 2.10: OD-CFU Standard Curve for *L. sanfranciscensis*. The curve was generated from the dataset using a generalized additive model in using the mgcv package in R ® [37].
Figure 2.11: pH Profile of Doughs with *L. sanfranciscensis* Added. The solid curve represents the best fit curve of the pH profile for Champaign, King Arthur, and BreadTopia sourdoughs, as obtained and presented in Figure 2.7. The dotted points represent pH reading of doughs with *L. sanfranciscensis* added in place of sourdough starter. The 50 mL inoculation size most closely matched the pH profile of the sourdoughs.
Figure 2.12: HPLC Readings of *L. sanfranciscensis* in Dough. Acetate and ethanol levels were not detected. Glucose and fructose levels were steadily depleted as lactate accumulated.
Figure 2.13: Data on *L. sanfranciscensis* for Parameter Estimation. Top Right: Log population density plotted in order to obtain an estimate for $\mu$, the specific growth rate, which was taken to be the maximum value of the slope of the best fit curve fitted through the data. Bottom Left: pH values plotted against lactate concentration, used to generate the function $f_2$, the best fit curve relating pH to lactate concentration values. Bottom Right: Curve obtained from [6], depicting the effect of pH on the growth rate of *L. sanfranciscensis*. 
Figure 2.14: Model of Population Growth, pH, Lactate Production, and Maltose Consumption. Different values of $k_s$ were chosen and the resulting dynamics are plotted in color. The black lines represent experimental data.
2.8 References


56


Chapter 3

Invasion By a Rare Species in a Three-Species Ecosystem

3.1 Abstract

This chapter considers the general problem of invasion, in which a rare species in an ecosystem suddenly becomes abundant as a result of a discrete mutation in ecological parameters describing the structure of the system. We first consider the most general case of first-order ecosystem dynamics, and then show that a 3-species Lotka-Volterra (LV) system has certain properties that make it amenable to studying this problem. Simulations were performed with 15 ecosystem structures, denoted topologies, that were tested for their invasibility. Invasibility was defined by the minimum change to the system parameters required to cause a stable equilibrium point to move from a position at which a given species is rare, to a position at which that species is abundant. Each community structure was ranked in this manner from least to most invasible by using a distance metric, performed for 100 simulations per topology. Results showed that the magnitude of invasibility is highly sensitive to changing the growth rates, but the direction of invasion is invariant for each topology. Thus, constraining the ecosystem structure doesn’t constrain how far the system parameters have to move in order for a rare species to become abundant, but rather constrains the direction the parameters have to move in.

3.2 Background and Problem Statement

Invasion biology attempts to predict, across multiple scales and environments, what makes an ecological community vulnerable to invasion by one or more foreign species [1]. Early research on invasions focused on the diffusion and dispersal processes that allow invaders to arrive in an environment for the first time—the classic scenario being arrival at an island [1–3]. Although this is an important driver of invasions in practice, the more interesting problem is to assume that there is some arrival event and ask what is the fate of the initial invader propagule over time. A recent meta-analysis on plant invasions [4] has in fact
suggested that, while ecological communities have little intrinsic ability to prevent an arrival event itself, they can be very effective in intrinsically limiting the population level of an invader once it has landed. Thus we can think of a large portion of the invasibility problem as a problem not of resistance to invasion, but of containment of the invader [4].

In microbial ecosystems the phenomenon of dormancy makes the problem of containment particularly pronounced. It is not uncommon, in naturally occurring microbial ecosystems, that upwards of 20% of all bacteria recovered from an environmental sample are metabolically inactive [5]. Although some proportion of these are dead or near-dead, many have simply entered into a reversible state of low metabolic activity (i.e. dormancy) as a survival strategy resulting from unfavorable environmental cues. This allows, among other things, the community to sustain extremely low population levels of certain dormant species, levels much lower than would be the case if all individuals were forced to either grow or die. As a result, invasion events are just as likely to occur via a newly landed individual as they are via a dormant individual which suddenly changes its ecological strategy or otherwise responds to a sudden environmental change. Even without dormancy, it is generally the case that there are a small number of abundant species, but many rare species, in any given naturally occurring ecosystem [6].

In this chapter, therefore, we consider invasion by one or more rare populations in a community. We assume the system is at some attractor in which the population levels of one or more rare species are stably infinitesimally small. We assume, further, that rare populations are kept rare not because of a choice to enter into dormancy, but because some property of the system (i.e. an environmental quantity or ecological interaction) prevents the population from growing. We consider an invasion event to represent some sudden, discrete change in one or more of these properties of the system, such that the community is caused to move towards a new attractor at which at least one of the previously rare populations is now abundant.

We aim to develop a notion of robustness to invasion, which describes the magnitude of the smallest discrete change in system properties that can turn a rare population into an abundant one.

3.3 System overview

We consider an ecological community consisting of \( n \) interacting populations, with each population size \( x_i \) modeled according to continuous-time, non-autonomous, first-order dynamics [7]:

\[
\frac{dx_i(t)}{dt} = f_i(x(t), a(t))
\]  (3.1)
Here, $\alpha = \{\alpha_1, \ldots, \alpha_k\}$ represents $k$ ecological parameters that fully describe the system, and $f$ is a vector-valued function governing the system’s dynamics.

In any real ecosystem the parameter set $\alpha$ will vary over both time and space. However, even the spatially homogeneous system $f = f(x(t), \alpha(t))$ is such a complex one that, to make use of it, one must in practice either impose certain geometrical constraints on $\alpha(t)$ or constrain how $\alpha(t)$ is allowed to change with time. It is common to consider time-invariant systems which make no restrictions on the properties of $\alpha$, or time-varying systems which impose certain symmetries or inequalities on the $\alpha_i$ terms. The generalized Lotka-Volterra system \[7\] is an example of the former, whereas the competitive non-autonomous Lotka-Volterra system is an example of the latter \[8\].

We can take advantage of a key insight from Allesina \[9\], that, for a given system, rather than attempting to measure each $\alpha_i$ with high spatiotemporal granularity, we are best off looking for "families" of $\alpha$ that share certain properties. Successful analytical approaches include generating random matrices and searching for ensembles with similar behavior \[10\], or subjecting a given $\alpha$ to a small perturbation and asking if its properties hold \[11–13\]. Here we take a hybrid approach in which we impose a strong restriction on the time-behavior of $\alpha$ and a weak restriction on its geometric properties.

First, we suppose that $\alpha(t)$ is constrained to be a discrete function of time, that "mutates" every so often but remains constant otherwise. That is,

$$\left\| \frac{d\alpha(t)}{dt} \right\| = \begin{cases} \infty & \text{at jump discontinuities} \\ 0 & \text{else} \end{cases}$$

Second, we suppose that mutations in $\alpha$ always preserve the sign of each $\alpha_i$. That is, we allow changes in the magnitude of ecological interactions, but we disallow changes in their sign. Thus we can have predator and prey that each innovate in their effectiveness, but we do not allow predator and prey to switch roles. In Section 3 we will present two additional constraints on the sign of terms in $\alpha$ that are sensible in the context of microbial ecology.

To make this set of assumptions useful, we further assume that, for any fixed $\alpha$ and starting point $x_0$, there exists some attractor that the system approaches asymptotically, and moreover that, with high probability, $\alpha$ mutates at a frequency slower than the time it takes in general for $x(t)$ to approach one of its attractors. This effectively reduces the problem of solving for $x(t)$ from a dynamical systems problem to a geometry problem. To see why this is so, observe that if, with high probability, $x(t)$ in general has enough time to get close to one of its attractors before there is a mutation in $\alpha$, and if mutations in general send the system towards some new attractor, then, the effect of a mutation from $\alpha$ to $\tilde{\alpha}$ can be understood simply in terms of the difference in properties between the original and
resultant attractor.

This allows us ask, can a discrete-time mutation from $\alpha$ to $\tilde{\alpha}$ cause the system to move from one stable steady state $x^*$ to another stable steady state $\tilde{x}^*$, in which $x^*$ and $\tilde{x}^*$ have entirely different properties? We will seek a functional form of $f$ that allows us to answer this question simply by knowing the two values of $\alpha$ and $\tilde{\alpha}$.

The following definitions will be of use:

**Definition 3.1.** A *mutation* of a matrix $A$ is any matrix $\tilde{A}$ satisfying $\text{sign}(A) = \text{sign}(\tilde{A})$ in all dimensions.

**Definition 3.2.** The concepts *absent*, *rare*, and *abundant* are defined such that, for some population $x_i$,

$$
\begin{align*}
  x_i = 0 & \Rightarrow x_i \text{ is absent} \\
  0 < \frac{x_i}{\|x\|} < \epsilon_0 & \Rightarrow x_i \text{ is rare} \\
  \frac{x_i}{\|x\|} > \epsilon_1 & \Rightarrow x_i \text{ is abundant}
\end{align*}
$$

for small $\epsilon_0, \epsilon_1$ satisfying $0 < \epsilon_0 < \epsilon_1$. This is motivated by the fact that, in most microbial ecosystems, $n$ is very large but some $x_i$ are very small.

**Definition 3.3.** An ecosystem is said to be *invasible* as follows. Suppose there exists some stable equilibrium point $x^*$ satisfying $f(x^*, \alpha) = 0$, and suppose that at least one population in $x^*$ is rare. We say $x^*$ is *invasible* with respect to $\alpha$ if there exists some $\tilde{\alpha}$ and $\tilde{x}^*$ such that:

- One or more species that were rare at $x^*$ are abundant at $\tilde{x}^*$.
- $\tilde{x}^*$ is a stable equilibrium point satisfying $f(\tilde{x}^*, \tilde{\alpha}) = 0$
- $x^*$ falls within the attractor basin of $\tilde{x}^*$ as long as $f = f(x(t), \tilde{\alpha})$
- The distance $\|\tilde{\alpha} - \alpha\|$ is small

The above conditions are necessary and sufficient for an ecosystem, located at some stable equilibrium point $x^*$ containing at least one rare population, to, upon a discrete-time mutation of the system parameter set from $\alpha$ to $\tilde{\alpha}$, begin a trajectory towards a new equilibrium point $\tilde{x}^*$ which is characterized by at least one of the previously rare populations now being abundant. If $\alpha$ and $\tilde{\alpha}$ are close to one another, then, evidently, we say that the point $x^*$ is invasible with respect to $\alpha$.

Since we are only concerned with evaluating the invasibility of a given point as defined above, we can simplify Equation 3.1 to the autonomous case:

$$
\frac{dx(t)}{dt} = f(x(t), \alpha) \tag{3.3}
$$
Note that the definition of invasion presented here only holds for invasion events that move the system from one stable equilibrium point to another, although in theory the definition could expand to include transitions within a broader class of attractors.

The simplest viable system of this form in Equation 3.3 is the generalized Lotka-Volterra (LV) system [7] defined by the governing equation set:

$$f_i = x_i(t) \left( r_i - \sum_{j=1}^{n} a_{ij} x_j(t) \right)$$

in which \( r_i \) is the intrinsic growth rate of population \( i \), and the \( a_{ij} \) terms describe the effect of the size of population \( j \) on the growth rate of population \( i \). We define the growth rate vector \( \mathbf{r} = \{ r_1, \ldots, r_n \} \) and community matrix \( \mathbf{A} = [a_{ij}] \), which together make the parameter set \( \alpha = \{ r_1, \ldots, r_n, a_{11}, \ldots, b_{nn} \} \).

It is sensible to impose the restraints:

$$r_i \geq 0 \quad (3.5)$$
$$x_i \geq 0 \quad (3.6)$$
$$a_{ii} = a_{jj} > 0 \quad (3.7)$$

thus disallowing negative intrinsic growth rates and population size, and forcing population size to be equally self-limiting in all species.

LV systems can, in general, describe any type of first-order pairwise interaction between species. The signs of any two opposing entries \( a_{ij} \) and \( a_{ji} \) in the community matrix describe the type of interaction between populations \( i \) and \( j \), encompassing a wide array of ecological interactions but remaining agnostic to the actual mechanism:

$$+ + \quad \text{mutualism or symbiosis}$$
$$+ - \quad \text{predator-prey}$$
$$+ 0 \quad \text{commensalism}$$
$$- - \quad \text{competition}$$
$$- 0 \quad \text{amensalism}$$
$$0 0 \quad \text{neutralism}$$

In general, the ordered list of signs of all off-diagonal terms in \( \mathbf{A} \) determines the community structure, or topology, of the system. In a 3-species system, which is the smallest system in which a nontrivial invasion event can occur, there are six off-diagonal terms in the community matrix and thus \( 3^6 = 739 \) total possible topologies, which reduces to 138 unique
topologies when isomorphisms are taken into account. All 138 topologies were ranked in order of increasing number of nonzero interaction terms, the first 15 of which are shown in Figure 3.1. Blue arrows denote a positive interaction, red denotes a negative interaction, and gray denotes no interaction. All topologies were ranked according to increasing number of nonzero interaction terms. The first 15 are pictured here.

The generalized LV system has the property that, if all $x_i \neq 0$,

$$f = 0 \Rightarrow r - Ax = 0 \quad (3.8)$$

Thus if $A^{-1}$ exists, the point $x^* = A^{-1}r$ is unique and is a steady state solution of the system. Moreover, LV systems have the property that

$$J_{ij} = \frac{\partial f_i}{\partial x_j} = \begin{cases} -a_{ij}x_i & \text{if } i \neq j \\ r_i - 2a_{ii}x_i & \text{if } i = j \end{cases} \quad (3.9)$$

where $J_{ij}$ represents entries in the Jacobian matrix $J$. Classical stability analysis allows us to define the stability of some fixed point $x^*$ by analyzing its Jacobian. We present this in Definition 3.4.

**Definition 3.4.** We call a fixed point $x^*$ satisfying $f(x^*, \alpha) = 0$ stable if the real parts of all eigenvalues $\lambda_i$ of the Jacobian matrix $J = [\partial f_i / \partial x_j]$ evaluated at $x^*$ are negative. We call $x^*$
unstable otherwise. Thus,

\[
\text{stability}(x^*, \alpha) = \begin{cases} 
\text{stable} & \text{if all Re}(\lambda_i) < 0 \\
\text{unstable} & \text{else}
\end{cases}
\]

### 3.4 Classification Task

The following classification task allows each topology to be ranked in terms of invasibility. Consider two Boolean-valued functions, \( h_2(\alpha) \) and \( h_3(\alpha) \), defined on the parameters describing a 3-dimensional LV system. We say that a parameter set \( \alpha \) is \textit{invasible} if, for some mutation \( \tilde{\alpha} \), we have:

- \( h_2(\alpha) = \text{true} \)
- \( h_3(\tilde{\alpha}) = \text{true} \)
- the distance \( \|\alpha - \alpha^*\| \) is small

The classification task is then, for each topology, to find regions in \( \alpha \)-space which satisfy \( h_2 \) and \( h_3 \) respectively (these regions are mutually exclusive). The significance of this is that, in any LV system, it is straightforward to define the functions \( h_2 \) and \( h_3 \) such that satisfying the above three properties also satisfies Definition 3.3 of invasibility. Thus, calculating a measure of distance between \( h_2 \) and \( h_3 \) regions gives a measure of the invasibility of each topology.

In a 3-species LV system, we define \( h_2 \) and \( h_3 \) as follows:

**Definition 3.5.** For \( \alpha = \{r, A\} \), \( h_2 \) evaluates to true if and only if:

- \( \text{stability}(x^*) = \text{stability}(A^{-1}r) = \text{true} \)
- Exactly one \( x_i \) is rare and the other two are abundant

**Definition 3.6.** For \( \alpha = \{r, A\} \), \( h_3 \) evaluates to true if and only if:

- \( \text{stability}(x^*) = \text{stability}(A^{-1}r) = \text{true} \)
- All three \( x_i \) are abundant
- The matrix \( A \) is dissipative, that is, there exists a positive-valued diagonal matrix \( D \) such that \( DA \) is positive definite in the sense of quadratic forms. When \( n = 3 \) this is true if the system of equations, in which \( B = A^{-1} \),

\[
\begin{align*}
(a_{21} t + a_{12})^2 &< 4 a_{11} a_{22} t \\
(b_{21} t + b_{12})^2 &< 4 b_{11} b_{22} t
\end{align*}
\]
has a positive solution $t$ that satisfies $0 < t < \infty, a_{ii} > 0, b_{ii}$. This is proved in [11].

### 3.5 Method and Results

Numerical simulations were performed as follows. First, we limited our study to the first 44 topologies, in which there are only one, two, or three nonzero interaction terms in the community matrix. Second, we limited our study to mutations that occur in these terms only, rather than in the intrinsic growth rate terms or the self-inhibition terms. For each topology, we held the intrinsic growth rate terms fixed, and generated one million points from a multivariate half-normal distribution representing different values of the community interaction terms in a given topology. Then, each point was classified either as a "2-node" (i.e. points for which $h_2$ evaluates to true) or "3-node" (i.e. points for which $h_3$ evaluates to true). This was repeated one hundred times for each topology, each time choosing a different set of parameters that were held constant for $r_1, r_2, r_3,$ and $a_{ii}$.

We found that approximately 15% of all points were 3-nodes, whereas only 0.1% of all points generated were 2-nodes. Furthermore, by numerically examining the condition number, i.e. the expected value of the ratio change in output to a random small change in input for a given function, of each function occurring in the definitions of $h_2$ and $h_3$, we found that all functions in question are well conditioned (i.e. the condition number is always less than $10^3$ for a sample size of 10000 points generated for each function [14]). This meant that $h_2$ and $h_3$ both evaluate to true in continuous regions in $\alpha$-space, which greatly simplified our task. We verified that this was the case by generating one million points per topology, and noting that 2-nodes and 3-nodes always occurred in distinct regions in $\alpha$-space. We can show this is the case graphically, as is shown in Figure 3.2 for 2-nodes of topology 8 (blue points are 2-nodes, red points are non 2-nodes).
Next, we defined a measure of the distance between regions containing 2-nodes and regions containing 3-nodes. For each point in the 2-node region, we found the closest point in the 3-node region. We then defined the "invasion distance" as the mean of that distribution. For each topology, we obtained one hundred point clouds of 2-nodes and 3-nodes, each point cloud corresponding to a fixed set of values for $r_1, r_2, r_3$, and $a_{ii}$. For each point cloud, one invasion distance number was obtained. Thus, one hundred invasion distance numbers were obtained for each topology, corresponding to one hundred sets of fixed values for $r_1, r_2, r_3$, and $a_{ii}$.

Surprisingly, the variance of invasion distances within a given topology—i.e. the variance across those 100 trials for the topology—was not significantly larger than the variance of invasion distances between topologies. To illustrate this, consider these plots of 2-nodes (green) and 3-nodes (blue), as shown in Figure 3.3.

Figure 3.2: 2-Nodes in Topology 8. Each dimension represents one of the two nonzero interaction terms in the community matrix. Blue points represent 2-nodes; red points represent non 2-nodes. Note that blue points occur in a continuous region with no red points, indicating that the 2-node region is continuous and lacks holes.
Figure 3.3: Variance in Invasion Distance Within and Between Topologies. Each dimension represents one of the two nonzero interaction terms in the community matrix. The top two figures show 2-nodes (blue) and 3-nodes (green) of two instances of points in Topology 8. The bottom two figures show the same, for two instances in Topology 12.

While the shape of the green and blue point clouds are similar within a given topology (i.e. both rows of figures), the distance between green and blue points can vary significantly within a topology. When the between-variance and within-variance were quantified, they were found to be within a factor of 2 of one another. A sanity check was performed by rescaling the invasion distances against the magnitude of $r_1$, $r_2$, and $r_3$. No significant relationship between invasion distance and either of these variables was found for any of the topologies. As an example, Figure 3.4 shows invasion distance plotted on log scale against $r_1$ for topology 6. No significant trend can be observed.
Nevertheless, although the invasion distance was highly variant within any given topology, the invasion direction—defined as the direction of shortest path from 2-node to 3-node—was, in comparison, invariant. Invasion direction as defined this way varies significantly between topologies, but not significantly within a topology. This can be visualized using an angular histogram showing the invasion direction for topologies 6 and 8, as depicted in Figure 3.5.
3.6 Discussion and Future Work

The preceding results showed that while the magnitude of invasion distance doesn’t depend significantly on the topology, the direction of potential invasion is highly dependent on the topology. This is a reasonable result intuitively. The concept of community structure as used in this chapter constrains the direction of interactions between species in a community, but not their magnitude. So, it is reasonable that community structure, in turn, constrains the direction that system parameters must move in to cause an invasion, more than it does the magnitude.

To qualify the meaning of invasion distance and invasion direction, consider the example of a simple ecosystem consisting of a population of gazelles, a population of lions, and several other populations of other plant and animal species. Suppose the system is characterized by a community matrix with off-diagonal terms describing pairwise interactions between species, such as the effect of the gazelle population on the lion population (call this \(a_{12}\)) and the effect of the lion population on the gazelles (call this \(a_{21}\)). Now, suppose one were to ask the question, how must the interaction terms change in order for the gazelle population to become more abundant in the fraction of the carrying capacity of the ecosystem that they have assimilated? Now, if the system’s attractors are fixed point attractors, "more abundant" simply means moving the fixed point of interest to a new location at which the gazelle population is greater in size. If the system has more complex attractors such as limit cycles, then "more abundant" means moving the centroid of the attractor of interest to a new location at which the gazelle population is greater in size. In either case, the "direction" and "magnitude" of changes to the system’s parameters can be thought of as follows. Consider a Cartesian plane consisting of the axes \(a_{12}\) and \(a_{21}\). Suppose that each of these parameters is set to a given value, and that the system is at an attractor that is stable as long as \(a_{12}\) and \(a_{21}\) do not change. Then, suppose that one or both of these parameters change suddenly, representing either a mutation, a change in ecological strategy, or a change in the environmental conditions. If one were to ask, to what new values would \(a_{12}\) and \(a_{21}\) need to change in order for the gazelle population at the new resultant attractor to be more abundant, the meaning of invasion distance and direction become clear. The distance is simply the distance in the Cartesian plane between the new and the old values. The direction is the unit direction vector pointing from the original parameter values to the new ones.

An analytical approach, rather than a computational one, would be useful in elucidating these ideas further. By defining the gradient of the fixed point steady state solution \(\mathbf{x}^*\) with respect to the interaction terms \(a_{ij}\), it is possible to recover the invasion direction without
performing any simulations. Taking the gradient with respect to \( a_{11}, \ldots, a_{33} \) as follows:

\[
\nabla \left( \frac{x_i^*}{\|x^*\|} \right)
\]

(3.10)

gives a vector that points in the direction, in parameter space, that causes \( x_i \) to become more abundant. We can verify that the direction of this vector aligns with the invasion direction as previously defined, as depicted for example in Figure 3.6.

Future work will examine the gradient field defined above analytically, and examine how its properties vary across all 138 topologies. Possible questions include whether there are any correlations between the invasion direction (a vector) and the identity of a given topology (also expressible as a vector), and whether there are any other relationships between properties of the gradient field (such as any constraints on what quadrants the streamlines can point towards) and properties of the topology (such as the number of nonzero interaction terms). The result of this work will be to elucidate further trends in how invasibility and community structure are related.
Figure 3.6: Correspondence of Gradient Field With Invasion Direction. On the left are depicted 2-nodes (blue) and 3-nodes (green) from the analysis in section 3.5. On the right are plots of the gradient field describing the invasion direction. Arrows correspond to the direction that parameters must change in order to move from the blue points to the green points. Each dimension represents one of the two nonzero interaction terms in the community matrix.

3.7 References


Chapter 4

The Emergence of Trade-Offs in a Phage-Bacterium System

4.1 Abstract

This chapter introduces a mathematical model for a phage-bacterium ecosystem. From an invasion biology standpoint, either the phage or the bacterium can be considered as an invader, depending on the ecosystem they inhabit and also on one’s point of view. A model was developed to describe an ecosystem containing three populations: phage-infected bacteria, uninfected bacteria, and an environmental reservoir of free bacteriophages. Five system parameters were used as inputs to the model: intrinsic host death rate, virus degradation rate, environmental transmission rate, lysis rate, and burst size. The model was set up to capture logarithmic growth of bacteria, as well as both lytic and lysogenic life stages of the phage. All pairwise relationships between the five system parameters were investigated, by examining three payoff functions plotted as contour surfaces over each pair of parameters. The payoff functions used were $R_0$, the basic epidemiological reproduction number; $P_1$, the maximum phage population size over a simulated time period; and $P_2$, the average phage population size over a simulated time period. The resulting data showed that while $R_0$ could only capture monotonic relationships between system parameters and the payoff, $P_1$ and $P_2$ could capture more complex relationships such as intermediate maxima. For three of the system parameters, payoff relationships were always monotonic, such that a high burst size, low virus degradation rate, and low host death rate always resulted in the highest payoff. In contrast, for lysis rate and environmental transmission rate, phage payoffs $P_1$ and $P_2$ always reached a maximum for intermediate parameter values. When lysis rate and environmental transmission rate were plotted against one another, a significantly more complex payoff surface emerged both for $P_1$ and $P_2$. This result had an interesting implication for the classic "trade-off" hypothesis in the theory of virulence evolution, which maintains that virulence (i.e. lysis rate) and transmission rate are coupled. While no coupling relationship was assumed in our model, the relationship between lysis rate and transmission rate was markedly the most complex out of all pairwise
relationships considered, thus validating the substantial amount of interest historically in this relationship as opposed to other relationships between system parameters. Finally, conditions were found for a stable steady-state solution representing a desirable outcome in a phage therapy context. These conditions were overlayed upon the phage payoff surfaces, and the location of the desired stability conditions was compared with the location of the optimum phage payoff. The stability region and payoff optima were found to always be co-located with respect to the virus degradation rate, whereas for the host death rate and the lysis rate, the stability region was located at higher parameter values than the payoff optima.

4.2 Background

Trade-offs between life history traits are ubiquitous in biology and serve to reduce the dimensionality of many problems in evolutionary biology by constraining the set of phenotypic options available to an organism [1,2]. Classic examples include the trade-off of speed for stamina, offspring size for number, and reproduction for survival [1–4]. Trade-offs can arise probabilistically (e.g. the tallest individual in a population will rarely be the lightest in pigment) [5], or as the result of direct or indirect coupling between traits. Direct coupling can be the outcome of fundamental physics, e.g. in the trade-off between speed versus size or between the rate and yield of ATP production [6], or the outcome of an allocation of limited resources, for example the trade-off between flower size and number [7] or fertility versus longevity [1]. Indirect coupling occurs when two traits both are functions of a third underlying trait; in the special case in which this underlying trait is a gene, this coupling is called pleiotropy [3]. Although trade-offs generally are understood to imply a negative correlation between traits, recent studies have used the term to describe any scenario in which traits are directly or indirectly coupled [2,8].

In the theory of virulence evolution, which attempts to explain how parasites evolve, trade-offs have been a source of significant debate over the last several decades [9]. The classic but contested notion in virulence theory—the "trade-off hypothesis"—is that there is a saturating relationship between the host-to-host transmission rate ($\beta$) and the parasite-induced host death rate ($\alpha$), which is frequently defined as synonymous with virulence. The hypothesis states that any increase in $\beta$, which increases parasite propagation, must be won at the cost of an increase in $\alpha$, which decreases parasite propagation between hosts [9–11]. The standard explanation for this is that both $\alpha$ and $\beta$ are increasing functions of parasite density within the host: greater numbers of parasites increase the probability of host-to-host transmission but also, separately, increase the probability of an earlier host death [11,12]. Thus the parasite must choose between a low transmission rate lasting for a long lifespan
within the host, or a high transmission rate lasting a short lifespan within the host—one key assumption being that deceased hosts quickly lose their transmission ability after death.

An abundance of papers cite this relationship and employ, as part of mathematical models, some increasing function for $\beta(\alpha)$, sometimes parameterized from empirical data but most often allowed to be arbitrary [12–16]. The theory of adaptive dynamics says that if $\beta(\alpha)$ is a concave function, as it is sometimes assumed to be, then there exists a single, evolutionarily stable virulence level which necessarily has an intermediate (i.e. nonzero and noninfinite) value [9]. This has been cited as one possible explanation for the fact that virulence levels found in nature are not as high as they could be [11], and indeed the trade-off hypothesis in this form (positing an increasing, saturating relationship between virulence and transmissibility) has eclipsed the older idea that parasites should always evolve towards avirulence, in order to, so to speak, "farm" their hosts [9].

Nevertheless, many critiques have been leveled against the trade-off hypothesis, including: a lack of evidence especially for whether the trade-off relationship is saturating [9,11,17]; evidence that in some systems transmission rate and virulence are not correlated [10]; the possibility that a trade-off surface exists in more than simply two dimensions [11]; and finally the need to incorporate system-specific biological details, such as varying transmission modes, processes such as superinfection, and host properties such as host heterogeneity, immune response, and spatial structure [9–11,18]. Moreover, the classic trade-off hypothesis assumes that transmission occurs directly via host-to-host contact, rather than via an environmental reservoir of virions as is the case with most phage-bacterium systems. While there is a good mechanistic explanation for a trade-off via direct transmission (high virulence leads to earlier host death), there is no similar explanation for a trade-off via environmental transmission. Indeed in models that include both environmental and host-to-host transmission, authors have repeatedly incorporated a trade-off only with respect to host-to-host transmission, leaving the relationship between virulence and environmental transmission rate uncoupled [8, 19].

In phage-bacterium systems, numerous trade-offs besides the classic transmissibility-virulence trade-off have been proposed, and at least three are worth mentioning. First, for lysogenic phages, is a natural relationship between burst size and lysis rate: the greater the residence time a phage has inside its host, the more virus particles it can produce [20]. Second, for temperate phages, is a coupling between lysis rate and intrinsic host death rate: the worse the environment for the host, the less the phage benefits by remaining in lysogeny, due to the decreased reproductive capacity of the host [21]. Third, in systems where the hosts can recover, is a coupling between virulence and host recovery rate, proposed to be an increasing relationship based on modeling by [22] and shown to be an inverse relationship based on empirical data by [23]. Other potential relationships have been investigated.
empirically by [1].

4.3 Objectives

In this study, a phage-bacterium system was modeled in which no trade-offs are a priori assumed. Instead of using trade-offs to predict evolutionary outcomes, as is normally done, the objective here was to determine whether any trade-offs emerge naturally by maximizing a set of payoff functions for the success of the phage. Five parameters in a phage-bacterium system were modeled: intrinsic host death rate, virus degradation rate, transmission rate from an environmental reservoir, lysis rate, and burst size. Transmission was assumed to occur only via the reservoir rather than directly from host to host. Of the three relevant trade-offs proposed in the literature—environmental transmission rate vs. virulence, burst size vs. lysis rate, and intrinsic host death rate vs. lysis rate—only one, that between burst size and lysis rate, has a tenable a priori mechanistic explanation, based on the fact that it takes time to produce virions within the host [20]. Thus, an experiment was designed to assess the existence not just of these three relationships proposed previously, but of all pairwise relationships between the five system parameters, \( \binom{5}{2} = 10 \) in total.

The principle behind utilizing payoff functions is as follows. While observational and laboratory studies can probe for the existence of fundamentally coupled relationships between life history traits of a virus and host, theoretical models must either load these relationships from experimental data, choose them arbitrarily, or not include them at all [1, 8, 9, 13–17, 19, 21, 22, 24–29]. Theoretical models can, however, identify well-defined payoffs, such as the basic phage reproduction number or the time averaged gain in phage population size, and delineate what constraints on the underlying system parameters are necessary for the payoff in question to reach a maximum. There is a subtle difference though between trade-offs arising from fundamentally coupled system parameters and trade-offs arising from this method of payoffs. Suppose there are two system parameters \( X \) and \( Y \), and a payoff function \( Z \). A coupled relationship would mean that for \( X \) to take on a certain value, \( Y \) must necessarily be constrained in some way. A payoff-induced constraint would mean that for \( Z \) to take on a certain value, \( X \) and \( Y \) must necessarily be jointly constrained in some way. Thus the payoff function serves in a similar fashion to indirect coupling between \( X \) and \( Y \), but instead of the traits being a function of some underlying entity, the payoff is a function of the underlying traits.

By examining three different payoff functions layered upon a dynamical model of a phage-bacterium system, implications for the classic trade-off hypothesis and for other potential trade-offs were elucidated. The explanatory power of the payoffs themselves was also evaluated.
Finally, a brief analysis relating to phage therapy was explored. Phage therapy attempts to control unwanted bacterial populations in eukaryotic hosts, by using bacteriophages as therapeutic agents. One of the practical challenges is successfully engineering or selecting for phage populations with the right transmission rate, degradation rate, and lysis rate parameters required to be effective. An important consideration for this is that the optimum parameters for phage therapy do not necessarily coincide with the optimum parameters with respect to various phage payoffs. Thus, a simple steady state solution representing a positive outcome for phage therapy was overlaid a subset of the payoff surfaces, in order to evaluate the degree to which the phage therapy optima and phage payoff optima coincide.

### 4.4 Methods

#### 4.4.1 Development of Mathematical Model

The phage-bacterium system was modeled using a system of nonlinear coupled ordinary differential equations. The model is functionally similar to the familiar $S$–$I$–$V$ models typically utilized to describe phage–bacterium ecosystems and for modeling phage therapy [30–32], but incorporates logistic growth behavior rather than unlimited exponential growth to allow for outcomes where the phage becomes "pseudolysogenic" [33–35]. The model incorporates vertical transmission from infected parent host cell to both daughter cells, and assumes that horizontal transmission occurs via an environmental reservoir rather than directly from host to host. The model further assumes that phage particles are released through lysis rather than constantly being shedded by infected cells.

The ecosystem is comprised of three populations: susceptible bacteria, phage-infected bacteria, and bacteriophages (viruses). These populations are denoted $S$, $I$, and $V$, respectively, and their behavior over time was modeled according to Equations 4.1–4.3:

$$
\frac{dS}{dt} = r_S S \left(1 - \frac{S + I}{K}\right) - \beta_S S - \rho SV \quad (4.1)
$$

$$
\frac{dI}{dt} = r_I I \left(1 - \frac{S + I}{K}\right) - \beta_I I + \rho SV - \alpha I \quad (4.2)
$$

$$
\frac{dV}{dt} = \delta \alpha I - \beta_V V - \rho SV \quad (4.3)
$$

In this model, the growth of both susceptible and infected bacterial populations is treated...
as logistic growth with a growth rate of $\tilde{r}_S$ and $\tilde{r}_I$ respectively, and carrying capacity $K$. The uninfected and phage-infected bacteria are assumed to be in neutral competition for the same resources, but not directly harming or benefiting each other. This neutral competition is modeled by the negative feedback term $\frac{S + I}{K}$ in the model, which incorporates a nutrient-limited growth constraint on the bacterial population.

The uninfected host population is depleted via a linear intrinsic death rate ($\bar{\mu}_S$), which can include natural and induced (e.g. through antibiotics) host cell death. Additionally, infection with phages reduces the population of susceptible bacteria at a rate proportional to $\tilde{\rho}SV$. This transmission process is modeled per the law of mass action, with the rate of infections per unit time proportional to the concentrations of free phages ($V$) and the available uninfected population ($S$), with $\tilde{\rho}$ as the constant of proportionality.

The population of infected bacteria increases via the addition of newly infected bacteria as well as via its own logistic growth. The infected population decreases through the phage induced lysis rate ($\tilde{\alpha}$) and also through the host intrinsic death rate describing natural and induced attrition ($\bar{\mu}_I$). Both of these effects are proportional to the population size.

Each time an infected bacterium is lysed, a burst of free virions is produced, of size $\delta$, the burst size. These virions are immediately released into the environment. The virion (phage) population in the system decreases both through linear degradation ($\bar{\mu}_V$) and through new infections of susceptible bacteria ($\tilde{\rho}SV$).

To improve the computational tractability of the model, rate variables were normalized with respect to $\tilde{r}_S$. The relative reproduction and degradation rates of the bacteria were assumed to be unaffected by phage infection ($\tilde{r}_S = \tilde{r}_I$ and $\bar{\mu}_S = \bar{\mu}_I$). The population variables were normalized with respect to $K$ (by setting $K = 1$), so that the quantities $S$ and $I$ could be interpreted in terms of a fraction of the carrying capacity $K$. Finally, dimensional time $t$ was replaced by nondimensional time $\tau$, by substituting $\tilde{r}_S dt = d\tau$. The resulting system is thus simplified to Equations 4.4–4.6:

$$\frac{dS}{d\tau} = S (1 - S - I) - \mu_S S - \rho SV \quad (4.4)$$
$$\frac{dI}{d\tau} = I (1 - S - I) - \mu_S I + \rho SV - \tilde{\alpha} I \quad (4.5)$$
$$\frac{dV}{d\tau} = \delta\tilde{\alpha} I - \mu_V V - \rho SV \quad (4.6)$$

in which the terms $\rho$, $\alpha$, $\mu_S$ and $\mu_V$ are the $\tilde{r}_S$-normalized counterparts of barred terms $\tilde{\rho}$, $\tilde{\alpha}$, $\bar{\mu}_S$ and $\bar{\mu}_V$ in Equations 4.1–4.3. The parameters in this simplified model are presented in Table 4.1. Since lysis rate is the only source of virulence that is measureable in the model, the terms lysis rate and virulence are treated in this model as synonymous. The terms infection rate and environmental transmission rate are also treated as synonymous, since
both describe infection from an environmental reservoir.

Table 4.1: Parameters and State Variables in the Phage-Bacterium Model. Parameters are held constant over time, whereas state variables are time-varying. A dimensionless unit is denoted by #.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
<th>Unit</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_S$</td>
<td>$\frac{\rho_s}{r_s}$</td>
<td>#</td>
<td>Host intrinsic death rate</td>
</tr>
<tr>
<td>$\mu_V$</td>
<td>$\frac{\rho_V}{r_s}$</td>
<td>#</td>
<td>Virus degradation rate</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>$\frac{\delta}{r_s}$</td>
<td>#</td>
<td>Lysis rate (virulence)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$\frac{\rho}{r_s}$</td>
<td>#</td>
<td>Infection rate</td>
</tr>
<tr>
<td>$\delta$</td>
<td>$\delta$</td>
<td>#</td>
<td>Burst size</td>
</tr>
<tr>
<td>$S$</td>
<td>$S$</td>
<td>#</td>
<td>Susceptible host population size</td>
</tr>
<tr>
<td>$I$</td>
<td>$I$</td>
<td>#</td>
<td>Infected host population size</td>
</tr>
<tr>
<td>$V$</td>
<td>$V$</td>
<td>#</td>
<td>Phage population size</td>
</tr>
<tr>
<td>$\tau$</td>
<td>$\tau$</td>
<td>#</td>
<td>Non-dimensional time</td>
</tr>
</tbody>
</table>

4.4.2 Implementation of Payoff Functions

Basic Reproduction Number

Three different payoff functions were developed. The first payoff considered is the basic reproduction number, $R_0$, a concept used in epidemiology to describe the fitness of a parasite spreading through a population of susceptible hosts [9,36–38]. For systems in which transmission occurs exclusively from host to host (i.e. no environmental transmission), the biological definition of $R_0$ is simple: it is defined as the average number of secondary infections caused by a single infected host introduced into a very large disease-free population, over a time period equal to the infectious lifespan of the single introduced host [39]. If $R_0 < 1$, the outbreak will die off; if $R_0 > 1$, it will persist in the population. For systems in which transmission also occurs via an environmental reservoir of parasites, the "size" of the outbreak includes not just the number of infected individuals, but also the size of the environmental reservoir at a given moment in time. This can be captured by introducing a
matrix $K$, called the Next Generation Matrix, which allows the outbreak to be compartmentalized into different disease populations, or types. Each disease type is associated with a unique biological meaning such as age group, infection route, or life history stage (e.g. free-living or host-associated) of the parasite \[37, 40, 41\]. In our model the two disease types are $I$, the population of infected hosts, and $V$, the population of free virions. The matrix $K$ is defined such that each element $K_{ij}$ denotes the average number of secondary infections of type $i$ that are produced by a single individual of type $j$ introduced into a large disease-free population, over a time period equal to the lifespan of the single initial individual. The spectral radius of $K$, equal to the magnitude of its largest eigenvalue, is then defined to be equal to the basic reproduction number $R_0$ \[9, 36–38, 40, 41\]. Intuitively, this is equivalent to treating $K$ as a linear operator on the vector of all disease populations in the system, such that $R_0$ is the largest possible scale factor accompanying the action of the system of governing equations on the disease populations at an initially disease-free steady state. The matrix $K$ can be derived analytically for any system governed by maps or ordinary differential equations (ODEs); the derivation for ODEs is presented in Appendix C. In our system, $R_0$ is obtained via this method (see Appendix C) and presented in Equation 4.7:

$$R_0 = \frac{\mu V - \mu S \mu V + \rho \mu S (1 + \alpha \delta - \mu S)}{\mu V (\alpha + \mu S)(\mu V + \rho \mu S)}$$ \hspace{1cm} (4.7)

**Numerical Payoff Functions**

In addition to the basic reproduction number $R_0$, two other payoffs were formulated, designed to more realistically capture the reproductive payoff to the phage over the course of an outbreak. While $R_0$ is defined specifically to capture the system’s linearized behavior over a short period of time after the introduction of a single bacteriophage, the payoffs $P_1$ and $P_2$ are computed by actually solving the system numerically over a time period of interest, using $S_0 = 0.1$, $I_0 = 0$, and $V_0 = 10$ as initial conditions. The numerical solution $V(\tau)$ was then used as an input into the payoff function. The first payoff, $P_1$, measures the maximum free phage population occurring over the time period of interest. The second payoff, $P_2$, measures the average free phage population over that same period. In both cases the payoff measures free virus particles but not virions within infected host cells. Both are defined in Equations 4.8–4.9, in which $\tau$ is nondimensional time and the interval $[0, k]$ is the time period of interest:
\[ P_1 = \max(V(\tau)|_{0,k}) \]  
\[ P_2 = \frac{1}{k} \int_{0}^{k} V(\tau) d\tau \]  

**Simulations**

The payoffs \( R_0, P_1, \) and \( P_2 \) were computed for a multidimensional grid of parameter values. For each parameter \( (\mu_S, \mu_V, \alpha, \delta, \) and \( \rho) \), a range of feasible values was taken from the literature, as shown in Table 4.2. All 10 pairwise relationships between parameters were investigated as follows. For a given pair of parameters, a two-dimensional grid of \( 10^4 \) evenly spaced points spanning both relevant parameter ranges was generated. Then, each payoff function was plotted as a two-dimensional contour surface over this grid, holding the other three parameters in the system constant. In order to evaluate whether the geometries of the payoff surfaces were invariant to changes in the other three parameters, each contour plot was redrawn over 100 times, each time with the three extra parameters set to different values. These values were drawn from a second, three-dimensional grid of evenly spaced points spanning the relevant parameter ranges of the three extra parameters, with the exact number of points depending on if the endpoints in the parameter ranges were included or excluded from the grid, as shown in Table 4.2. The result was that for each pairwise relationship between two parameters, three series of plots were generated, one for each payoff function. Each plot series was plotted over the same axes (corresponding to the two pairwise parameters investigated), and each plot in the series had different values set for the three additional parameters not captured in the plot axes, as explained above.

Each plot series was then analyzed by characterizing the payoff surfaces visually, one axis at a time. In almost all cases the payoff surfaces were simple to describe. With respect to a given axis of a plot, payoff surfaces were categorized as follows: monotonically increasing, monotonically decreasing, taking on an intermediate maxima or minima, or nontrivial. The final, nontrivial category was a catch-all for all surfaces that didn’t fit neatly into the other three. After categorizing both axes of each plot series in this way, the full set of categorizations was analyzed for patterns.

**Steady State Solutions**

Finally, by solving for \( \frac{dS}{d\tau} = \frac{dI}{d\tau} = \frac{dV}{d\tau} = 0 \), steady state solutions to Equations 4.4–4.6 were obtained. It was recognized that one of these solutions, the solution \( S = V = I = 0 \), represents a desirable outcome in a phage therapy context. Other nontrivial solutions contained either a nonzero \( S \) population or a nonzero \( I \) population at steady state, thus
Table 4.2: Parameter Ranges. Adapted from measured and estimated values in [31,42–45].

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_S$</td>
<td>(0,0.1)</td>
<td>Bacterial death rate of up to 10% the growth rate</td>
</tr>
<tr>
<td>$\mu_V$</td>
<td>(1,10)</td>
<td>Phages have a degradation rate greater than or equal to bacterial reproduction rates</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>(0,2)</td>
<td>Lysis rate of up to two times the bacterial growth rate</td>
</tr>
<tr>
<td>$\delta$</td>
<td>(0,200)</td>
<td>Number of infective viral particles released per lysis event</td>
</tr>
<tr>
<td>$\rho$</td>
<td>(0,5)</td>
<td>Rate of new infections per unit phage and target host numbers, allowed to be up to five times the bacterial growth rate</td>
</tr>
</tbody>
</table>

limiting their usefulness for phage therapy. A classical stability analysis on the $S = V = I = 0$ solution was performed, and it was deduced that for the solution to be a stable node, the following conditions must be met:

$$
\delta > 0 \quad (4.10)
$$

$$
\rho > 0 \quad (4.11)
$$

$$
\mu_V > 0 \quad (4.12)
$$

$$
0 < \mu_S < 1 \quad (4.13)
$$

$$
\alpha = 1 - \mu_S + \mu_V \quad (4.14)
$$

The first four of these conditions are met by default, simply by constraining the system parameters to the ranges reported in the literature and presented in Table 4.2. Thus the fifth condition is the one which, in practice, determines the stability of the solution in question. This condition, $\alpha = 1 - \mu_S + \mu_V$, was plotted as a line overlayed on each payoff surface. Where possible, the location of the payoff maxima was compared against the location of this line.

### 4.5 Results

For all series of plots, the payoff surfaces were consistent in their shape, in other words invariant to changes in the value of the three extra system parameters. Moreover, the payoffs $P_1$ and $P_2$ were consistent with one another in every plot series. Three of the
system parameters (host death rate, virus degradation rate, and burst size) were found to vary strictly monotonically for all three payoffs. Payoffs $P_1$ and $P_2$ were always greatest with the smallest possible host death rate, the smallest possible virus degradation rate, and the largest possible burst size. Payoff $R_0$ was greatest with the largest possible host death rate, the smallest possible virus degradation rate, and the largest possible burst size. For the remaining two parameters, lysis rate and environmental transmission rate, the payoff surfaces were more complex. Firstly, there was a disagreement between the $R_0$ payoff surface and the $P_1$ and $P_2$ payoff surfaces. While $P_1$ and $P_2$ showed that the highest payoff to the phage is achieved when both lysis rate and transmission rate are set to intermediate values (i.e. highest payoff at an intermediate maxima), $R_0$ could only ever capture a monotonically increasing or monotonically decreasing payoff surface. Secondly, when lysis rate and environmental transmission rate were plotted against one another, a substantially more complex payoff surface resulted, one which could not be described by a monotonic relationship or an intermediate maxima alone. Finally, when the phage therapy stability condition $\alpha = 1 - \mu_s + \mu_v$ was overlayed on the payoff surfaces $P_1$ and $P_2$, it always appeared at higher lysis rate parameter values than the payoff optima.

Figure 4.1 shows two contour plots of the payoff surface $P_1$, drawn over axes representing transmission rate and lysis rate. Each plot has different parameter values for the three parameters (host death rate, virus degradation rate, burst size) not plotted. While the location of the payoff surfaces in the two plots is different, their shape is invariant. This invariance holds across all plots in the plot series. A similar invariance holds for other the plot series (i.e. payoff surfaces plotted against other parameter axes) too.

Figures 4.2–4.3 show all payoffs plotted, portraying one representative figure per series. For each plot, the relationship between the payoff and both parameter axes is described. The keywords "high", "low", "intermediate", and "nontrivial" are used to describe relationships that are monotone increasing, monotone decreasing, taking on an intermediate maxima, and nontrivial, respectively. These relationships are summarized and presented together in Table 4.3.
Transmission rate ($\rho$) vs. Lysis rate ($\alpha$)

$\{\mu_S \rightarrow 0.1, \mu_V \rightarrow 10.0, \delta \rightarrow 200.\}$

Stable Cure Region: $\alpha \approx 10.9$

Transmission rate ($\rho$) vs. Lysis rate ($\alpha$)

$\{\mu_S \rightarrow 0.1, \mu_V \rightarrow 10.0, \delta \rightarrow 40.0\}$

Stable Cure Region: $\alpha \approx 10.9$

Figure 4.1: Invariance of Payoff Surfaces. The payoff $P_1$ is plotted over axes representing transmission and lysis rates. The shape of the payoff surface is invariant towards changes in the three extra parameters (host death rate, virus degradation rate, and burst size).
$R_0 \sim (\text{high } \delta, \text{ low } \mu_V)$

Burst size ($\delta$) vs. Virus death rate ($\mu_V$)

{\(\mu_S \rightarrow 0.08, \rho \rightarrow 1, \alpha \rightarrow 0.4\)}

Stable Cure Region: $\mu_V = -0.52$

$R_0 \sim (\text{low } \mu_V, \text{ high } \mu_S)$

Virus death rate ($\mu_V$) vs. Host death rate ($\mu_S$)

{\(\rho \rightarrow 3, \alpha \rightarrow 0.8, \delta \rightarrow 120\)}

Stable Cure Region: $\mu_V = -0.2 + \mu_S$

Figure 4.2: Payoff $R_0$ plotted over all pairs of parameters. Above each plot, the relationship between the payoff and both parameter axes is described. The keywords "high", "low", "intermediate", and "nontrivial" are used to describe relationships that are monotone increasing, monotone decreasing, taking on an intermediate maxima, and nontrivial, respectively. (Continued on next page.)
Figure 4.2: (Continued.)
Figure 4.2: (Continued.)
Figure 4.3: Payoffs $P_1$ (left) and $P_2$ (right) plotted over all pairs of parameters. Above each plot, the relationship between the payoff and both parameter axes is described. Relationships are either monotone increasing ("high"), monotone decreasing ("low"), taking on an intermediate maxima ("intermediate"), or undescrivable in such simple terms ("nontrivial").
Figure 4.3: (Continued.)
Figure 4.3: (Continued.)
Figure 4.3: (Continued.)
Figure 4.3: (Continued.)
Table 4.3: Relationship between Payoff Functions and System Parameters. The keywords "high", "low", "intermediate", and "nontrivial" are used to denote relationships that are monotone increasing, monotone decreasing, taking on an intermediate maxima, and nontrivial, respectively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Payoff</th>
<th>Plotted Against:</th>
<th>μ&lt;sub&gt;S&lt;/sub&gt;</th>
<th>μ&lt;sub&gt;V&lt;/sub&gt;</th>
<th>ρ</th>
<th>α</th>
<th>δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host Death Rate (μ&lt;sub&gt;S&lt;/sub&gt;)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;1&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>P&lt;sub&gt;2&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>R&lt;sub&gt;0&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Virus Degradation Rate (μ&lt;sub&gt;V&lt;/sub&gt;)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;1&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>P&lt;sub&gt;2&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>R&lt;sub&gt;0&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Infection Rate (ρ)</td>
<td></td>
<td></td>
<td>Intermediate</td>
<td>Intermediate</td>
<td>Nontrivial</td>
<td>Intermediate</td>
<td>Intermediate</td>
</tr>
<tr>
<td>P&lt;sub&gt;1&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>Intermediate</td>
<td>Intermediate</td>
<td>Nontrivial</td>
<td>Intermediate</td>
<td>Intermediate</td>
</tr>
<tr>
<td>P&lt;sub&gt;2&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R&lt;sub&gt;0&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>Nontrivial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lysis Rate (α)</td>
<td></td>
<td></td>
<td>Intermediate</td>
<td>Intermediate</td>
<td>Nontrivial</td>
<td>Intermediate</td>
<td>Intermediate</td>
</tr>
<tr>
<td>P&lt;sub&gt;1&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>Intermediate</td>
<td>Intermediate</td>
<td>Nontrivial</td>
<td>Intermediate</td>
<td>Intermediate</td>
</tr>
<tr>
<td>P&lt;sub&gt;2&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>Nontrivial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R&lt;sub&gt;0&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>Low</td>
<td>Low</td>
<td>Nontrivial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burst Size (δ)</td>
<td></td>
<td></td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;1&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P&lt;sub&gt;2&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R&lt;sub&gt;0&lt;/sub&gt;</td>
<td></td>
<td></td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.6 Discussion and Further Work

To conclude, there was a surprising consistency between payoffs $P_1$ and $P_2$ over all plot series. Payoffs were always highest when host death rate and virus degradation rate were low, and when burst size was high. The basic reproduction number $R_0$ was able to capture these relationships consistently as well, with the exception of host death rate, for which $R_0$ was highest at high rather than low parameter values. For lysis rate and environmental transmission rate, however, the highest payoffs $P_1$ and $P_2$ were achieved for intermediate parameter values. $R_0$ was unable to capture this, as it was only able to express monotone increasing or monotone decreasing relationships between the payoff and underlying system parameters. When lysis rate and environmental transmission rate were plotted against each other, significantly more complex payoff surfaces were generated by the functions $P_1$ and $P_2$. This suggests that lysis rate and transmission rate are coupled jointly to the payoffs $P_1$ and $P_2$, whereas the other system parameters are coupled independently (i.e. changing one parameter doesn’t affect the payoff gained by the others). These findings validate the significant attention given to the relationship between lysis rate (virulence) and transmission rate in the literature on virulence evolution.

A significant limitation of the use of $R_0$, besides its demonstrated limited expressivity, is that when used in systems with multiple disease states, it only describes direct transitions from one disease state to another. That is, $R_0$ cannot accurately describe transmission patterns which involve chains of infection that have multiple steps. In our model, one important chain of infection is the transition from $V$ to $I$ and back to $V$ (i.e. infection followed by lysis). To express this in a single "reproduction number", another quantity, the Type Reproduction Number, has been recently developed by Heesterbeek and Roberts [40, 41], which can be utilized to describe a separate reproduction number for each disease type in the system. Future work could include an additional plot series using the Type Reproduction Number for $V$ in our system.

An additional outcome of our model was a limited result concerning phage therapy. The stability condition $\alpha = 1 - \mu_S + \mu_V$ was always located at higher parameter values of lysis rate than was optimum to achieve maximum payoff $P_1$ or $P_2$. The implication is that bacteriophages benefit by evolving towards lower lysis rates than would be beneficial therapeutically in combating unwanted bacteria in eukaryotic hosts, which supports the classical idea that parasites in general evolve towards avirulence.

One final direction for future work is to examine the interplay between phage payoffs and payoffs to the infected and uninfected hosts. A short summary of initial results on potential payoffs to $S$ and $V$ is presented in Appendix B.
4.7 References


Appendix A

Additional Sourdough Results

Growth of *L. sanfranciscensis* on Different Sugars

An additional experiment was run with *L. sanfranciscensis*, to evaluate its preferential uptake of different sugars. sMRS broth was prepared with six different sugar levels: no sugar, fructose, glucose, maltose, both fructose and glucose, and both fructose and maltose. When a single sugar was used, the formulation was 20 g per liter of broth. When two sugars were used, 6.66 g and 13.33 g were used. Fructose was always used in the smaller amount, since in theory it is used as an electron acceptor by *L. sanfranciscensis* rather than as a carbon source. The level of yeast extract and malt extract were kept to a minimum in order to allow maximum separation of growth rates between the cultures with sugars added and the culture with none. *L. sanfranciscensis* was first grown overnight in sMRS broth with both maltose and fructose (the medium with the theoretically highest growth rate). Then, after inoculating into the broths with different sugars, growth was monitored turbidimetrically. Figure A.1 shows that while there was negligible growth in the culture with no sugars, all of the other cultures demonstrated very similar growth rates.

![Figure A.1: Growth of *L. sanfranciscensis* on Different Sugars](image-url)

99
**Volumetric Rise of S. cerevisiae**

Rapid-rise yeast (*S. cerevisiae*) is perhaps the fastest-growing baker’s yeast that is sold as a consumer product. Sourdough yeasts are typically slower growing, but resist freeze-drying, flash-freezing, or other packaging methods that would make them amenable to be sold commercially. To test if baker’s yeast could potentially be used in a synthetic sourdough starter (i.e. *L. sanfranciscensis* plus *S. cerevisiae*), we added dry granules of rapid-rise yeast to a series of doughs, and measured the associated volumetric rise of the dough. This was compared with baseline dough rise data taken on the doughs made from Champaign, King Arthur, and BreadTopia sourdough starters in Section 2.6. As shown in Figure A.2, even an inoculation size of 0.01% baker’s yeast into the dough, by weight, rises noticeably faster than the sourdoughs (which are very vigorous sourdoughs), thus not giving the sourdough LAB enough time to acidify the dough. This reinforces the logic behind the industry practice of adding baker’s yeast to doughs right at the end of fermentations when ample acidification has already occurred.

![Figure A.2: Comparison of Leavening Activity of Baker’s Yeast to Wild Sourdough Yeasts.](image)

This figure shows the volumetric rise of doughs with baker’s yeast added as a percentage of the weight of the dough, compared to the best fit curve from the type I sourdoughs used in Section 2.6. Even the smallest baker’s yeast inoculation size results in significantly faster leavening than the sourdoughs.
Appendix B

Additional Phage Results

Additional payoff surfaces measuring the payoffs to $S$ (susceptible bacteria) and $I$ (infected bacteria) were defined as follows:

$$P_{\text{Inf1}} = \max(I(\tau)|_{0,k})$$  \hspace{1cm} (B.1)
$$P_{\text{Susc1}} = \max(S(\tau)|_{0,k})$$  \hspace{1cm} (B.2)
$$P_{\text{Inf2}} = \frac{1}{k} \int_{0}^{k} I(\tau)d\tau$$  \hspace{1cm} (B.3)
$$P_{\text{Susc2}} = \frac{1}{k} \int_{0}^{k} S(\tau)d\tau$$  \hspace{1cm} (B.4)

Tables B.2–B.1 summarize the relationship between these payoffs and all five parameters, deduced via visual inspection of the contour surfaces. The keywords "high", "low", and "intermediate" are used to describe relationships that are monotone increasing, monotone decreasing, and taking on an intermediate maxima, respectively.

Remarkably, the relationship between system parameters and the two payoffs for infected bacteria were consistent across all plots. A low host death rate, high virus degradation rate, and low burst size always gave rise to a higher payoff to infected bacteria. The relationship between system parameters and the two payoffs for susceptible bacteria, however, were not consistent at all. Nevertheless, the relationship between payoff to the infected bacteria and system parameters has interesting implications, when compared with the relationship between system parameters and the payoffs to the viruses. Firstly, both virus and infected bacteria benefited when the host intrinsic death rate is low. This is reasonable since a higher host death rate is obviously bad for bacteria whether infected or not, and it is bad for the phage since it gives the phage less control over the total effective host death rate, which includes the background death rate as well as the virus induced lysis rate. The virus can increase the effective host death rate but not decrease it; thus, a low intrinsic host death rate affords the phage greater flexibility. The second parameter, virus death rate, gave rise to the highest virus payoff at low values, and to the highest infected bacteria payoff at high values. That a low virus death rate would be beneficial to viruses is obvious. High virus
death rates are beneficial to infected bacteria because it increases the value of the infected bacteria to the virus—the greater the virus death rate in the environment, the greater the incentive to keep infected hosts alive. For the final parameter, burst size, a similar argument can be made. High burst size is obviously beneficial to the phage, but low burst size is beneficial to infected bacteria because it increases the relative value of keeping infected bacteria alive (i.e. reproducing rather than lysing).

Table B.1: Relationship between Payoff Functions ($P_{\text{Inf1}}, P_{\text{Inf2}}$) and System Parameters. The keywords "high", "low", and "intermediate" are used to denote relationships that are monotone increasing, monotone decreasing, and taking on an intermediate maxima, respectively. A dash is used when the relationship was nontrivial, inconsistent across the plot series, or not possible to discern.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Payoff</th>
<th>Plotted Against:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_s$</td>
<td>$\mu_v$</td>
</tr>
<tr>
<td>Host Death Rate ($\mu_s$)</td>
<td>$P_{\text{Inf1}}$</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{Inf2}}$</td>
<td>—</td>
</tr>
<tr>
<td>Virus Degradation Rate ($\mu_v$)</td>
<td>$P_{\text{Inf1}}$</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{Inf2}}$</td>
<td>High</td>
</tr>
<tr>
<td>Infection Rate ($\rho$)</td>
<td>$P_{\text{Inf1}}$</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{Inf2}}$</td>
<td>Low</td>
</tr>
<tr>
<td>Lysis Rate ($\alpha$)</td>
<td>$P_{\text{Inf1}}$</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{Inf2}}$</td>
<td>Low</td>
</tr>
<tr>
<td>Burst Size ($\delta$)</td>
<td>$P_{\text{Inf1}}$</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{Inf2}}$</td>
<td>Low</td>
</tr>
</tbody>
</table>
Table B.2: Relationship between Payoff Functions ($P_{\text{Suscl}}$, $P_{\text{Susc2}}$) and System Parameters. The keywords "high", "low", "intermediate", and "nontrivial" are used to denote relationships that are monotone increasing, monotone decreasing, and taking on an intermediate maxima, respectively. A dash is used when the relationship was nontrivial, inconsistent across the plot series, or not possible to discern.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Payoff</th>
<th>Plotted Against:</th>
<th>$\mu_S$</th>
<th>$\mu_V$</th>
<th>$\rho$</th>
<th>$\alpha$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host Death Rate ($\mu_S$)</td>
<td>$P_{\text{Suscl}}$</td>
<td>High</td>
<td>—</td>
<td>High</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{Susc2}}$</td>
<td>Low</td>
<td>Low</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Virus Degradation Rate ($\mu_V$)</td>
<td>$P_{\text{Suscl}}$</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{Susc2}}$</td>
<td>Low</td>
<td>Low</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Infection Rate ($\rho$)</td>
<td>$P_{\text{Suscl}}$</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{Susc2}}$</td>
<td>High</td>
<td>High</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Lysis Rate ($\alpha$)</td>
<td>$P_{\text{Suscl}}$</td>
<td>High</td>
<td>High</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{Susc2}}$</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Burst Size ($\delta$)</td>
<td>$P_{\text{Suscl}}$</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{Susc2}}$</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>—</td>
</tr>
</tbody>
</table>
Appendix C

Next Generation Matrix Methods

Derivation of $R_0$ and $T$

In a system with multiple disease types, the Next Generation Matrix $K$ can be used to define the basic reproduction number $R_0$ [1–3]. Suppose the governing equations have the form

$$\frac{dx_i}{dt} = f_i(x)$$

(C.1)

and suppose that $x_1, \ldots, x_k$ are "disease" states, e.g. free phage particles or infected hosts, and that $x_{k+1}, \ldots, x_m$ are "disease-free" states, e.g. uninfected hosts. The system’s Disease-Free Equilibrium (DFE), if it exists, can be found by solving for:

$$x_1 = \cdots = x_k = 0$$

(C.2)

$$\frac{dx_{k+1}}{dt} = \cdots = \frac{dx_m}{dt} = 0$$

(C.3)

giving the solution

$$x^* = (0, \ldots, 0, x_{k+1}^*, \ldots, x_m^*)$$

(C.4)

Next, a subsystem of equations, termed the disease subsystem, is defined as:

$$\begin{bmatrix}
\frac{dx_1}{dt} \\
\vdots \\
\frac{dx_k}{dt} \\
\end{bmatrix} = \begin{bmatrix}
f_1(x) \\
\vdots \\
f_k(x) \\
\end{bmatrix} = \mathcal{F} - \mathcal{V}$$

(C.5)

The subsystem is separated into two parts, $\mathcal{F}$ and $\mathcal{V}$. The first part, $\mathcal{F}$, describes the rate of appearance of new individuals into a given disease compartment from a non-disease compartment (e.g. birth of an infected daughter cell), whereas the second part, $\mathcal{V}$, describes the rate of transfer of individuals from one disease compartment to another (e.g. transfer of a phage particle from the environment to inside an infected host), or from a disease
compartment back to a non-disease compartment (e.g. recovery) or to death. This is done in order to separate the generation of secondary infections that increase the total size of the outbreak ($F$), from other events in the disease transmission cycle ($V$). The Jacobian of the subsystem is then evaluated at the DFE, and split into two components $F$ and $V$ corresponding to $F$ and $V$:

$$J_{DFE} = \left[ \begin{array}{ccc} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k(x)}{\partial x_1} & \cdots & \frac{\partial f_k(x)}{\partial x_k} \end{array} \right]_{x^*} = F - V$$  \hspace{1cm} (C.6)

Finally, the Next Generation Matrix $K$ and the basic reproduction number $R_0$ can be calculated as follows:

$$K = F V^{-1}$$ \hspace{1cm} (C.7)

$$R_0 = \text{spectral radius} (K)$$ \hspace{1cm} (C.8)

where the spectral radius of a matrix is defined the maximum modulus (absolute value) of all eigenvalues of the matrix in question.

**Derivation for a Phage-Bacterium Model**

In our system (Equations 4.4–4.6), the Disease-Free Equilibrium (DFE) is given by:

$$0 = S^*(1 - S^*) - \mu S S^*$$  \hspace{1cm} (C.9)

$$\Rightarrow S^* = 1 - \mu S$$ \hspace{1cm} (C.10)

The disease subsystem of equations, which includes infected hosts ($I$) and free phage particles ($V$), is given by:

$$\begin{bmatrix} \frac{dI}{dt} \\ \frac{dV}{dt} \end{bmatrix} = \begin{bmatrix} I(1 - S - I) + \rho S V - \mu S I - \alpha I \\ \delta \alpha I - \mu V V - \rho S V \end{bmatrix}$$ \hspace{1cm} (C.11)

and can be broken down into the matrices $F$ and $V$, according to Table C.1:

$$\begin{bmatrix} \frac{dI}{dt} \\ \frac{dV}{dt} \end{bmatrix} = \begin{bmatrix} I(1 - S - I) + \rho S V \\ 0 \end{bmatrix} - \begin{bmatrix} \mu S I + \alpha I \\ -\delta \alpha I + \mu V V + \rho S V \end{bmatrix} = F - V$$ \hspace{1cm} (C.12)
Table C.1: Placement of Terms into Matrices $\mathcal{F}$ and $\mathcal{V}$

<table>
<thead>
<tr>
<th>Term</th>
<th>Interpretation</th>
<th>Transition</th>
<th>Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I(1 - S - I)$</td>
<td>Logistic growth of infected hosts</td>
<td>New appearance</td>
<td>$\mathcal{F}$</td>
</tr>
<tr>
<td>$\rho SV$</td>
<td>New infection</td>
<td>New appearance from $S$</td>
<td>$\mathcal{F}$</td>
</tr>
<tr>
<td>$-\mu_S I$</td>
<td>Death of infected hosts</td>
<td>Transition from $I$ to death</td>
<td>$\mathcal{V}$</td>
</tr>
<tr>
<td>$-\alpha I$</td>
<td>Lysis</td>
<td>Transition from $I$ to $V$</td>
<td>$\mathcal{V}$</td>
</tr>
<tr>
<td>$\delta \alpha I$</td>
<td>Lysis</td>
<td>Transition from $I$ to $V$</td>
<td>$\mathcal{V}$ $^\dagger$</td>
</tr>
<tr>
<td>$\mu_V V$</td>
<td>Death of free phages</td>
<td>Transition from $V$ to death</td>
<td>$\mathcal{V}$</td>
</tr>
<tr>
<td>$-\rho SV$</td>
<td>New infection</td>
<td>Transition from $V$ to $I$</td>
<td>$\mathcal{V}$</td>
</tr>
</tbody>
</table>

$^\dagger$: Could also reasonably be placed in $\mathcal{F}$ instead

The full Jacobian evaluated at the DFE is:

$$ J_{DFE} = \begin{bmatrix} (1 - S^* - \mu_S - \alpha) & (\rho S^*) \\ (\delta \alpha) & (\mu_V - \rho S^*) \end{bmatrix} $$  \hspace{1cm} (C.13)

with components:

$$ F = \begin{bmatrix} 1 - S^* & \rho S^* \\ 0 & 0 \end{bmatrix} $$  \hspace{1cm} (C.14)

$$ V = \begin{bmatrix} \mu_S + \alpha & 0 \\ -\delta \alpha & \mu_V + \rho S^* \end{bmatrix} $$  \hspace{1cm} (C.15)

Thus, solving for $K = FV^{-1}$ and plugging in $S^* = 1 - \mu_S$, gives:

$$ K = \begin{bmatrix} \frac{1 + \mu_S}{\alpha + \mu_S} & \frac{\rho \mu_S}{\mu_V + \rho \mu_S} \\ 0 & 0 \end{bmatrix} $$  \hspace{1cm} (C.16)

Finally, $R_0$ is the spectral radius of $K$,

$$ R_0 = \frac{\mu_V - \mu_S \mu_V + \rho \mu_S (1 + \alpha \delta - \mu_S)}{(\alpha + \mu_S)(\mu_V + \rho \mu_S)} $$  \hspace{1cm} (C.17)
References

