

UNIVERSITY OF CALIFORNIA

SANTA CRUZ

**MATHEMATICAL MODELING OF CHOLERA:  
FROM BACTERIAL LIFE HISTORIES TO HUMAN EPIDEMICS**

A dissertation submitted in partial satisfaction of the  
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

by

**Leah R. Johnson**

June 2006

The Dissertation of Leah R. Johnson  
is approved:

---

Professor Marc S. Mangel, Chair

---

Professor Josh Deutsch

---

Professor Fitnat Yildiz

---

Lisa C. Sloan  
Vice Provost and Dean of Graduate Studies

Copyright © by

Leah R. Johnson

2006

# Table of Contents

List of Figures	vi
Abstract	ix
Dedication	x
Acknowledgments	xi
<b>1 Introduction</b>	<b>1</b>
1.1 <i>Vibrio Cholerae</i> and Cholera . . . . .	2
1.1.1 Epidemiology of Cholera . . . . .	3
1.1.2 <i>Vibrio Cholera</i> Life History and Microbiology . . . . .	5
<b>2 A Model of Human Cholera with a Reservoir of <i>Vibrio Cholerae</i></b>	<b>9</b>
2.1 Modeling Infection Dynamics . . . . .	9
2.2 Cholera and The Aquatic Reservoir . . . . .	11
2.2.1 Modeling Cholera including a Reservoir with an SIR type Model . . . . .	12
2.3 The Dynamics of the Bacterial Population . . . . .	18
<b>3 Bacterial Life Histories and Aging</b>	<b>26</b>

3.1	Fitness Models of Bacteria and Single-Celled Organisms . . . . .	28
3.1.1	The Baseline Model: Bacterial Population without Aging . . . . .	29
3.2	Effects of Limited Reproduction on the Fitness of Simple Organisms . . . . .	32
3.3	Effects of Limited Reproduction and Increasing Doubling Times on the Fitness of Simple Organisms . . . . .	42
3.3.1	Specification of $T(a)$ : Multiplicative changes to doubling time . . . . .	44
3.4	Discussion . . . . .	49
<b>4</b>	<b>Microcolony and Colony Formation</b>	<b>55</b>
4.1	Community Formation as a Survival Strategy for Bacteria . . . . .	56
4.1.1	<i>Vibrio cholerae</i> Communities . . . . .	57
4.2	Approaches for Modeling Bacterial Communities . . . . .	58
4.3	A Mathematical Model for Community Formation . . . . .	61
4.4	Microcolony Formation . . . . .	62
4.4.1	Model Assumptions . . . . .	63
4.5	Null Model, and Models with Stopping, Clumping, and Direct Interactions . . .	66
4.5.1	The Null Model, $M_0$ . . . . .	66
4.5.2	Model $M_1$ : Stopping and Clumping . . . . .	67
4.5.3	Model $M_2$ : Direct Interactions . . . . .	69
4.5.4	Model Simulations: $M_1$ and $M_2$ . . . . .	72
4.6	Model $M_3$ : Direct Interactions, Births and Deaths . . . . .	74
4.6.1	Model Simulations: $M_3$ . . . . .	77
4.7	Model $M_4$ : Colony Spreading . . . . .	78
4.7.1	Model Simulations: $M_4$ . . . . .	84
4.8	Discriminating between models . . . . .	87

4.9	Fitness of bacteria in microcolonies . . . . .	93
4.10	Discussion . . . . .	98
<b>5</b>	<b>Conclusions and Future Work</b>	<b>103</b>
5.1	Future Work . . . . .	105
<b>A</b>	<b>Definitions</b>	<b>108</b>
<b>B</b>	<b>Mathematical Background</b>	<b>111</b>
B.1	Exponential and Logistic Population Growth . . . . .	111
<b>C</b>	<b>Susceptible - Infected - Recovered Type Disease Models</b>	<b>116</b>
C.1	The SIS Model . . . . .	116
C.2	The SIR Model (without demographic processes) . . . . .	120
C.3	The SIR Model (with demographic processes) . . . . .	124
	<b>Bibliography</b>	<b>130</b>

# List of Figures

1.1	Microscopic views of <i>Vibrio cholerae</i> . . . . .	6
2.1	Sample Population Trajectories for the Model of Cholera Disease Dynamics . .	19
2.2	Sample Population Trajectories for Endemic Cholera with Environmental Forcing (a). . . . .	22
2.3	Sample Population Trajectories for Endemic Cholera with Environmental Forcing (b). . . . .	23
2.4	Sample Population Trajectories for Endemic Cholera with Environmental Forcing (c). . . . .	24
2.5	Sample Population Trajectories for Endemic Cholera with Environmental Forcing (d). . . . .	25
3.1	The intrinsic rate of natural increase, $r$ , for a bacterial population as a function of doubling time, $T$ , and mortality, $m$ . . . . .	30
3.2	$r$ as a function of allocation strategy, $\rho$ , without aging. . . . .	32
3.3	Intrinsic rate of natural increase for an aging population, $\tilde{r}$ , as a function of the maximum reproductive age, $a_{max}$ . . . . .	35
3.4	Comparison of analytic approximation and numerical solutions of $\tilde{r}$ vs. $a_{max}$ . . . . .	36

3.5	$\Delta\tilde{r}$ vs $a_{max}$ . . . . .	37
3.6	Comparison of $\tilde{r}$ with $r$ as a function of allocation strategy for various values of $a_{max}$ . . . . .	38
3.7	Close-ups of maxima from Figures 3.6 a and d. . . . .	39
3.8	Mortality of aging bacteria as a function of the maximum reproductive age, $a_{max}$ . . . . .	43
3.9	Model with increasing doubling times. . . . .	44
3.10	$\tilde{r}$ vs. $a_{max}$ and $s$ for slowing reproduction with age . . . . .	46
3.11	$\tilde{r}$ vs. $a_{max}$ for slowing reproduction with age . . . . .	47
3.12	$\Delta\tilde{r}$ vs. $a_{max}$ for slowing reproduction with age . . . . .	48
3.13	$\partial_s\tilde{r}$ vs. $s$ for slowing reproduction with age . . . . .	48
3.14	$\tilde{r}$ vs. $\rho$ for slowing reproduction with age . . . . .	50
3.15	Experiments to predict $\rho^*$ . . . . .	53
4.1	Steps in Biofilm Formation . . . . .	62
4.2	Schematic of transitions between stage states for cell $i$ . . . . .	65
4.3	Sample of simulated cell movement. . . . .	68
4.4	Distribution of colony sizes without interactions between cells . . . . .	73
4.5	Distribution of colony locations and sizes with and without interactions between cells . . . . .	75
4.6	Histogram of final colony sizes for model $M_3$ with no movement along the surface. . . . .	79
4.7	Histogram of final colony sizes for model $M_3$ with $A = 10$ . . . . .	80
4.8	Histogram of final colony sizes for model $M_3$ with $A = 0$ . . . . .	81
4.9	Histogram of final colony sizes for model $M_3$ with $A = -10$ . . . . .	82
4.10	Alternative histograms of final colony size for model $M_3$ . . . . .	83

4.11 Cell distribution and colony membership before and after implementation of the spreading heuristic – attractive example. . . . .	86
4.12 Cell distribution and colony membership before and after implementation of the spreading heuristic – repulsive example. . . . .	87
4.13 Histogram of final colony sizes for model $M_4$ . . . . .	88
4.14 Fitness of bacteria as a function of groups size (a). . . . .	99
4.15 Fitness of bacteria as a function of groups size (b). . . . .	100
4.16 Diagram of a Flowcell. . . . .	102
B.1 Schematic of birth and death rate lines ( $r_b, r_d$ ) for values of $D_0, B_0, b$ and $d$ . . .	113
B.2 Exponential vs. Logistic Population Growth . . . . .	115
C.1 Phase planes for the SIS model. . . . .	118
C.2 Sample population trajectories for the SIS model. . . . .	121
C.3 Phase plane for the SIR model, without vital statistics. . . . .	122
C.4 Sample population trajectories for the SIR model, without vital statistics. . . .	123
C.5 Phase planes for the SIR model with vital statistics. . . . .	125
C.6 Sample population trajectories for the SIR model, with vital statistics. . . . .	129

## **Abstract**

Mathematical Modeling of Cholera:

From Bacterial Life Histories to Human Epidemics

by

Leah R. Johnson

Cholera is an epidemic disease caused by the bacteria *Vibrio cholerae*. Although it has been studied for over one-hundred years, over 100,000 people die annually from cholera. In order to understand the patterns of infection in a human population we need to understand what factors influence transmission of the disease as well as the dynamics of the bacteria in their aquatic environment. In this dissertation, I explore the dynamics of cholera on three scales. First, I introduce a fairly simple model for cholera in a human population coupled to an environmental reservoir of bacteria. This model demonstrates the need to understand more fully how *V. cholerae* survives and evolves in the aquatic environment. Next I explore the life histories of bacteria, in particular how aging and environmental factors influence bacterial fitness. Finally I examine the implications of various types of inter-cellular interactions during surface colonization for the structure and composition of bacterial communities using an individual based model, as well as examining under what conditions living in communities of various sizes would be optimal.

*In memory of my Granny, Laverta Johnson,  
who always believed in the importance of education  
for her children and grandchildren.*

## Acknowledgments

This dissertation would not have been possible to complete without the support and input of many people. First of all is my advisor, Marc Mangel. Without his support, encouragement, and enthusiasm, switching my area of concentration over half way through my graduate school career would not have been possible. I have not only learned about an entirely new field of study, but have also learned a lot about pacing myself, planning my time, and approaching problems one step at a time.

I would also like to thank to other two members of my dissertation committee, Josh Deutsch and Fitnat Yildiz. Josh has served as my official liason to the Physics department, and has been very supportive, while allowing me the freedom to explore areas of research that are not typical of in a physics department. Fitnat has welcomed me into her lab meetings, which has allowed me to learn about experimental approaches for studying *V. cholerae*, what aspects of the system are well understood, and what are the open questions. Working with the lab has also been helpful in keeping my mathematical models grounded in reality.

A large part of my ability to change directions in my research has been due to the generous financial support of the UCSC Physics Department. They supported me for three years through a GAANN Fellowship, and then nominated me for a Presidents Dissertation Year Fellowship, which allowed me to concentrate on finishing the final portions of this dissertation.

I would also like to thank the Applied Mathematics and Statistics faculty and students for being so supportive and enthusiastic over the last few years. The atmosphere of the department makes collaboration and learning fun and exciting. Herbie Lee has been especially helpful in answering my statistics questions, as well as providing me with opportunities to hone my own teaching skills through tutoring assignments. My fellow students in the Mangel lab have been the best academic family anyone could ask for – always willing to hash through ideas,

read drafts of documents, and listening to practice talks. My writing group – Yasmin, Kate, and Christine – have been especially supportive in this final year, and this document would not have come together as quickly without their insightful and thorough comments.

I would not have finished as quickly, and had as much fun here if it had not been for Bobby. He has been my rock, a constant friend and supporter. He has helped keep me focused, while encouraging me to explore new places and opportunities. I love you, and am looking forward to our next adventure.

Finally, I would not have made it to this point at all without the support of my family. For the past 20 years of my educational life, they have believed in me, encouraged me, and would never let me quit. I never would have made it here with out you.

# Chapter 1

## Introduction

Epidemiological experimentation can be difficult, expensive, and possibly even unethical. Disease dynamics are a result of complex systems interacting in very complicated ways. Since experimental manipulation or observation of these or other complex biological systems can be very difficult, if not impossible, mathematical modeling is an important tool for understanding their dynamics [2, 3, 42, 43, 44]. Mathematical modeling enables us to characterize the general and specific behavior of these systems analytically and to understand which aspects contribute the most to the observed dynamics. This information can be useful in analyzing data from epidemics, designing experiments, or in making policy decisions for preventive measures.

Modeling of diseases is generally approached from a macroscopic or big picture perspective, *i.e.* models consider how an infection spreads through and affects a host population. However, for many diseases the observed dynamics likely arise from factors outside of the host population. The spread of a disease can depend on the biology and evolution of the infective agent, the biology and behavior of a vector population, environmental factors, and the biology and behavior of the host population. Understanding how these factors interact is key to the

description and prediction of disease behavior in a real population. This is especially true of cholera, for which transmission dynamics are coupled to dynamic environmental reservoirs of *Vibrio cholerae*, the bacteria that cause cholera. Therefore we need to understand not only the dynamics of the disease in a human population, but also the dynamics of the bacteria in their aquatic environment.

This dissertation is divided into five chapters. In this chapter, I review what is currently known about the epidemiology and biology of cholera and *V. cholerae*. Definitions of medical terminology are included in Appendix A. In Chapter 2, I introduce a fairly simple model for cholera in a human population coupled to an environmental reservoir of bacteria. Some mathematical background necessary for this disease model is presented in Appendix B. The model from Chapter 2 demonstrates the need to understand more fully how the causative agent of cholera, *Vibrio Cholerae*, survives and evolves in the aquatic environment. Chapter 3 explores the life histories of bacteria, in particular how aging and environmental factors influence bacterial fitness. In Chapter 4 I examine the implications of various types of inter-cellular interactions during surface colonization for the structure and composition of bacterial communities using an individual based model (IBM). I conclude in the final chapter with a discussion of important results and directions for future work.

## 1.1 *Vibrio Cholerae* and Cholera

Cholera has long been, and continues to be, a world health issue. Despite study of this disease for over one-hundred years, approximately 120,000 people die from cholera annually [22]. War and extreme environmental events can contribute to the disease's ability to ravage communities. Because it continues to be a threat to much of the world, it is important to continue to try to understand the disease dynamics and how interactions with environmental and

human factors contribute to the epidemic behavior observed during current cholera outbreaks. Additionally, the large amount of data available on the disease makes it an ideal system to study how extrinsic and intrinsic factors contribute to the dynamics of a disease.

### 1.1.1 Epidemiology of Cholera

#### History

Some of the earliest recorded evidence of a cholera-like disease comes from the ancient Greek writings of Hippocrates around 470-400 B.C. and in the *Sushrute Samhita*, a Sanskrit text from around 500–400 B.C. Modern accounts of epidemics across Asia record outbreaks starting from 1438 and continuing to the present time. It was not until 1817 that the modern pandemics began, and cholera started to spread further abroad. [16]

The First Pandemic (1817-1823) began in Kishnagur, India in May of 1817. By the end of the pandemic, cholera had spread over  $100^\circ$  in longitude, and  $67^\circ$  in latitude without reaching Europe. The Second Pandemic (1826-1837) spread from India, across Asia and Europe, and by 1834 had even reached most major cities in the United States and Canada. The Third and Fourth Pandemics (1846-1863 and 1865-1875 respectively) spread much like the Second Pandemic, though with more virulence. In 1854, during the Third Pandemic, 23,000 people were killed in England and Wales, and even more in southern Europe [16].

During the 19th and early 20th centuries a total of six cholera pandemics occurred, ending in 1923. The Seventh Pandemic, effecting mostly the southern hemisphere, began in 1961 and continues to this day [22, 74].

## Symptoms, Transmission, and Environmental Connections

Cholera is a water transmitted disease caused by the bacteria *Vibrio cholerae*. Cholera is characterized by severe watery diarrhea caused by the production of cholera toxin (an enterotoxin) by *V. cholerae* bacteria in the small intestine. The bacteria are transmitted via contaminated drinking water or food. Pathogenic *V. cholerae* can survive refrigeration and freezing in food supplies [74]. The dosage of bacteria required to cause an infection in healthy volunteers via oral administration of living vibrios is greater than  $\sim 1000$  organisms. After consuming an antacid, however, cholera developed in most volunteers after consumption of only  $\sim 100$  cholera vibrios. Experiments also show that vibrios consumed with food are more likely to cause infection than those from water alone [25].

Cases tend to be clustered by location as well as season, with most infections occurring in children ages 1-5 years. Increased protection against the disease can be gained by improving sanitation and hygiene [22]. Most cases of cholera currently occur in developing countries. Cholera is currently endemic in India and Bangladesh near the Bay of Bengal as well as in coastal regions of South America [53]. Cases in these regions tend to have seasonal cycles, generally associated with fluctuations in water temperature, zooplankton levels, monsoon cycles [53], and possibly El Niño events [15, 75]. For instance, in Bangladesh there is a small peak in the number of cases in the spring, and then a larger peak in the fall, following the monsoon season. These epidemics tend to coincide with dry weather and higher water temperatures; cases are reduced in the winter. In Calcutta, cases tend to peak from April through June, and in South America in January and February (austral summertime) [53].

### 1.1.2 *Vibrio Cholera* Life History and Microbiology

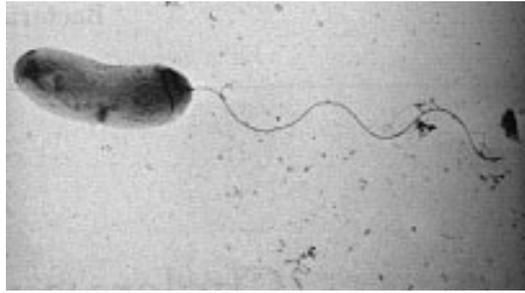
Scientific study of cholera began over 100 years ago. In 1849 Dr. John Snow was the first to hypothesize that cholera was transmitted by contaminated water, though this theory was not generally accepted until 1871 [16]. In 1854, Pacini first described the “comma bacillus” responsible for cholera. He studied the etiology of cholera for 20 years, and published a number of papers. However, as the idea that cholera was contagious was still under dispute, his data were ignored [53]. It was not until 1883 (a year after Pacini’s death) that the waterborne bacteria now called *Vibrio cholerae* was “discovered” by Robert Koch, and the bacterium was finally accepted to be the cholera pathogen [53, 74].

*Vibrio cholerae* is a large and very diverse species. It is divided into about 200 serogroups, denoted by O1, O2, etc., of which only O1 and O139 contain pathogenic members. Most of the strains isolated from cholera patients belong to the O1 serogroup [74]. This serogroup is divided into three serotypes, namely Inaba, Ogawa, and Hijokima. Additionally, strains from this serogroup can be divided into two biotypes, Classical and El Tor. These two biotypes can contain members from all of the serotypes mentioned above [82].

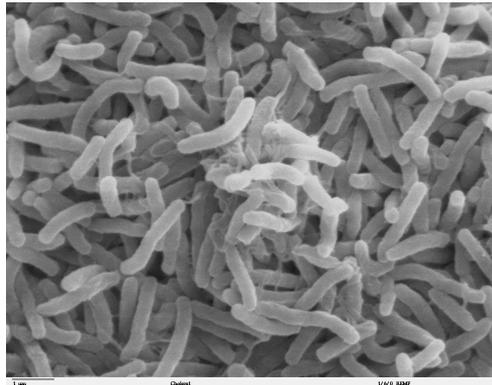
Not all members of the O1 serotype cause cholera, while some non-O1 strains can cause diarrhea, though not epidemic or endemic disease. Members of the O139 serotype cause disease clinically indistinguishable from cholera and has caused epidemics. However, as of 1996 only *V. cholerae* O1 is reportable to the World Health Organization (WHO) as cholera [82].

*Vibrio cholerae* is a motile gram-negative, rod-shaped bacteria with a single polar flagellum (Figure 1.1). Vibrios survive well in alkaline conditions, but are sensitive to acid and die in solutions with pH less than 6. They can grow anaerobically, but can reach higher population densities in an aerated solution. [25]

It was once thought that the bacteria only survived for a short time outside of an in-



(a)



(b)

Figure 1.1: Microscopic views of *Vibrio cholerae* (a) Single *V. cholerae* bacterium [University of Wisconsin] (b) Scanning electron microscope image of *V. cholerae* bacteria [46].

ected host. However, we now know that *V. cholerae* (both pathogenic and non-pathogenic) are naturally abundant in many aquatic ecosystems around the world. They are often found associated with aquatic organisms such as zooplankton, phytoplankton, macrophytes, and sometime with prawns, oysters, and crabs [11, 17]. The ability of the bacteria to connect to and live inside aquatic organisms may enable them to survive in harsher environments [12, 34].

*V. cholerae* have other methods for surviving environmental fluctuations. They undergo a phase transition, likely in response to environmental stresses. In particular, there are two morphologically different colonial variants: “smooth” or “rugose”. The smooth colonial morphotype exhibits a smooth, translucent appearance, and is larger in size. Rugose colonies,

on the other hand, are smaller, more wrinkled and more opaque than smooth variants. Research thus far has shown that the rugose variant possesses an increased ability to form biofilms and to deal with environmental stresses than smooth variants. However, they are smaller than their smooth counterparts, and may be susceptible to increased predation by bacterivores. Bacteria in either form can also transition into the other form, with the rugose to smooth transition occurring more frequently [94].

*V. cholerae* also are able to switch between free-living and surface-attached forms. When attached to a surface they may also form biofilms [17, 94]. Biofilm formation appears to provide some additional protection for bacteria in extreme environmental conditions, and can even make them more resistant to chlorination and antibiotics [18, 66]. This is true of *V. cholerae* as well. In particular, rugose forms of *V. cholerae* bacteria exhibit an increased ability to form biofilms, which may contribute to their ability to survive in harsher environments.

If environmental conditions are particularly harsh, *V. cholerae* may be able to transition into a viable but nonculturable (VNC or VBNC) state. In a VBNC state, a bacterium's metabolism is reduced, it loses its flagellum, changes into a smaller spherical form, and can no longer reproduce in a standard culture. These forms may still be able to cause infection, and convert back to a culturable form. However, research in the area is inconclusive, and the role that these forms may have in *V. cholerae* survival and in the disease dynamics of cholera still needs to be investigated [22, 23, 74].

Considering that non-pathogenic *V. cholerae* are more abundant and appear to survive better in aquatic ecosystems, horizontal transmission of virulence genes between toxigenic and non-toxigenic vibrios is a strong possibility. One way that this may occur is via infection of the non-toxigenic vibrios by bacteriophages [22]. These viruses may be able to penetrate the O1 and O139, and insert its DNA into the DNA of the cell, resulting in a bacteria that is

pathogenic. While a particular phage with this characteristic has been identified, the degree to which this process contributes to the appearance and maintenance of pathogenic *V. cholerae* in aquatic ecosystems is unknown [17, 22, 23, 53, 74].

## Chapter 2

# A Model of Human Cholera with a Reservoir of *Vibrio Cholerae*

### 2.1 Modeling Infection Dynamics

One of the first examples of mathematical modeling of epidemic dynamics was conducted by Sir Ronald A. Ross. In 1902 he was awarded the Nobel Prize in Medicine and Physiology for his work confirming that mosquitos are the vectors that transmit malaria [64]. He later developed mathematical models to describe mosquito/malaria dynamics and was able to predict how limiting (or eliminating) the mosquito population could help control malaria. His work pioneered the use of mathematics for studying disease dynamics [77, 78, 79, 80].

Since then, modeling has become an important tool in epidemiology [2, 3, 42, 43, 44]. It provides a way to understand primary dynamics of infections, and to conduct theoretical experiments that are not possible in practice. Modeling can also be used to predict which control procedures or treatments could be most effective in stopping (or preventing) the spread

of a disease [3].

One method for modeling macroscopic disease dynamics in a population is to use coupled differential equations. In this type of model, differential equations are used to describe how the composition of a population exposed to a disease changes over time. Generally the population is divided into at least two or three disjoint sub-populations, such as Susceptible, Infected, Removed. The “Susceptible” class is comprised of individuals who can incur the disease or are not yet transmitting the disease. Members of the “Infected” class have contracted the disease and are also transmitting it to others. Those in the “Removed” class were infected, but are no longer actively transmitting the disease, due to death, or recovery with either permanent or temporary immunity. In a population where a disease was not present, and had never occurred, the entire population would be in the susceptible class. Upon exposure to a contagion, an initial group would become infected. These individuals would interact with the susceptible group in some fashion, and infect some additional number of the susceptibles.

Individuals can leave the infected class by a number of different methods, depending on the type of disease one is interested in modeling. An infected individual can recover and return to the susceptible class, or can recover with permanent or temporary immunity to the recovered class, or can be removed by death.

Each of these variations can be addressed with models of differing complexity and parameterization in order to capture different behaviors. Three of the standard models, SIS (Susceptible, Infected, Susceptible), SIR (Susceptible, Infected, Recovered) without demographics, and SIR with demographics are reviewed in Appendix C. The SIS model addresses the case when individuals recover from a disease and then can be re-infected immediately afterwards, which results in a disease either immediately dying out, or reaching an endemic state with a constant, non-zero number of infected individuals. The SIR model without demographics as-

sumes that individuals recover from a disease and then are permanently immune. Since there are no births in this model, the pool of Susceptible individuals is not refreshed, so even if an epidemic can occur, the infection will always eventually die out. The SIR model with demography can exhibit either epidemic or endemic behavior, since births allow the pool of Susceptibles to be refreshed, and deaths of Recovered individuals keeps the mixing between Susceptibles and Infecteds relatively high.

In this chapter I develop a model for human cholera coupled with a bacterial reservoir. I begin by reviewing factors that could contribute to the survival of *Vibrio cholerae* in an aquatic reservoir, and to the transmission of these bacteria to a human population. I then introduce and analyze an SIR type model that includes demography and a reservoir of bacteria.

## 2.2 Cholera and The Aquatic Reservoir

*Vibrio Cholerae* bacteria live in and are transmitted by contact with water. Under certain environmental conditions, both toxigenic, and non-toxigenic *V. cholerae* can survive and persist in aquatic sources, even in the absence of humans. Because of this, it is important to understand how the bacteria survive in an aquatic environment, what the transmission dynamics are, and how this effects cholera as a human disease.

Human contact with contaminated water tends to have seasonal cycles, especially in areas without adequate water treatment and sewer systems. For instance, in areas where heavy rains from monsoons cause flooding, water used for drinking, cooking, and washing can become contaminated when sewage areas overflow. In other areas, extreme temperatures and long dry seasons cause ponds and water holes to dry up. With fewer sources of water available, contamination and disease can increase. Any of these phenomena could increase contact with bacteria and cause an increase in cholera incidences.

Environment and weather can also have a direct effect on the abundances of *V. cholerae* in water. Increased rainfall can decrease salinity or decrease concentrations of bacteria, which in turn can effect the survival of the bacteria or the transmission rates to humans. Factors such as temperature fluctuations, amount of sunlight, and nutrient content in the water, could all impact survival of *V. cholerae* and, as a result, impact incidences of cholera in humans.

The impacts of these seasonal effects on the survival and replication of *V. cholera* in an aquatic reservoir are not entirely understood. Models with parameters that take into account the ecology of *V. cholerae* as well as fluctuations in human contact with water sources could indicate which sets of variables are most important in the dynamics of cholera as a human disease.

### **2.2.1 Modeling Cholera including a Reservoir with an SIR type Model**

Since *V. cholerae* exist in a free living stage outside of human hosts, it is important to include this stage in a model for cholera. One SIR type model for cholera that includes an aquatic reservoir of *V. cholerae* interacting with an infectable population is found in [14]. This work makes a good start in understanding how the aquatic reservoir can impact disease progression in a population that contacts the reservoir. An important addition is to consider how different models of the human population influence the dynamics. Another is to consider how different types of *V. cholerae* bacteria in the environment might interact. This kind of modeling might provide information on what circumstances are required for multiple infective forms of *V. cholerae* (for instance Classical and El Tor) to persist simultaneously. It may also be useful to use these kinds of models to see under which circumstances both rugose and smooth bacteria will persist.

I begin by considering an SIR model with demography. Suppose that there are  $J$

types of *V. cholerae* present in the reservoir, denoted by  $C_j$ , the concentration of the  $j$ th type of bacteria in the water. Infections are caused only by interactions of susceptibles with contaminated water, rather than being passed directly from infected to susceptible. Any susceptible individual can contract any one type of the bacteria, where the probability of becoming infected when the bacteria have concentration  $C_j$  is  $\lambda(C_j)$ . The individuals then join the  $j$ th infected class, denoted by  $I_j$ , then recover or die, never returning to the susceptible class. Infected individuals, in turn, contribute new bacteria to the reservoir directly.

I consider logistic growth for the human population, described in Appendix B.1 instead of the exponential growth assumed in the typical SIR type models explored in Appendix C. However, if we assume that recovered individuals do not lose immunity, and do not pass this immunity on to their offspring, we need to break apart the death and birth contributions for this group, so that recovereds die in the recovered group, but produce offspring into the susceptible group. For this we can use equations (B.5) and (B.6) for total births and deaths, respectively, derived in Appendix B.1:

$$\begin{aligned} \text{total births} &= \left[ D_0 + r - \frac{r}{2K} \right] N \\ \text{total deaths} &= - \left[ D_0 + \frac{r}{2K} \right] N. \end{aligned}$$

Here  $N$  denotes the population size,  $D_0$  is the density independent death rate of the population,  $r$  is the *per capita* growth rate, and  $K$  is the carrying capacity (see also Table 2.1).

Additionally, suppose that individuals in the infected population can only recover or die from the disease, but do not give birth or die from other causes while they are infected. The susceptible population will follow logistic growth, plus logistic births from the recovered class, minus loss to disease. Recovered individuals contribute births into the susceptible population,

Symbol	Description	Units
$r$	per capita birth rate	day <sup>-1</sup>
$K$	carrying capacity	number of individuals
$D_0$	density independent death rate	day <sup>-1</sup>
$a$	water exposure rate	day <sup>-1</sup>
$\gamma$	recovery rate from cholera	day <sup>-1</sup>
$\mu$	death rate due to cholera	day <sup>-1</sup>
$\delta$	bacteria contribution to reservoir per infected	bacteria/(ml-day-individual)
$n$	bacterial reproduction rate	day <sup>-1</sup>
$m$	bacterial death rate	day <sup>-1</sup>

Table 2.1: Parameters used in the model expressed in Equations (2.1)–(2.4)

and otherwise experience logistic deaths. The following system of differential equations models this scenario:

$$\frac{dS}{dt} = rS \left(1 - \frac{N}{K}\right) + \left[r + D_0 - \frac{rN}{2K}\right] R - a \left(\sum_{j=1}^J \lambda_j(C_j)\right) S \quad (2.1)$$

$$\frac{dI_j}{dt} = a\lambda_j(C_j)S - (\gamma_j + \mu_j)I_j \quad (2.2)$$

$$\frac{dR}{dt} = \sum_j \gamma_j I_j - \left[D_0 + \frac{rN}{2K}\right] R \quad (2.3)$$

$$\frac{dC_j}{dt} = \delta_j I_j + (n_j - m_j)C_j \quad (2.4)$$

$$N(t) = S(t) + \sum_{j=1}^J I_j(t) + R(t).$$

The parameters used in the model are described in Table 2.1. The system extends the standard SIR model by modeling the human population dynamics with logistic instead of exponential growth. Another major addition is the set of  $J$  reservoirs of *V. cholerae* outside of the host population, denoted by  $C_j$  for  $j = 1, \dots, J$ . These populations experience dynamics separate from those of the host population, and are currently modeled using exponential growth.

### First Model - one strain of bacteria

For only one strain of bacteria, the model in (2.1)–(2.4) can be written somewhat more simply:

$$\frac{dS}{dt} = rS \left(1 - \frac{N(t)}{K}\right) + \left[r + D_0 - \frac{rN(t)}{2K}\right] R - a\lambda(C)S \quad (2.5)$$

$$\frac{dI}{dt} = a\lambda(C)S - (\gamma + \mu)I \quad (2.6)$$

$$\frac{dR}{dt} = \gamma I - \left[D_0 + \frac{rN(t)}{2K}\right] R \quad (2.7)$$

$$\frac{dC}{dt} = \delta I + (n - m)C \quad (2.8)$$

$$N(t) = S(t) + I(t) + R(t).$$

To analyze the stability of the system around the steady state, as in Section C.3, I linearize the system around the steady state  $(s_0, i_0, r_0, c_0)$ , and define  $x = S - s_0$ ,  $y = I - i_0$ ,  $z = R - r_0$  and  $w = C - c_0$ . The linearized system of equations becomes:

$$\begin{aligned} \frac{dx}{dt} = & \left[ r \left(1 - \frac{n_0}{K}\right) - \frac{r}{K}(s_0 + r_0/2) - a\lambda(c_0) \right] x - \frac{r}{K}(s_0 + r_0/2)y \\ & + \left[ \frac{-r}{K}(s_0 + r_0/2 + n_0/2) + r + D_0 \right] z - as_0 \left[ \frac{\partial \lambda(c_0)}{\partial C} \right] w \end{aligned} \quad (2.9)$$

$$\frac{dy}{dt} = a\lambda(c_0)x - (\gamma + \mu)y + as_0 \left[ \frac{\partial \lambda(c_0)}{\partial C} \right] w \quad (2.10)$$

$$\frac{dz}{dt} = -r \frac{r_0}{2K}x + \left( \gamma - r \frac{r_0}{2K} \right) y - \left[ D_0 + \frac{r}{2K}(n_0 + r_0) \right] z \quad (2.11)$$

$$\frac{dw}{dt} = \delta y + (n - m)w. \quad (2.12)$$

I choose a Holling type 2 functional response to model transmission:

$$\lambda(C) = \frac{C}{C + B}, \quad (2.13)$$

Inserting this into (2.9)–(2.12), yields the following coefficient matrix for the linearized system:

$$\begin{bmatrix} r - \phi_1 - \frac{ac_0}{c_0+B} & -\frac{r}{K}(s_0 + \frac{r_0}{2}) & -\phi_2 + r + D_0 & -as_0 \left[ \frac{2c_0+B}{(c_0+B)^2} \right] \\ \frac{ac_0}{c_0+B} & -(\gamma + \mu) & 0 & as_0 \left[ \frac{2c_0+B}{(c_0+B)^2} \right] \\ -\frac{rr_0}{2K} & (\gamma - \frac{rr_0}{2K}) & -[D_0 + \frac{r}{2K}(n_0 + r_0)] & 0 \\ 0 & \delta & 0 & (n - m) \end{bmatrix} \quad (2.14)$$

where  $n_0 = s_0 + i_0 + r_0$ ,  $\phi_1 = \frac{r}{K}(n_0 + s_0 + \frac{r_0}{2})$ , and  $\phi_2 = \frac{r}{K}(2s_0 + r_0 + n_0)$ .

The easiest steady state point to find (and analyze) is the disease free steady state,  $(K, 0, 0, 0)$ , for which (2.14), simplifies considerably:

$$\begin{bmatrix} -r & -r & D_0 - r/2 & -\frac{aK}{B} \\ 0 & -(\gamma + \mu) & 0 & \frac{aK}{B} \\ 0 & \gamma & -D_0 - r/2 & 0 \\ 0 & \delta & 0 & (n - m) \end{bmatrix}. \quad (2.15)$$

The four eigenvalues for this system are given by:

$$L_1 = -r \quad (2.16)$$

$$L_2 = -D_0 - \frac{r}{2} \quad (2.17)$$

$$L_{3,4} = -\frac{1}{2}(\mu + \gamma + m - n) \pm \frac{1}{2}\sqrt{(\mu + \gamma + n - m)^2 + 4\delta\frac{aK}{B}}. \quad (2.18)$$

The first two eigenvalues are always negative. The first term in (2.18) is negative if either of 2 conditions is met. The first condition is that  $n - m < 0$ , so that the bacterial reproduction rate is less than the death rate and so the net rate of change in the bacterial population is negative. The second condition is  $n < \mu + \gamma + m$ , so that bacterial reproduction

is unable to balance both the natural deaths of the bacteria in the reservoir with the reduction in the amount bacteria being contributed by the human population due to death and recovery from the disease.

Next we consider the discriminant in equation (2.18), *i.e.*,  $(\mu + \gamma + n - m)^2 + 4\delta\frac{aK}{B}$ . First, notice that every term in the discriminant is always positive, so the eigenvalues will not have any imaginary parts, and the stability of the disease free steady state depends upon the magnitude of this term. If  $n > m$ , so that the bacterial reproduction rate is greater than the death rate, then the second term in (2.18) will always be larger than the first, so there will always be at least one positive eigenvalue, meaning that the disease free steady state will be unstable. That is, when  $n > m$ , the bacterial population will grow exponentially as soon as the bacteria is introduced into the environment and will continue to infect the population regardless of the rate at which infecteds contribute bacteria to the reservoir. However, if the bacterial reproduction rate is smaller than the death rate,  $n < m$ , and additionally

$$\frac{\delta a K}{B} < (\mu + \gamma)(m - n) \quad (2.19)$$

then the second term in (2.18) will be smaller than the first, all of the eigenvalues will be negative, and the disease free steady state will be stable. This then gives the condition for a transition between the system being in a disease free state to an endemic or epidemic state. So the critical carrying capacity,  $K_c$ , for an endemic state to occur is given by:

$$K_c = \frac{B(\mu + \gamma)(m - n)}{\delta a}. \quad (2.20)$$

This implies that the human population must have a minimum size for a endemic state to persist. This minimum size is determined by the biological parameters for the survival of the

bacteria, contact between humans and bacteria, as well those determining the infectiousness and virulence of the disease. Conversely, we can use this information to determine if a population of a given size could support endemic cholera, given these parameters. We can re-write this criteria, so we see an endemic state if:

$$\frac{\delta a K}{B(\mu + \gamma)(m - n)} \geq 1, \quad (2.21)$$

from which we conclude that the basic reproductive rate,  $R_0$  is given by:

$$R_0 = \frac{\delta a K}{B(\mu + \gamma)(m - n)}. \quad (2.22)$$

Sample trajectories for this system appear in Figure 2.1. If initially either  $C_0 > 0$  or  $I_0 > 0$  in a population where  $r$  is positive, as we've assumed, there will always be a cholera outbreak in a population where  $K \geq K_c$ . Notice that the transition between a disease free and epidemic or endemic state depends partly upon the population carrying capacity,  $K$ . However, the rest of the variables that determine this transition are disease parameters, *i.e.*, the biology of the disease, including the biology of the bacteria in the reservoir, determines how the disease will progress in a population.

### 2.3 The Dynamics of the Bacterial Population

We can now see that the values of the various model parameters involving the bacterial population and the contact between the bacterial and human populations are key in determining the possibility of endemic infection. Variation in these parameters, such as some kind of oscillatory forcing caused by seasonal effects, can influence the size and shape of outbreaks. For

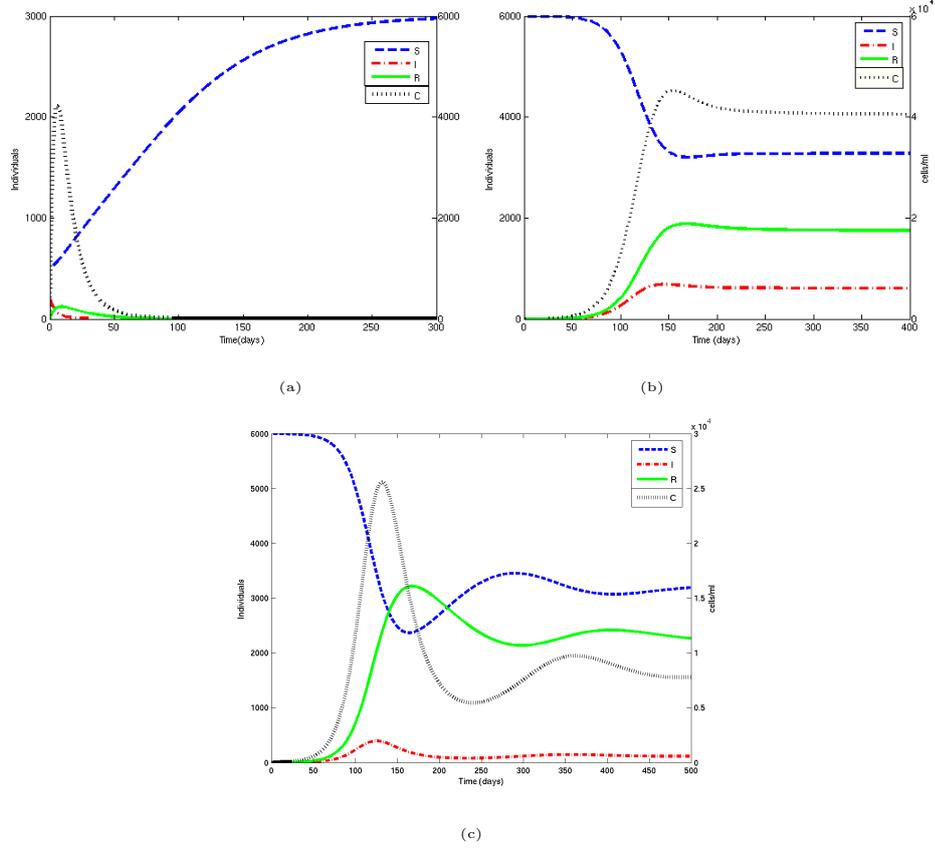


Figure 2.1: Sample population trajectories for the model (2.5)–(2.8). (a) Disease Free:  $r = 0.2$ ,  $K = 3000$ ,  $D_0 = 0.06$ ,  $a = 1$ ,  $B = 10^6$ ,  $\gamma = 0.2$ ,  $\mu = 0.01$ ,  $\delta = 10$ ,  $n = 0.3$ ,  $m = 0.45$ , with initial conditions  $(1000, 100, 0, 0)$  (b) Endemic:  $r = 0.02$ ,  $K = 6000$ ,  $D_0 = 0.06$ ,  $a = 1$ ,  $B = 10^6$ ,  $\gamma = 0.2$ ,  $\mu = 0.01$ ,  $\delta = 10$ ,  $n = 0.3$ ,  $m = 0.45$ , with initial conditions  $(6000, 1, 0, 0)$  (c) Endemic with oscillations:  $r = 0.002$ ,  $K = 6000$ ,  $D_0 = 0.01$ ,  $a = 1$ ,  $B = 10^6$ ,  $\gamma = 0.2$ ,  $\mu = 0.01$ ,  $\delta = 10$ ,  $n = 0.3$ ,  $m = 0.45$ , with initial conditions  $(6000, 1, 0, 0)$

this system, environmental forcing on a seasonal time scale would likely cause oscillations in either contact rates or bacterial survival. Longer time scale variations, such as due to climate change or evolution of the bacteria, could also occur, but I will ignore these effects for now and focus on the possible effects of seasonal oscillations. If both the contact rate and bacterial survival parameters are varying simultaneously, these variations could have the same oscillatory period, they can be offset or have different periods altogether. How these oscillations combine could have major implications for the behavior of the system. They would definitely cause the

populations to oscillate in time, but since this a non-linear system, the behavior is likely to be complicated.

To explore this more concretely, I first assume an oscillation of the form

$$s(\theta) = 1 + \cos(\theta) \tag{2.23}$$

where  $\theta = \frac{t}{t_e}$  with  $t_e$  the characteristic time of environmental oscillations. We can then define the contact rate so that

$$a \rightarrow as(\theta), \tag{2.24}$$

and the bacterial survival/reproduction rate,

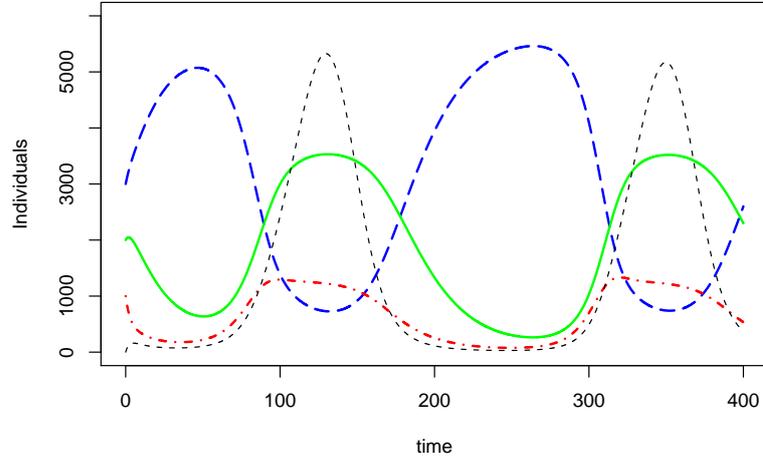
$$n - m \rightarrow (n - m)s(\theta). \tag{2.25}$$

In Figures 2.2 through 2.5 I show examples of the kinds of behavior the system exhibits for two values of  $t_e$  with various combinations of oscillations in the contact rate and bacterial survival. Notice that longer periods of oscillations in bacterial survival generally results in endemic infection peaks that are slightly lower and more spread out, but also tends to increase the maximum size that the bacterial population can achieve (Figures 2.2). Oscillations in the contact rate tends to keep the overall bacterial population lower, with infection outbreaks that have steep initial edges and longer tails (Figures 2.3). When both oscillations are present, these trends still occur. However, when the oscillations are synchronized, the dynamics of the contact rate tend to dominate (Figures 2.4), whereas when the oscillations are offset, the forcing in the bacterial survival influences the dynamics of the system more (Figures 2.5).

The characteristics of *Vibrio cholerae* as both an infectious agent and a member of

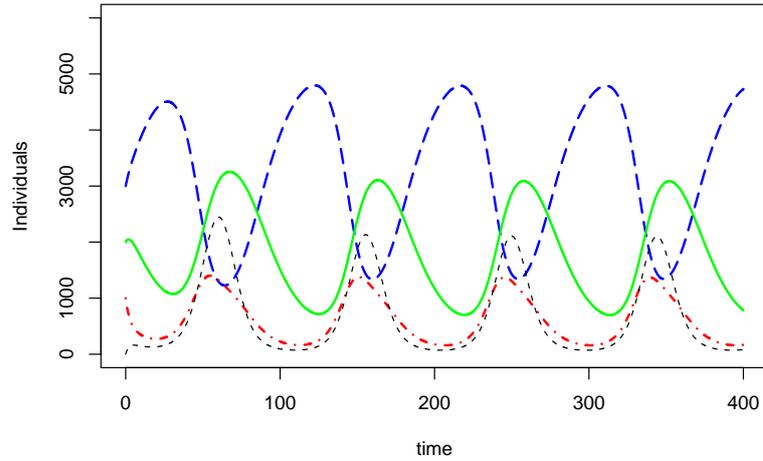
the aquatic community greatly affect the dynamics of the disease cholera. The reactions of the model system to forcing in different parameters indicate that to understand the dynamics of the disease in a human population it is important to understand the underlying causes of the observed seasonal oscillations. In other words, we need to understand how the environment impacts both the survival and reproduction of bacteria, as well as how environmental factors may influence the likelihood of humans to have contact with critical doses of bacteria. Understanding these factors will help us better understand and predict outbreaks, and help to determine what methods would be best control these outbreaks. The following chapters will focus on exploring factors influencing fitness and survival of bacteria on both individual and population levels.

Endemic cholera: time varying bacterial survival:  $\theta=t/35$



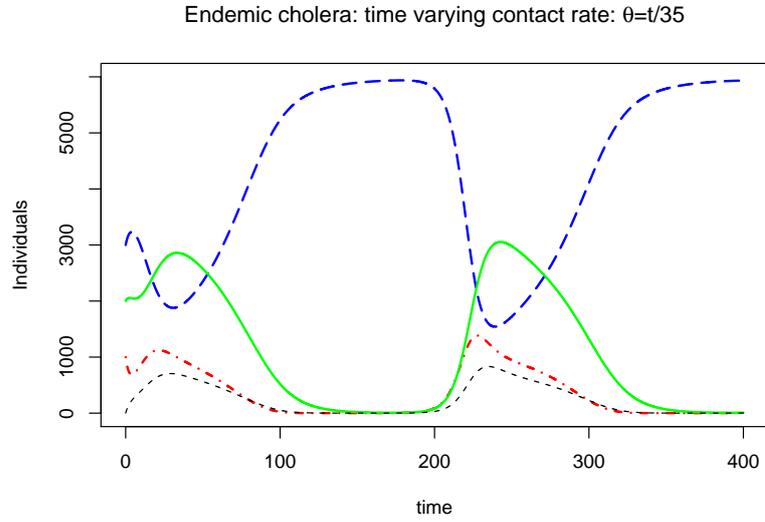
(a)

Endemic cholera: time varying bacterial survival:  $\theta=t/15$

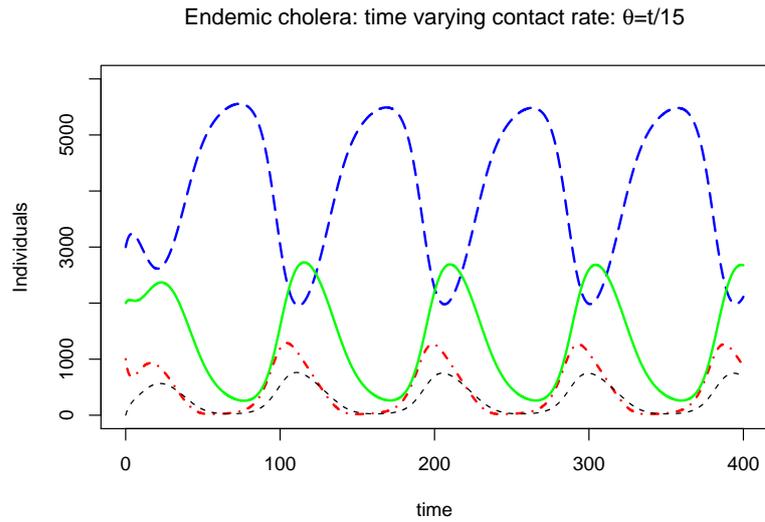


(b)

Figure 2.2: Sample population trajectories for the model (2.5)–(2.8) with forcing as described in 2.23. Parameters are for the endemic case:  $r = 0.02$ ,  $K = 6000$ ,  $D_0 = 0.06$ ,  $a = 1$ ,  $B = 10^6$ ,  $\gamma = 0.2$ ,  $\mu = 0.01$ ,  $\delta = 10$ ,  $n = 0.3$ ,  $m = 0.45$ , with initial conditions  $(3000, 2000, 1000, 0)$ . Lines are the same as in Figure 2.1 with the bacterial density corresponding to  $y \cdot 100$  cells/mL (a) Oscillations in bacterial survival only, with  $\theta = \frac{t}{35}$  (b) Oscillations in bacterial survival only, with  $\theta = \frac{t}{15}$



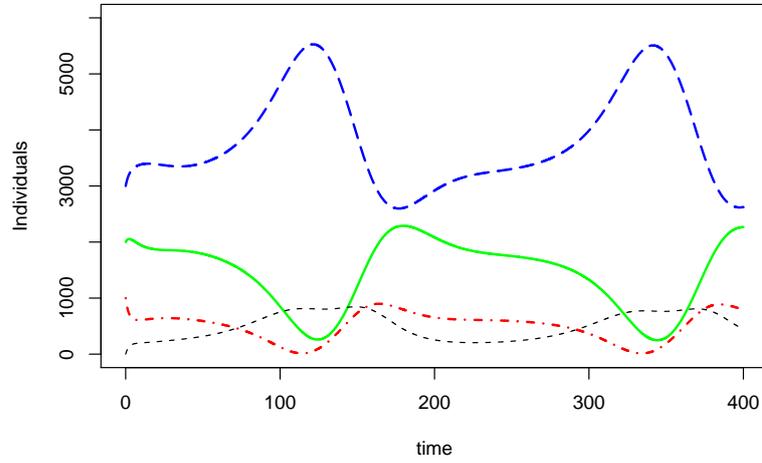
(a)



(b)

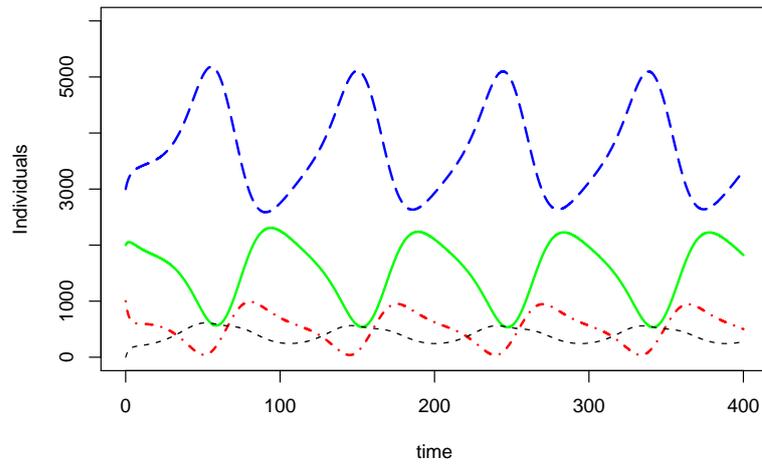
Figure 2.3: Sample population trajectories for the model (2.5)–(2.8) with forcing as described in 2.23. Parameters are for the endemic case:  $r = 0.02$ ,  $K = 6000$ ,  $D_0 = 0.06$ ,  $a = 1$ ,  $B = 10^6$ ,  $\gamma = 0.2$ ,  $\mu = 0.01$ ,  $\delta = 10$ ,  $n = 0.3$ ,  $m = 0.45$ , with initial conditions (3000, 2000, 1000, 0). Lines are the same as in Figure 2.1 with the bacterial density corresponding to  $y * 100$  cells/mL (a) Oscillations in contact rate only, with  $\theta = \frac{t}{35}$  (b) Oscillations in contact rate only, with  $\theta = \frac{t}{15}$

Endemic cholera: time varying contact rate & bacterial survival:  $\theta=t/35$



(a)

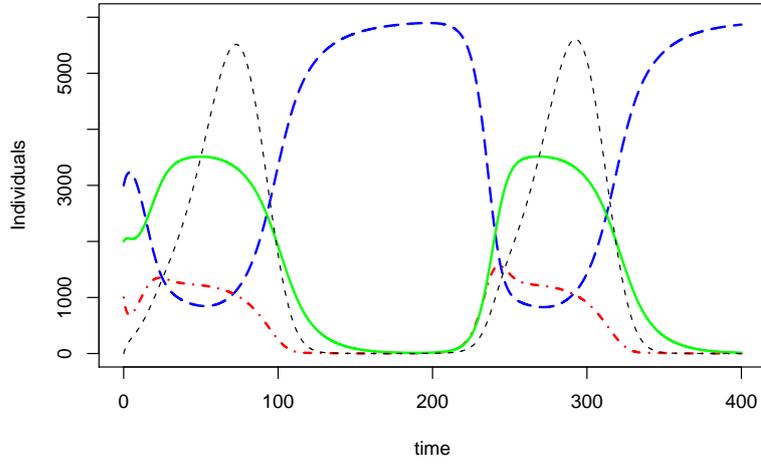
Endemic cholera: time varying contact rate & bacterial survival:  $\theta=t/15$



(b)

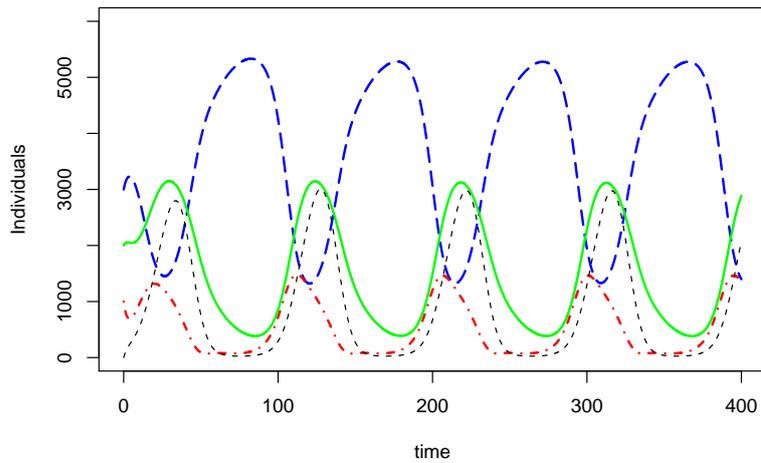
Figure 2.4: Sample population trajectories for the model (2.5)–(2.8) with forcing as described in 2.23. Parameters are:  $r = 0.02, K = 6000, D_0 = 0.06, a = 1, B = 10^6, \gamma = 0.2, \mu = 0.01, \delta = 10, n = 0.3, m = 0.45$ , with initial conditions  $(3000, 2000, 1000, 0)$ . Lines are the same as in Figure 2.1 with the bacterial density corresponding to  $y * 100$  cells/mL (a) Oscillations in both bacterial survival and contact rate, with  $\theta = \frac{t}{35}$  (b) Oscillations in both bacterial survival and contact rate, with  $\theta = \frac{t}{15}$

Endemic cholera: time varying contact rate & offset bacterial survival:  $\theta=t/35$



(a)

Endemic cholera: time varying contact rate & offset bacterial survival:  $\theta=t/15$



(b)

Figure 2.5: Sample population trajectories for the model (2.5)–(2.8) with forcing as described in 2.23. Parameters are:  $r = 0.02$ ,  $K = 6000$ ,  $D_0 = 0.06$ ,  $a = 1$ ,  $B = 10^6$ ,  $\gamma = 0.2$ ,  $\mu = 0.01$ ,  $\delta = 10$ ,  $n = 0.3$ ,  $m = 0.45$ , with initial conditions  $(3000, 2000, 1000, 0)$ . Lines are the same as in Figure 2.1 with the bacterial density corresponding to  $y * 100$  cells/mL (a) Oscillations in both bacterial survival and contact rate, with  $\theta = \frac{t}{35}$  and with the oscillations offset by  $\frac{2i}{2}$  (b) Oscillations in both bacterial survival and contact rate, with  $\theta = \frac{t}{15}$  and with the oscillations offset by  $\frac{2i}{2}$

## Chapter 3

# Bacterial Life Histories and Aging

Living organisms exhibit an astonishing variety of lifespans and life histories, characterized by diverse patterns of aging and senescence. A giant sequoia tree can live for thousands of years, producing only a few offspring over its lifetime, while a dandelion survives only for a season, but produces thousands of seeds, often resulting in hundreds of new plants the next year. Within a single genus, such as rockfish (*Sebastes*), species exhibit lifespans ranging from 10 to 200 years [55]. Even bacteria and single celled organisms appear to age. New research by Stewart *et al.* [84] indicates that even *Escherichia coli*, which appear to divide symmetrically, actually divide into one “old” and one “young” cell, with the old cell reproducing more slowly with each generation.

Current theories of aging seek to combine principles of evolution with theories from physiology, microbiology, and genetics [73]. For instance, the mutation accumulation theory of aging hypothesizes that lethal genetic mutations which affect organisms late in life will not be

selected against, because the force of selection decreases with age. Over time, these negative, late-acting mutations can accumulate, resulting in increased mortality as organisms age [60]. The antagonistic pleiotropy theory takes this a step further and hypothesizes that these late acting mutations may be selected for if they benefit an organism earlier in life [91, 32]. On the other hand, the reliability theory of aging and longevity hypothesizes that over time organisms wear out and eventually fail due to the loss of irreplaceable parts [26, 27].

Another theory is the disposable soma theory of aging [21, 24, 45]. This theory predicts that because organisms have a finite amount of energy to use for all life functions, there is a trade-off between repairing and maintaining the soma or reproducing. If energy is used to maintain the soma, there might not be enough energy to reproduce, and vice versa. We therefore expect that the optimal allocation strategy, which would maximize the representation of an organism's genes in future generations, will not be one that allows an organism to maintain the soma indefinitely.

These theories make similar predictions about how aging and lifespan will evolve in response to extrinsic mortality. For instance, if an organism experiences high natural mortality, natural selection will result in greater investment in offspring than in soma. High natural mortality is also predicted to encourage earlier maturation, so an organism will be less likely to die before having an opportunity to reproduce. Organisms with low natural mortality are predicted to maintain the soma for longer, produce offspring less frequently, and experience longer lives. However, disposable soma theory also predicts how natural mortality is expected to influence life span and reproductive schedules, and gives insight into responses to aging in terms of allocation of resources for repair and maintenance.

## 3.1 Fitness Models of Bacteria and Single-Celled

### Organisms

Single-celled organisms, as well as any organisms where the soma and the germ line are not separate, were once expected to be immortal [91]. Later it was observed that some asymmetrically dividing single celled organisms, such as the yeast *Saccharomyces cerevisiae* [63] and the bacterium *Caulobacter crescentus* [1] age and die. Therefore, it was hypothesized that the distinction between organisms that age and those that do not depends upon asymmetry in reproduction [68]. Recent research by Stewart *et al.* [84] indicates that even bacteria that appear to divide symmetrically, such as *Escherichia. coli*, actually produce functionally asymmetric cells during cell division. They identify one of the cells as the aging parent cell that produces an offspring that is “rejuvenated,” and find evidence that these older cells reproduce more slowly as they age, and may even stop reproducing [84].

Quantitative models can be useful for exploring how evolutionary trade-offs shape aging and senescence in these simple organisms. General understanding of the trade-offs that shape bacterial life histories is important for understanding *V. cholerae* specifically. In this chapter I utilize a simple mathematical model to explore the effect of senescence, in the form of finite reproductive lifespan and increasing doubling times, on bacterial fitness and resource allocation. I first introduce a baseline model without aging to provide a point of reference with which to compare the model with aging. I then present two models of aging in simple organisms. The first considers only limited reproduction. The second includes both limited reproduction as well as increasing doubling times.

### 3.1.1 The Baseline Model:

#### Bacterial Population without Aging

Various approaches are available for modeling life-history strategies [9, 76, 83]. I use the Euler-Lotka equation to explore the effects of life history choices on fitness of single-celled organisms. My selected measure of fitness is the intrinsic rate of natural increase, for populations living in a constant environment with age dependent reproduction and mortality schedules, denoted by  $r$ . The Euler-Lotka Equation in continuous time is given by

$$1 = \int_0^{\infty} e^{-rx} l_x b_x dx. \quad (3.1)$$

Here the probability of surviving to age  $x$  is denoted as  $l_x$  and the number of offspring born to an individual of age  $x + dx$  is  $b_x dx$ . Although this model assumes exponential growth,  $r$  is an appropriate measure of fitness even for populations whose growth is limited by density dependent effects, as long as the carrying capacity is not strategy dependent [13].

Before examining a model with aging, I review a baseline model that assumes infinite reproductive potential. Kirkwood [45] proposed a simple model of bacterial fitness for cells that divide perfectly symmetrically, based upon (3.1), with appropriate choices of  $b_x$  and  $l_x$  for a clonally reproducing population. First, let  $l_x$  be an exponentially decreasing survival probability,  $l_x = e^{-mx}$ , where  $m$  is the constant extrinsic mortality rate. Since so little is known, this is a good first approximation to extrinsic mortality for these simple organisms. The birth rate depends upon the doubling time,  $T$ . If an individual bacteria survives to time  $T$ , it divides. Since the division is perfectly symmetric, we cannot tell the difference between the two resulting cells. We therefore consider both of the cells to be identical offspring, and the original bacteria is essentially “dead” (rather like a semelparous organism). If the offspring

have the same doubling time as the original cell, an appropriate “birth” rate would therefore be  $b_x = 2\delta(x - T)$ , where  $\delta(x - T)$  is the Dirac delta function<sup>1</sup>.

With these expressions for  $l_x$  and  $b_x$  the Euler-Lotka equation (3.1) becomes:

$$2e^{-(m+r)T} = 1. \quad (3.2)$$

In Figure 3.1, I show this functional relationship between fitness,  $r$ , mortality rate,  $m$ , and doubling time,  $T$ . As  $m$  and  $T$  increase,  $r$  decreases. As  $T \rightarrow 0$ ,  $r \rightarrow \infty$  regardless of the value of  $m$ . As  $m$  and  $T$  increase,  $r$  decreases.

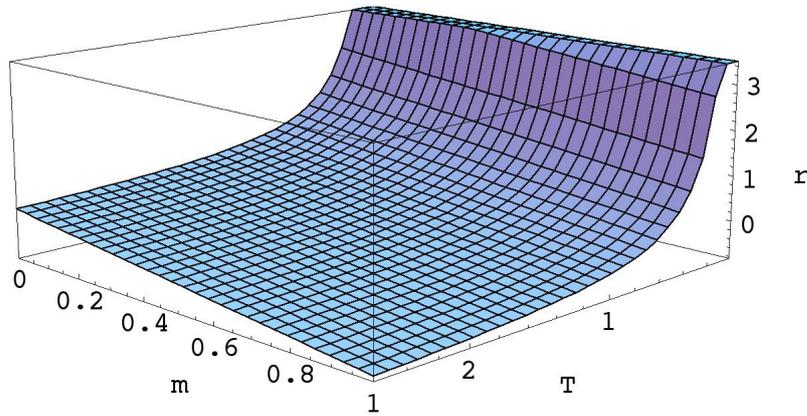


Figure 3.1: The intrinsic rate of natural increase,  $r$ , as a function of the doubling time,  $T$ , and the mortality,  $m$ .

Trade-offs between mortality and reproduction can be explored by examining how

---

<sup>1</sup>The Dirac delta function is defined as a unit impulse at some point  $x_0$  such that:

$$\begin{aligned} \delta(x - x_0) &= 0, & x &\neq x_0 \\ \int_{-\infty}^{\infty} \delta(x - x_0) dx &= 1, \end{aligned}$$

and given an arbitrary function  $f(x)$ :

$$\int_{-\infty}^{\infty} f(x)\delta(x - x_0)dx = f(x_0).$$

resource allocation impacts the mortality rate and doubling time of the bacteria. I denote the fraction of resources allocated for growth and reproduction by  $\rho$ , and the fraction allotted for maintenance/repair and survival by  $1 - \rho$ . Following Kirkwood [45], I can parameterize the mortality,  $m$ , and doubling time,  $T$ , in terms of  $\rho$  as:

$$T(\rho) = \frac{T_0}{\rho} \quad (3.3)$$

$$m(\rho) = \frac{m_0}{1 - \rho}. \quad (3.4)$$

Here,  $T_0$  can be thought of as the minimum possible time it would take for the bacteria to reproduce if all of its resources are allocated to growth;  $m_0$  is the minimum mortality of the bacteria if all resources are allocated to survival. Solving for  $r$  in (3.2) with the expressions for  $T$  and  $m$  in (3.3)–(3.4) yields:

$$r = \frac{\rho}{T_0} \ln 2 - \frac{m_0}{1 - \rho}. \quad (3.5)$$

Maximizing (3.5) with respect to  $\rho$  gives the optimal resource allocation,  $\rho^*$ :

$$\rho^* = 1 - \left( \frac{T_0 m_0}{\ln 2} \right)^{\frac{1}{2}}, \quad (3.6)$$

and the corresponding value of  $r_{max}$ ,

$$r_{max} = \frac{1}{T_0} \left( \ln 2 - 2(m_0 T_0 \ln 2)^{\frac{1}{2}} \right). \quad (3.7)$$

If  $T_0$  and  $m_0$  are constants,  $r(\rho)$  will follow curves similar to those depicted in Figure 3.2. We can see from Eqn. (3.6) and Figure 3.2 that if there is any mortality (*i.e.*  $m_0 \neq 0$ ), the optimal strategy will never be to dedicate all resources to reproduction, *i.e.*  $\rho^* < 1$ . Also, from Eqn. (3.6) we see that optimal allocation strategy is influenced equally by the minimum

mortality,  $m_0$ , and minimum doubling time,  $T_0$ , in particular by the dimensionless quantity  $m_0T_0$ . In an environment with low mortality, it can be optimal to invest most resources in reproduction ( $\rho^* \rightarrow 1$ ), as shown in Figure 3.2a, even if the generation time is relatively long (lowest curve). However, if the mortality is high and/or the doubling time is long, the optimal resource allocation may be to use more resources for survival ( $\rho^* \rightarrow 0$ ) (Figure 3.2b). However, the maximum fitness,  $r_{max}$ , is more sensitive to the value of the minimum doubling time,  $T_0$ , than to the minimum mortality rate,  $m_0$ , so we expect that there would be stronger selection to reduce  $T_0$  than  $m_0$ .

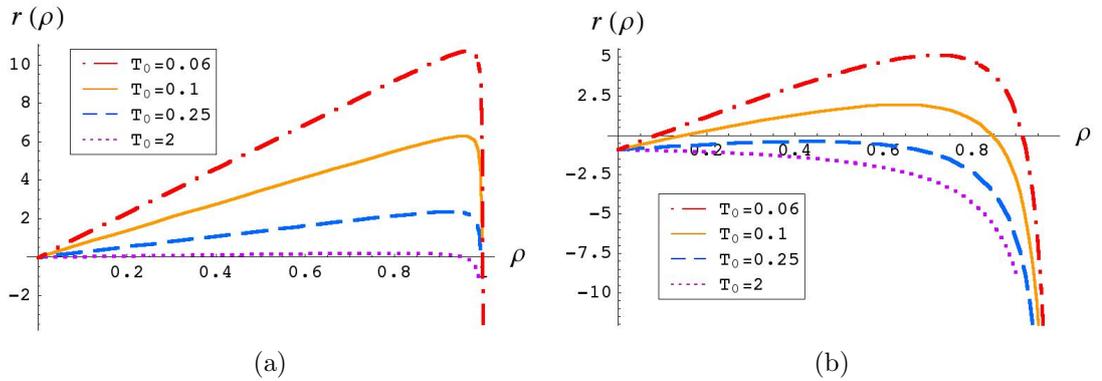


Figure 3.2:  $r$  as a function of  $\rho$  for four values of  $T_0 = (2, 0.25, 0.1, 0.06)$  for (a)  $m_0 = 0.015$  (b)  $m_0 = 0.9$ . High values of  $T_0$  correspond to the lowest curves in each plot, and small  $T_0$  to higher curves.

## 3.2 Effects of Limited Reproduction on the Fitness of Simple Organisms

Stewart *et al.* [84] found that over time *E. coli* reproduce more slowly, and may eventually stop reproducing altogether. This is also the case for organisms such as *Saccharomyces cerevisiae* [63] and the fission yeast *Schizosaccharomyces pombe* [4]. In addition to slowing repro-

duction with age, both of these yeasts have short replicative lifespans ( $\sim 30$  and  $10$  replications respectively).

In this section, I explore how the fitness of unicellular organisms, measured by the intrinsic rate of natural increase and denoted by  $\tilde{r}$ , is affected by finite reproductive life span. For simplicity, I assume that the only effect of aging is limited reproduction, and that mortality is not constant across all age classes. Additionally I assume that the doubling time is constant.

Starting from the Euler-Lotka Eqn. (3.1) I assume, as in the baseline model, an exponentially decreasing survival rate. I modify the birth rate,  $b_x$ , to take into account a functional asymmetry in cell division. The cell is able to divide and produce a single offspring in a given fixed doubling time  $T$ , then can live to divide again later. We define cellular “age”,  $a = 1, 2, \dots, a_{max}$ , as the number of times the cell has doubled, where  $a_{max}$  is the maximum number of times it can split. Additionally, I assume that the offspring is “rejuvenated”, in that its initial age is set to zero, instead of to the maternal age. The birth rate is then a sum of delta functions spaced a distance  $T$  apart:

$$b_x = \sum_{a=1}^{a_{max}} \delta(x - aT). \quad (3.8)$$

This, together with the previous expression for  $l_x$ , inserted into (3.1) gives

$$\begin{aligned} 1 &= \int_0^{\infty} e^{-\tilde{r}x} e^{-mx} \sum_{a=1}^{a_{max}} \delta(x - aT) dx \\ &= \sum_{a=1}^{a_{max}} e^{-(\tilde{r}+m)aT}. \end{aligned} \quad (3.9)$$

Evaluating the sum in Eqn. (3.9) results in

$$e^{(\tilde{r}+m)T} = 2 - e^{-(\tilde{r}+m)Ta_{max}}. \quad (3.10)$$

This equation has two solutions for  $a_{max} \geq 1$ . A trivial solution exists when  $\tilde{r} = -m$ . All other solutions depend upon the maximum age,  $a_{max}$ , the mortality rate,  $m$ , and the doubling time,  $T$ . Since Eqn. (3.10) does not have an exact closed form solution, I explore the relationships between fitness and other parameters using numerical solutions or analytic approximations. In Figure 3.3, I show the numerical solution to Eqn. (3.10) for various combinations of  $a_{max}$ ,  $m$ , and  $T$ . The value of  $\tilde{r}$  varies considerably depending upon the combination of these three parameters. Variation of  $T$  and  $a_{max}$  have the most impact upon  $\tilde{r}$ , as I show in Figure 3.3a, whereas variations in  $m$  act to shift  $\tilde{r}$  up or down a fixed amount when  $T$  is held constant (Figure 3.3b). Small perturbations of  $a_{max}$  when  $a_{max}$  is small has a much larger impact upon  $r$  than perturbations of  $a_{max}$  when  $a_{max}$  is large. When  $a_{max} \rightarrow \infty$ , Eqn. (3.10) reduces to Eqn. (3.2), and  $\tilde{r} \rightarrow r$ . In other words, infinite reproductive lifespan and perfectly symmetrical cell division are equivalent.

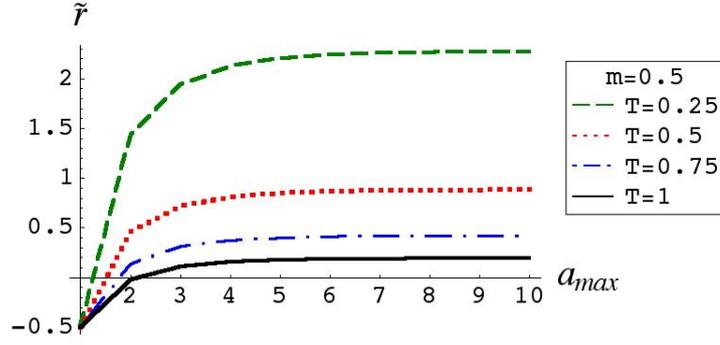
Since  $\tilde{r} \rightarrow r$  as  $a_{max}$  increases, I can approximate the second solution of Eqn. (3.10) in terms of  $r$  for large values of  $a_{max}$  using Newton's method. The iterative formula is given by

$$x_{n+1} \approx x_n - \frac{f(x_n)}{f'(x_n)},$$

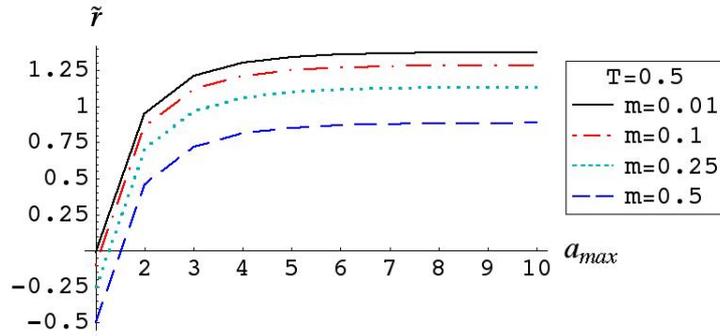
where  $f'(x_n)$  denotes the first derivative of  $f(x)$  at the point  $x_n$ . Choosing  $x_0 = r$ , the first order approximation ( $x_1$ ) to Eqn. (3.10) is simply:

$$\tilde{r} \approx r - \frac{1}{T} \frac{2^{-a_{max}}}{\ln 2 + a_{max} 2^{-a_{max}}}. \quad (3.11)$$

As can be seen in Figure 3.4, this approximation is remarkably good over a wide range of values of  $a_{max}$ , so further approximations are not necessary. This approximation also gives us some idea of the values of  $a_{max}$  that are "large". Where the approximation is valid,  $a_{max}$



(a)



(b)

Figure 3.3:  $\tilde{r}$  vs  $a_{max}$  for (a)  $m = 0.5$ ,  $T = (0.25, 0.5, 0.75, 1)$  and (b)  $T = 0.5$  and  $m = (0.5, 0.25, 0.1, 0.01)$  (c) Effects of  $a_{max} = (2, 3, 5, 10)$  on  $\tilde{r}(T)$  for  $m = 0.9$

must be large enough to make the difference in fitness between the aging bacteria and the immortal bacteria,  $r - \tilde{r}$ , very small, corresponding to values of  $a_{max} \geq 10$ .

Using this approximation, I can look at how fitness changes when an additional replication is added,

$$\Delta\tilde{r} = \tilde{r}(a_{max} + 1) - \tilde{r}(a_{max}), \quad (3.12)$$

to learn more about how selective pressure might influence the value of  $a_{max}$ . This is similar to the approach [32, 9, 10] used to explore how natural selection shapes fecundity and mortality as

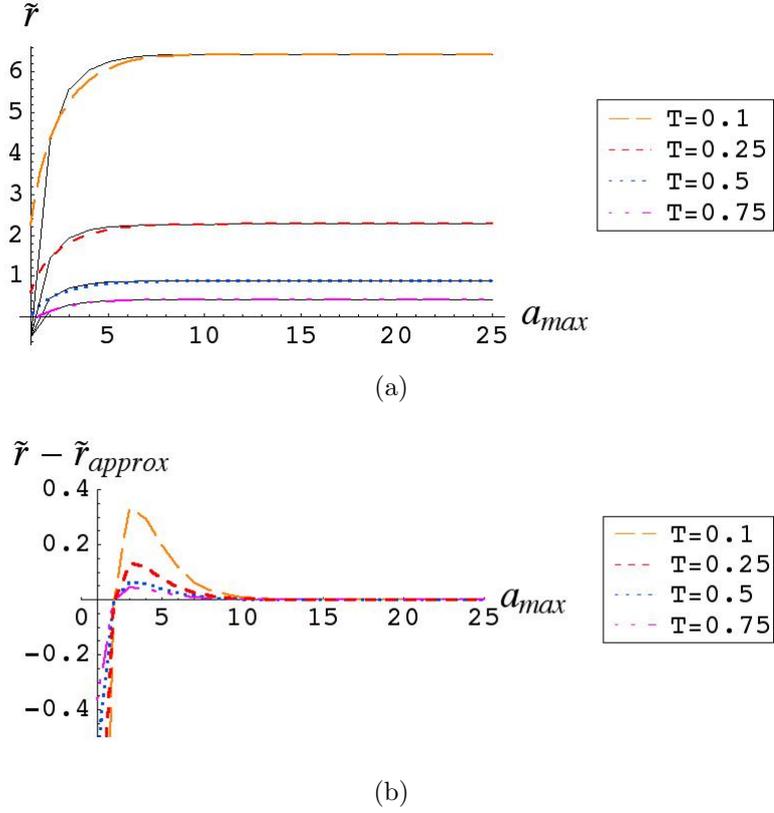


Figure 3.4: (a) Analytic approximation (dashed lines) and numerical solution (solid lines) of  $\tilde{r}$  vs  $a_{max}$  for  $m = 0.5$  and  $T = (0.1, 0.25, 0.5, 0.75)$  (b) The difference between the numerical solution and analytic approximation,  $\tilde{r} - \tilde{r}_{approx}$ , for the same values as in (a)

a function of age. The numerical solution is shown in Figure 3.5. We can also use Eqn. (3.11) to find an analytic approximation.

$$\Delta\tilde{r} \approx \frac{2^{-a_{max}-1}}{T} \left( \frac{2}{\ln 2 + a_{max}2^{-a_{max}}} - \frac{1}{\ln 2 + (a_{max} + 1)2^{-a_{max}-1}} \right) \quad (3.13)$$

From both the numerical solution and analytic approximation, we can see that the value of additional replications declines rapidly as the maximum number of possible replications,  $a_{max}$ , increases (Figure 3.5). This indicates that the selective pressure to increase  $a_{max}$  above

10 or 20 should be small.

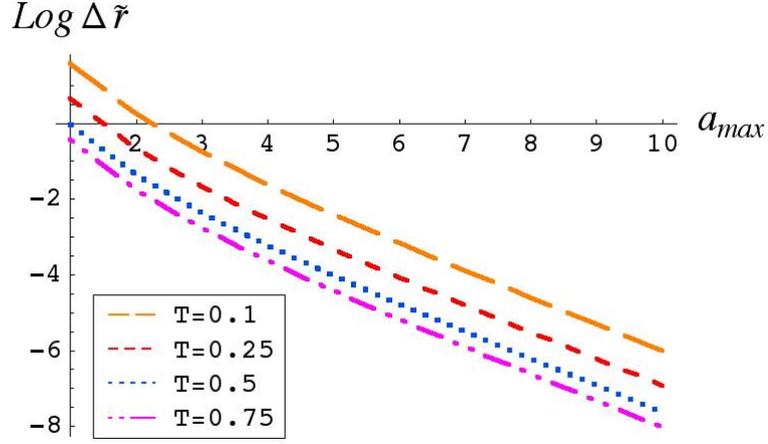


Figure 3.5: Fitness gained by the organisms by increasing the maximum number of doublings from  $a_{max}$  to  $a_{max} + 1$ , *i.e.*  $\Delta\tilde{r} = \tilde{r}(a_{max} + 1) - \tilde{r}(a_{max})$ .

Since the fitness of the aging and immortal bacteria are very close when  $a_{max}$  is large, we would expect that the allocation strategies would be similar as well. I denote the allocation strategy of an aging bacteria by  $\tilde{\rho}$ . Substituting (3.3) and (3.4) into (3.10) gives a relationship between  $\tilde{r}$ ,  $\tilde{\rho}$ ,  $m_0$ ,  $T_0$ , and  $a_{max}$ :

$$\exp\left(\left(\tilde{r} + \frac{m_0}{1 - \tilde{\rho}}\right)\frac{T_0}{\tilde{\rho}}\right) = 2 - \exp\left(-\left(\tilde{r} + \frac{m_0}{1 - \tilde{\rho}}\right)\frac{T_0}{\tilde{\rho}}a_{max}\right). \quad (3.14)$$

Unlike the baseline model without aging, Eqn. (3.14) does not have an exact closed form analytic solution. In this case a numerical solution was obtained using standard root finding algorithms in Mathematica. In Figure 3.6, I show a numerical solution for the fitness of the aging bacteria,  $\tilde{r}$ , as a function of fraction of resources allocated for reproduction,  $\tilde{\rho}$ , for various values of  $a_{max}$ ,  $m_0$ , and  $T_0$ . The fitness measure,  $\tilde{r}$ , increases dramatically between the lowest values of  $a_{max}$ . However, for higher values of  $a_{max}$ ,  $\tilde{r}$  hardly varies. Thus there is a

lower return in fitness for more opportunities to reproduce. Additionally, notice that as  $a_{max}$  increases, the peak in  $\tilde{r}$  becomes more pronounced, and shifts towards higher values of  $\tilde{\rho}$ . Thus, bacteria that have more opportunities to reproduce gain more fitness from focusing resources for reproduction than those that cannot.

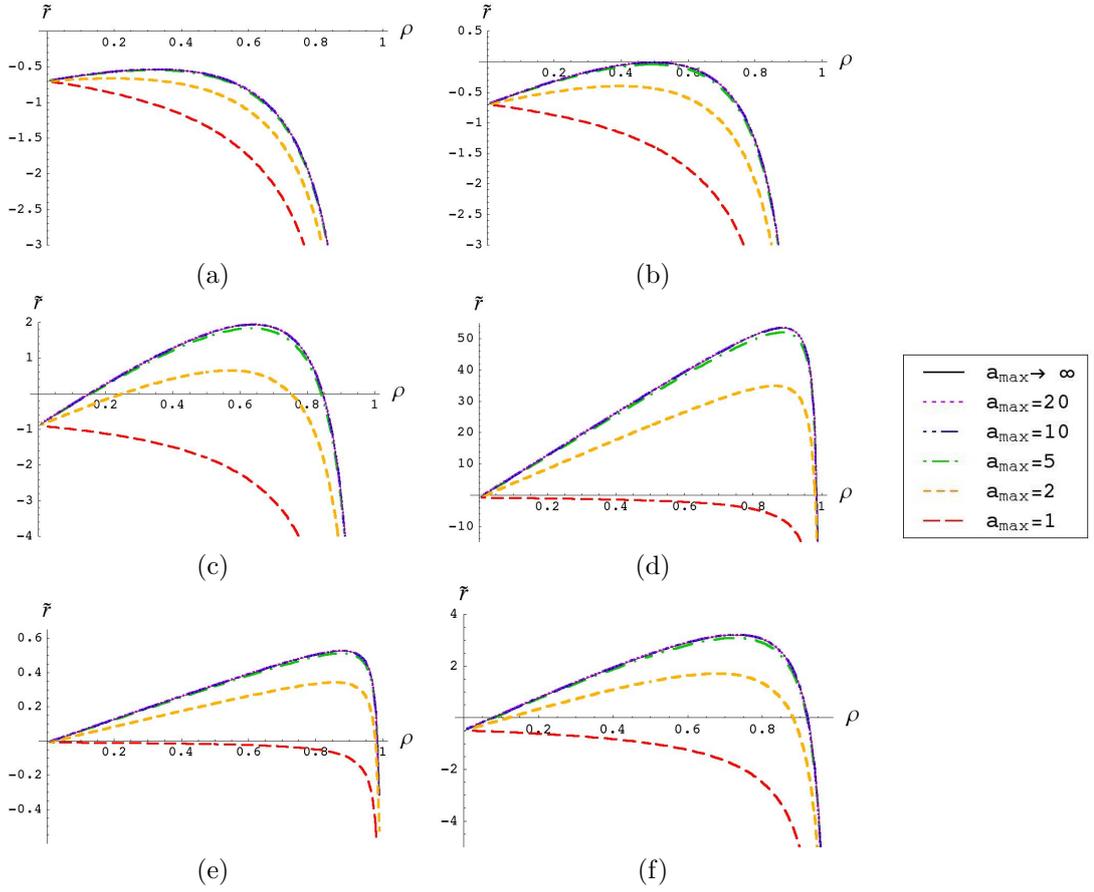


Figure 3.6: Effects of  $a_{max} = (1, 2, 5, 10, 20)$  on  $\tilde{r}(\tilde{\rho})$  and  $r(\rho)$  (solid black line) for (a)  $m_0 = 0.7, T_0 = 0.45$ ; (b)  $m_0 = 0.7, T_0 = 0.25$ ; (c)  $m_0 = 0.9, T_0 = 0.1$ ; (d)  $m_0 = 0.9, T_0 = 0.01$ ; (e)  $m_0 = 0.01, T_0 = 1$ ; and (f)  $m_0 = 0.5, T_0 = 0.1$ .

I can quantify how aging shifts the optimal allocation strategy away from the value determined in Eqn. (3.6) for immortal bacteria. If I differentiate Eqn. (3.14) with respect to  $\tilde{\rho}$ , and evaluate it at the optimal strategy,  $\tilde{\rho}^*$ , so that  $\frac{d\tilde{r}(\tilde{\rho}^*)}{d\tilde{\rho}} = 0$ , I find an implicit expression for

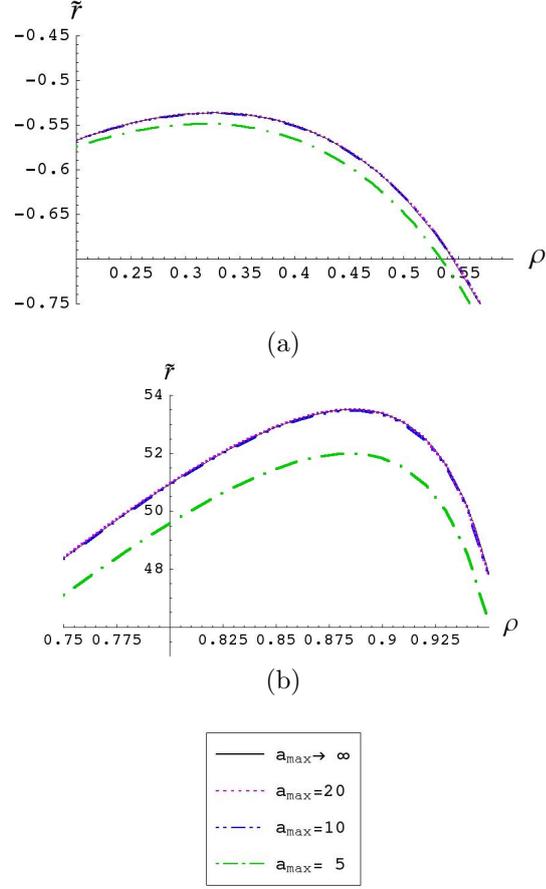


Figure 3.7: Close-ups of the maxima of  $\tilde{r}(\tilde{\rho})$  for  $a_{max} = (5, 10, 20)$  and  $r(\rho)$  (solid black line) for (a)  $m_0 = 0.7$ ,  $T_0 = 0.25$  (corresponding to Figure 3.6a ) (b)  $m_0 = 0.9$ ,  $T_0 = 0.01$ ; (corresponding to Figure 3.6d )

the maximum fitness,  $\tilde{r}(\tilde{\rho}^*) = \tilde{r}_{max}$ , as a function of  $\tilde{\rho}^*$ :

$$((2\tilde{\rho}^* - 1)m_0 - \tilde{r}_{max}(1 - \tilde{\rho}^*)^2) \left[ e^{(\tilde{r}_{max} + m(\tilde{\rho}^*))T(\tilde{\rho}^*)} + a_{max}e^{-(\tilde{r}_{max} + m(\tilde{\rho}^*))T(\tilde{\rho}^*)a_{max}} \right] = 0 \quad (3.15)$$

Since the sum of exponential terms in (3.15) can never be equal to zero, I conclude that

$$(2\tilde{\rho}^* - 1)m_0 - \tilde{r}_{max}(1 - \tilde{\rho}^*)^2 = 0,$$

and therefore  $\tilde{r}_{max}$  is related to  $\tilde{\rho}^*$  by

$$\tilde{r}_{max} = \frac{(2\tilde{\rho}^* - 1)}{(1 - \tilde{\rho}^*)^2} m_0. \quad (3.16)$$

It is straightforward to confirm that this expression holds for the case when  $a_{max} \rightarrow \infty$ , in which case we regain (3.6). As previously discussed, as the maximum cellular age,  $a_{max}$ , increases, the difference between the fitness of the immortal bacteria,  $r$ , and the fitness of the aging bacteria,  $\tilde{r}$ , decreases. In particular, for large values of  $a_{max}$  we saw that  $r \rightarrow \tilde{r}$ . We can also see in Figure 3.6 that when  $a_{max} \approx 10$  the peaks in the fitness curves are very close, so that  $r_{max} \approx \tilde{r}_{max}$ , and occur at about the same value of the allocation strategy, *i.e.*  $\tilde{\rho}^* \approx \rho^*$ .

This suggests that we can find approximate expressions for  $\tilde{\rho}^*$  and  $\tilde{r}_{max}$  by examining the difference between  $r_{max}$  and  $\tilde{r}_{max}$ :

$$r_{max} - \tilde{r}_{max} = r(\rho^*) - r(\tilde{\rho}^*) + r(\tilde{\rho}^*) - \tilde{r}(\tilde{\rho}^*). \quad (3.17)$$

Writing (3.17) in this way allows me to use the approximation from Eqn. (3.11) to determine the value of a small parameter,  $\epsilon$ , in the approximation. Substituting Eqns. (3.5), (3.11), and (3.16) into (3.17) and rearranging gives:

$$\frac{(2\tilde{\rho}^* - 1)}{(1 - \tilde{\rho}^*)^2} m_0 \approx \frac{\tilde{\rho}^* \ln 2}{T_0} + \frac{m_0}{1 - \tilde{\rho}^*} + \frac{\tilde{\rho}^*}{T_0} \epsilon$$

where  $\epsilon = \frac{2^{-a_{max}}}{\ln 2 + a_{max} 2^{-a_{max}}}$ . Solving this expression for  $\tilde{\rho}^*$  gives:

$$\tilde{\rho}^* \approx 1 - \left( \frac{m_0 T_0}{\ln 2 - \epsilon} \right)^{\frac{1}{2}} = \rho^* - \frac{1}{2} \left( \frac{\ln 2}{m_0 T_0} \right)^{\frac{1}{2}} \epsilon + O(\epsilon^2), \quad (3.18)$$

where  $O(\epsilon^2)$  denotes terms that are of order  $\epsilon^2$  or smaller. From this I also find an approximate

expression for  $\tilde{r}_{max}$ , similar to Eqn. (3.7):

$$\tilde{r}_{max} \approx \frac{1}{T_0} \left( \ln 2 - \epsilon - 2(m_0 T_0 (\ln 2 - \epsilon))^{\frac{1}{2}} \right). \quad (3.19)$$

This analytic approximation gives results that are very close to those obtained numerically. For instance, for  $T_0 = 0.05$  and  $m_0 = 0.5$ , the difference between the analytic approximation and numerical solution when  $a_{max} = 5$  is fairly large, as we might have guessed since this is when  $r - \tilde{r}$  is largest (Figure 3.4). In this case the numerical solution is  $(\tilde{\rho}^*, \tilde{r}_{max}) = (0.8077, 8.31959)$  compared to the analytic approximation  $(\tilde{\rho}^*, \tilde{r}_{max}) = (0.804836, 8.00324)$ . As  $a_{max}$  increases, the approximation improves. For instance, when  $a_{max} = 20$ , the numerical solution is  $(\tilde{\rho}^*, \tilde{r}_{max}) = (0.8101, 8.58945)$ , and the analytic approximation is  $(\tilde{\rho}^*, \tilde{r}_{max}) = (0.810086, 8.59738)$ , a difference of less than 0.1% in  $\tilde{r}_{max}$  and less than 0.002% in  $\tilde{\rho}^*$ .

The analytic approximation for  $\tilde{\rho}^*$  can be used to explore how the bacterial mortality rate under an optimal allocation strategy depends upon  $a_{max}$ . Substituting Eqn. (3.18), into (3.4), gives

$$\tilde{m}^* = m(\tilde{\rho}^*) \approx \left( \frac{m_0}{T_0} \right)^{1/2} (\ln 2 - \epsilon)^{1/2} \quad (3.20)$$

where the mortality under this optimal strategy for the aging bacteria is denoted by  $\tilde{m}^*$ .

In Figure 3.8a I show  $\tilde{m}^*$  as a function of  $a_{max}$ . For a given value of  $a_{max}$ , as  $T_0$  decreases, the optimal allocation strategy results in higher overall mortality rates. That is, an organism with shorter minimum doubling times is likely to live through most of its reproductions, so it gains more from shifting resources to decrease the doubling time further, letting the mortality rate rise. On the other hand, when  $T_0$  is large, the organism is better served by keeping the mortality rate low. This way it can survive through as many reproductions as possible, even though this increases the time between reproductions. Additionally, when  $a_{max}$

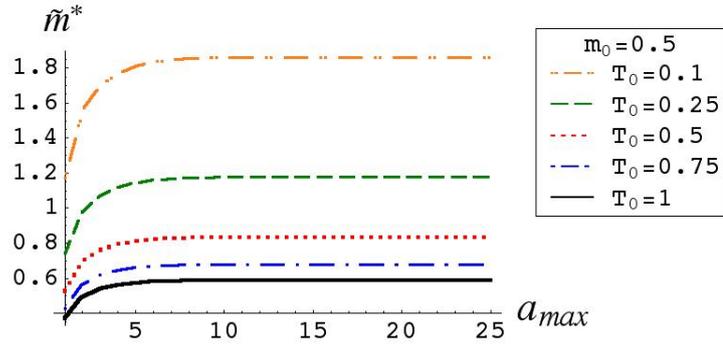
is small, it is to the organism's advantage to keep the mortality lower so it can survive through more of its possible reproductions. As  $a_{max}$  increases, however, it becomes less necessary to survive for all of the possible reproductions, and instead it is optimal to allow the mortality rate to increase so the doubling time can be decreased.

Also notice that as  $a_{max}$  increases,  $\tilde{m}^*$  approaches the mortality under the optimal strategy for bacteria that do not age ( $\tilde{m}^* \rightarrow m^*$ ), but the mortality rate for the bacteria with infinite reproductive capability is always higher than that of the aging bacteria (Figure 3.8 b). This seem counterintuitive, since we expect that it is to the organism's advantage to survive for as long as possible and get lots of opportunities to reproduce. However, the high value of the first few reproductions together with the trade-off between mortality and reproduction results in an optimal strategy that increases the mortality rate of the bacteria with infinite reproductive capability compared to the aging bacteria.

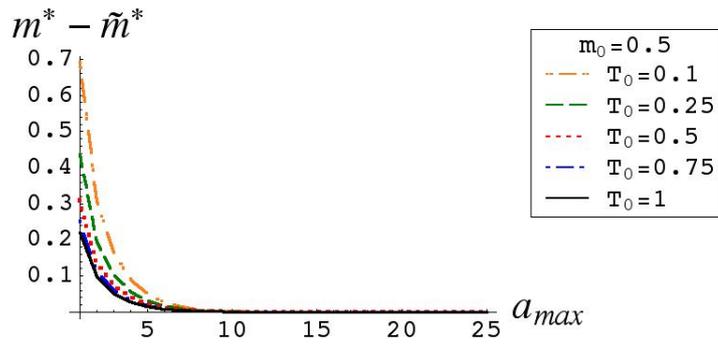
### 3.3 Effects of Limited Reproduction and Increasing Doubling Times on the Fitness of Simple Organisms

As discussed in Section 3.2, unicellular organisms such as bacteria and yeasts not only stop reproducing after a finite number of replications, but they also tend to reproduce more slowly after every replication. There are various theories as to the mechanisms that cause this [8, 35]. Here we are interested in exploring how slowed reproduction impacts fitness independent of specific molecular or genetic mechanisms, in hopes that we can gain insight into what trade-offs could shape the combination of increasing doubling times and limited lifespan in simple organisms.

Starting from the Euler-Lotka Eqn. (3.1) I again modify the birth rate,  $b_x$ . A cell of



(a)

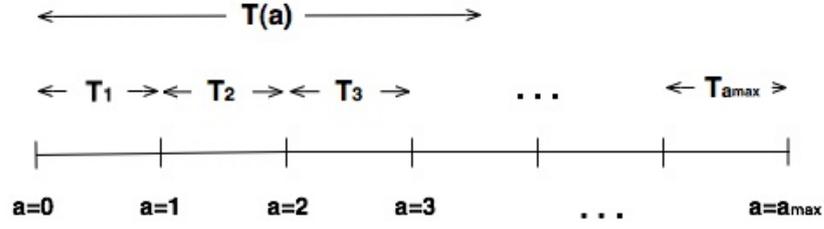


(b)

Figure 3.8: (a) Mortality under the optimal strategy for the aging bacteria,  $\tilde{m}^*$  as a function of  $a_{max}$  for  $m_0 = 0.5$ ,  $T_0 = (0.1, 0.25, 0.5, 0.75, 1)$ . (b) Difference in mortality rates for immortal and aging bacteria,  $m^* - \tilde{m}^*$ , as a function of  $a_{max}$  for the same values of  $m_0$  and  $T_0$  as in (a).

age  $a$  cell is able to divide and produce a single offspring in a given doubling time  $T_a$  where  $a = 1, 2, \dots, a_{max}$ . As before, this age corresponds to the number of times the cell has doubled, where  $a_{max}$  is the maximum number of times it can double. In absolute time, the cell then doubles at times  $T_1$ ,  $T_1 + T_2$ ,  $T_1 + T_2 + T_3$ , etc. This is shown schematically in Figure 3.9. The total time to the  $a^{th}$  doubling,  $T(a)$ , is given by:

$$T(a) = \sum_{i=1}^a T_i. \quad (3.21)$$



$$T(a) = T_1 + T_2 + T_3 + \dots + T_a$$

Figure 3.9: Schematic of the time to double  $a$  times for the model including increasing doubling times and finite lifespan.

The birth rate is then a sum of delta functions located at the times  $T(a)$ :

$$b_x = \sum_{a=1}^{a_{max}} \delta(x - T(a)). \quad (3.22)$$

This birth rate, together with the previous expression for  $l_x$ , inserted into (3.1) gives:

$$\begin{aligned} 1 &= \int_0^{\infty} e^{-\tilde{r}x} e^{-mx} \sum_{a=1}^{a_{max}} \delta(x - T(a)) dx \\ &= \sum_{a=1}^{a_{max}} e^{-(\tilde{r}+m)T(a)}. \end{aligned} \quad (3.23)$$

where  $\tilde{r}$  is again the intrinsic rate of natural increase for the aging population.

### 3.3.1 Specification of $T(a)$ : Multiplicative changes to doubling time

Before I can evaluate the expression in Eqn. (3.23) I must specify the functional form of  $T(a)$ . There are a number of possibilities. For instance, one option is to let the doubling

time increase additively. That is, the doubling time increases by a constant amount after each doubling. This would lead to a recursive relation for  $T_i$  of the form:

$$T_i = T_{i-1} + C.$$

However, there is evidence that the doubling times increase faster than this. Instead, I consider multiplicative increase, such as a 25% increase in doubling time after each doubling. I also assume that rejuvenated offspring have have ages reset to zero at birth, so they initially can reproduce with doubling time  $T$ . In this case the recursion for  $T_i$  is given by

$$T_i = sT_{i-1}$$

where  $s > 1$ . From 3.21 with  $T_1 = T$ ,  $T(a)$  for multiplicative increases is given by

$$T(a) = T \sum_{i=1}^a s^{i-1}.$$

In this case the birth rate is then

$$b(x) = \sum_{a=1}^{a_{max}} \delta \left( x - T \sum_{i=1}^a s^{i-1} \right)$$

and the Euler-Lotka Equation becomes

$$1 = \sum_{a=1}^{a_{max}} \prod_{i=1}^a \exp(-(m + \tilde{r})T s^{i-1}).$$

The product of exponentials can be re-written as

$$1 = \sum_{a=1}^{a_{max}} \exp\left(-\sum_{i=1}^a (m + \tilde{r}) T s^{i-1}\right).$$

Evaluating geometric sum in the exponent gives

$$1 = \sum_{a=1}^{a_{max}} \exp\left(- (m + \tilde{r}) T \left(\frac{1 - s^a}{1 - s}\right)\right). \quad (3.24)$$

Equation (3.24) cannot be solved analytically. The numerical solution of  $\tilde{r}$  as a function of both  $s$  and  $a_{max}$  for fixed  $T$  is shown in Figure 3.10, and  $\tilde{r}$  as a function of  $a_{max}$  for fixed  $T$  and five values of  $s$  is shown in Figure 3.11.

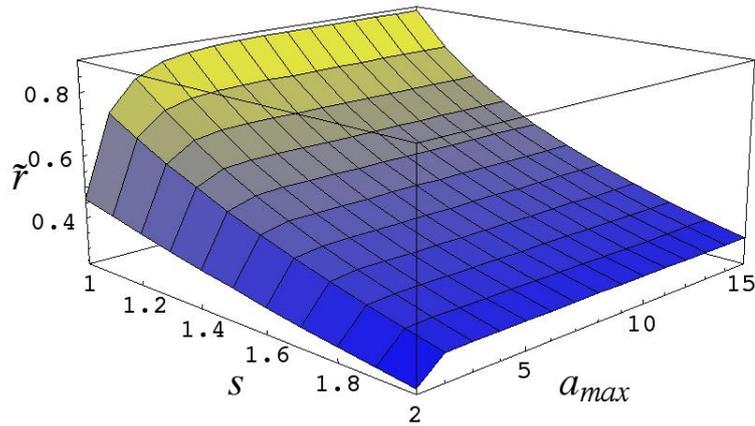


Figure 3.10:  $\tilde{r}$  vs.  $a_{max}$  and  $s$  for slowing reproduction with age.

We can see in Figures 3.10 and 3.11 that increasing  $s$  has a similar effect as increasing  $T$  and  $m$  simultaneously in the previous model (Section 3.2). That is, increasing  $s$  changes both the intercept at  $a_{max} = 1$  (similarly to the affect of increasing  $m$ ), as well as the slope and asymptotic value of  $\tilde{r}$  as  $a_{max} \rightarrow \infty$  (as we saw when  $T$  was varied).

In Figure 3.12 we can see more dramatically how increasing the ratio of sequential

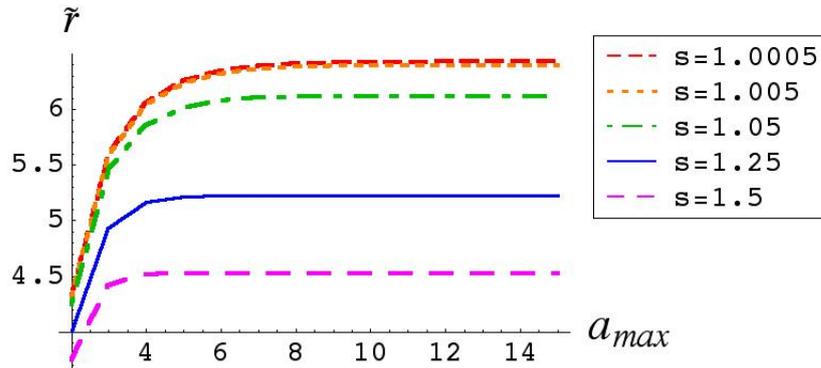


Figure 3.11:  $\tilde{r}$  vs.  $a_{max}$  and  $s$  for slowing reproduction with age

doubling times impacts the value of additional reproductions by examining  $\Delta r$ . In particular, for  $s$  close to 1, the results are similar to what we saw for the case of constant doubling times – the gain in fitness when  $a_{max}$  is incremented decreases exponentially. This is not surprising since it would take many replications for  $T_i$  to be significantly different from  $T$ . However, as  $s$  increases,  $\Delta \tilde{r}$  decreases faster than exponentially, so that a value of  $a_{max} \approx 5$  confers as much fitness as unlimited reproductive capacity when  $s \approx 1.5$ . In other words, reproducing many times is only advantageous as long as cellular maintenance is high enough to keep the doubling rate fairly constant over an organism’s lifetime. This model therefore predicts that there should be a correlation between doubling times and reproductive lifespans in these simple organisms. Organisms whose doubling times increase significantly as a function of reproductive age should double fewer times than those with doubling times that remain approximately constant.

I can also explore how selection might act on  $s$ . I use an approach similar the one used to look at the effect of selection on  $a_{max}$ . In Figure 3.13 I show a plot of  $\partial_s r$  as a function of  $s$ . When  $a_{max}$  is larger, and  $s$  is close to 1, small increases in  $s$  have a large negative affect on fitness. However, if  $s$  is already large, increasing it more doesn’t change the fitness much more,

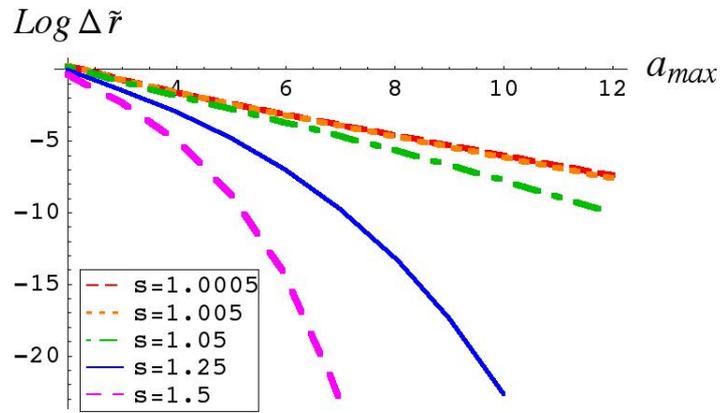


Figure 3.12:  $\Delta\tilde{r}$  vs  $a_{max}$  for slowing reproduction with age

especially if  $a_{max}$  is small. In other words, if  $a_{max}$  is large, there would be greater selective pressure to keep  $s$  closer to one, whereas for small  $a_{max}$  there is less advantage to having  $s$  near one, so selection pressures should be lower.

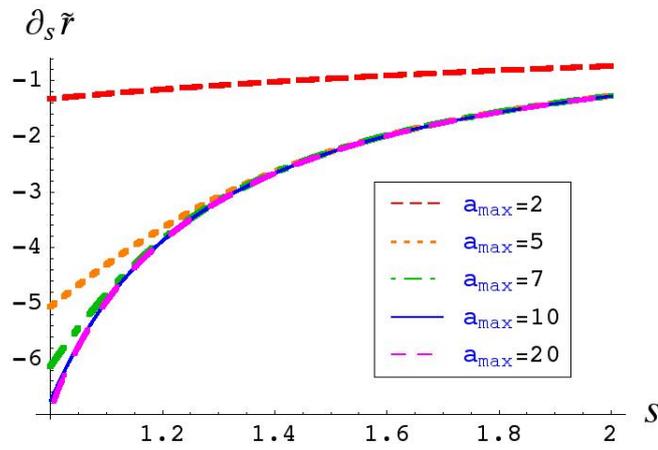


Figure 3.13:  $\partial_s \tilde{r}$  vs.  $s$  for slowing reproduction with age

Organisms experiencing increasing doubling times with age are likely to allocate resources differently from those with constant doubling times. If I parameterize the allocation

strategy as before, then (3.24) becomes

$$1 = \sum_{a=1}^{a_{max}} \exp \left( - \left( \frac{m_0}{1-\rho} + \tilde{r} \right) \frac{T_0}{\rho} \left( \frac{1-s^a}{1-s} \right) \right). \quad (3.25)$$

This allows allocation strategies to only effect mortality and minimum reproductive period, but not the rate at which the reproductive period increases. In Figure 3.14 I show  $\tilde{r}$  as a function of  $\rho$  for two values of  $a_{max}$  and five values of  $s$ . Notice that as  $s$  increases the maximum value of the fitness  $\tilde{r}_{max}$ , decreases fairly quickly, especially when  $a_{max}$  is larger (Figure 3.14 b). The optimal allocation strategy changes much more slowly (Table 3.1). As  $s$  increases,  $\tilde{\rho}^*$  decreases slightly. This indicates that slightly more resources are allocated for survival than for reproduction when  $s > 1$  than when  $s = 1$ , so the mortality rate under the optimal allocation strategy will actually decrease as  $s$  increases, while the baseline doubling time increases.

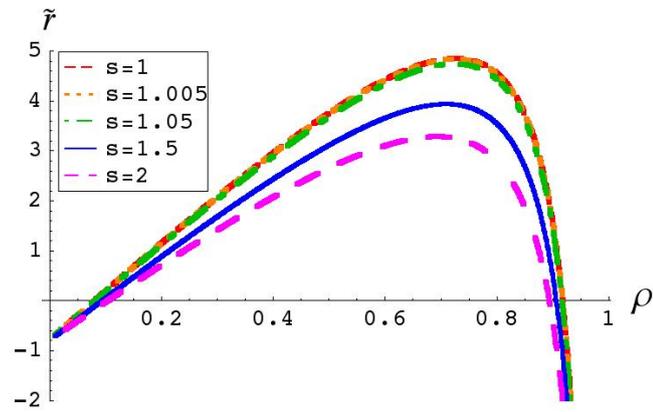
$s$	$\tilde{\rho}^*$	$s$	$\tilde{\rho}^*$
1.0	0.7265	1.0	0.772
1.005	0.726	1.005	0.7715
1.05	0.725	1.05	0.767
1.5	0.709	1.5	0.732
2.0	0.693	2.0	0.7065

(a)
(b)

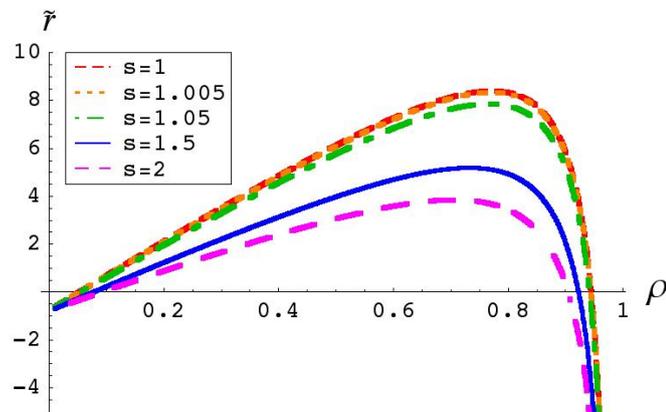
Table 3.1: Value of  $\tilde{\rho}^*$  for five values of  $s$ . (a)  $a_{max} = 2$ , (b)  $a_{max} = 20$ .

### 3.4 Discussion

In this chapter I have proposed a simple mathematical model to explore how aging, in the form of finite reproductive life span and increasing doubling times, affects fitness in unicellular organisms, particularly in bacteria. This allows me to explore, explicitly, predictions of the disposable soma theory of aging for systems that are fairly simple and easy to manipulate.



(a)



(b)

Figure 3.14:  $\tilde{r}$  vs.  $\rho$  for slowing reproduction with age with  $T_0 = 0.045$  and  $m_0 = 0.8$  (a)  $a_{max} = 2$  (b)  $a_{max} = 20$

The model has two major assumptions: (1) some kind of asymmetry in cell division, regardless of the source, that can enable us to differentiate between an “old” and a “young” cell; (2) that the young cell is some how rejuvenated. These assumptions make this model applicable to various unicellular organisms, not just bacteria.

This model allows us to quantify how senescence, specifically finite replicative ability and increasing doubling times, affects fitness of unicellular organisms, and provides a framework for quantifying how other traits associated with aging can affect fitness. Given the assumptions

about the form of  $l_x$  in the Euler-Lotka equation, as well as the relationships between mortality, doubling rate, and resource allocation, this model predicts that the organism's ability to manipulate the doubling time  $T$  has the greatest impact upon fitness. If the doubling time is short, even if the mortality is high, as is shown in Figure 3.6d, the bacteria's fitness is considerably higher than in a system with lower mortality, but longer doubling time (Figure 3.6e and 3.6f). Thus, we expect there to be strong selection for lower values of the minimum doubling time,  $T_0$ , but little selection for lower values of the minimum mortality rate,  $m_0$ .

We also find that bacteria experience surprisingly little loss of fitness when reproductive opportunities are reduced (Figures 3.3 and 3.6). For most combinations of the minimum mortality rate,  $m_0$ , and minimum doubling time,  $T_0$ ,  $a_{max} \approx 10$  or 20 is large enough to confer almost exactly the same amount of fitness to the bacteria as would an ability to reproduce indefinitely (Figure 3.6). Most of the bacteria's fitness is gained the first few times it doubles, and the amount gained in each subsequent doubling decreases rapidly, because after an organism has reproduced a few times, the lineages produced from the first few offspring are growing exponentially, so the fitness gained from producing a single new offspring now is much smaller in comparison. Therefore we expect that if there were a resource cost associated with increasing the maximum number of possible doublings then investing resources to survive to double the first few times would be better than investing additional resources to try to maintain cell integrity indefinitely.

When doubling time increases, this pattern is even more pronounced. Compared to an organism with similar values of  $T_0$  and  $m_0$  and constant doubling time, an organism with doubling time that increases with age has significantly lower fitness regardless of the value of  $a_{max}$ . In fact, when the ratio of subsequent reproductions,  $s$ , is large, the value of additional reproductions plummets, so that  $a_{max} \approx 5$  confers as much fitness as the ability to reproduce

indefinitely. This results in a correlation between the value of additional replications with the ratio of subsequent doubling times. When  $s$  is large, the fitness value of additional opportunities to reproduce is very small. However, as  $s \rightarrow 1$  additional replications can contribute much more to fitness. Therefore we might expect to see similar patterns in unicellular organisms. In particular, organisms with very small changes in doubling rate over time should reproduce more times, where as those organisms with more significant slowing in reproduction should experience fewer reproductions. This pattern is at least loosely observed. Bacteria, such as *E. coli*, reproduce a large number of times and have very small increases in doubling time over their reproductive lives. In contrast, yeasts experience fairly significant increases in doubling time, and only reproduce 10-30 times.

Studying the impacts of aging and resource allocation on fitness for bacteria is appealing for a number of reasons. Bacterial systems are relatively simple. They also are ideal for experimental manipulation. Metabolism and repair activity should be fairly straightforward to measure, giving indications of how bacteria allocate resources. Genetic manipulation of bacterial systems is also possible, which allows more direct measurement of the parameters in the model. Because generation times are short, bacterial systems also allow for replicate experiments at a reasonable cost. Thus, bacterial systems are ideal model systems for ecology [38].

It should be possible to test many of the model predictions experimentally (see Figure 3.15). The laboratory environment that the bacteria experiences could be manipulated in order to “set” values of  $m_0$  and  $T_0$ . For instance, temperature affects the rates at which chemical reactions occur and thus affects  $T_0$ . Other factors, such as nutrient levels, salinity, and oxygen levels, could also determine  $T_0$  and  $m_0$ . Then, assuming that the organisms act in a way which is approximately optimal, the optimal allocation strategy could be measured via gene

expression or some other proxy, and compared to the model predictions. For example, *E. coli* and other bacteria undergo drastic metabolic changes (stringent response) during amino acid starvation [56]. During the stringent response, genes responsible for proliferation and growth are down-regulated while genes responsible for survival and virulence are up-regulated. These responses are thought to be mediated by a small nucleotide, guanosine tetraphosphate (ppGpp) [56]. Measurement of gene expression, or levels of signaling molecules such as ppGpp, may be appropriate as proxies for  $\tilde{\rho}^*$  in order to test the predictions of the model, specifically quantitative predictions of allocation strategies, presented here. For the case with lengthening doubling times, confirmation of a correlation between number of reproductions and ratio of subsequent doubling times could be possible. Also, quantifying differences in allocation strategies between yeasts and bacteria, for instance, could also be used to test the validity of the model and model assumptions.

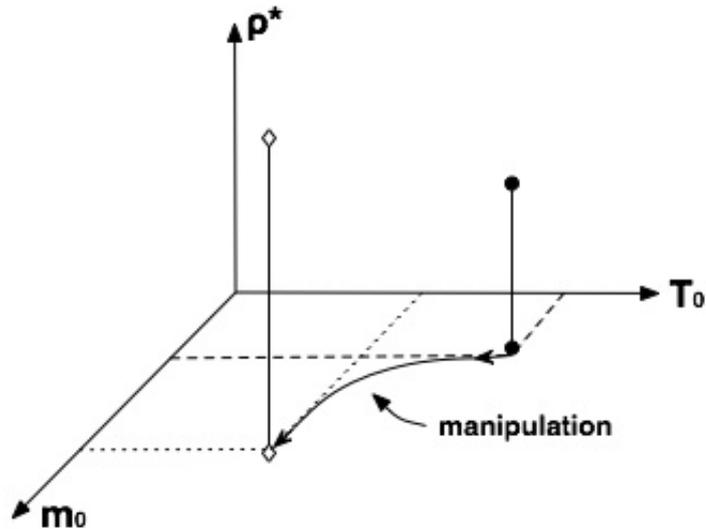


Figure 3.15: Schematic of an experiment to test predictions of our model. The experimental system starts out with particular values of  $m_0$  and  $T_0$ , and a corresponding value of  $\rho^*$ . After an experimental manipulation, the system has new values of  $m_0$  and  $T_0$ . A new value of  $\rho^*$  is then measured, and compared to model predictions.

Additionally, this simple model has the advantage of being fairly easy to analyze. Although numerical results for complex models are fairly easy to obtain, it can be difficult to understand and quantify the roles of model parameters. Good analytic approximations, such as those found in Section 3.2 for  $\tilde{r}$ ,  $\tilde{r}_{max}$ , and  $\tilde{\rho}^*$  for the model with limited reproduction, allow us to explore the effects of parameter variation in a very concrete way that might not be available for more complex models. Having a firm grasp on the roles of model parameters in a simple model also allows increased understanding of similar parameters in more complicated models such as the model explored in Section 3.3, especially when the models are nested, as are those presented in this chapter.

This simple model is only a first step in understanding how aging impacts bacterial fitness. This model predicts that the fitness of individuals that only reproduce a few times is comparable to the fitness of those that reproduce many times, and that selection for large values of  $a_{max}$  should be small. This raises two important questions. First, Why do these organisms reproduce as many times, and survive for as long, as they do? Second, What other factors influence replicative life span? Looking at more complicated versions of this model may give insight into the answers to these questions. For instance, the current model does not consider the effects of variable environmental conditions, density effects, or of increasing doubling times or mortality rates as a function of age. Also, here we assume no dependence of  $a_{max}$  or  $s$  on  $\rho$ , *i.e.* there is no cost associated with increasing  $a_{max}$  or decreasing  $s$ . These could be important factors in the development of bacterial life histories. However, in spite of these limitations, this model gives insight into why bacteria do not exhibit infinite lifespans, and suggests directions for further exploration.

## Chapter 4

# Microcolony and Colony Formation

Almost all bacteria spend a portion of their lives in some kind of surface-attached community. The ability to form communities allows bacteria to colonize and survive in many different environments. For pathogenic bacteria, such as *Vibrio cholerae*, that live both inside and outside a host, the ability to form communities may be especially important. Because of this, it is important to understand how bacteria colonize a surface and form communities.

In this chapter I introduce an Individual Based Model (IBM) [28] for microcolony formation on a surface. I begin with a review of the kinds of communities that bacteria form, and what is known about community forming strategies, particularly for *V. cholerae*. Next, I introduce an IBM that includes reproduction, death, and direct forces between cells. I then analyze simulated results and discuss how these could be compared to experimental data.

## 4.1 Community Formation as a Survival Strategy for Bacteria

Bacteria live in communities for many of the same reasons that other organisms live in communities: protection from predators or other external dangers, access to resources, and genetic diversity [36, 81]. Individuals in bacterial communities, such as in biofilms, experience increased resistance to antibiotics, thermal stress, or predation [31, 57]. These communities also allow bacteria to stay in favorable environments instead of being swept away. Bacterial fitness depends strongly on how quickly the cells can double. Since doubling rates of individuals in a community are generally lower than doubling rates of individuals outside of communities, living in a community often represents a trade-off between reproduction and survival. In order to keep doubling times small, bacteria also must keep their genomes small [51, 87]. Instead of keeping copies of genes for certain traits that might be useful later, but are not being used now, they tend to get rid of unused genes, but can pick up genetic material from outside the cell, in a phenomenon called *competence* [5, 61]. Living in communities therefore allows greater access to genes that a bacterium can then merge into its genome.

Bacterial communities can take various forms. The smallest communities are *microcolonies*. A microcolony is comprised of up to a thousand or so individuals, with little or no extra-cellular scaffolding. *Colonies* are larger communities, with thousands or tens of thousands of individuals, with the groups being fairly isolated from each other. Bacteria can colonize a surface and form a thin layer across a surface, rather like an extended colony. Some bacteria can also form *biofilms*. Biofilms are three-dimensional, surface-attached communities with extensive extra-cellular structures made of various polymers, called EPS. The bacteria secrete the polymers, microcolonies and individual bacteria then become imbedded in this EPS scaffolding,

allowing the community to remain cohesive while growing in three dimensions. The biofilms can grow to be fairly deep, and exhibit morphologies ranging from very thick and uniform in depth and density, to heterogeneous with mushroom-like structures and wide channels. Some bacteria can also form multi-species biofilms.

Community types and morphologies are determined by physical, biochemical, and genetic factors. Physical and biochemical factors – such as fluid flow, polymer tensile strength, and cell surface properties – influence the availability of nutrients within a community, the ability of a biofilm to hold together under shear stress, and determine whether or not cells can stick to a particular surface. Genetic factors constrain the behavioral options available to the bacteria, and determine responses to chemical or environmental signals. Current research indicates that gene expression of bacteria living in biofilms or other communities differs significantly from that of free-living cells, and may be mediated by a process called quorum sensing.

#### 4.1.1 *Vibrio cholerae* Communities

Genes that regulate the production of EPS are widely recognized as influencing biofilm formation [7, 18, 58, 94, 90]. Current research indicates that genes involved in flagellar motility and various chemotaxis and quorum sensing systems also appear to also influence the ability of many types of bacteria, including *V. cholerae*, to form biofilms, as well as influencing the morphology of the biofilms [19, 33, 67, 85]. Since motility is unlikely to be important to bacteria already in a biofilms (and bacteria in a biofilm appear to down-regulate motility genes) it has been hypothesized that aggregation of bacteria on a surface is an important step in the development of biofilms, particularly for *V. cholerae* [90]. However, data on the role of aggregation in initial stages of microcolony formation are inconclusive [47].

Bacterial aggregation is difficult to observe directly. This is partly because observa-

tions of densities are influenced by both aggregation and reproduction. Additionally, it can be very difficult to monitor individual bacteria on a surface. One reason for this is that many microscopic techniques that are able to clearly discern individual cells, such as confocal or scanning electron microscopy, often damage or kill cells. Mathematical models can be used to predict observable macroscopic patterns that indicate whether aggregation is occurring in the initial stages of biofilm development. Additionally, mathematical models allow exploration of how bacterial behaviors change macroscopic patterns, and help in linking these behaviors the function of various genes.

## 4.2 Approaches for Modeling Bacterial Communities

Models of organization of communities of micro-organisms, as well as larger organisms, generally take one of three approaches. To model very large groups of organisms, continuous partial differential equations (PDEs), such as those pioneered by Keller and Segel [39] to explore slime mold aggregation, are popular. For moderate numbers of individuals, cellular automata (CA) models are often used. For smaller groups of individuals, individual based models are popular. In this section, I briefly review these approaches and their application to biofilms.

### Continuous PDE models

The first approach to modeling movement of organisms is generally referred to as Eulerian methods. This type of approach ignores individual identity, instead focusing on densities of individuals in an area or volume. These are continuous models, focused on developing field equations that describe the flux of individuals in space. Examples of this approach include Keller and Segel's models of slime mold aggregation [39] and bacterial chemotaxis [40, 41], models of midge swarming [86], and a general approach to group formation in animals pro-

posed by Gueron and Levin [29]. The Eulerian approach is most useful for describing very large numbers of individuals with high density. It also has an advantage that there are many good analytical tools available. However, if the density of individuals is low, or one wishes to explore how individual strategies effect group dynamics, this type of approach is less useful [30].

### **Cellular automata**

Cellular automata (CA) [93] models are a class of discrete dynamical models. Here, the space occupied is divided into grids or boxes (often called cells, but for our purpose here will be referred to as *elements*) on a lattice. Time is usually discrete. Each element is characterized by some description of the element's state at some time  $t$ . The state of each element can be updated by a set of rules that may depend on the current state of the element as well as the states of neighbor elements.

For models of biofilms, the state of each element contains information about whether the cell is occupied by biomass and of what density and type. These models are often called biomass-based models (BbM) [50]. Generally, biomass can only move from an element to its nearest neighbor element. Some early CA models of biofilm growth allow biomass to spread by forming new biomass along the surface (*i.e.* in elements with empty neighbors), much like crystal growth [69, 92]. In real biofilms, nutrient limitation may result in more cells reproducing near the edges than in the depth of the biofilm. However, in general biofilms grow from within as well as at the surface. Although there has been progress in modeling this kind of growth within biofilms, most current methods are still fairly arbitrary and can result in strangely shaped colonies (e.g. rectangles and triangles) partly because of the discrete nature of the model [69].

Some of the most detailed biofilm models, developed by Picioreanu *et al.* [71, 70, 69, 88], involve a CA model for the biomass (living and dead cells and EPS) spreading. Addition-

ally, these methods include substrate transport via diffusion and convection, and can include modeling of detachment due to mechanical stress from fluid flow over the biofilm.

### **Individual Based Models (IBMs)**

Although both Eulerian and CA models can provide information about macroscopic properties of systems, they ignore variation of traits between individuals as well as interactions between individuals. Since both of these can greatly effect the dynamics of the larger system that the individuals compose, Individual Based Models (IBMs) can be used to explore these phenomena explicitly. IBMs comprise models where the behavior of each individual is modeled separately, and each individual follows a set of rules that determine their behavior. Some definitions have further requirements (e.g. see [28], which differentiates between Individual Based models and individual-oriented models).

These models are variants of  $N$ -body Newtonian dynamics problems, and are often called Lagrangian models. The rules that control individuals are a combination of forces and decision rules. Examples of forces include those that are physical or environmental (such as chemical gradients, gravity, or drag), and other such as attraction or repulsion between individuals [65, 30]. Rules can be hierarchical, such as those in the herding model in Gueron *et al.* [30], so that an individual's decisions only depend upon a subsection of the total population. Regardless of the types of decisions or forces chosen, the goal is to learn about how interactions and behaviors on small scales influence large-scale patterns. Additionally, once behaviors that result in "realistic" patterns have been identified, these can be compared to behaviors in populations, and used to evaluate which rules could be selected for in different situations.

BacSim [49, 50], a model of *E. coli* colony growth, is an example of an IBM model of biofilm growth. It features an IBM to model the growth and behavior of individual bacteria

(including uptake of substrate, growth, death, and reproduction) together with a simulation of the diffusion and reaction of substrate and other products. BacSim simulates spreading of the biomass by requiring that a minimum distance is maintained between cells. One of the primary results of the study by Kreft *et al.* [50] is that the initial seeding of a surface has a major impact on the development and morphology of simulated biofilms, likely because of the heterogeneity of substrate concentration in the biofilm. They concluded that “this stresses the primary importance of spreading and chance in the emergence of the complexity of the biofilm community.”

### 4.3 A Mathematical Model for Community Formation

Biofilm formation is generally thought to proceed as follows (see Figure 4.1):

1. individuals colonize surface
2. individuals  $\rightarrow$  microcolonies
3. microcolonies  $\rightarrow$  biofilms

Most current models of biofilms, regardless of approach, focus on growth of biofilms from a seeded surface, using physical and chemical factors, *i.e.*, they focus on the third step listed above. These models ignore the initial surface colonization events and microcolony formation. These also largely ignore biochemical interactions, such as quorum sensing, which allow bacteria to modify the community via chemical cues between individuals. Many bacteria with inactive or mutated chemotaxis, quorum sensing, and motility systems exhibit degraded ability to form biofilms. Since motility and chemotaxis are not as important to bacteria already in biofilms, changes in these systems may influence initial colonization events, which will, in turn, affect biofilm formation.

In the rest of this chapter I present a model that addresses three major questions. First, how do interactions between cells (mediated by processes on cellular and genetic levels) influence microcolony characteristics? Second, can we differentiate between different kinds of processes? Finally, what can we learn about the behavior of surface attached bacteria from “snapshots” of growing microcolonies?

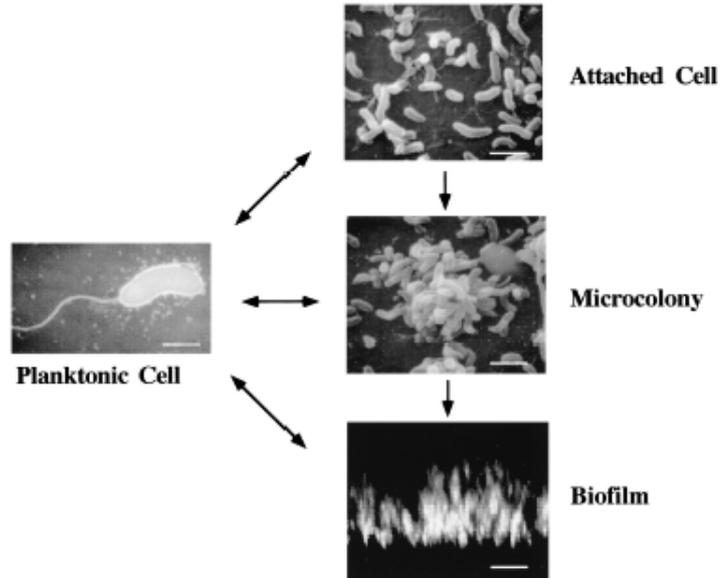


Figure 4.1: Steps in Biofilm Formation from [89]

## 4.4 Microcolony Formation

Before describing the details of the model, I will review what is known about each step of colony formation. I begin with just the steps involved in surface attachment and microcolony formation, and then present a mathematical model for this process.

Bacteria often move through a medium via a series of runs and tumbles due to flagellar action. A bacterium “runs” in a straight line for short distances, and then “tumbles” to change

the direction of travel. If there are no external stimuli, this motion is random, *i.e.*, directions of runs will be uncorrelated. During surface colonization, individuals first move around in a liquid medium until they encounter and attach to a surface. Once on a surface, they may move on a surface, by flagellar or twitching motility.

A bacterium's direction and speed of movement through a medium or on a surface is influenced by various stimuli. The collective response to stimuli are known as taxis. For instance, chemotaxis is movement in response to chemical gradients, and phototaxis is movement in response to light gradients. Since bacteria are very small, they explore gradients with movement. Instead of measuring differences in concentrations of a signal across the cell body, bacteria keep track of the temporal changes in signal strength. There is evidence that a bacterium experiencing temporal increases in an attractant decreases the frequency of tumbles, resulting in longer runs when moving up the chemical gradient. When the attractant level decreases, tumbling frequency increases, giving the bacterium more opportunities to find the desired direction.

Bacteria also respond to their environment by changing their physical, chemical, or morphological characteristics, such as metabolism or doubling rates. As mentioned in Chapter 1.1.2, *E. coli* and other bacteria can drastically alter metabolic processes during amino acid starvation, called the stringent response [56]. Bacteria may also produce and sense certain chemicals in the environment to identify that other bacteria are close by (quorum sensing) and modify their behavior accordingly.

#### 4.4.1 Model Assumptions

My goal is to explore a model with basic assumptions that can be used in conjunction with experimental data to infer processes that could generate observed patterns of surface

Symbol	Description
$\mathbf{X}_i$	cell position (center of cell)
$\mathbf{F}_i$	force on cell
$\mathbf{V}_i$	velocity of cell
$\tau_i$	time since last division
$T_i$	doubling time of cell
$a_i$	cellular age, in number of divisions
$\gamma_i$	cell is moving $\gamma_i = 1$ , or stopped $\gamma_i = 0$
$\alpha_i$	cell is alive $\alpha_i = 1$ , or dead $\alpha_i = 0$
$\phi_i$	cell overlaps other cells $\phi_i = 0$ , otherwise $\phi_i = 1$
$r_i$	cell radius
$C_i$	colony membership of the cell

Table 4.1: State Variables for the  $i^{th}$  cell, included in models explored in this chapter.

attached microcolonies in the initial stages of biofilm formation. I am most interested in how interactions among individuals affect microcolony sizes during initial stages of biofilm formation. I assume that the environment is homogeneous and external forces (such as drag) are minimal. I also assume that the density of organisms is sufficiently low so that growth and reproduction are not affected by diffusion limitation, and competition for nutrients is minimal.

The IBM begins with a description of the variables necessary to describe an individual's state. In general, at a time  $t$  a cell is characterized by a set of state variables. These variables include the cell's position, velocity, genotype, age, size, *etc.* I will denote the vector that describes the overall state of the cell as  $\mathbf{S}_i(t)$ . For the models presented here, the state of the cell is described by a combination of the variables listed in Table 4.1.

I denote the *stage-state* of the  $i^{th}$  cell at time  $t$  by  $s_{i,t} = \{a, \gamma, \alpha, \phi\}_{i,t}$ . The value of the stage-state is the component of the full state of the cell,  $\mathbf{S}_i(t)$ , that describes whether the cell is moving and whether it is alive. The cell can then transition between various stage-states at any time step. Figure 4.2 shows a graphical representation of the conditional transitions between the various stages and their probabilities. In a small increment of time from  $t$  to  $t + \Delta t$ , a moving cell stops with probability  $p_1$ . A moving cell can also die (and thus simultaneously

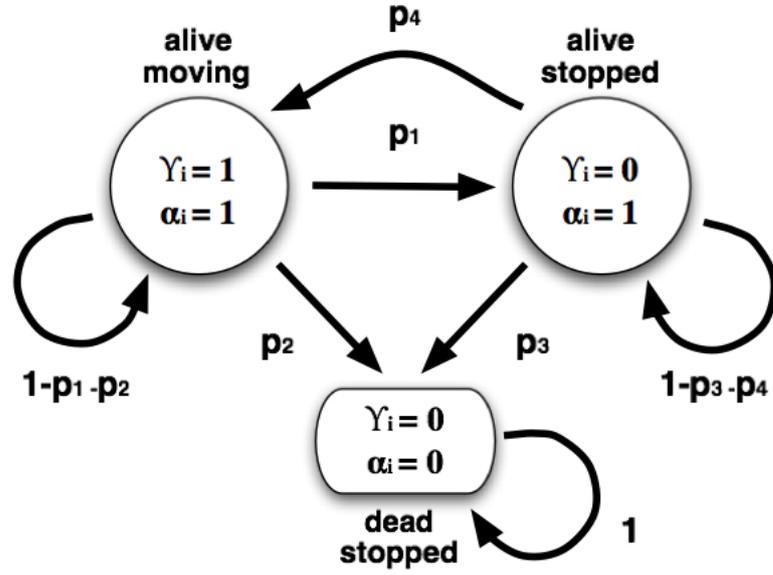


Figure 4.2: Schematic of transitions between stage-states for cell  $i$ , without interactions with other cells. Transition probabilities are denoted by  $p_1$  through  $p_4$

stop) with probability  $p_2$ . A stopped cell can die with probability  $p_3$ . Finally, a stopped but alive cell can move again with probability  $p_4$ . In other words:

$$Pr\{\gamma_{i,t+\Delta t} = 0, \alpha_{i,t+\Delta t} = 1 \mid \gamma_{i,t} = 1, \alpha_{i,t} = 1\} = p_1$$

$$Pr\{\gamma_{i,t+\Delta t} = 0, \alpha_{i,t+\Delta t} = 0 \mid \gamma_{i,t} = 1, \alpha_{i,t} = 1\} = p_2$$

$$Pr\{\gamma_{i,t+\Delta t} = 0, \alpha_{i,t+\Delta t} = 0 \mid \gamma_{i,t} = 0, \alpha_{i,t} = 1\} = p_3$$

$$Pr\{\gamma_{i,t+\Delta t} = 1, \alpha_{i,t+\Delta t} = 1 \mid \gamma_{i,t} = 0, \alpha_{i,t} = 1\} = p_4$$

Therefore, a cell that is moving and alive remains in this state with probability  $1 - p_1 - p_2$ , a stopped and alive cell remains in this state with probability  $1 - p_3 - p_4$ , and a dead cell stays

in this state with probability 1.

$$Pr\{\gamma_{i,t+\Delta t} = 1, \alpha_{i,t+\Delta t} = 1 \mid \gamma_{i,t} = 1, \alpha_{i,t} = 1\} = 1 - p_1 - p_2$$

$$Pr\{\gamma_{i,t+\Delta t} = 0, \alpha_{i,t+\Delta t} = 1 \mid \gamma_{i,t} = 0, \alpha_{i,t} = 1\} = 1 - p_3 - p_4$$

$$Pr\{\gamma_{i,t+\Delta t} = 0, \alpha_{i,t+\Delta t} = 0 \mid \gamma_{i,t} = 0, \alpha_{i,t} = 0\} = 1$$

We can also describe this system using a transition matrix, in order to connect the state of the  $i^{th}$  cell at time  $t + \Delta t$  to the state at time  $t$ :

$$\begin{bmatrix} \gamma_{i,t+\Delta t} = 1, \alpha_{i,t+\Delta t} = 1 \\ \gamma_{i,t+\Delta t} = 0, \alpha_{i,t+\Delta t} = 1 \\ \gamma_{i,t+\Delta t} = 0, \alpha_{i,t+\Delta t} = 0 \end{bmatrix} = \begin{bmatrix} 1 - p_1 - p_2 & p_4 & 0 \\ p_1 & 1 - p_3 - p_4 & 0 \\ p_2 & p_3 & 1 \end{bmatrix} \begin{bmatrix} \gamma_{i,t} = 1, \alpha_{i,t} = 1 \\ \gamma_{i,t} = 0, \alpha_{i,t} = 1 \\ \gamma_{i,t} = 0, \alpha_{i,t} = 0 \end{bmatrix}. \quad (4.1)$$

In general, the probabilities  $p_1$  through  $p_4$  could also depend upon the age of the cell (hence the inclusion of  $a$  in the stage-state vector). For the models in the rest of this chapter, I assume no age dependence, that the probability of stopping is  $p_1 = P_1$ , all cells die with equal probability,  $p_2 = p_3 = P_2$ , and no stopped cells can start moving again  $p_4 = 0$ .

## 4.5 Null Model, and Models with Stopping, Clumping, and Direct Interactions

### 4.5.1 The Null Model, $M_0$

I begin by describing a null model in which the individual cells neither interact nor reproduce. In this model,  $N$  cells, denoted by  $i = 1, \dots, N$ , are confined to a two dimensional

surface of area  $A$ . Here I am concerned with three components of the cell state: position,  $\mathbf{X}$ ; velocity,  $\mathbf{V}$ ; and radius,  $r$ . Here, all cells are alive, and cannot die or stop on their own, so  $P_1 = P_2 = 0$ . Each cell is modeled as a circle of radius  $r$ . First, I assume that all cells move at random, independently of the others (no interactions of any kind), and all cells are of equal size. I will call this model  $M_0$ , and model it as a Ornstien-Uhlenbeck process:

$$\begin{aligned} d\mathbf{X}_i &= \mathbf{V}_i dt \\ d\mathbf{V}_i &= -\eta\mathbf{V}_i dt + q d\mathbf{W}. \end{aligned}$$

Here  $\eta$  is a drag or dissipation coefficient. However, in this model and in the ones that follow I will assume that  $\eta = 0$ , so that the velocities of bacteria do not dissipate. This is reasonable as bacteria can propel themselves, and overcome viscous forces. However, instead of modeling this process explicitly, I will simply assume  $\eta = 0$ . In this model, a cell moves with a random velocity at each time step, with the random variable  $dW \sim N(0, 1)$  and the parameter  $q$  determining the change in the velocity at each time step. In this case, we expect to see behavior like classical cells in a box – cells will remain randomly distributed in the space without clumping developing over time. An example of the paths of the cells when they move randomly is shown in Figure 4.3.

#### 4.5.2 Model $M_1$ : Stopping and Clumping

Next, I add a simple interaction to  $M_0$ : if two cells bump into each other (*i.e.* if  $|\mathbf{X}_i - \mathbf{X}_j| \leq 2r$ ) they stick to each other and to the surface, and stop moving. This is indicated

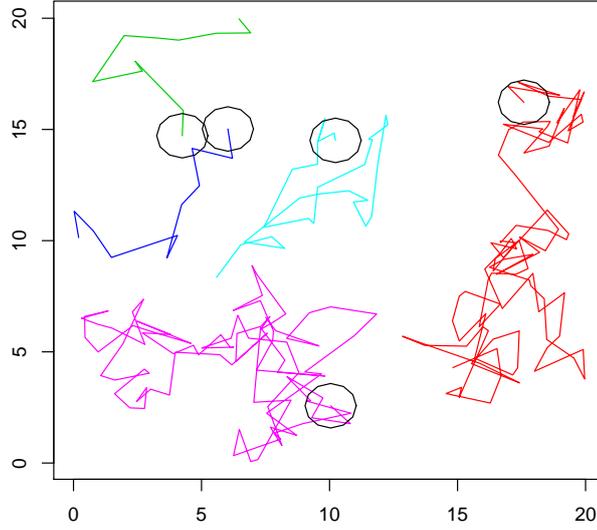


Figure 4.3: Sample of simulated cell movement.

using the state variable  $\phi_i$  which is defined as

$$\phi_i = \prod_{\substack{j=0 \\ j \neq i}}^N H(|\mathbf{X}_i - \mathbf{X}_j| - 2r), \quad (4.2)$$

where  $H(x)$  is the Heaviside step function<sup>1</sup>, so that  $\phi_i = 1$  if the  $i$ th cell is not overlapping any other cells, and  $\phi_i = 0$  otherwise.

I also allow individual cells to stop and stick to the surface with non-zero probability  $P_1$ . The state variable  $\gamma_{i,t}$  indicates if the cell is moving:  $\gamma_{i,t} = 1$  when the cell is moving, and  $\gamma_{i,t} = 0$  when the cell has stopped. Then  $\gamma_{i,t} = 1 \rightarrow \gamma_{i,t+\Delta t} = 0$  with probability  $P_1$ . The

---

<sup>1</sup>The Heaviside function is a discontinuous step function, defined as

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0. \end{cases}$$

transitions are then:

$$\begin{bmatrix} \gamma_{i,t+\Delta t} = 1 \\ \gamma_{i,t+\Delta t} = 0 \end{bmatrix} = \begin{bmatrix} 1 - p_1 & 0 \\ p_1 & 1 \end{bmatrix} \begin{bmatrix} \gamma_{i,t} = 1 \\ \gamma_{i,t} = 0 \end{bmatrix}. \quad (4.3)$$

This is model  $M_1$ :

$$\begin{aligned} d\mathbf{X}_i &= \gamma_{i,t}\phi_i\mathbf{V}_i dt \\ d\mathbf{V}_i &= -\eta\mathbf{V}_i dt + qd\mathbf{W} \end{aligned}$$

In this case, the cells can clump. The spatial distribution, densities, and spatial extents of the clumps depend on the parameters chosen for the model.

### 4.5.3 Model $M_2$ : Direct Interactions

Next, I add an interaction between cells. Generally, cells interact indirectly, e.g. via the sensing of chemicals produced by other cells (chemotaxis or quorum sensing), and bacteria tend to move up the gradient of a chemical signal. However, modeling the diffusion of chemicals that are being produced and consumed by the bacteria can increase the computational requirements significantly. Instead, I begin by looking at a model with direct interactions in the form of forces between cells as a proxy for the behavior we expect to see due to chemotaxis, similar to looking at particles undergoing Brownian motion in a field of force [48]. In this way chemotaxis produces an effective force between individuals. This approach is fairly common [52, 62]. Lee *et al.* (2001) examined a one dimension, continuous ODE model of chemotaxis, and showed that a non-local approximation for processes such as chemotaxis is valid if the rate of diffusion of a chemotactic signal is much faster than the rate of diffusion of the organisms, which is the case

with many bacteria. For example, the ratio of the effective diffusion of bacteria cells to the rate of diffusion of chemoattractants for myxobacteria is  $\sim 0.0001$  [52]. Lee *et al.* argue that this kind of approach allows us to learn about general properties of a system without needing to take into account the dynamics of faster scale processes. Additionally, solution of the diffusion equations governing the chemotactic chemicals has the potential to be very computationally intensive. So instead of modeling chemotaxis explicitly, I assume that the cells in this model interact via non-local or “direct” interactions. I call this model  $M_2$ :

$$\begin{aligned} d\mathbf{X}_i &= \gamma_{i,t}\phi_i\mathbf{V}_i dt \\ d\mathbf{V}_i &= \frac{1}{m}F_i dt + q d\mathbf{W}. \end{aligned}$$

Here  $F_i$  is the force on the  $i$ th cell due to all other cells,

$$F_i = \sum_{\substack{i=0 \\ i \neq j}}^N f_{ij}(|\mathbf{X}_i - \mathbf{X}_j|), \quad (4.4)$$

where  $f_{ij}$  is the force on  $i$  due to  $j$ , which depends only on the distance between the cells. This force can be attractive or repulsive, and differs in strength depending upon the functional form chosen for  $f_{ij}$ . There are various possible functional forms of non-local forces that have been used to model interactions between various organisms. Many of these options are gradient type forces [62]. Lee *et al.* (2001) showed that in one dimension an “effective interaction force” for chemotactic cells can take the form of a decaying exponential. Also typical are inverse power forces, with the form [62]:

$$F(x) = \frac{A}{x^n} - \frac{R}{x^m}.$$

I use a force of this type for the interaction between cells in this model. The value of the power law used is influenced by geometric considerations. The concentration of a chemical dispersing from a point source (such as a bacterium) in three dimensions would fall off as the square of the distance to the source. I also expect that a cell will be influenced more by nearby cells than by distant cells. This is analogous to a physical force like a gravitational force. I therefore choose a functional form for the effective force between the cells that is inversely proportional to the square of the distance between cells. I define the force between cells  $i$  and  $j$  to be

$$f_{ij} = \frac{A_{i,j}}{\mathbf{R}_{ij}^2} \hat{\mathbf{R}}_{ij}, \quad (4.5)$$

where  $A_{i,j}$  is the force constant between cell  $i$  and  $j$  and  $\mathbf{R}_{ij} = \mathbf{X}_i - \mathbf{X}_j$  is the vector between cells  $i$  and  $j$ . The unit vector from  $i$  to  $j$  is denoted by  $\hat{\mathbf{R}}_{ij}$ . If  $A_{ij} > 0$  the force is repulsive, and if  $A_{ij} < 0$  it is attractive.

This model could also be generalized to include attraction towards or repulsion from fixed sources of chemical signals. For instance, imagine there was an external point source of a chemical located at  $\mathbf{X}_s$ , and the effective force between cell  $i$  and this source, denoted as  $F_{ext}$ , has a form similar to Eqn. 4.5

$$F_{ext} = \frac{B}{|\mathbf{X}_i - \mathbf{X}_s|^2} \hat{\mathbf{R}}_{is}$$

where  $B$  is the force constant and  $\hat{\mathbf{R}}_{is}$  is the unit vector from cell  $i$  to the chemical source. Then the total force on the  $i^{th}$  cell (Eqn. 4.4) would become:

$$F_i = F_{ext} + \sum_{\substack{i=0 \\ i \neq j}}^N f_{ij}(|\mathbf{X}_i - \mathbf{X}_j|).$$

However in the remaining simulations and models, I assume that the only sources of attraction

or repulsion in the system are other cells.

#### 4.5.4 Model Simulations: $M_1$ and $M_2$

All simulations presented in this section consist of 5000 cells initially distributed randomly in a 500 x 500 square space (in units of the cell radius). All simulations began with the same initial distribution. I also set  $q = 1$ , and assume reflecting boundaries. When cells run into each other or stop on their own they form a “colony”, defined here as simply a set of cells that are touching. At each time step the simulation proceeds as follows:

1. Move all of the cells that are not already stopped.
2. Stop any of the cells that have collided with other moving cells or with stopped cells and add them to the appropriate colony.
3. If cell  $i$  overlaps with cell  $j$ , but they are in different colonies (*i.e.*  $C_i \neq C_j$ ), then merge the colonies.
4. Consider stopping each moving cell w.p.  $P_1$ .

Under this model ( $M_1$ ), the final distribution of colony sizes for three values of the stopping probability,  $P_1$ , are shown in Figure 4.4. Increasing the stopping probability mainly acts to shift more cells from colonies of size 2 into colonies of size 1 (as can be seen when comparing Figure 4.4a & b). However, as  $P_1$  gets much larger, the proportion of individual cells stopping without running into other cells increases. This results in the formation of more colonies of size 1, and a decrease in the number of colonies of size  $> 2$ . (Figure 4.4c).

In Figure 4.4 I show realizations of the final distributions of colony sizes for three values of the stopping probability for model  $M_1$ . Figure 4.5 shows the final distributions of colonies and colony sizes when cells experience attractive forces, (Figure 4.5 a), no forces

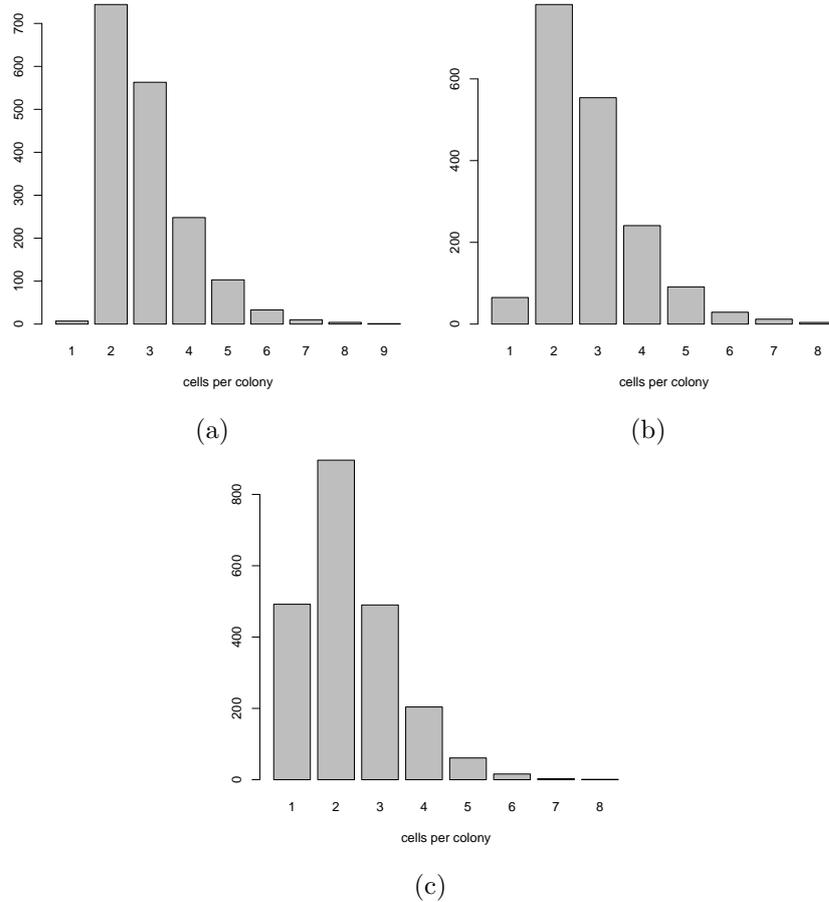


Figure 4.4: Distribution of colony sizes for model  $M_1$  (stopping and clumping, but no interactions between cells) for three values of the stopping probability,  $P_1$  (a)  $P_1 = 0.0001$  (b)  $P_1 = 0.001$  (c)  $P_1 = 0.01$ , after all cells have stopped moving.

(Figure 4.5 b), or repulsive forces (Figure 4.5 c) for model  $M_2$ . For the attractive case the force constant  $A = -10$ , and for the repulsive case  $A = 10$ . Notice that each of these models exhibits distinctive distributions of colony size (numbers of cells in a colony). In particular, for  $M_2$  when there are attractive forces between cells, the colonies are much more likely to be large, and the distribution of group sizes exhibits a longer and heavier right hand tail. When the cells repulse the colonies are smaller than when they only interact randomly. We can confirm that the three cases explored here (*i.e.* attractive, no force, and repulsive) do in fact result in distinct

distributions by performing a simple  $\chi^2$  test. If I assume that the case of motion without forces is the “expected” distribution, I can test two hypotheses. My first null hypothesis,  $H_0$ , is that the distribution from the attractive case is the same as that of the the distribution found when there were no forces present. The alternative hypothesis,  $H_1$ , is that they are not equal. Since the  $\chi^2$  test requires at least five counts in any bin to give a good result, I re-bin each distribution into 7 bins:  $\{1, 2, 3, 4, 5, 6, 7+\}$ . This corresponds to 6 degrees of freedom in my test. I choose my significance level to be 0.001 which corresponds to a  $\chi^2$  value of 22.547. For the  $\chi^2$  test comparing the attractive distribution to the no force (expected) distribution, I find  $\chi^2 = 2448.036$  leading to a rejection. For the  $\chi^2$  test comparing the repulsive distribution to the no force (expected) distribution, I find  $\chi^2 = 1029.437$ , leading to another rejection. Thus as the 0.001 level, I reject the hypotheses that the attractive and repulsive distributions are the same as the random case.

These three cases can be viewed as representative of the types of clumping that might be observed in a bacterial system. For instance, in a system where bacteria attract via chemotaxis, we might see distributions of microcolonies similar to those shown in Figure 4.5a. However, if the chemotaxis system is disrupted, for instance by some chemical additive or by genetic manipulation, then the new system might look more like what is shown in Figure 4.5b. If on the other hand, the system is manipulated so that the chemotaxis system is reversed, so cells move away from each other, the system would appear more like what we see in Figure 4.5c.

## 4.6 Model $M_3$ : Direct Interactions, Births and Deaths

The models presented in Sections 4.5.2 and 4.5.3 are useful over fairly short time periods, *i.e.* periods much less than the average doubling time of a cell, so that the size of the population is unlikely to change. For periods longer than this, it is important to include

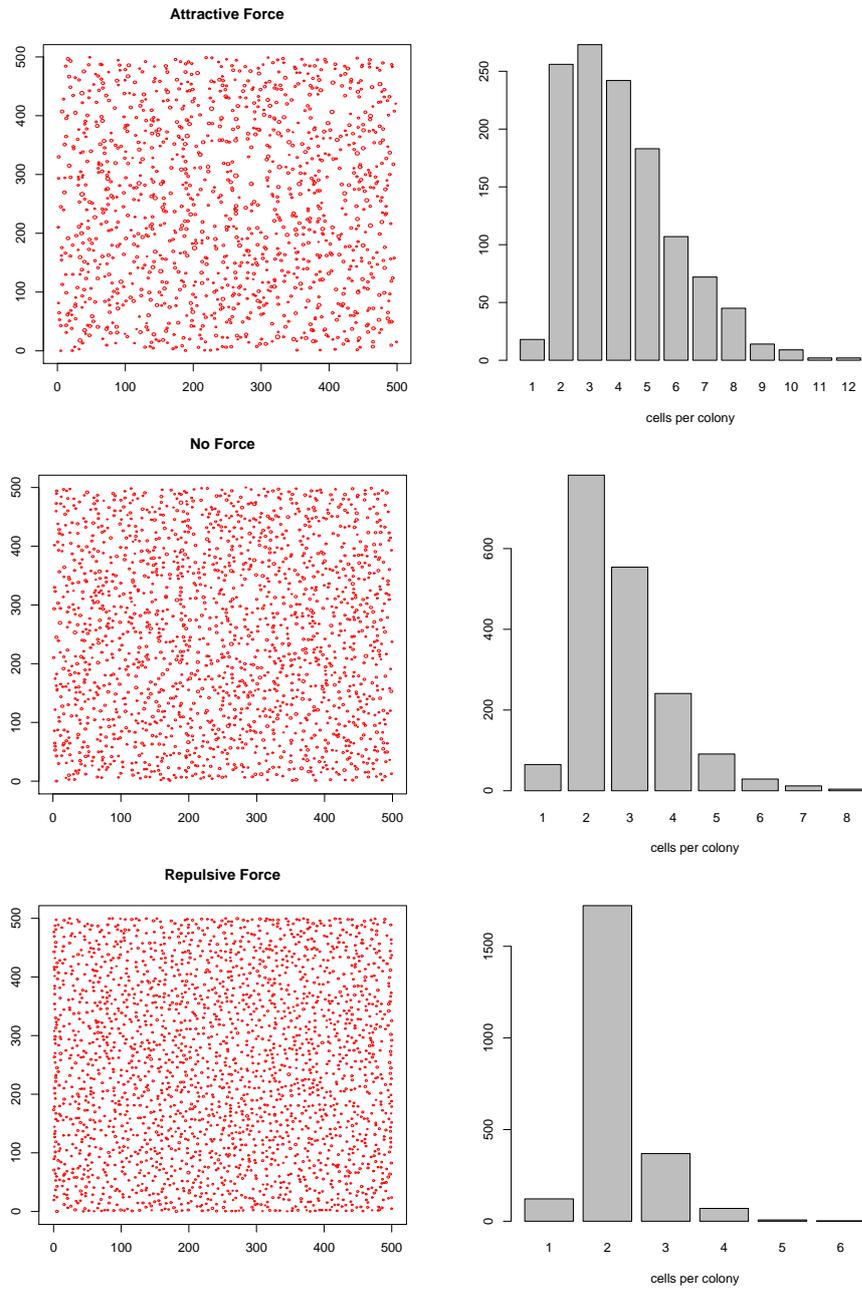


Figure 4.5: Distribution of colony sizes for models  $M_1$  (no interactions between cells) and  $M_2$  for three values of  $P_1 = 0.001$ . (TOP) Attractive forces between cells; (MIDDLE) no forces between cells; (BOTTOM) repulsive forces between cells.

population dynamics, in the form of births and deaths. In order to do this I use four more state variables:  $\alpha_i$ ,  $T_i$ ,  $a_i$ , and  $\tau_i$  (see Table 4.1). The first,  $\alpha_i$ , indicates whether or not the  $i^{th}$  cell is alive. Cells transition from  $\alpha_{i,t} = 1 \rightarrow \alpha_{i,t+\Delta t} = 0$  with probability  $P_2$ :

$$p(\alpha_{i,t+\Delta t} = 0 | \alpha_{i,t} = 1) = P_2.$$

If  $\alpha_i = 0$ , then  $\gamma_i = 0$ , *i.e.* the cell also stops moving. Additionally, a living cell can double in a doubling period  $T_i$ . The number of times a cell has divided is the cellular age,  $a_i$ . Finally, the number of time steps since the last doubling of a cell will be denoted by  $\tau_i$ . At every time step,  $\tau_i$  is incremented. When the time since the last doubling,  $\tau_i$ , reaches the doubling time,  $T_i$ , *i.e.*,  $\tau_i = T_i$ , the cell doubles, the cellular age  $a_i$  is incremented, and the time since the last double is reset, so  $\tau_i = 0$ . For model  $M_3$  then, the state of the  $i^{th}$  cell at time  $t$  is given by:

$$\mathbf{S}_i(t) = \{\mathbf{X}, \mathbf{V}, s_{i,t}, T_i, \tau_i\}_t$$

where  $s_{i,t}$  is the stage state of the cell described in Section 4.4.1.

The equations for the change in position and velocity over time from model  $M_2$  need to be modified slightly, in order to introduce deaths into the system:

$$\begin{aligned} d\mathbf{X}_i &= \alpha_{i,t} \gamma_{i,t} \phi_i \mathbf{V}_i dt \\ d\mathbf{V}_i &= \frac{1}{m} F_i dt + q d\mathbf{W}. \end{aligned}$$

Again  $F_i$  is the force on the  $i$ th cell due to all other cells, *i.e.*

$$F_i = \sum_{\substack{i=0 \\ i \neq j}}^N f_{ij}(|\mathbf{X}_i - \mathbf{X}_j|),$$

where the force on  $i$  due to  $j$ ,  $f_{ij}$ , is the same as in model  $M_2$

#### 4.6.1 Model Simulations: $M_3$

All simulations presented in this section consisted of cells initially distributed randomly in a 500 x 500 square space (in units of the cell radius). Parameter settings for the simulations are shown in Table 4.2. At each time step, simulation proceeds as follows:

1. Move all of the cells that are not already stopped.
2. Stop any of the cells that have collided with other moving cells or with stopped cells and add them to the appropriate colony.
3. If cell  $i$  overlaps with cell  $j$ , but they are in different colonies, (*i.e.*  $C_i \neq C_j$ ), then merge the colonies.
4. Consider killing any living cell w.p.  $P_2$ .
5. Consider stopping any living, moving cell w.p.  $P_1$ .
6. If any remaining living cells have  $t_i = T_i$ , create a new cell a distance  $2r + \epsilon$  from the dividing cell, update  $a_i$  and reset  $t_i$ . Increment  $t_i$  for any cells that don't divide.

In Figures 4.6 through 4.10 I show visualizations of the frequencies of colony sizes cells experiencing attractive forces, for 3 different doubling times (200 steps, 300 steps or 400 steps). Force constants were set to  $A = -10$  and  $A = 10$  for the attractive and repulsive cases, respectively.

Parameter	Value
$N_{initial}$	250
$t_{final}$	1000
$P_1$	0.001
$P_2$	0.00005
# replicates	50

Table 4.2: Parameter settings for simulations of model  $M_3$

Except for the case of repulsive forces, where cells continue moving for longer and form smaller colonies, the spacing between the major peaks of the distribution corresponds to  $2^d$  with  $d = \left\lfloor \frac{T_i}{t_{tot}} \right\rfloor$ , where  $t_{tot}$  is the total run time,  $T_i$  is the doubling time, and  $\lfloor \cdot \rfloor$  denotes the *floor* function, which returns the largest integer smaller than or equal to the argument. This value is also usually the mode, or at least the second largest peak. So for  $T_i = 200$  the largest peak is at  $2^5 = 32$  and other major peaks are spaced at  $\sim 32$  individuals apart (e.g. in Figure 4.6 bottom panel). The other simulations show similar patterns, with variability being caused by the interactions between individual cells, and stochastic deaths. There is additional variability in the simulation with  $T_i = 300$  and  $T_i = 400$  since these correspond to non-integer values of  $\left( \frac{T_i}{t_{tot}} \right)$ . In these cases, all cells reproduce at least  $d$  times, but a proportion will reproduce  $d + 1$  times. However, the pattern of major peaks is approximately  $2^d$  times the distribution of colony sizes for model  $M_2$  (interactions without births and deaths) as shown in Figure 4.5.

## 4.7 Model $M_4$ : Colony Spreading

With the previous model, I explored how colony size varied depending upon interactions between moving cells and birth and death rates in the population. However, this model did not allow for interactions between individuals within a colony, resulting in significant over-

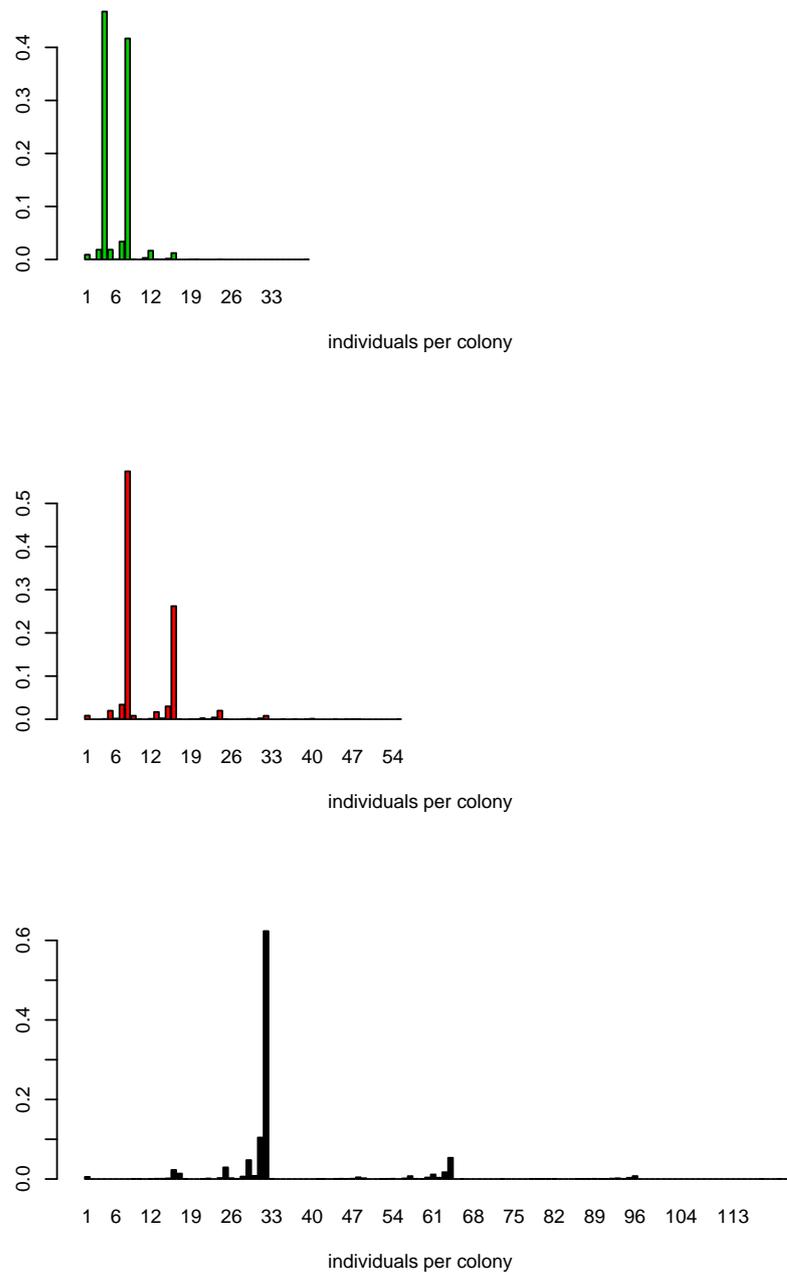


Figure 4.6: Histogram of final colony sizes for model  $M_3$  with the stopping probability  $P_1 = 1$  (so there is no movement along the surface), for three settings of the doubling time. TOP:  $T_i = 400$ , MIDDLE:  $T_i = 300$ , BOTTOM:  $T_i = 200$

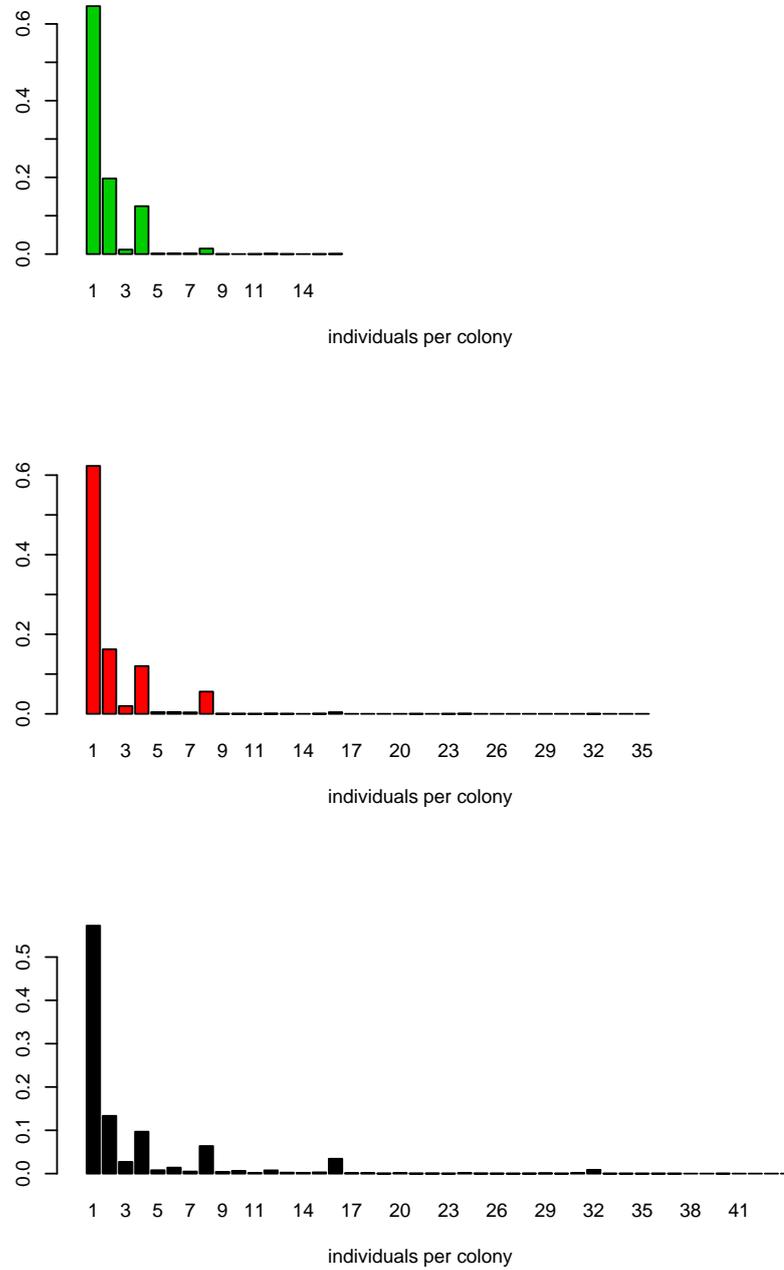
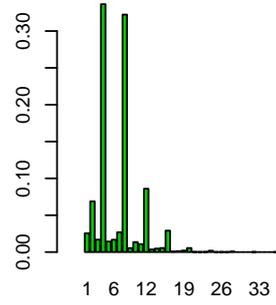
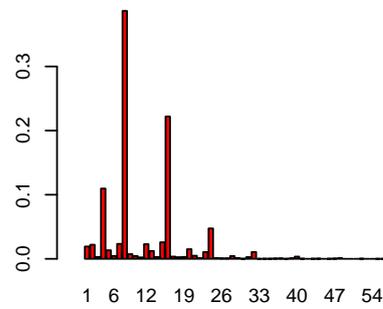


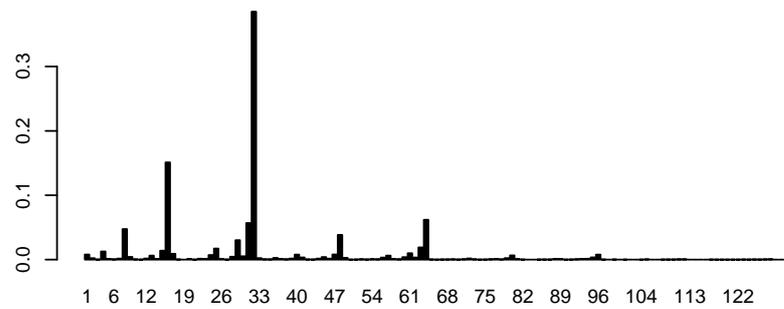
Figure 4.7: Histogram of final colony sizes for model  $M_3$  with  $A = 10$  (repulsive forces between cells), for three settings of the doubling time. TOP:  $T_i = 400$ , MIDDLE:  $T_i = 300$ , BOTTOM:  $T_i = 200$



individuals per colony

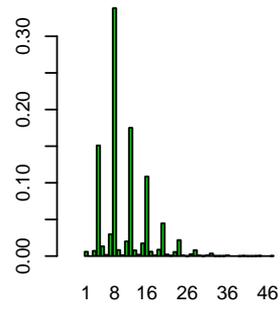


individuals per colony

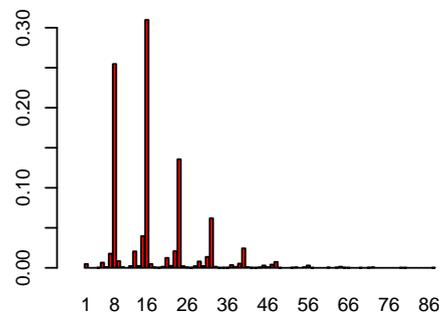


individuals per colony

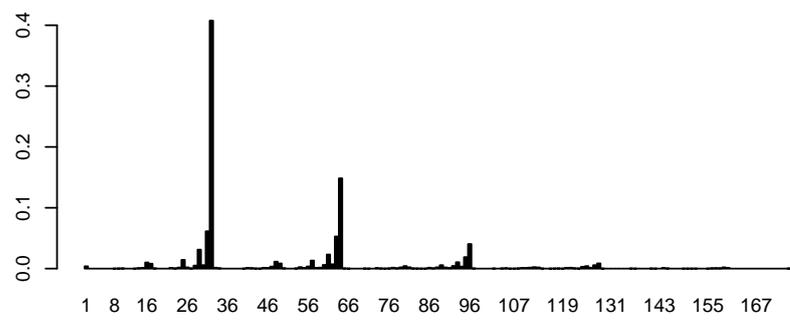
Figure 4.8: Histogram of final colony sizes for model  $M_3$  with  $A = 0$  (no forces between cells), for three settings of the doubling time. TOP:  $T_i = 400$ , MIDDLE:  $T_i = 300$ , BOTTOM:  $T_i = 200$



individuals per colony



individuals per colony



individuals per colony

Figure 4.9: Histogram of final colony sizes for model  $M_3$  with  $A = -10$  (attractive forces between cells), for three settings of the doubling time. TOP:  $T_i = 400$ , MIDDLE:  $T_i = 300$ , BOTTOM:  $T_i = 200$

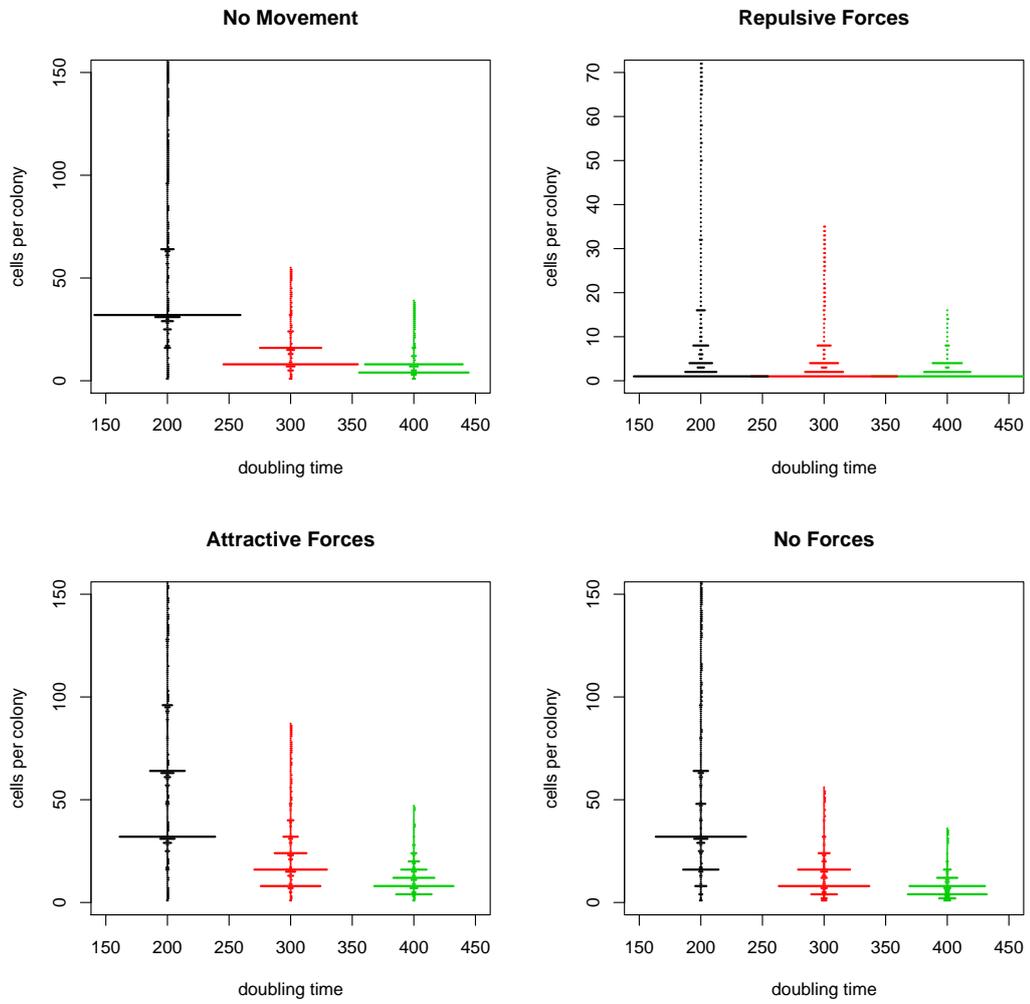


Figure 4.10: Alternative histograms of final colony size. These plots show the same data as in Figures 4.6 - 4.9 in a more compact form. The width of the horizontal bars correspond to the counts in the histogram.

lap of cells. Since physical cells cannot occupy the same space at the same time, it is important to include at least a simple mechanism to space cells out as they are reproducing within a colony.

This spacing could be approached in many ways [69]. Explicitly balancing forces between cells in a colony would be computationally expensive. Placing new cells at the perimeter of the colony is an alternative, but later exploration of phenotypic or genotypic variation within a colony using this framework would be impossible. Instead, I use a simple movement heuristic where overlapping cells will move away from each other, *i.e.*, cells shove each other out of the way. This method is similar to that used in BacSim [50]. The algorithm proceeds as follows:

1. Check if the  $i$ th cell overlaps with any other cells, *i.e.* if  $\phi_i = 0$ .
2. Calculate the unit vector from the  $i$ th cell to each of the overlapping cells.
3. Sum these vectors, and move the  $i$ th cell a distance  $D$  along this new direction.
4. Repeat 1-3 over all the cells in a colony.

The distance to move the cell,  $D$ , will influence how quickly the colony re-adjusts itself. In general, if this distance is more than the distance required to move the  $i$ th cell away from its nearest neighbor, the colony will up being disjoint. The choice of  $D$  will therefore determine the density and compactness of the microcolonies. If reproduction stopped, the microcolonies would end up arranged such that cells would be just touching, as long as  $D$  is less than or equal to the minimum overlap distance between cells. Also, if a cell from one colony overlaps with a cell from another colony at any time, these two colonies merge to form one larger colony.

#### 4.7.1 Model Simulations: $M_4$

Simulations of  $M_4$  run similarly to simulation of  $M_3$ . At every time step

1. Move all of the cells that are not already stopped.
2. Stop any of the cells that have collided with other moving cells or with stopped cells and add them to the appropriate colony.
3. If cell  $i$  overlaps with cell  $j$ , but they are in different colonies, (*i.e.*  $C_i \neq C_j$ ), then merge the colonies.
4. Consider killing any living cell w.p.  $P_2$ .
5. Consider stopping any living, moving cell w.p.  $P_1$ .
6. If any remaining living cells have  $t_i = T_i$ , create a new cell a distance  $2r + \epsilon$  from the dividing cell, update  $a_i$  and reset  $t_i$ . Increment  $t_i$  for any cells that don't divide.
7. Within each colony, re-space cells using the shoving algorithm.

Examples of the cell distributions on the surface before and after implementing the shoving heuristic appear in Figures 4.11 and 4.12. For these four simulations generating these distributions, the value of the shoving distance  $D$  was set to  $\frac{1}{4}$  of the distance between the  $i$ th cell and its nearest neighbor. We can see that colonies are more likely to merge when their members spread themselves out. This should increase the proportion of large colonies that are observed. For cases where the cell density is low, we can expect this effect to be small. However for larger populations, this effect could be significant. For the repulsive case, the spreading also increases the magnitude of the edge effects when the area of interest is small or cell density is high (Figure 4.12).

In order to explore this effect, I ran simulations with the same parameter values as for model  $M_3$  (Table 4.2). In Figure 4.13 I show plots of the distribution of final colony sizes aggregated over 50 runs. Notice that these distributions are very similar to those without

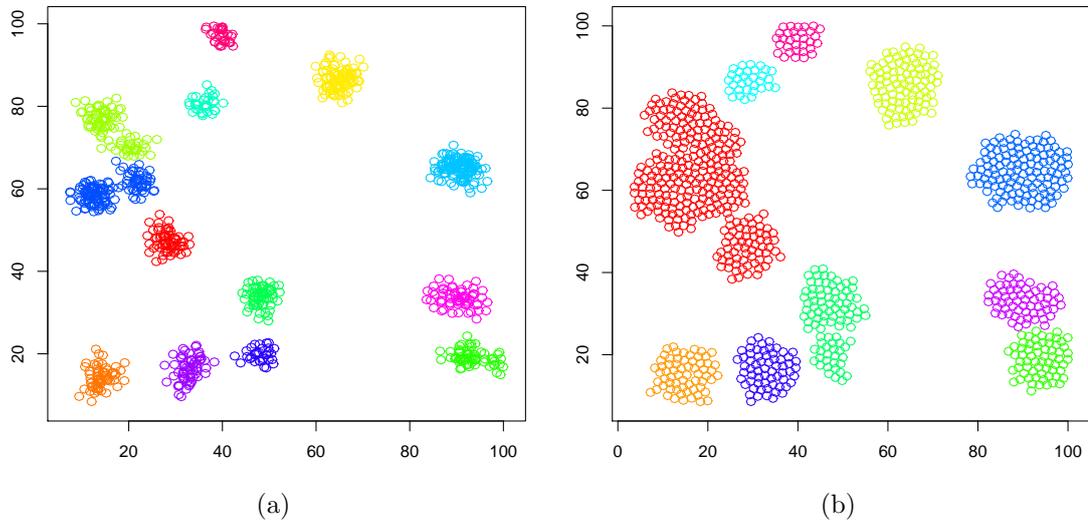


Figure 4.11: Example of cell distribution and colony membership (cells in a colony are the same color) before and after implementation of the spreading heuristic. Both simulations were run starting from the same initial cell distribution of 30 cells. Simulation parameters:  $T_i = 200$ ,  $force = -10$ ,  $prob_{death} = 0.00001$ ,  $prob_{stop} = 0.0005$ ,  $T_{final} = 1000$ .

shoving (Figure 4.10) except that the tails are slightly longer and heavier. For long doubling time (*i.e.*  $T = 400$ ) the effect is minimal, since the final population size is smaller.

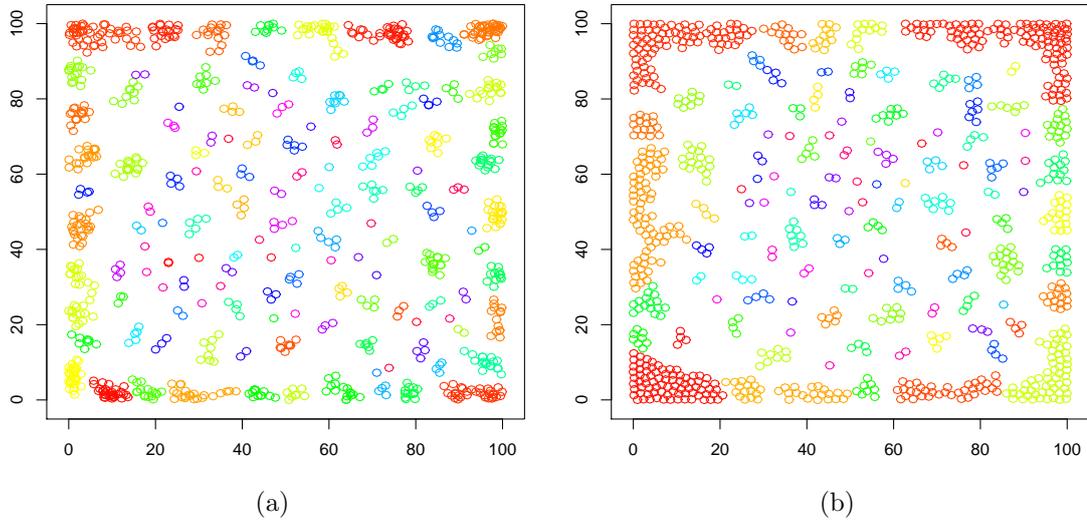


Figure 4.12: Example of cell distribution and colony membership (cells in a colony are the same color) before and after implementation of the spreading heuristic, but with repulsive forces between the cells. Both simulations were run starting from the same initial cell distribution of 30 cells. Simulation parameters:  $T_i = 200$ ,  $\text{force}=10$ ,  $\text{prob}_{\text{death}} = 0.00001$ ,  $\text{prob}_{\text{stop}} = 0.0005$ ,  $T_{\text{final}} = 1000$ .

## 4.8 Discriminating between models

In order for this model formulation to be useful in approaching data, it must be possible to discriminate between distributions of colony sizes with different parameter settings. In Section 4.5.4 I approached this issue by using a  $\chi^2$  hypothesis test. This method cannot be used for discriminating between the distributions obtained for models  $M_3$  or  $M_4$  because the number of counts of colony sizes is too low (and often zero) in many bins, rendering the  $\chi^2$  test invalid. The data cannot be re-binned without loss of information about the structure of the distribution. Instead, a different measure of the distance between the histograms is necessary. A simple choice is to measure what is known as the Minkowski Distance. If  $H_1(n; T)$  and  $H_2(n; T)$  are the histograms of interest, with  $n$  the number of bins in the histogram, and  $T$  the

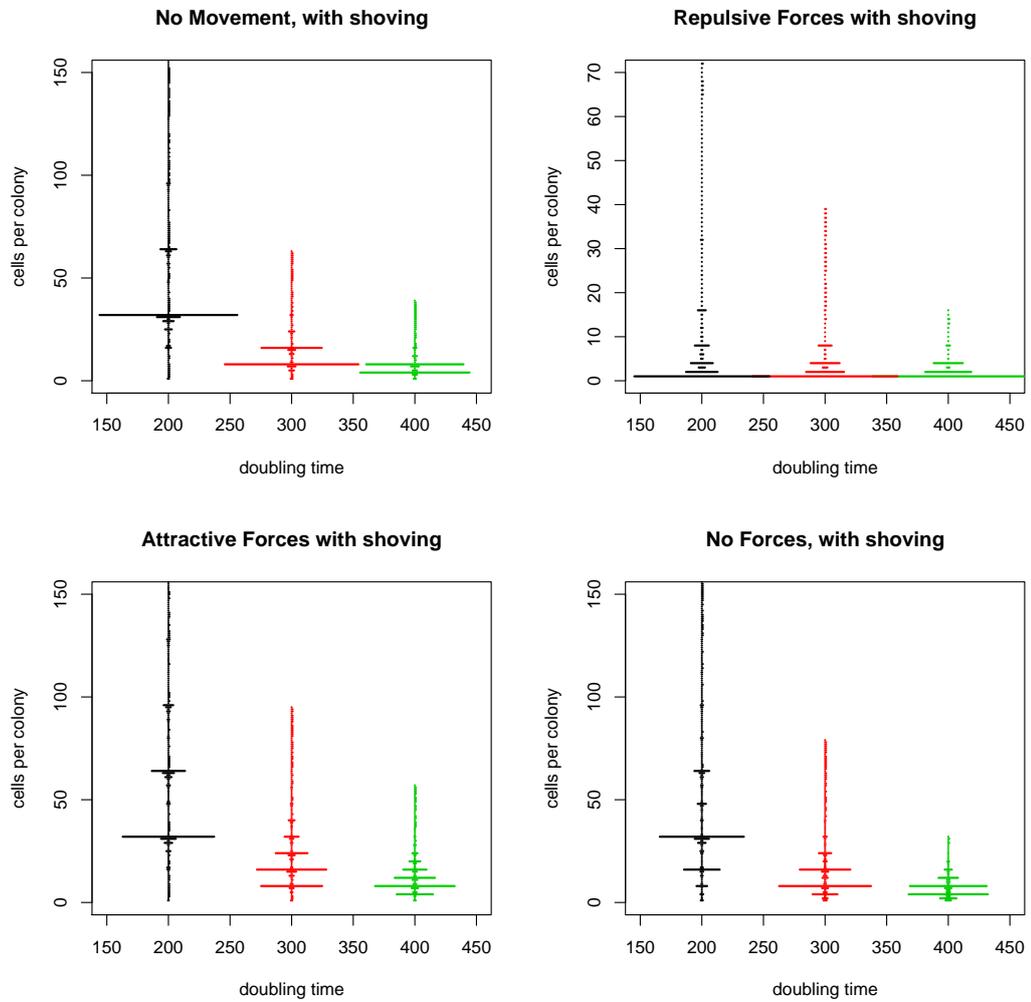


Figure 4.13: Histogram of final colony sizes for model  $M_4$ . Parameter settings are the same as for the simulations for model  $M_3$ .

total number of counts in the histogram, then the  $L_p$  Miknowski distance is

$$D(H_1, H_2) = \left( \sum_{x=1}^n |H_1(x; T) - H_2(x; T)|^p \right)^{(1/p)}. \quad (4.6)$$

When  $p = 2$ , this distance gives the Euclidian distance. This is the measure I use in the remainder of this section. Although this distance gives a measure of the absolute distance between two histograms, it does not provide a sense of relative distance. That is, how large should  $D(H_1, H_2)$  be before  $H_1$  and  $H_2$  are said to characterize different distributions? Without knowing something about how the distributions of distances between histograms generated via independent random evolution of a fixed model and parameterization, I cannot draw any conclusions about the closeness of distributions generated by different models, or what model and parameterization new data is most similar. One way to do this takes its inspiration from likelihood-free Bayesian inference [6, 72] and from a leave-one-out cross validation Monte Carlo approach. I start with  $n$  realizations of a simulation with a particular set of parameter values, with the results summarized in  $n$  histograms denoted by  $H_i$  with  $i = 1, \dots, n$ . I then calculate the distance,  $d_i$  from the  $i$ th histogram to the aggregated histogram of the  $n - 1$  remaining realizations, denoted by  $H_{a,-i}$ , *i.e.*,

$$d_i = D(H_i, H_{a,-i}).$$

I call these  $d_i$  the intra-model distances. Let  $H_a$  denote the aggregate histogram of all  $n$  realizations of the simulation. We now observe new data, denoted as  $H_{data}$ . I want some way to discern if  $H_{data}$  came from the model that generated  $H_a$ . The first step is to calculate the

distance from this new histogram to the aggregate  $H_a$ :

$$d_{data} = D(H_{data}, H_a).$$

I can now perform a simple hypothesis test of the null hypothesis that  $H_{data}$  was generated from the model that generated the aggregate  $H_a$ . I set the rejection region so that if  $d_{data}$  is larger than the 95% quantile of the distribution of the intra-model distances,  $\{d_i\}_{i=1}^n$ , (or similarly the 0.95n order statistic), I reject the hypothesis that  $H_{data}$  is a realization of the model that generated the aggregate distribution  $H_a$ . As long as the realizations of the simulations By choosing the 95% quantile as my rejection threshold in this way, the probability of incorrectly rejecting the null hypothesis when the data is from the model that generated the aggregate, *i.e.* the probability of making a type I error, is approximately 0.05.

Model	Max	95% Quantile	75% Quantile	Mean
Attractive ( $A = -10$ )	16.59	12.94	11.10	9.67
Repulsive ( $A = 10$ )	74.49	48.03	33.92	27.15
Random ( $A = 0$ )	35.63	20.75	15.11	12.95
No Movement	17.65	17.23	13.02	10.52

Table 4.3: Maximum, 95% Quantile, 75% Quantile, and Mean of the distributions of distances  $d_i$  for 4 parameter settings of model  $M_4$  (reference models). Parameters are as in 4.2, with doubling time  $T = 300$ , except for the no movement case which has stopping probability  $P_1 = 1$ .

To demonstrate this concept, and explore how similar different models are, I take  $n = 50$  realizations of final colony size from model  $M_4$  for each of the following parameter sets: attractive ( $A = -10$ ), repulsive ( $A = 10$ ), no force ( $A = 0$ ), and no movement ( $A = 0$  and  $P_1 = 1$ ), all with doubling time  $T = 300$ . The aggregate distributions are shown in Figure 4.13. I will refer to these as the reference models or distributions. In Table 4.3 I show the maximum, 95% quantile, 75% quantile, and mean of the distributions of Euclidian distances ( $\{d_i\}_{i=1}^n$ ) for these cases. The value of the 95% quantile serves as a rejection threshold for hypothesis tests

comparing new data with the reference models.

Model	Attractive	Repulsive	Random	No Movement
Attractive ( $A = -10$ )	<b>6.81</b>	894.0	49.16	63.61
Repulsive ( $A = 10$ )	102.5	<b>23.35</b>	166.3	189.5
Random ( $A = 0$ )	27.74	877.8	<b>16.76</b>	40.76
No Movement	43.52	1020	58.20	<b>16.80</b>

Table 4.4: Distances from new individual simulations to the aggregate distributions ( $d_{data}$ ). The new data sets are simulated using the same parameter settings as the four reference models.

I next conduct four additional simulation runs, one for each of the attractive, repulsive, random, and no movement cases. I then calculate the Euclidian distance from each of these individual runs to the reference histograms ( $d_{data}$ ). These appear in Table 4.4. The rows correspond to reference histograms, and columns to individual runs. The distance from the individual run to the reference distribution generated from the same set of parameters is indicated in bold face. Notice that the bold value is the lowest distance in each column. These bold distances are all lower than the 95% quantile, so we cannot reject the hypothesis that the individual run is from the corresponding reference model. Distances to all other reference distributions are higher than the respective 95% quantile. In these cases, the hypothesis that the individual run is from the same distribution as the reference model is rejected.

Model	$A = 5$	$A = 0.5$	$A = -2$	$A = -5$	$A = -8$
Attractive ( $A = -10$ )	582.5	40.17	65.57	22.47	58.80
Repulsive ( $A = 10$ )	131.0	183.4	156.9	135.1	129.3
Random ( $A = 0$ )	559.6	68.00	92.15	25.34	81.63
No Movement	682.4	97.07	110.1	35.83	99.00

Table 4.5: Distances from new individual simulations to the aggregate distributions ( $d_{data}$ ). The new data sets are simulated using the same parameter settings as the four reference models, except for the force constant  $A$ .

The first set of comparisons presented above indicates that we should be able to correctly classify that an individual is from a model with particular parameter values when

comparing the run to a reference model with the same parameter values. We can also see that we correctly reject that the distributions are the same when the parameter values are very different. However, there is still the possibility that we would not be able to distinguish between models that have similar (but not exactly the same) parameters. To explore this, I ran ten additional simulations, this time with at least one of either the force constant ( $A$ ) or the doubling time ( $T$ ) different from any of the reference models. First, I looked at variation in the force constant only. In Table 4.5 I show distances between the reference models and individual runs with five different values of  $A$ , ranging from  $A = 5$  (mild repulsion) through  $A = -8$  (moderate attraction). Notice that all of these distances are in the rejection region. Moreover, most of them are also significantly larger than maximum of the within model distances,  $d_i$ .

Model	$A = -10,$ $T = 250$	$A = 10,$ $T = 250$
Attractive ( $A = -10$ )	42.65	1169
Repulsive ( $A = 10$ )	102.0	38.76
Random ( $A = 0$ )	61.39	1152
No Movement	78.20	1344

Table 4.6: Distances from new individual simulations to the aggregate distributions ( $d_{data}$ ). The new data sets are simulated using the same parameter settings as the attractive and repulsive reference models, except with  $T = 250$ .

Next, I changed the doubling time, so that  $T = 250$  and calculated the distances from these new runs to the reference distributions (Table 4.6). Here again, all of the distances fall outside of the non-rejection region, although the distance from the individual attractive ( $A = -10$ ) run is closest to the attractive reference distribution, and the individual repulsive ( $A = 10$ ) run is closest to the repulsive reference distribution.

Finally, I calculated distances between individual runs and the reference distributions with both  $A$  and  $T$  different from those used to generate the aggregate distributions (Table 4.7). Again, all of these distances are large enough that the hypothesis that the individual run was

Model	$A = -5,$ $T = 250$	$A = 5,$ $T = 250$	$A = 60,$ $T = 450$
Attractive ( $A = -10$ )	64.21	649.3	578.4
Repulsive ( $A = 10$ )	139.1	200.3	51.02
Random ( $A = 0$ )	88.89	622.7	574.6
No Movement	107.33	775.6	663.8

Table 4.7: Distances from new individual simulations to the aggregate distributions ( $d_{data}$ ). The new data sets are simulated using the same parameter settings as the four reference models, except for  $A$  and  $T$  (as indicated in the table).

generated from the same model that generated the aggregate distribution would be rejected.

The method outlined here may be useful in model fitting, for instance by using  $D$  as a measure of likelihood for particular parameter sets. Another approach would be to use the distributions of intra-model distances to set acceptance thresholds to use with a likelihood-free Bayesian approach [6, 72], in order to approximate posterior distributions for each of the parameters. Either way, this process provides a way to compare models with data and draw conclusions about the behavior of bacteria on a surface.

## 4.9 Fitness of bacteria in microcolonies

This analysis of models of microcolony formation presented neglects the question: why do bacteria employ a strategy that results in a particular microcolony size? Since bacteria cannot plan to form microcolonies or biofilms, the strategies must be related to the fitness of the bacteria. For instance, if bacteria repel each other, there must be an advantage to small group size, perhaps or greater access to nutrients and increased reproductive rates. On the other hand, aggregation would provide some advantage that favors large groups, such as protection from predation.

Mortality rates are often size dependent [54, 59]. For instance, a traditional model for

a size dependent mortality rate,  $m(L)$ , is

$$m(L) = m_1 + \frac{m_2}{L} \quad (4.7)$$

where the length of the organism is  $L$ ,  $m_1$  is some constant mortality, and  $m_1 + m_2$  gives the mortality rate of the organism at length  $L = 1$ . For a bacterial microcolony, I use this form for the mortality of a bacterium in the group, where  $L$  is the diameter of the group (assuming the group is approximately circular). Measuring the group size as a multiple of the diameter of a cell,  $m_1 + m_2$  is the mortality rate of a single bacterium,  $m_1$  is the mortality rate of bacteria in a large group, and  $m(L)$  gives the average mortality of the bacteria in a group (*i.e.* I allow all bacteria to experience the same mortality rate, ignoring edge effects, *etc.*). Living in a microcolony or biofilm may improve survival, but since resources are limited, living in a group is likely to decrease the rate of reproduction, or increase the doubling time. I assume a simple linear relationship between group size and doubling time,  $T$ , such that

$$T(L) = T_1 + T_2 L^\gamma. \quad (4.8)$$

I examine the case of  $\gamma = 1$ . In this case,  $T_1 + T_2$  is the doubling time of a single bacterium. I can then use Eqns. (4.7) and (4.8) together with the results from Chapter 3 to explore how the fitness of the bacteria is influenced by colony size. Recall that fitness of an aging bacterium is approximately (Eqn. 3.11)

$$\begin{aligned} \tilde{r}(L) &\approx r - \frac{\epsilon}{T} \\ &= \frac{\ln 2}{T} - m - \frac{\epsilon}{T} \end{aligned} \quad (4.9)$$

where  $m$  is the mortality rate,  $T$  is the doubling time, and  $\epsilon = \frac{2^{-a_{max}}}{\ln 2 + a_{max} 2^{-a_{max}}}$ , where  $a_{max}$  is the maximum number of times a bacterium can double. Using Eqns. (4.7) and (4.8) for  $m$  and  $T$  in Equation (4.9) gives

$$\tilde{r} \approx \frac{b}{T_1 + T_2 L} - m_1 - \frac{m_2}{L} \quad (4.10)$$

where  $b = \ln 2 - \epsilon$ . The optimal group length  $L^*$  will maximize the fitness, so that  $\frac{d\tilde{r}(L^*)}{dL} = 0$ . Taking this derivative and solving for  $L^*$ , gives

$$L^*_{\pm} = \frac{-T_1 \pm T_1 \left( \frac{b}{m_2 T_2} \right)^{\frac{1}{2}}}{T_2 - \frac{b}{m_2}}.$$

For this to make sense as a length,  $L^*$  must be positive. The first of these solutions

$$L^*_{+} = \frac{-T_1 + T_1 \left( \frac{b}{m_2 T_2} \right)^{\frac{1}{2}}}{T_2 - \frac{b}{m_2}}$$

cannot yield positive group lengths. The second solution

$$\begin{aligned} L^*_{-} &= \frac{-T_1 - T_1 \left( \frac{b}{m_2 T_2} \right)^{\frac{1}{2}}}{T_2 - \frac{b}{m_2}} \\ &= \frac{T_1 + T_1 \left( \frac{b}{m_2 T_2} \right)^{\frac{1}{2}}}{\frac{b}{m_2} - T_2} \end{aligned}$$

corresponds to a positive solution if  $T_2 < \frac{b}{m_2}$ . If I define  $\omega = \frac{b}{m_2 T_2}$ , this positive solution can be re-written as

$$L^* = \frac{T_1}{T_2} \left( \frac{\omega^{\frac{1}{2}} + 1}{\omega - 1} \right) \quad (4.11)$$

where  $\omega > 1$ . Notice that the optimal group size does not depend upon the size independent mortality rate  $m_1$ , since this only shifts the fitness by a constant.

If the group of bacteria is approximately circular, and the area covered by the groups is approximate the number of cells in the group,  $N$ , multiplied by the area covered by a single cell, then the diameter of the group scales as  $N^{\frac{1}{2}}$ . The optimal number of individuals in a group is then approximately

$$N^* \approx (L^*)^2 = \left( \frac{T_1}{T_2} \left( \frac{\omega^{\frac{1}{2}} + 1}{\omega - 1} \right) \right)^2. \quad (4.12)$$

Very large values of  $N^*$  will therefore be optimal if at least one of two conditions holds. The first is that  $T_2 \ll T_1$ , *i.e.*, larger group sizes only have a small effect on the doubling rate, so that the ratio of  $T_1$  to  $T_2$  will be large. The optimal group size will also be large when  $\omega \approx 1$  which corresponds to  $m_2 \approx \frac{b}{T_2}$ . If both conditions hold, the optimal group size will be especially large. In contrast, if the doubling time increases quickly with group size, *i.e.*  $T_2$  is large, the group must provide significant protection for large group sizes to be optimal. Therefore, large group sizes will tend to when occur when  $m_2 T_2 \approx b$ . If  $m_2$  is due to predation, for instance, then the threat would not be constant in time, and the optimal group size would then depend on predators being present in the environment. This requires that bacteria be able to sense that there is danger and modify their behavior accordingly. *V. cholerae* appear to exhibit this behavior. Matz *et al.* [57] studied biofilm formation in *V. cholerae* in the presence of predatory protozoa. They found that bacteria living in biofilms are protected from protozoa compared to those living in planktonic state, and that bacteria in the biofilm produce compounds which inhibit the growth of the protozoa. They also found that if there is predation of the planktonic state, biofilm formation is significantly enhanced compared to growth without predation. The bacteria appear to regulate these various responses to the protozoa via quorum sensing.

The calculation for optimal group size here gives a way to estimate how strong the predation must be in order for the biofilm state to be preferred, and perhaps estimate what

levels of predation and what group sizes are necessary to initiate quorum sensing. On the other hand, if the mortality rates of individual bacteria and large groups can be measured and the doubling time for planktonic bacteria is known (as in the experiments by Matz *et al.*), the values of  $T_2$  required to select for large group sizes could be calculated. These calculations also indicate what types of environments (presence/absence of predators, nutrient limitation, *etc.*) are likely to be correlated with the different types of surface movements explored in the IBM introduced in this chapter. For instance, if  $T_2$  is large, and the risk of predation is low ( $m_2$  is small), then repulsive forces, which results in smaller group sizes, would be predicted to be prevalent. However, attraction between individuals, which would allow larger groups to be formed faster, would be optimal if the risk of predation is high.

To make this more concrete, I calculate approximate values of the parameters using the experimental data from [57]. In experiments lasting for 72 hours, Matz *et al.* found that with a planktonic predator, the population of planktonic bacteria was reduced by  $> 94\%$ . On the other hand, the biofilm population was unaffected by a surface predator. I assume doubling time for planktonic bacteria ( $T_1 + T_2$ ) is 0.34 hours (about 20 minutes). Since  $T_2$  is unlikely to be very small (as biofilms tend to decrease the diffusion of nutrients to many of the members) and the system favors very large groups when predators are present (as was indicated by increased biofilm formation in the presence of the planktonic predator), I also assume that  $T_2 \approx \frac{\ln 2}{m_2}$  – *small*, so that  $L^*$  is very large. The estimated values of the parameters are then:  $m_1 \approx 0.0001$ ,  $m_2 \approx 2.079$ ,  $T_1 \approx 0.0068$ , and  $T_2 \approx 0.3332$ , corresponding to an optimal group size of  $L^* = 3200$  or  $N^* \approx 10^7$ . In Figure 4.14 a & b I show  $\tilde{r}$  as a function of  $L$  for these parameters, with the optimal group size indicated on the curve. Notice that for this case, the fitness is always less than one, and the population will decline, whereas if  $m_2$  is reduced a small amount (Figure 4.14 c) the fitness would be positive. This is because of the value of the

constant mortality  $m_1$ . If this mortality were zero, then the population in a large group would be stationary ( $\tilde{r} = 0$ ).

In Figure 4.15 I show plots of  $\tilde{r}$  as a function of  $L$  varying  $m_2$  (Figure 4.15 a) and varying  $T_2$  (Figure 4.15 b) with the other parameter values set to those estimated here. For the optimal group size to be reduced to  $N^* \approx 1$ , the predation mortality must be decreased to about  $m_2 = 2.02$ , a reduction of only  $\approx 3\%$ . In otherwords, this model predicts that a small reduction in mortality of planktonic bacteria could result in individual bacteria being more prevalent (top three lines in Figure 4.15 a). If the mortality,  $m_2$ , remains constant, a decrease in the size dependent mortality rate to  $T_2 \approx 0.25$  gives a smaller optimal groups size (in this case  $N^* \approx 4$ ), so small groups should be more prevalent. Notice that when the optimal group size is small the fitness surface is sharply peaked around the optimal value, so there would be little variability in clump sizes. On the other hand, as the optimal group size increases, the fitness surface flattens out. This is especially obvious in Figure 4.14 b, where the difference in fitness between the maximum fitness at  $L^* = 3200$  and the fitness of a groups of size  $L = 2500$  is on the order of  $10^{-9}$ . This indicates that there would be more variability in clump sizes under these conditions, since even group with sizes that are far from the optimal will have similar fitness.

## 4.10 Discussion

Individual Based Models have traditionally been used to explore how the behavior of individuals combine to create large scale or group patterns. Instead of encoding a particular large scale pattern explicitly into the dynamics of the system, IBMs seek to identify individual behaviors that can generate emergent properties or patterns.

The formation of bacterial communities, such as biofilms, has major implications for

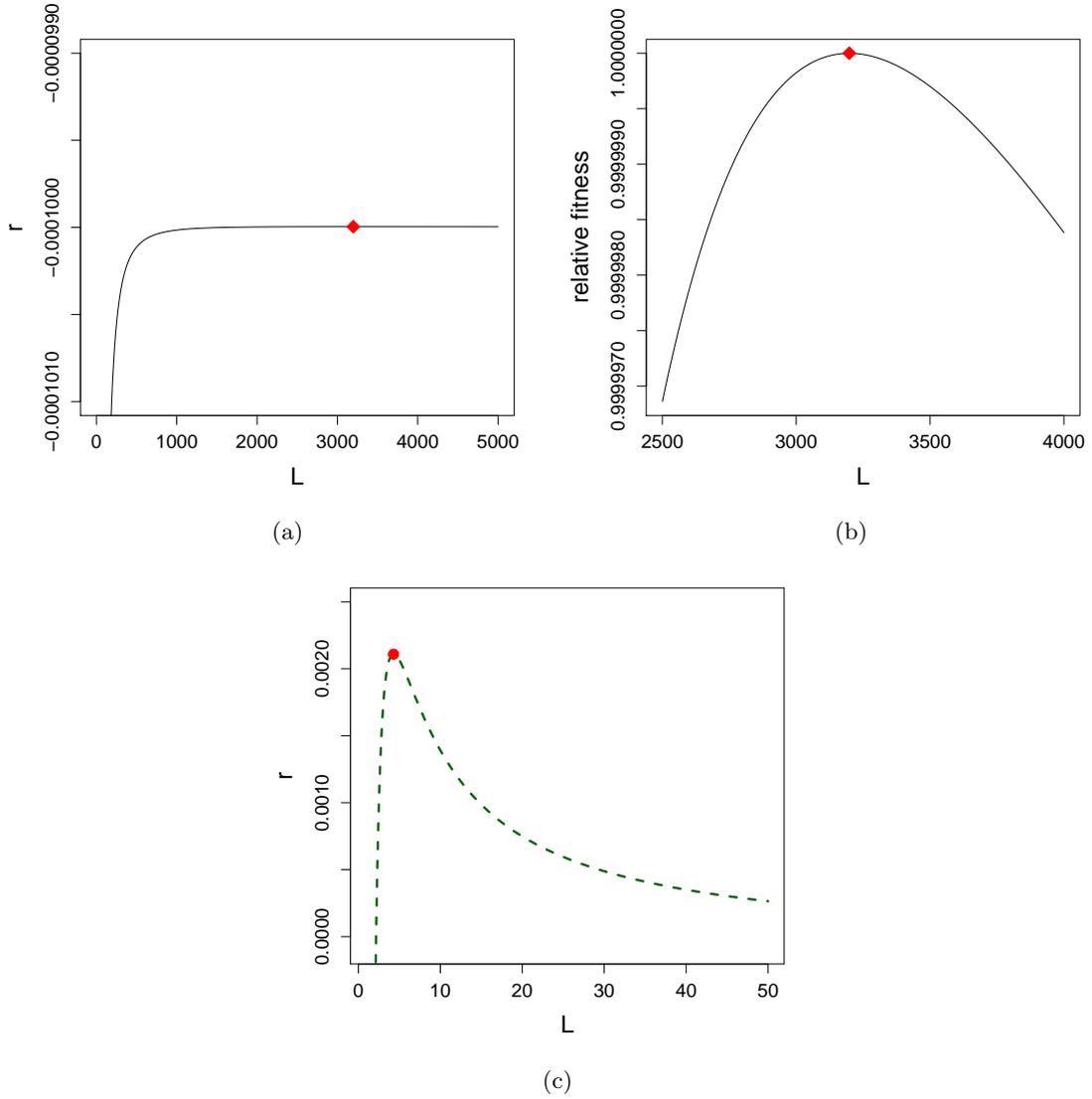


Figure 4.14: (a) Fitness of bacteria.  $\tilde{r}$  as a function of groups size,  $L$  for parameters estimated from data in Matz, *et al.*  $m_2 = 2.079$  with  $T_2 = 0.3334$ ,  $m_1 = 0.0001$ ,  $T_1 = 0.0068$ ,  $b = \log 2$ .  $L^*$  is indicated with a red point. (b) Close-up of (a) near the optimal group size,  $L^*$ . The y-axis here indicates relative fitness Relative fitness ( $\tilde{r}/\tilde{r}_{max}$ ) (c)  $r$  vs.  $L$  for parameters estimated from data in Matz, *et al.*, except with  $m_2 = 2.06$ .

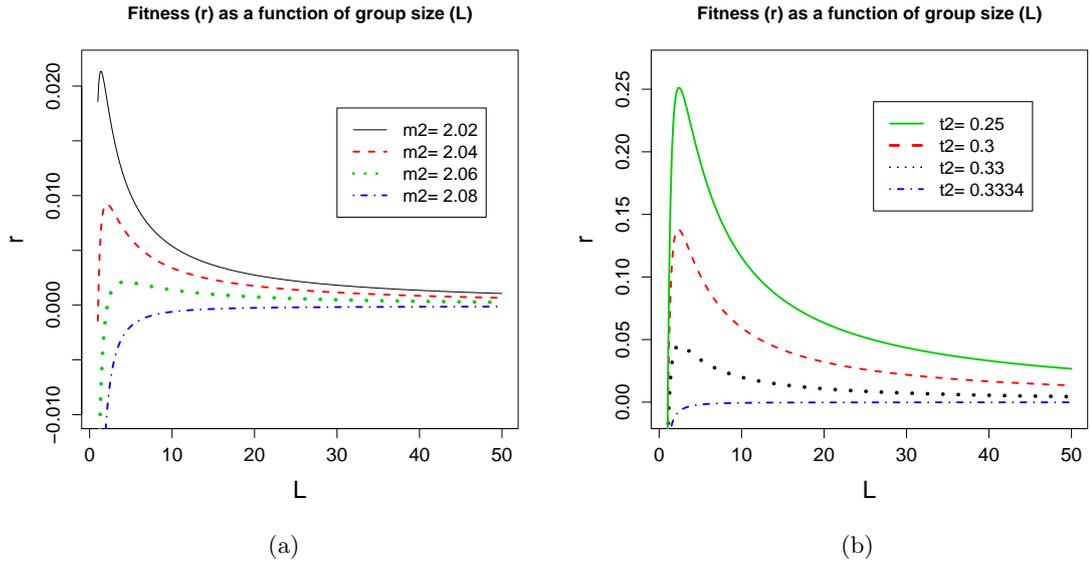


Figure 4.15: Fitness of bacteria as a function of groups size for (a) 4 values of  $m_2$  with  $T_2 = 0.3334$  and (b) 4 values of  $T_2$  with  $m_2 = 2.08$ . Values of other parameters are:  $m_1 = 0.0001$ ,  $T_1 = 0.0068$ ,  $b = \log 2$ .

survival and fitness of bacteria. To understand patterns of group formation in these organisms we must gather information about how behavior of individual bacteria influences these patterns. The IBM presented here is one way to approach this problem. Two features of this model are unique among IBMs of group formation. First of all, this model allows individuals to stop movement indefinitely, allowing the creation of static groups. Also, IBMs of group formation generally assume constant population sizes. However, a major factor in group formation in bacteria is reproduction. The model I introduced here is unique in that it explicitly includes reproduction.

This model provides useful qualitative predictions of how the size of microcolonies depends on the behavior of the cells. For instance, the doubling time of the cell is very important in determining the spacing between major peaks in the distribution. More specifically, the spacing between peaks is proportional to the expected number of doublings in the observation period. Intercellular interactions tend to influence the mean colony size, maximum colony size,

and heaviness of distribution tails. For example, repulsive forces result in smaller mean colony sizes, with short tails, whereas attractive forces increased the likelihood of larger colonies, as well the maximum observed colony size.

The results of this model also indicate that although the patterns of colony size are complicated, different behaviors result in patterns that are statistically different. The patterns are surprisingly distinct. That enables us to conclude that groups of individuals are exhibiting different behavior simply by comparing the patterns of group sizes. The distinctness of the patterns means that this model would be useful for making quantitative predictions, and for inferring from empirical data the strategies employed by bacteria in the initial stages of biofilm formation.

When studying bacterial biofilm formation, we are often interested in genes that are important for the development of the biofilms. In laboratory experiments, genes that are thought to be involved in attachment or biofilm formation are deleted or mutated. Bacteria with these modifications are then grown *in vitro* to see if they can still form biofilms or produce certain substances. However, often the particular functions of the genes involved are unknown. Even if the function is known, why that gene is important or how it influences community structure may be unknown. For instance, genes that are believed to influence chemotaxis and motility in *V. cholerae* are also important to biofilm formation. However, how these genes mediate inter-cellular interactions and biofilm formation is well not understood. As a first step, we are interested in learning if these genes may be playing a part in the initial stages of biofilm formation. The model presented here provides a framework for exploring these kinds of questions.

For instance, imagine a flow cell experiment (Figure 4.16). The flow cell is first inoculated with bacteria, which are allowed a short amount of time to form initial attachments

to the surface. After a short time has elapsed, the flow cell is flushed so that any bacteria not already attached to the surface are removed. Initial images are taken to determine the approximate initial cell density. Then fresh medium is allowed to flow through the flow cell. The remaining bacteria reproduce and/or move around on the surface. Images can be taken at later times, and the approximate number of cells per colony can be recorded. These data can then be compared with model predictions. This would provide information about which of the models presented best matches the results of the experiment as well as what type of behavior the bacteria are exhibiting, *i.e.* if they are attracting via chemotaxis, or moving along the surface at all. Experiments involving bacteria with certain mutations, such as one that completely removes the ability of a bacteria to produce a flagellum or a mutation that alters a chemotaxis system, could be used to pin down estimates of parameters such as the doubling time or effective force of attraction/repulsion.

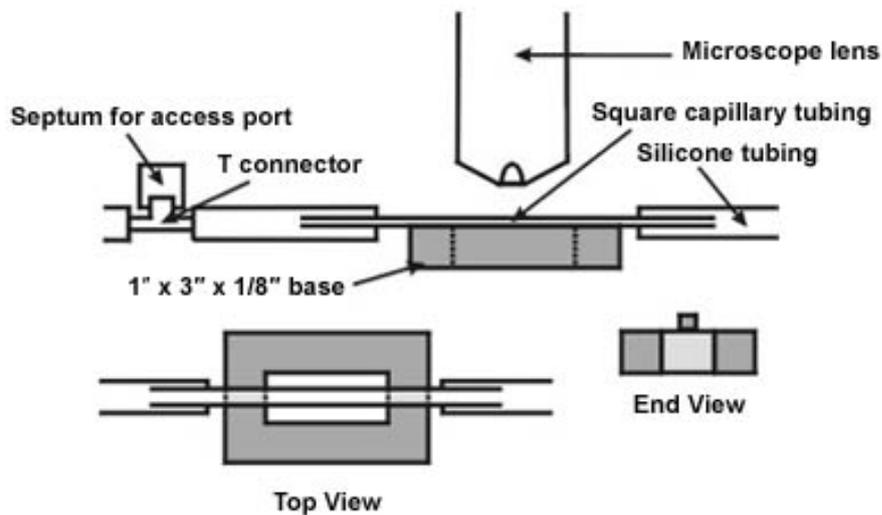


Figure 4.16: Diagram of a flowcell used to grow biofilms. Nutrients and bacteria enter the cell from the left, flow through the square tubing (center of flowcell) and exit through the tubing to the right. From BiofilmsOnline ([www.biofilmsonline.com](http://www.biofilmsonline.com))

## Chapter 5

# Conclusions and Future Work

Cholera is a pandemic disease that affects hundreds of thousands of people every year. Because the bacterium that causes cholera, *Vibrio cholerae*, is always present in aquatic ecosystems, eradication of the bacteria is not feasible. Instead, by learning about how the bacteria survives in different environments we may be able to find better ways of treating drinking water, predicting future outbreaks, and treating infected individuals.

In this dissertation I explored the dynamics of cholera on three different scales. In Chapter 2 I presented a model of the dynamics of cholera in a human population, with an emphasis on how the dynamics of a reservoir of bacteria influence patterns of infection in humans. This model emphasized how the patterns of infection are dependent upon the dynamics that influence survival and transmission of the bacteria. It highlights the need to better understand the mechanisms that influence bacterial survival and reproduction in the aquatic environment, and provides motivation for the models developed in Chapters 3 & 4.

New research has provided the insight that symmetrically dividing bacteria age [84], but the possible consequences of aging in these organisms had not been explored. In Chapter 3

I introduced a simple life history model for bacteria to explore the effects aging and of trade-offs between survival and reproduction on fitness. Based on the Euler-Lotka equation, this model provides a framework for exploring life-history trade-offs in bacteria that is simple and flexible. Under this model, I found that bacteria with finite reproduction experience very little loss of fitness compared with bacteria with the ability to reproduce indefinitely. However, when doubling times also increase over time, fitness is much lower. In this case, the more quickly the doubling time increases, the less fitness the organism gains from later reproductions. Aging also changes the optimal allocation strategies. In general, aging organisms tend to allocate a higher proportion of resources to survival than would be optimal for an organism with infinite reproductive capacity.

In Chapter 4 I focused on the formation of microcolonies on a surface. These microcolonies are interesting for two reasons. First of all, they may represent an important survival strategy for bacteria on a surface, in and of themselves. Most importantly, microcolony formation is a precursor to biofilm formation. Biofilms are already recognized as important for survival and persistence of bacteria in harsh environments, as bacteria in these communities exhibit resistance to stressors such as antibiotics and chlorination. Mechanistic models of biofilm formation such as that of Kreft et al. [50] indicate that the way in which the surface is seeded can effect the morphology of simulated biofilms. Experimental studies of *V. cholerae* indicate that genes which are important for chemotaxis also influence biofilm formation, perhaps by influencing aggregation on a surface. Understanding aggregation and microcolony formation could therefore help clarify factors influencing biofilm formation as well as understanding how groups may influence the fitness of bacteria in an aquatic environment, since group formation may be an important strategy for dealing with extrinsic sources of mortality, such as predation. The IBM model presented in Chapter 4 demonstrates that different behaviors involved in group

formation, such as attraction or repulsion between individuals, result in patterns of group sizes that are statistically distinct. The ability to distinguish between patterns could be helpful in elucidating what genes regulate behaviors necessary for biofilm formation, by enabling identification of behaviors associated with the genes. I also provided a recipe for comparing data with model hypotheses.

## 5.1 Future Work

Although I have addressed many interesting questions in this dissertation, the results have raised a new questions and indicated areas where more work could be done.

In Chapter 2 I explored a model of cholera in a human population with a reservoir of environmental *V. cholerae*. Although the full formulation of this model, as expressed in Section 2.2.1, allows the possibility of multiple types of bacteria, I have currently only explored the dynamics of the system with one strain. Because *V.cholerae* bacteria can live in either a planktonic or surface attached state, and the likelihood that an individual will become infected may be different depending upon which type they are exposed to, analyzing the model with two types of bacteria may give useful insights into the relative importance of the two survival strategies in influencing patterns of infection in the human population. It may also provide indications of what types of interventions (e.g. treatment of water with chlorine vs. filtration) would be most effective in preventing outbreaks. Recently, a similar model to the one I presented here explored how phage predation on *V. cholerae* may effect outbreaks of cholera [37]. A similar extension with protozoa predation, instead of phage predation, could be informative, since the populations of both planktonic and surface attached bacteria are influenced by protazoan predation, and the bacteria switch strategies in response to predation, as discussed in Section 4.9.

The model for bacterial life histories presented in Chapter 3 is quite general, and could easily be extended to explore other aspects of bacterial life histories, e.g. optimal groups sizes (as in Section 4.9). A limitation of this model is that it assumes constant environment. Bacteria that persist in aquatic environments, such as *V. cholerae* experience significant variability in their environment. It is possible that environmental stochasticity may influence the optimal strategies. Other approaches, such as state variable models, could be used to explore the effects of this kind of phenomenon.

Ideally, the next step for the IBM explored in Chapter 4 would be to compare the simulated distributions with empirical data. As part of this, it would be useful to estimate empirically the probability of making a type II error (*i.e.* accepting the hypothesis that the data is generated by a particular model, when the data was actually generated by another model). This would require simulating large sets of data from each of the reference distributions, which would then be compared with reference distributions as described in Section 4.8

Fairly simple extensions and modifications of the models explored here could be used to explore other kinds of bacterial communities. For instance, by adjusting the shoving parameters, the clumpiness and density of the communities can be manipulated, which may give insight into what behaviors are involved in determining smooth vs. rugose colony morphology. Allowing shoving to shape the surface attached groups in three dimensions could also be useful in this regard. One way to do this would be to combine the IBM explored here with BacSim (or some other biofilm modeling tool) to explore how colony and biofilm structures are influenced by the behaviors of individuals during initial surface colonization.

Another important extension would be to explore how different functional forms for the interaction forces between individuals influence the distributions of colony sizes. In particular, modeling chemotaxis explicitly together with an analysis of what functional form of direct forces

is the best approximation for chemotaxis would be useful.

The model could also be easily expanded to include multiple types of bacteria, or genotypic and phenotypic variation between individuals. For instance, one study on *Pseudomonas aeruginosa*, conducted by Klausen, *et al.* (2003) looked at aggregation and biofilm formation when individuals were marked with one of two florescent markers (either blue or green). They concluded that because bacterial colonies were not 50-50 blue-green, the bacteria were not aggregating on the surface. The IBM presented here could be used to explore how the size and composition of groups varies depending on how individuals interact with other individuals (of the same and different colors) more explicitly, in order to draw more concrete conclusions about the behavior of these bacteria.

# Appendix A

## Definitions

These definitions are from the Dorland's Medical Dictionary online, unless otherwise noted.

- **Bacteriophage** - virus that lyses bacteria; virus capable of producing transmissible lysis of bacteria; the virus particle attaches to the bacterial cell wall and viral nucleoprotein enters the cell, resulting in the synthesis of virus and its liberation on physical disruption of the cell. Bacterial viruses are usually specific for bacterial species, but they may be strain-specific or may infect more than one species of bacteria
- **Biotype** (or biovar) - a variant strain of a bacterial species, differing in identifiable physiologic characteristics.
- **EPS** (Extracellular polymeric substances) - Polysaccharides, proteins, lipid, etc. excreted by bacteria and used to construct microbial aggregates, such as biofilms.
- **Etiology** - 1. the study or theory of the factors that cause disease and the method of their introduction to the host. 2. the causes or origin of a disease or disorder.

- **Flagellum** (plural: flagella) [L. whip] - a long, mobile, whiplike projection from the free surface of a cell, serving as a locomotor organelle.
- **Gram-negative** - losing the stain or decolorized by alcohol in Gram's method of staining, a primary characteristic of bacteria having a cell wall composed of a thin layer of peptidoglycan covered by an outer membrane of lipoprotein and lipopolysaccharide.
- **Macrophytes** - a member of the macroscopic plant life especially of a body of water. (Merriam-Webster Dictionary)
- **Motile** - having spontaneous but not conscious or volitional movement.
- **Pathogen** - any disease-producing microorganism.
- **Pathogenesis** - the development of morbid conditions or of disease; more specifically the cellular events and reactions and other pathologic mechanisms occurring in the development of disease.
- **Pathogenic** - giving origin to disease or to morbid symptoms.
- **Phytoplankton** - the minute plant (vegetable) organisms which, with those of the animal kingdom, make up the plankton of natural waters.
- **ppGpp** - Guanosine tetraphosphate, the effector molecule of the stringent response. (From [56])
- **Serogroup** - 1. a group of bacteria containing a common antigen, possibly including more than one serotype (q.v.), species, or genus. A serogroup is a tentative and unofficial designation, used in the classification of certain genera of bacteria.
- **Serotype** (or serovar) - 1. the type of a microorganism as determined by the kinds and combinations of constituent antigens present in the cell. 2. to distinguish organisms on

the basis of their constituent antigens. 3. a taxonomic subdivision of bacteria based on the kinds and combinations of constituent antigens present in the cell, or a formula expressing the antigenic analysis on which such a subdivision is based.

- **Stringent response** - Global changes in metabolism due as a response to amino acid starvation. Response may be modulated by ppGpp (see above). (From [56])
- **Toxigenic** - producing or elaborating toxins.
- **Vibrio** - Any of various short, motile, S-shaped or comma-shaped bacteria of the genus *Vibrio*, especially *V. cholerae*.
- **Zooplankton** - the minute animal organisms which, with those of the vegetable kingdom (phytoplankton), make up the plankton of natural waters.

## Appendix B

# Mathematical Background

### B.1 Exponential and Logistic Population Growth

A population is a group of organisms that live together and reproduce. The size of the population can be regulated by four factors: births ( $B$ ), deaths ( $D$ ), immigration ( $I$ ), and emigration ( $E$ ). A population of size  $N$  will therefore always change according to:

$$\frac{dN}{dt} = B - D + I - E.$$

The manner in which these factors interact determine the way the population size will change in time. Simple models generally assume a closed population, that is they ignore immigration and emigration and focus only on births and deaths.

One of the most straightforward and familiar of the deterministic population growth models is exponential growth. It is often most suitable for modeling populations of organisms with many more resources available to them than they are able to use. Here the net birth and

death rates,  $B$  and  $D$ , will be proportional to the total size of the population:

$$\frac{dN}{dt} = B - D = r_b N - r_d N = (r_b - r_d)N. \quad (\text{B.1})$$

In (B.1)  $r_b$  and  $r_d$  are per capita birth and death rates, respectively. We can define a per capita growth rate,  $\sigma = r_b - r_d$ , so that B.1 becomes

$$\frac{dN}{dt} = \sigma N. \quad (\text{B.2})$$

This has solutions

$$N = N_0 e^{\sigma t},$$

where  $N_0$  is the initial population size. As  $t \rightarrow \infty$ , populations experiencing exponential growth increase without bound if  $\sigma > 0$  and will decay to zero if  $\sigma < 0$ .

Consider a slightly different model of population growth. Instead of having a constant per capita growth rate, suppose that the birth and death rates are density dependent, resulting in a maximum possible population size, called the carrying capacity. This kind of growth is commonly called logistic growth. Because of the density dependence, this type of growth may be a more appropriate model than exponential growth for many populations, since no population experiences unlimited resources or space.

First, let's write the net birth and death rates,  $r_b$  and  $r_d$  in terms of a linear combination of density independent birth and death rates, denoted by  $B_0$  and  $D_0$ , and density

dependent rates,  $bN$  and  $dN$ . That is:

$$r_b = B_0 - bN$$

$$r_d = D_0 + dN.$$

As the population increases, the birth rate decreases while the death rate increases, as shown in Figure B.1 (a)-(c). It is generally required that  $B_0 > D_0$  for the population to exist. Notice in the figure that different combinations of the variables shift the location of the meeting point between the two lines.

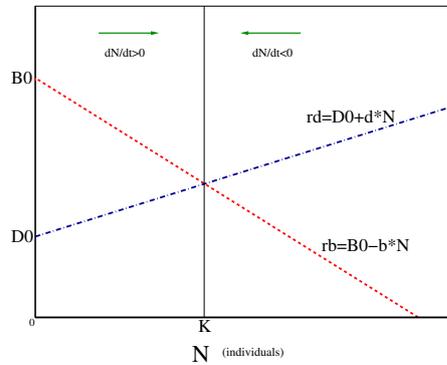


Figure B.1: Schematic of birth and death rate lines ( $r_b, r_d$ ) for values of  $D_0, B_0, b$  and  $d$ .

If we take the exponential growth model from (B.2), where  $\sigma$  is the growth rate, and substitute  $(r_b - r_d)$  for  $\sigma$ , we have

$$\frac{dN}{dt} = (r_b - r_d)N = (B_0 - bN)N - (D_0 + dN)N, \quad (\text{B.3})$$

where the first term models births in the population, and the second term deaths. Rearranging

gives:

$$\frac{dN}{dt} = (B_0 - D_0)N \left( 1 - \frac{b+d}{B_0 - D_0} N \right). \quad (\text{B.4})$$

Setting  $r = B_0 - D_0$  and  $K = \frac{B_0 - D_0}{b+d}$  and letting  $b = d$ , birth and death contributions to  $\frac{dN}{dt}$  can be written in terms of  $r$ ,  $K$ , and  $D_0$ .

$$\text{total births} = \left[ D_0 + r - \frac{r}{2K} \right] N \quad (\text{B.5})$$

$$\text{total deaths} = - \left[ D_0 + \frac{r}{2K} \right] N \quad (\text{B.6})$$

Substituting  $r$  and  $K$  into (B.4) gives the classic representation of the logistic growth equation:

$$\frac{dN}{dt} = rN \left( 1 - \frac{N}{K} \right), \quad (\text{B.7})$$

where  $K$  is the carrying capacity, and  $r$  is the per capita growth rate. Any population with growth dynamics that obey (B.7) is said to be undergoing logistic population growth. Notice that as  $K \rightarrow \infty$ , we recover exponential growth. Equation (B.7) can be solved exactly by making the substitution  $N = 1/u$ , or using partial fractions, yielding:

$$N = \frac{K}{C e^{-rt} + 1}. \quad (\text{B.8})$$

Here  $C$  is a constant of integration determined from the initial condition  $N(t = 0) = N_0$ :

$$C = \frac{K - N_0}{N_0}. \quad (\text{B.9})$$

Sample trajectories of a populations undergoing exponential and logistic growth are shown in Figure B.2 (a) and (b), respectively. Notice that the behavior of these two populations

are very different. Because of this, when we choose one of these models we are making a significant statement about the behavior allowed by the system. These population models will form the backbone of the disease models in the next chapter. We will couple them together in order to describe how disease can effect a population.

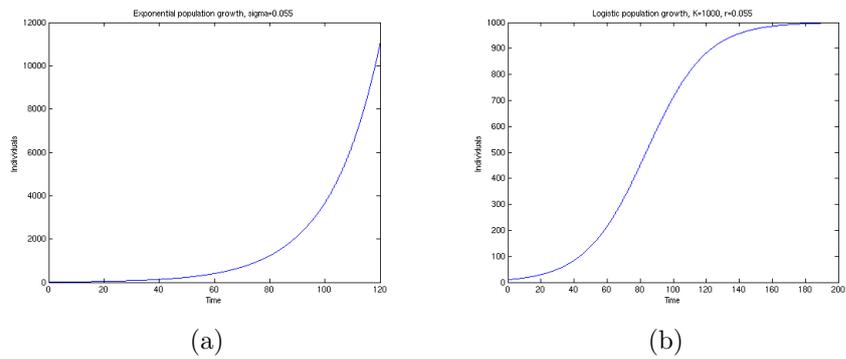


Figure B.2: (a) Exponential population growth with  $\sigma = 0.055$  (b) Logistic growth for a population with carrying capacity  $K = 1000$  and per capita growth rate  $r = 0.055$ .

## Appendix C

# Susceptible - Infected -

# Recovered Type Disease Models

### C.1 The SIS Model

The first model is the SIS (Susceptible - Infected - Susceptible) model. Suppose there is a fixed population of size  $N$ , which is initially entirely composed of susceptibles, whose population is denoted by  $S$ . Then some number,  $I_0$ , become infected at  $t = 0$ . If the population is mixing randomly, new infecteds (denoted by  $I$ ) are predicted to appear at a rate proportional to both the number of infected and the number of susceptibles. Infected individuals recover at some rate and become susceptible again. The following equations define the behavior of this

system:

$$\frac{dS}{dt} = -\lambda IS + \gamma I \quad (\text{C.1})$$

$$\frac{dI}{dt} = \lambda IS - \gamma I \quad (\text{C.2})$$

$$I + S = N. \quad (\text{C.3})$$

Here,  $\lambda$  is the daily contact rate, which is the average number of contacts per infective per unit time that actually result in disease, and  $\gamma$  is the rate at which infected individuals recover from the disease, and become susceptible. The incidence, or number of cases per unit time,  $\lambda IS = \lambda I(N - I)$  is modeled as a mass action law here (as a product of  $S$  and  $I$ ), so that transmission only depends on the relative sizes of the two classes. We assume that the latent period is zero, so that the proportion of individuals infective at time  $t_0$  who are still infective at time a later time  $t_0 + t$  is  $\exp(-\gamma t)$ , i.e. the average period of infectivity is  $1/\gamma$ .

There are various methods that can be used to explore the behavior of the system in (C.1-C.3). Two possibilities are via a phase plane analysis, or, since this is a fairly simple system, we can solve it exactly. We start by looking at the phase plane. First we find the steady states, or equilibrium points, where  $\frac{dS}{dt} = \frac{dI}{dt} = 0$ . By looking at (C.1) and (C.2) we can see that this occurs when  $I = 0$  or when  $S = \frac{\gamma}{\lambda}$ . So there are two cases to consider, one in which  $N < \gamma/\lambda$  (Figure C.1 a) and another where  $N > \gamma/\lambda$  (Figure C.1 b).

Consider the first case where  $N < \gamma/\lambda$  or  $\frac{\lambda N}{\gamma} < 1$ , depicted graphically in Figure C.1 (a). Due to the constraint (C.3), the pair  $(S, I)$  is always to the left of the line  $S = \gamma/\lambda$ , where  $\frac{dS}{dt} > 0$ , indicating that the number of susceptibles is always increasing (as indicated by the blue open arrow), and  $\frac{dI}{dt} < 0$  so the number of infecteds must decrease (indicated by the solid red arrow). The result is that the population progresses down the line  $S + I = N$  (as indicated

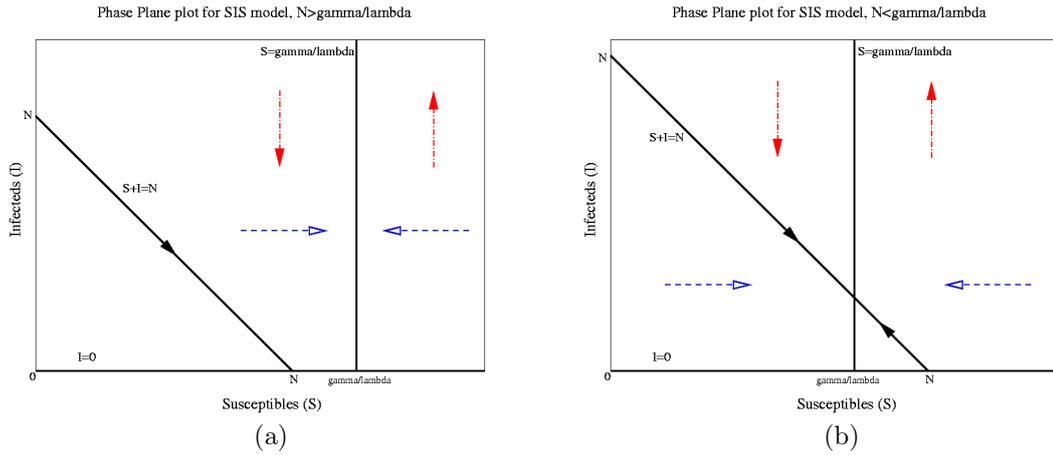


Figure C.1: Phase planes for the SIS model. Solid red arrows indicate the direction of change of  $I$ , and the open blue arrows indicate the direction of change in  $S$ . (a)  $N < \gamma/\lambda$  The constant population constraint,  $S + I = N$ , never intersects the line  $S = \frac{\gamma}{\lambda}$ , so only the disease free equilibrium is accessible. (b)  $N > \gamma/\lambda$ . The constant population constraint intersects the line  $S = \frac{\gamma}{\lambda}$  at  $I = N - \frac{\gamma}{\lambda}$ . Here, the disease free equilibrium is accessible, but unstable, while the endemic equilibrium is stable.

by the black arrow on this line), until there are no infecteds left. The disease free equilibrium,  $(I = 0, S = N)$ , is in this case a stable equilibrium point, and perturbations away from this point will return to the disease free state.

In the second case, depicted graphically in Figure C.1 (b), where  $N > \gamma/\lambda$  or  $\frac{\lambda N}{\gamma} > 1$ , there are two steady states accessible to the system; the first is the disease free equilibrium at  $(0, N)$ , and the second is at the point  $(N - \gamma/\lambda, \gamma/\lambda)$ . If the system starts out with  $S_0 > \gamma/\lambda$ , then  $\frac{dI}{dt} > 0$  and  $\frac{dS}{dT} < 0$  which means that the number of infecteds will increase while susceptibles decrease until the population approaches  $(N - \gamma/\lambda, \gamma/\lambda)$ . Starting from  $S_0 < \gamma/\lambda$ , on the other hand,  $\frac{dI}{dt} < 0$  and  $\frac{dS}{dT} > 0$ , with the result that infecteds decrease and susceptibles increase, converging again at  $(N - \gamma/\lambda, \gamma/\lambda)$ . This is the stable, endemic equilibrium. Starting from the disease free equilibrium, introduction of even one infected will result in the population moving to the endemic state.

In summary, a phase plane analysis reveals that the value of a single ratio determines the kind behavior the system can display. This ratio is often denoted as  $R_0$  and is called the

basic reproductive rate or number. It corresponds to the average number of secondary infections caused by an infective. For the system in (C.1-C.3):

$$R_0 = \frac{\lambda N}{\gamma}. \quad (\text{C.4})$$

If this rate is larger than one, the disease persists. However, if it is smaller than one, the disease will die out.

Phase plane analysis can be a powerful tool for providing insight into the dynamics of complicated models. However, simple models like (C.1-C.3) can be solved exactly. Since the population size is fixed, we can rewrite (C.2) in terms of  $I$  and  $N$  alone:

$$\frac{dI}{dt} = \lambda I(N - I) - \gamma I. \quad (\text{C.5})$$

This can be rearranged:

$$\frac{dI}{dt} = (\lambda N - \gamma)I \left[ 1 - \frac{I}{N - \gamma/\lambda} \right], \quad (\text{C.6})$$

which is the logistic equation, discussed in Section B.1. The solution is therefore:

$$I = \frac{N - \gamma/\lambda}{1 + C e^{-(\lambda N - \gamma)t}}. \quad (\text{C.7})$$

The constant of integration,  $C$  is given by:

$$C = \frac{(N - \frac{\gamma}{\lambda}) - I_0}{I_0}. \quad (\text{C.8})$$

where  $I(0) = I_0$  is the initial number of infected individuals.

Many interesting things can be learned from the analytic solution. For example, we can see that if  $(\lambda N - \gamma) > 0$  or  $\frac{\lambda N}{\gamma} > 1$  then as  $t \rightarrow \infty$ ,  $I$  approaches a constant value, specifically  $(N - \gamma/\lambda)$ . On the other hand, if  $\frac{\lambda N}{\gamma} < 1$  then as  $t \rightarrow \infty$ ,  $I$  decays to zero. So here, as with the phase plane analysis, we find that we have two possible steady states, one of which corresponds to the infection dying off ( $\frac{\lambda N}{\gamma} < 1$ ) and the other to an endemic equilibrium ( $\frac{\lambda N}{\gamma} > 1$ ) where the infection persists. Again, in the case where there is no immunity to a disease, the dynamics depend only on the ratio  $R_0 = \frac{\lambda N}{\gamma}$ . Sample trajectories for these two cases are shown in Figure C.2. In other models (we shall see) that the basic reproductive rate can have a more complicated form, but retains the same meaning.

## C.2 The SIR Model (without demographic processes)

The SIR (Susceptible-Infected-Recovered) model is an extension of the SIS model. Here, infected individuals are removed from the infected class into a “recovered” class, instead of returning to the susceptible class. Recovered individuals cannot become reinfected. So in a population with constant size  $N$ , the SIR model is depicted by the following system of differential equations:

$$\frac{dS}{dt} = -\lambda IS \tag{C.9}$$

$$\frac{dI}{dt} = \lambda IS - \gamma I \tag{C.10}$$

$$\frac{dR}{dt} = \gamma I \tag{C.11}$$

$$S + I + R = N. \tag{C.12}$$

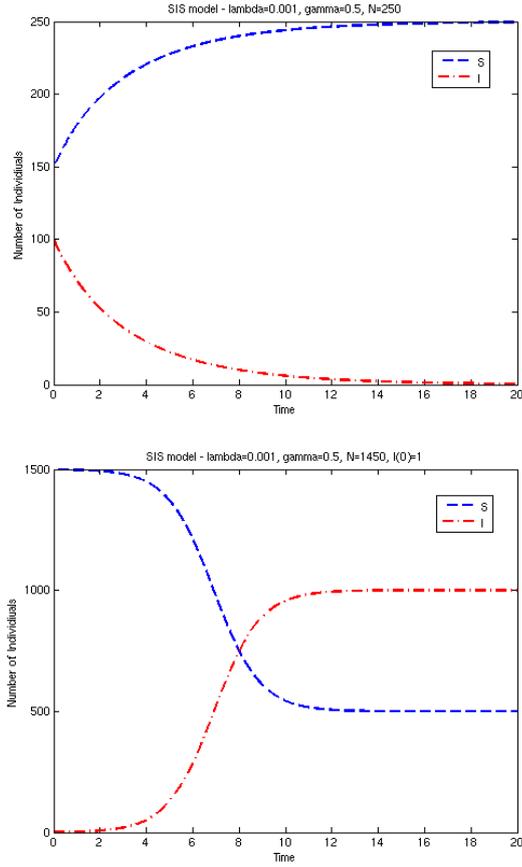


Figure C.2: Sample population trajectories for the SIS model. (a)  $\lambda = 0.001, \gamma = 0.5, N = 250$  (b)  $\lambda = 0.001, \gamma = 0.5, N = 1450, I(0) = 1$ .

Here  $\lambda$  and  $\gamma$  have the same interpretation as in the SIS model. However, since we have a closed population where individuals cannot become re-infected, any outbreak will end with the disease dying out. The question is: Does an epidemic occur, or does the disease die out immediately? This system cannot be solved analytically like the SIS model. However, insights into the dynamics of the model can be obtained by analyzing the phase space plot, or a numerical solution can be sought.

Looking at (C.9), we can see that  $\frac{dS}{dt} = 0$  when  $I = 0$  or  $S = 0$ , otherwise,  $\frac{dS}{dt} < 0$ .

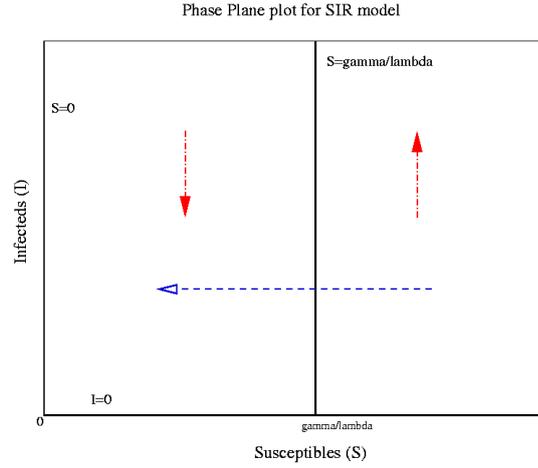


Figure C.3: Phase plane for the SIR model, without vital statistics. Solid red arrows indicate the direction of change of  $I$ , and the open blue arrows indicate the direction of change in  $S$ . Here the line  $S = \gamma/\lambda$  separates the areas of phase space in which epidemics occur ( $S_0 > \gamma/\lambda$ ) from the area where the disease dies out immediately ( $S_0 < \gamma/\lambda$ ).

From (C.10) we conclude that  $\frac{dI}{dt} = 0$  when  $I = 0$  or  $S = \gamma/\lambda$ . Therefore the whole  $S$ -axis is a line of equilibrium. If the system starts at some point ( $S_0 > 0, I_0 > 0$ ) what happens? The phase plane plot, Figure C.3, indicates that the number of infecteds will increase when  $S_0 > \gamma/\lambda$  since  $\frac{dI}{dt} > 0$  (solid red arrow). Susceptibles then decrease (open blue arrow) until the susceptible population declines to  $\gamma/\lambda$  since  $\frac{dS}{dt} < 0$ . After crossing this threshold,  $\frac{dI}{dt} < 0$  and the infected population starts to decline (red solid arrow), until reaching equilibrium at ( $S_f < \gamma/\lambda, I_f = 0$ ). In other words, if  $S_0 > \gamma/\lambda$  or  $\frac{\lambda S_0}{\gamma} > 1$  an epidemic will occur. However, if initially  $\frac{\lambda S_0}{\gamma} < 1$ , the infected population will decrease, and no epidemic occurs, regardless of how many infecteds are introduced into the population (see sample trajectories, Figures C.4(a,b)).

The basic reproductive rate for the SIR model is:

$$R_0 = \frac{\lambda S_0}{\gamma}. \quad (\text{C.13})$$

This is very similar to what we found for the SIS model, except here  $R_0$  only depends on the initial number of susceptibles,  $S_0$ , not the total population. However,  $R_0$  plays a different role in the SIR model, acting as a threshold between epidemic and non-epidemic states, whereas in the SIS model it separated the endemic and non-endemic states. In fact, in the SIR model, endemic states are not possible. In addition, at the end of an epidemic, unless a new group of susceptible individuals is introduced by births or immigration, a disease with the same parameters cannot invade the population.

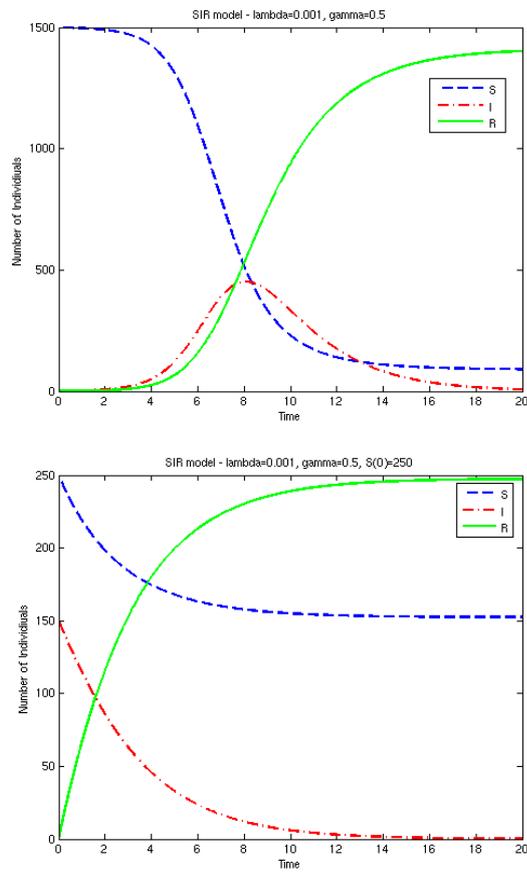


Figure C.4: Sample population trajectories for the SIR model, without vital statistics. (a)  $\lambda = 0.001, \gamma = 0.5, S(0) = 1499, I(0) = 1$  (b)  $\lambda = 0.001, \gamma = 0.5, S(0) = 450, I(0) = 150$

### C.3 The SIR Model (with demographic processes)

When modeling a disease that effects a population over a large amount of time, it is important to include population dynamics such as births and deaths. These kinds of dynamics can have a significant effect on the way a disease behaves in a population. A population can include various demographic dynamics. It may experience exponential or logistic growth (as discussed in Section ??). Certain portions of the population, such as infecteds or recovereds, may not be able to reproduce because of some effect of the the disease. All individuals entering the population may enter into the susceptible class, or some portion of new births could be born with immunity. In addition, disease fatality may also occur which can be an important additional dynamic.

An SIR model that includes these “demographic processes” might look something like this:

$$\frac{dS}{dt} = -\lambda IS + B - \mu S \quad (\text{C.14})$$

$$\frac{dI}{dt} = \lambda IS - \gamma I - \mu I - mI \quad (\text{C.15})$$

$$\frac{dR}{dt} = \gamma I - \mu R \quad (\text{C.16})$$

$$B = \text{const.} \quad (\text{C.17})$$

Here,  $\lambda$  and  $\gamma$  retain the same interpretation as before. However  $\mu$  is added to account for a daily death removal rate, which corresponds to a negative exponential age structure with average life expectancy  $1/\mu$ , and  $m$  is added to account for the mortality due to the disease. In this model, the birth rate (or general rate of influx of susceptible individuals via births or immigration) is constant. We can choose to let  $B = \mu K$  where  $K$  is the initial population size. In this case, when the total population  $N = K$  and there are no deaths due to disease,

the net birth and death rates will cancel, and the total population size will remain constant. When deaths due to disease are added, the effective per capita growth rate will be increased. Unlike the the SIR model without demography, the susceptible population in this model can be replenished, since all births are assumed to be into the susceptible class, and moreover the total population size,  $N$ , can vary (due to deaths from disease). Because of this, the behavior of the system in (C.14-C.17) can be very different from the SIR model without demography.

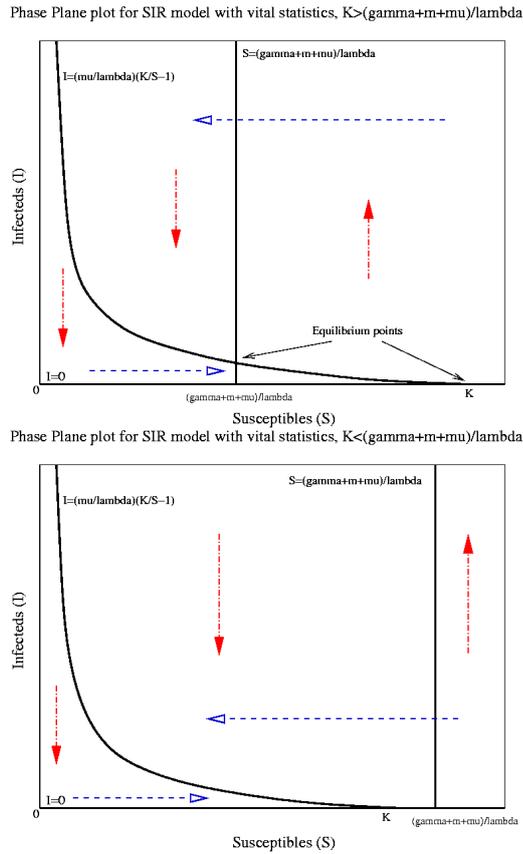


Figure C.5: Phase planes for the SIR model with vital statistics. Solid red arrows indicate the direction of change of  $I$ , and the open blue arrows indicate the direction of change in  $S$ . (a)  $\lambda K > \mu + \gamma + m$  The curve  $I = \frac{\mu}{\lambda} \left( \frac{K}{S} - a \right)$  intersects the line  $S = \frac{\gamma + \mu + m}{\lambda}$  at the point  $(I, S) = \left( \frac{\mu K}{\mu + \gamma + m} - \frac{\mu}{\lambda}, \frac{\mu + \gamma + m}{\lambda} \right)$ . Here, the disease free equilibrium is accessible, but unstable, while the endemic equilibrium is stable. (b)  $\lambda K < \mu + \gamma + m$  The curve  $I = \frac{\mu}{\lambda} \left( \frac{K}{S} - a \right)$  never intersects the line  $S = \frac{\gamma + \mu + m}{\lambda}$ , so only the disease free equilibrium is accessible, and stable.

The phase planes for this system are shown in Figure C.5. Analysis of these is similar to the SIS and SIR phase planes and so will not be discussed here in detail. Instead, we can learn about the steady state behavior of the system can by examining the stability of the equilibrium solutions. We want to start by linearizing the system in (C.14-C.17) around the equilibrium points. Near the equilibrium point  $(S_0, I_0)$ , the r.h.s of (C.14) and (C.15) can be approximated by a Taylor expansion. For example:

$$\frac{dS}{dt} \approx \left[ \frac{dS}{dt} \right]_{(S_0, I_0)} + \frac{\partial}{\partial S} \left[ \frac{dS}{dt} \right]_{(S_0, I_0)} (S - S_0) + \frac{\partial}{\partial I} \left[ \frac{dS}{dt} \right]_{(S_0, I_0)} (I - I_0),$$

and similarly for  $\frac{dI}{dt}$ . Since  $(S_0, I_0)$  is an equilibrium point,  $\left[ \frac{dS}{dt} \right]_{(S_0, I_0)} = \left[ \frac{dI}{dt} \right]_{(S_0, I_0)} = 0$ . Letting  $x = S - S_0$  and  $y = I - I_0$  gives the linearized system of equations (ignoring (C.16) for now since  $I$  and  $S$  don't depend on  $R$ ):

$$\frac{dx}{dt} = -(\lambda I_0 + \mu)x - \lambda S_0 y \quad (\text{C.18})$$

$$\frac{dy}{dt} = \lambda I_0 x + (\lambda S_0 - \mu - \gamma - m)y \quad (\text{C.19})$$

The matrix of coefficients for the system near the equilibrium states is:

$$\begin{bmatrix} -\lambda I_0 - \mu & \lambda S_0 \\ \lambda I_0 & \lambda S_0 - \mu - \gamma - m \end{bmatrix}. \quad (\text{C.20})$$

This system of equations will have solutions like  $e^{Lt}$  with eigenvalues denoted by  $L$  if:

$$\begin{vmatrix} -\lambda I_0 - \mu - L & \lambda S_0 \\ \lambda I_0 & \lambda S_0 - \mu - \gamma - m - L \end{vmatrix} = 0 \quad (\text{C.21})$$

If the eigenvalues,  $L$ , have positive real parts, the equilibrium is unstable. If the real

parts are all negative, the equilibria are stable. Imaginary parts indicate oscillatory behavior.

The system in (C.14-C.17) has two possible equilibria:  $(I, S) = (0, K)$  and  $(I, S) = (\frac{\mu K}{\mu + \gamma + m} - \frac{\mu}{\lambda}, \frac{\mu + \gamma + m}{\lambda})$  corresponding to disease free and endemic equilibria, respectively. For the first case, disease free equilibrium, (C.21) becomes

$$\begin{vmatrix} -\mu - L & \lambda K \\ 0 & \lambda K - \mu - \gamma - m - L \end{vmatrix} = 0,$$

which has solutions:

$$\begin{aligned} L_1 &= -\mu \\ L_2 &= \lambda K - \mu - \gamma - m. \end{aligned}$$

So, for this case, the equilibrium is stable if the second eigenvalue  $L_2$  is negative, i.e. when  $\lambda K < \mu + \gamma + m$  (corresponding to the phase plane plot shown in Figure C.5b) and unstable otherwise (corresponding to Figure C.5a).

The endemic equilibrium point,  $(I, S) = (\frac{\mu K}{\mu + \gamma + m} - \frac{\mu}{\lambda}, \frac{\mu + \gamma + m}{\lambda})$ , is actually only available to the system if  $\lambda K > \mu + \gamma + m$ , since  $I$  cannot be negative (see Figure C.5). At this equilibrium point, (C.21) looks like:

$$\begin{vmatrix} -\frac{\lambda \mu K}{\mu + \gamma + m} - L & -(\mu + \gamma + m) \\ \mu(\frac{\lambda K}{\mu + \gamma + m} - 1) & -L \end{vmatrix} = 0.$$

The eigenvalues for this system are given by:

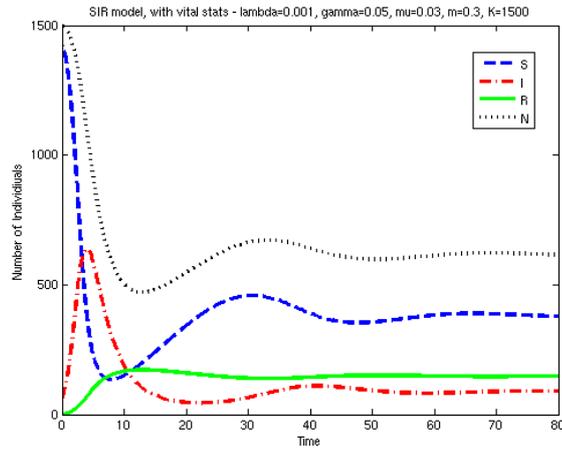
$$L_{1,2} = -\frac{\lambda \mu K}{2(\mu + \gamma + m)} \pm \frac{1}{2} \sqrt{\left(\frac{\lambda \mu K}{\mu + \gamma + m}\right)^2 - 4\mu(\lambda K - \mu - \gamma - m)}. \quad (\text{C.22})$$

When  $\lambda K > \mu + \gamma + m$ , the second term in the discriminant must always be negative. This means that the second term in (C.22) will either be imaginary, or real but smaller than the first term, so the real parts of both eigenvalues will always be negative, and the equilibrium will be stable. However, even when  $\lambda K > \mu + \gamma + m$ , there are certain circumstances when the system can “miss” the endemic equilibrium. Some trajectories will hit the line  $I = 0$  before the equilibrium is achieved (if  $\frac{\lambda K}{\mu + \gamma + m} \approx 1$ , or if  $\frac{\lambda K}{\mu + \gamma + m} \gg 1$ ).

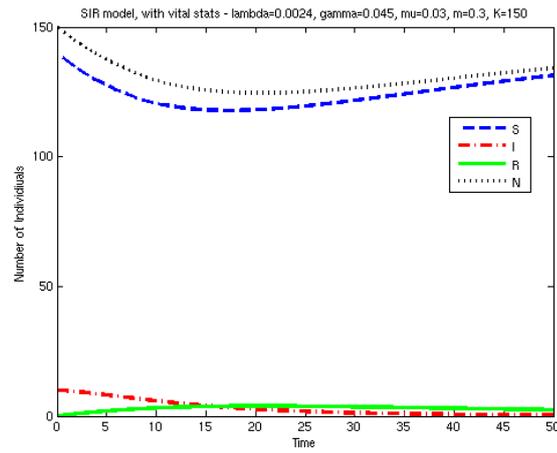
Again we have found existence of a ratio that determines the kind of steady state behavior that will be exhibited by the system. For the model in (C.14-C.17), the basic reproductive number,  $R_0$  is given by:

$$R_0 = \frac{\lambda K}{\mu + \gamma + m}, \quad (\text{C.23})$$

which describes a threshold between endemic and non-endemic states, similar to the SIS model. We expect to see a disease die out if  $R_0 < 1$  (Figure C.6a), and become endemic if  $R_0 > 1$  (Figure C.6a). Notice that in this system, an endemic disease can actually act to control a population size, holding it well below maximum capacity. There can also be small, mini-epidemics where the number of infecteds increase temporarily.



(a)



(b)

Figure C.6: Sample population trajectories for the SIR model, with vital statistics. (a)  $\lambda = 0.001, \gamma = 0.05, \mu = 0.03, m = 0.3, K = 1500$  (b)  $\lambda = 0.0024, \gamma = 0.045, \mu = 0.03, m = 0.3, K = 150$

# Bibliography

- [1] M. Ackerman, S. C. Stearns, and U. Jenal. Senescence in a bacterium with asymmetric division. *Science*, 300:1920, 2003.
- [2] R. M. Anderson. The Kermack-McKendrick Epidemic Threshold Theorem. *Bulletin of Mathematical Biology*, 53(1/2), 1991.
- [3] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: dynamics and control*. Oxford University Press, New York, 1991.
- [4] M. G. Barker and R. M. Walmsley. Replicative Ageing in the Fission Yeast *Schizosaccharomyces pombe*. *Yeast*, 15:1511–1518, 1999.
- [5] D. H. Bartlett and F. Azam. Chitin, Cholera, and Competence. *Science*, 310(5755):1775–1777, 2005.
- [6] M. A. Beaumont, W. Zhang, and D. J. Balding. Approximate Bayesian Computation in Population Genetics. *Genetics*, 162(4):2025–2035, 2002.
- [7] S. S. Branda, Å. Vika, L. Friedmanb, and R. Kolter. Biofilms: the matrix revisited. *TRENDS in Microbiology*, 13(1):20–26, 2005.
- [8] M. J. Carrier, M. Kogut, and A. R. Hipkiss. Changes in intracellular proteolysis in *Escherichia coli* during prolonged growth with a low concentration of dihydrostreptomycin. *FEMS Microbiology Letters*, 22:223–227, 1984.
- [9] B. Charlesworth. *Evolution in age-structured populations*. Cambridge University Press, Cambridge, UK, 1980.
- [10] B. Charlesworth. Fisher, Medawar, Hamilton and the Evolution of Aging. *Genetics*, 156:927–931, 2000.
- [11] D. A. Chiavelli, J. W. Marsh, and R. K. Taylor. The Mannose-Sensitive Hemagglutinin of *Vibrio cholerae* Promotes Adherence to Zooplankton. *Appl. Environ. Microbiol.*, 67(7):3220–3225, 2001.
- [12] M. A. Chowdhury, A. Huq, B. Xu, F. J. Madeira, and R. R. Colwell. Effect of alum on free-living and copepod-associated *Vibrio cholerae* O1 and O139. *Appl. Environ. Microbiol.*, 63(8):3323–3326, 1997.
- [13] C. W. Clark and M. Mangel. *Dynamic State Variable Models in Ecology*. Oxford University Press, Oxford , UK, 2000.

- [14] C. Codeço. Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. *BMC Infectious Diseases*, 1(1):1, 2001.
- [15] R. R. Colwell. Global climate and infectious disease: The cholera paradigm. *Science*, 274(5295):2025–2031, December 1996.
- [16] G. C. Cook. *The Asiatic cholera: an historical determinant of human genomic and social structure*, pages 18–53. In Drasar and Forrest [20], 1996.
- [17] K. L. Cottingham, D. A. Chiavelli, and R. K. Taylor. Environmental microbe and human pathogen: the ecology and microbiology of *vibrio cholerae*. *Frontiers in Ecology and the Environment*, 1(2):80–86, 2003.
- [18] M. E. Davey and G. A. O’Toole. Microbial Biofilms: from Ecology to Molecular Genetics. *Microbiol. Mol. Biol. Rev.*, 64(4):847–867, 2000.
- [19] D. G. Davies, M. R. Parsek, J. P. Pearson, B. H. Iglewski, J. W. Costerton, and E. P. Greenberg. The Involvement of Cell-to-Cell Signals in the Development of a Bacterial Biofilm. *Science*, 280(5361):295–298, April 1998.
- [20] B. S. Drasar and B. D. Forrest, editors. *Cholera and the Ecology of Vibrio cholerae*. Chapman and Hall, London, 1996.
- [21] F. Drenos and T. B. L. Kirkwood. Modeling the disposable soma theory of aging. *Mechanisms of Aging and Development*, 126:99–103, 2005.
- [22] S. M. Faruque, M. J. Albert, and J. J. Mekalanos. Epidemiology, Genetics, and Ecology of Toxigenic *Vibrio cholerae*. *Microbiol. Mol. Biol. Rev.*, 62(4):1301–1314, 1998.
- [23] S. M. Faruque and G. B. Nair. Molecular ecology of toxigenic *vibrio cholerae*. *Microbiology and Immunology*, 46(2):59–66, 2002.
- [24] C. E. Finch and T. B. L. Kirkwood. *Chance, Development, and Aging*. Oxford University Press, Oxford, UK, 2000.
- [25] R. A. Finkelstein. *Cholera, Vibrio cholerae O1 and O139, and Other Pathogenic Vibrios*, chapter 24. Univ of Texas Medical Branch, 1996.
- [26] L. A. Gavrilov and N. S. Gavrilova. *The Biology of Life-Span: A Quantitative Approach*. Harwood, Chur, Switzerland, 1991.
- [27] L. A. Gavrilov and N. S. Gavrilova. The reliability theory of aging and longevity. *Journal of Theoretical Biology*, 213:527–545, 2001.
- [28] V. Grimm and S. F. Railsback. *Individual-based Modeling and Ecology*. Princeton University Press, Princeton, NJ, 2005.
- [29] S. Gueron and S. A. Levin. The Dynamics of Group Formation. *Mathematical Biosciences*, 128:243–264, 1995.
- [30] S. Gueron, S. A. Levin, and D. I. Rubenstein. The Dynamics of Herds: From Individuals to Aggregations. *Journal of Theoretical Biology*, 182:85–98, 1996.
- [31] M. W. Hahn and M. G. Höfle. Grazing of protozoa and its effect on populations of aquatic bacteria. *FEMS Microbiology Ecology*, 35:113–121, 2001.

- [32] W. D. Hamilton. *Live Now, Pay Later: The moulding of senescence by natural selection*, volume 1 of *Narrow Roads of Gene Land: the collected papers of W. D. Hamilton*, pages 85–128. W.H. Freeman, Oxford, UK, 1996.
- [33] A. Heydorn, B. Ersboll, J. Kato, M. Hentzer, M. R. Parsek, T. Tolker-Nielsen, M. Givskov, and S. Molin. Statistical Analysis of *Pseudomonas aeruginosa* Biofilm Development: Impact of Mutations in Genes Involved in Twitching Motility, Cell-to-Cell Signaling, and Stationary-Phase Sigma Factor Expression. *Appl. Environ. Microbiol.*, 68(4):2008–2017, 2002.
- [34] M. Islam, B. Drasar, and R. Aack. *Ecology of Vibrio Cholerae: role of aquatic flora and fauna*, pages 187–227. In Drasar and Forrest [20], 1996.
- [35] S. M. Jazwinski and J. Wawryn. Profiles of Random Change During Aging Contain Hidden Information about Longevity and the Aging Process. *Journal of Theoretical Biology*, 213:599–608, 2001.
- [36] K. K. Jefferson. What drives bacteria to produce a biofilm? *FEMS Microbiology Letters*, 236:163–173, 2004.
- [37] M. A. Jensen, S. M. Faruque, J. J. Mekalanos, and B. R. Levin. Modeling the role of bacteriophage in the control of cholera outbreaks. *PNAS*, 103(12):4652–4657, 2006.
- [38] C. M. Jessup, R. Kassen, S. E. Forde, B. Kerr, A. Buckling, P. B. Rainey, and B. J. M. Bohannan. Big questions, small worlds: microbial model systems in ecology. *TRENDS in Ecology and Evolution*, 19(4):189–197, 2004.
- [39] E. F. Keller and L. A. Segel. Initiation of Slime Mold Aggregation Viewed as an Instability. *Journal of Theoretical Biology*, 26:399–415, 1970.
- [40] E. F. Keller and L. A. Segel. Model for Chemotaxis. *Journal of Theoretical Biology*, 30:225–234, 1971.
- [41] E. F. Keller and L. A. Segel. Traveling Bands of Chemotactic Bacteria: A Theoretical Analysis. *Journal of Theoretical Biology*, 30:235–248, 1971.
- [42] W. O. Kermack and A. G. McKendrick. Contributions to the Mathematical Theory of Epidemics - I. *Proceedings of the Royal Society*, 115A, 1927.
- [43] W. O. Kermack and A. G. McKendrick. Contributions to the Mathematical Theory of Epidemics - II. The Problem of Endemicity. *Proceedings of the Royal Society*, 138A, 1932.
- [44] W. O. Kermack and A. G. McKendrick. Contributions to the Mathematical Theory of Epidemics - II. Further Studies of the Problem of Endemicity. *Proceedings of the Royal Society*, 141A, 1933.
- [45] T. B. L. Kirkwood. *Repair and its Evolution: Survival verses Reproduction*, pages 165–189. Sinauer Associates, INC., Sunderland, MA, 1981.
- [46] T. Kirn, M. Lafferty, C. Sandoe, and R. Taylor. Delineation of pilin domains required for bacterial association into microcolonies and intestinal colonization. *Molecular Microbiology*, 35(4):896–910, 2000.

- [47] M. Klausen, A. Heydorn, P. Ragas, L. Lambertsen, A. Aaes-Jørgensen, S. Molin, and T. Tolker-Nielsen. Biofilm formation by *Pseudomonas aeruginosa* wild type, flagella and type IV pili mutants. *Molecular Microbiology*, 48(6):1511–1524, 2003.
- [48] H. A. Kramers. Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions. *Physica*, 7(4):284–288, 1940.
- [49] J.-U. Kreft, G. Booth, and J. W. T. Wimpenny. BacSim, a simulator for individual-based modeling of bacterial colony growth. *Microbiology*, 144:3275–3287, 1998.
- [50] J.-U. Kreft, C. Picioreanu, J. W. T. Wimpenny, and M. C. M. van Loosdrecht. Individual-based modeling of biofilms. *Microbiology*, 147:2897–2912, 2001.
- [51] N. Lane. *Power, Sex, Suicide : Mitochondria and the Meaning of Life*. Oxford University Press, Oxford, UK, 2005.
- [52] C. Lee, M. Hoopes, J. Diehl, W. Gilliland, G. Huxel, E. Leaver, K. McCann, J. Umbanhowar, and A. Mogilner. Non-local Concepts and Models in Biology. *Journal of Theoretical Biology*, 210:201–219, 2001.
- [53] E. K. Lipp, A. Huq, and R. R. Colwell. Effects of Global Climate on Infectious Disease: the Cholera Model. *Clin. Microbiol. Rev.*, 15(4):757–770, 2002.
- [54] K. Lorenzen. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. *Journal of Fish Biology*, 49:627–647, 1996.
- [55] M. S. Love, M. Yoklavich, and L. Thorsteinson. *The Rockfishes of the Northeast Pacific*. University of California Press, Berkeley, CA, 2002.
- [56] L. U. Magnusson, A. Farewell, and T. Nyström. ppGpp: a global regulator in *Escherichia coli*. *TRENDS in Microbiology*, 13:236–242, 2005.
- [57] C. Matz, D. McDougald, A. M. Moreno, P. Y. Yung, F. H. Yildiz, and S. Kjelleberg. Biofilm formation and phenotypic variation enhance predation-driven persistence of *vibrio cholerae*. *PNAS*, 102(46):498–506, 16819–16824.
- [58] C. Mayer, R. Moritz, C. Kirschner, W. Borchard, R. Maibaum, J. Wingender, and H.-C. Flemming. The role of intermolecular interactions: studies on model systems for bacterial biofilms. *International Journal of Biological Macromolecules*, 26:3–16, 1999.
- [59] M. D. McGurk. Natural mortality of marine pelagic fish eggs and larvae: role of spacial patchiness. *Marine Ecology*, 43:227–242, 1986.
- [60] P. B. Medawar. *An Unsolved Problem of Biology*, pages 28–54. Dover Publications, New York, 1981.
- [61] K. L. Meibom, M. Blokesch, N. A. Dolganov, C.-Y. Wu, and G. K. Schoolnik. Chitin Induces Natural Competence in *Vibrio cholerae*. *Science*, 310(5755):1824–1827, 2005.
- [62] A. Mogilner, L. Edelstein-Keshet, L. Bent, and A. Spiros. Mutual interactions, potentials, and individual distance in a social aggregation. *Journal of Mathematical Biology*, 47(4):353–389, 2003.

- [63] R. K. Mortimer and J. R. Johnston. Life span of individual yeast cells. *Nature*, 183:1751–1752, 1959.
- [64] *Nobel Lectures, Physiology or Medicine 1901-1921*. Elsevier Publishing Company, Amsterdam, 1967.
- [65] A. Okubo. Dynamical aspects of animal grouping: Swarms, schools, flocks, and herds. *Advances in Biophysics*, 22:1–94, 1986.
- [66] G. O’Toole, H. B. Kaplan, and R. Kolter. Biofilm formation as microbial development. *Annual Review of Microbiology*, 54(1):49–79, 2000.
- [67] M. R. Parsek and E. Greenberg. Sociomicrobiology: the connections between quorum sensing and biofilms. *TRENDS in Microbiology*, 13(1):27–33, 2005.
- [68] L. Partridge and N. Barton. Optimality, mutation and the evolution of ageing. *Nature*, 362:305311, 1993.
- [69] C. Picioreanu and M. C. M. van Loosdrecht. *Use of mathematical modelling to study biofilm development and morphology*, pages 5–15. IWA, London, 2003.
- [70] C. Picioreanu, M. C. M. van Loosdrecht, and J. Heijnen. *Multidimensional modelling of biofilm structure*. Atlantic Canada Society for Microbial Ecology, Halifax, Canada, 1999.
- [71] C. Picioreanu, M. C. M. van Loosdrecht, and J. J. Heijnen. Mathematical Modeling of Biofilm Structure with a Hybrid Differential-Discrete Cellular Automaton Approach. *Biotechnology and Bioengineering*, 58(1):101–116, 1998.
- [72] V. Plagnol and S. Tavaré. Approximate Bayesian Computation and MCMC. In H. Niederreiter, editor, *Proceedings of MCQMC2002*. Springer Verlag, 2003.
- [73] C. L. Rauser, L. D. Mueller, and M. R. Rose. Evolution of late life. *Aging Research Reviews*, 5:14–32, 2005.
- [74] J. Reidl and K. E. Klose. *Vibrio cholera* and cholera: out of the water and into the host. *FEMS Microbiology Reviews*, 26:125–139, june 2002.
- [75] X. Rodo, M. Pascual, G. Fuchs, and A. Faruque. ENSO and choera: A nonstationary link related to climate change? *PNAS*, 99(20):12901–12906, 2002.
- [76] D. A. Roff. *Life History Evolution*. Sinauer Associates, Inc., Sunderland, MA, 2002.
- [77] R. Ross. Some *a priori* pathometric equations. *British Medical Journal*, i:546–547, 1915.
- [78] R. Ross. An application of the theory of probabilities to the study of *a priori* pathometry, i. *Proceedings of the Royal Society, A.*, 92:204–230, 1916.
- [79] R. Ross. An application of the theory of probabilities to the study of *a priori* pathometry, ii. *Proceedings of the Royal Society, A.*, 93:212–225, 1917.
- [80] R. Ross and H. P. Hudson. An application of the theory of probabilities to the study of *a priori* pathometry, iii. *Proceedings of the Royal Society, A.*, 93:225–240, 1917.
- [81] D. B. Roszak and R. R. Colwell. Survival strategies of bacteria in the natural environment. *Microbiological Reviews*, 51(3):365–379, 1987.

- [82] B. Said and B. Drasar. *Vibrio cholerae*, pages 1–17. In Drasar and Forrest [20], 1996.
- [83] S. C. Stearns. *The Evolution of Life Histories*. Oxford University Press, Oxford, UK, 1992.
- [84] E. J. Stewart, R. Madden, G. Paul, and F. Taddei. Aging and Death in an Organism That Reproduces by Morphologically Symmetric Division. *PLoS Biol.*, 3(2):e45, 2005.
- [85] P. Suntharalingam and D. G. Cvitkovitch. Quorum sensing in streptococcal biofilm formation. *TRENDS in Microbiology*, 13(1):3–6, 2005.
- [86] Y. Tyutyunov, I. Senina, and R. Arditi. Clustering due to Acceleration in the Response to Population Gradient: A Simple Self-Organization Model. *The American Naturalist*, 164(6):722–735, 2004.
- [87] J. D. van Elsas, S. Turner, and M. J. Bailey. Horizontal gene transfer in the phytosphere. *New Phytologist*, 157:525–537, 2002.
- [88] M. C. M. van Loosdrecht, J. J. Heijnen, H. Eberl, J.-U. Kreft, and C. Picioreanu. Mathematical modelling of biofilm structures. *Antonie van Leeuwenhoek*, 81:245–256, 2002.
- [89] P. Watnick and R. Kolter. Biofilm, City of Microbes. *J. Bacteriol.*, 182(10):2675–2679, 2000.
- [90] P. I. Watnick and R. Kolter. Steps in the development of a *Vibrio cholerae* El Tor biofilm. *Mol Microbiol*, 34(3):586–586, 1999.
- [91] G. Williams. Pleiotropy, natural selection, and the evolution of senescence. *Evolution*, 11:398411, 1957.
- [92] J. W. T. Wimpenny and R. Colasanti. A Unifying hypothesis for the structure of microbial biofilms based on cellular automaton models. *FEMS Microbiology Ecology*, 22:1–16, 1997.
- [93] S. Wolfram. *A New kind of Science*. Wolfram Media, Inc., Champaign, IL, 2002.
- [94] F. H. Yildiz and G. K. Schoolnik. *Vibrio cholerae* O1 El Tor: Identification of a Gene Cluster Required for the Rugose Colony Type, Exopolysaccharide Production, Chlorine Resistance, and Biofilm Formation. *Proc. Natl. Acad. Sci. USA*, 96(7):4028–4033, March 1999.