UC Merced
UC Merced Electronic Theses and Dissertations

Title
Mathematical Investigations of Heritable Microbial Phenotypes

Permalink
https://escholarship.org/uc/item/3px1z3vz

Author
Santiago, Fabian S

Publication Date
2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Mathematical Investigations of Heritable Microbial Phenotypes

UNIVERSITY OF CALIFORNIA, MERCED

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

Doctor of Philosophy in Applied Mathematics

Fabian Santiago

Committee in charge:

Professor Suzanne S. Sindi, Chair
Professor Miriam Barlow
Professor Camille Carvalho
Professor Erica Rutter

2021

Chapter 3 © 2021 Illinois State University, Normal, IL USA
All other material © Fabian Santiago, 2021
All rights reserved.

The dissertation of Fabian Santiago is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

(Professor Miriam Barlow)

(Professor Camille Carvalho)

(Professor Erica Rutter)

(Professor Suzanne S. Sindi, Chair)

University of California, Merced

2021

DEDICATION

Esta disertación está dedicada a mi familia y a la memoria de mi bisabuelo Rutilo Fabián Díaz quien siempre dijo que nuestro futuro está en la educación; orgullosamente Oaxaqueños de ancendencia Zapoteca.

TABLE OF CONTENTS

LIST OF NOTATIONS AND ABBREVIATIONS

US	United States
Sup35	Suppressor Protein 35
GdnHCl	Guanadene Hydrocholride
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
AM	Adaptive Metropolis
ATP	Asymmetric Transmission of Propagons
t	Time
$a_i(t)$	Number of Propagons at Time t After i Cell Divisions
η	Propagon Amplification Dynamics Model
$\boldsymbol{\theta}$	Propagon Replication Dynamics Parameters
$Y_i(t,a)$	Propagon Density at Time t After i Cell Divisions
$\alpha_i(t)$	Rate of Cell Division at Time t After i Cell Divisions
$\beta_i(t)$	Rate of Cell Death at Time t After i Cell Divisions
ρ	Transmission bias
λ	Propagon Replication Rate
M	Maximal Number of Cell Generations
$\Upsilon(a)$	Initial Aggregate Density
$y_i(t,a)$	Normalized Propagon Density at Time t After$i$ Cell Divisions
$n_i(t)$	Number of Cells at Time t Having Gone Through i Division
$N(t)$	Total Number of Cells at Time t
R	Domain for the Number of Aggregates a
N_0	Initial Number of Cells
L	Likelihood Function
$P(\cdot)$	Probability
$Z(t,a;\boldsymbol{\theta})$	Total Propagon Density at Time t
Z_A	Asymmetric Division of ATP Model
Z_S	Symmetric Division of ATP model
K	Number of Parameters
m	Number of Data Points

AIC	Akaike Information Criterion
AIC^c	Bias Corrected AIC
Δ_i	AIC Difference Between the Minimum AIC and Model i
W	Model Weight
J	Proposal Distribution
π	Target Distribution
V	Covariance Matrix
$\alpha(\boldsymbol{\theta}_{new}\|\boldsymbol{\theta}_{i-1})$	Acceptance Probability
I_p	Identity Matrix (p Dimensional)
σ	Standard Deviation
μ	Mean
$\boldsymbol{q}$	Vector of Values Sorted in Descending Order
IQR	Inner Quartile Range
T_Δ	Time Since Experiment Began
$\xi(\boldsymbol{q})$	Numerically Stable Recursive Log Sum of Entries in $\boldsymbol{q}$
CFSE	Carboxyfluorescein Succinimidyl Ester
LxW	Lax-Wendroff
DLSPM	Division and Label Structured Population Model
ESBL	Extended Spectrum β-lactamase
MIC	Minimum Inhibitory Concentration
DHMMC	Dignity Health Mercy Medical Center
FDR	False Discovery Rate
Ec	*Escherichia coli*
Kp	*Klebsiella pneumoniae*
Φ	Coefficient of Correlation
χ^2	Chi-Square Statistic

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my advisor Professor Suzanne Sindi for her mentorship. As a first-generation college student her faith in my abilities, her support for my ideas in research, and her being one of my greatest advocates are what made my work possible and the process worth it. No words can fully describe the impact her mentorship has had on me as a researcher and my life in general.

A special thanks to the members of my committee who have been an immense source of support. Professor Miriam Barlow who introduced me to bacterial antibiotic resistance research, for helping me become a more well-rounded researcher, and for being another one of my greatest advocates. Professors Camille Carvalho and Erica Rutter, for their support, encouragement and fantastic feedback on my work.

My fellow collaborators greatly impacted my work and development as a researcher. Especially Evin Doscher, Dr. Jourjina Alkhouri, and Candace Cole of the Barlow Lab. Dr. Omar DeGuchy through his support, collaboration, and friendship. Dr. Juan C. Meza, whose instruction on clear scientific reasoning continues to shape my research. Professor Roummel Marcia and Professors Arnold Kim who created opportunities and funding for me to work with undergraduate students.

This experience would have been lonely without the friendship and solidarity of my graduate student cohort. Matea Santiago, Omar DeGuchy, Alex J. Quijano, Jessica Taylor, and Michael Kelly, our camaraderie, support, and passion for mathematics surpassed my hopes of working alongside enthusiastic colleagues in graduate school. Additionally, a warm thanks to the applied mathematics faculty and graduate students at UC Merced. My hope is that the sense of community and active peer support in the department continues to be passed on to future students.

My pursuit of higher education would never have been possible without the endless support provided by my family. My parents Juliana Fabian and Aureliano Santiago, my brothers Raymundo Santiago and Bertín Santiago, my sisters Janet Santiago, Itzel Santiago and Vivian Santiago, thank you for your truly unconditional love and support. Thank you to my wife Matea Santiago; your intelligence, humility, kindness, and support are everything to me. I am very fortunate to continue to see the personal growth of the confident young woman I first met in a linear algebra class at Sonoma State University. You constantly inspire me to be a better person.

There have been many friends, peers, teachers, and professors whose support I will never forget. Professor Suzanne Rivoire, who gave me my first opportunity to do research and ultimately opened the door for me to pursue a graduate education. Matthew Hardwick, who first told me to talk to professor Rivoire about a research opportunity and who insisted that I email her right then and there. Jacob Combs, our crazy adventures helped me through some serious lows and allowed me to get to this point. Thank you for your friendship, sincerity, and for showing me that although we were new to research, we still had a lot to contribute.

VITA

2021	Ph.D. in Applied Mathematics. Advisor: Suzanne S. Sindi. University of California, Merced. Merced, CA.
2015	B.S. in Applied Mathematics. Sonoma State University. Rohnert Park, CA.
2013	A.A. in Mathematics. Santa Rosa Junior College. Santa Rosa, CA

PUBLICATIONS

C. Guzman-Cole, F. Santiago, S. Garsevanyan, S. Sindi, and M. Barlow. "Distribution of β-Lactamase Genes in Clinical Isolates from California Central Valley Hospital Deviates from the United States Nationwide Trends." Antibiotics (2021), 10(5), 498.

F. Santiago, K. B. Flores, and S. S. Sindi. "Numerical approaches to division and label structured population models." Letters in Biomathematics. Vol. 7, no. 1, Nov. 2020, pp. 153–170.

F. Santiago, E. Doscher, J. Kim, M. Camps, J. Meza, S. S. Sindi, and M. Barlow, "Growth rate assays reveal fitness consequences of β-lactamases." PloS one 15.1 (2020): e0228240.

O. DeGuchy, F. Santiago, M. Banuelos, and R. F. Marcia. "Deep Neural Networks for Low-Resolution Photon-Limited Imaging." Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing, 2019.

J Combs, J, Nazor, R. Thysell, F. Santiago, M. Hardwick, L. Olson, S. Rivoire, C. Hsu, S. W. Poole. "Power signatures of high-performance computing workloads." Proceedings of the 2nd Workshop on Energy-Efficient Supercomputing (E2SC). New Orleans, LA, Nov. 2014.

C. Hsu, J. Combs, J. Nazor, F. Santiago, R. Thysell, S. Rivoire, S. W. Poole. "Application power signature analysis." Proceedings of the 10th Workshop on High Performance, Power-Aware Computing (HPPAC). Phoenix, AZ, May 2014.

ABSTRACT OF THE DISSERTATION

Mathematical Investigations of Heritable Microbial Phenotypes

by

Fabian Santiago
Doctor of Philosophy in Applied Mathematics in

University of California Merced, 2021

Professor Suzanne S. Sindi, Chair

This dissertation presents mathematical models, numerical methods, and data driven investigations of heritable cellular phenotypes. First, mathematical models and numerical methods for the dynamics of protein misfolding (prions) in yeast cells are presented (Chapters 2 and 3). This prion phenotype, in *Saccharomyces cerevisiae* yeast cells can be inherited by daughter cells through transmission of protein aggregates during cell division, a non-Mendelian form of inheritance. Prions are not harmful to yeast; this allows for their use as a biological model to gain insight into the mechanisms that govern prion replication and transmission. Second, data driven mathematical approaches to evaluate new biological techniques used to study the evolution of bacterial antibiotic resistance are presented (Chapters 4 and 5). Bacterial antibiotic resistance is a global human health problem. In the U.S. more than 35,000 people die from antibiotic-resistant infections and around 3 million get an antibiotic-resistant infection every year [30]. In this work, we used two repositories of infectious isolates collected from patients at Dignity Health Mercy Medical Center in Merced, California, USA, and a nationwide database compiled from clinical isolate genomes reported by the National Center for Biotechnology Information since 2013.

Chapter 1

Introduction and Organization of the Dissertation

The dissertation consists of two parts: in *part one* we investigate a misfolded protein based heritable phenotype found in yeast cells (Chapters 2 and 3), and in *part two* we investigate bacterial antibiotic resistance and the genetic heritability of resistance (Chapters 4 and 5). Together with biological experimentation, mathematics plays a critical role in understanding these biological phenomena (See Figure 1.1). The following provides an introduction for each part of the dissertation and summarizes the work in each chapter.

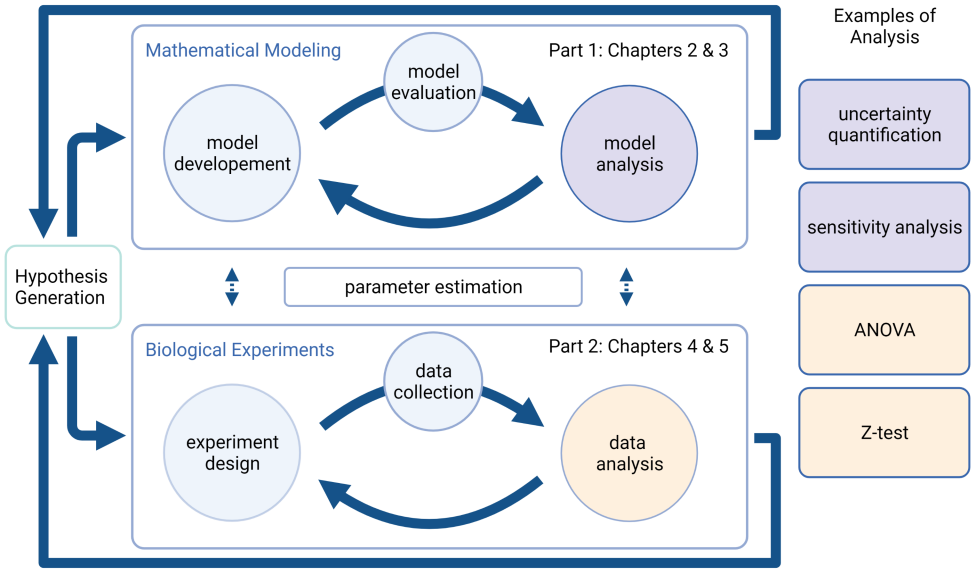


Figure 1.1: **Experimentation and Mathematical Modeling Workflow.** The flow diagram (adapted from [115]) shows that mathematical modeling, model analysis, and data analysis are critical components to understanding biological phenomena.

Part 1: Yeast Prion Proteins

There are many neurodegenerative diseases that affect humans and animals that are due to misfolded (abnormal) proteins called prions [36, 113]. Alzheimer's disease, the sixth-leading cause of death in the United States and the fifth-leading cause of death for people that are 65 and older [6], is an example of such a disease. In 2019 it was reported that approximately 55 million people in the US had Alzheimer's disease, and it is projected that this number almost double by 2050 [6]. This type of protein misfolding disorder has also been observed in animals such as scrapie in sheep, chronic wasting disease in deer and Bovine Spongiform Encephalopathy in cattle [126, 112]. These diseases are difficult to treat and their progression is fatal. While the details differ, all share two key commonalities: (1) a misfolded form of a protein appears in an otherwise healthy cell; (2) this misfolded form of the protein spreads to other cells.

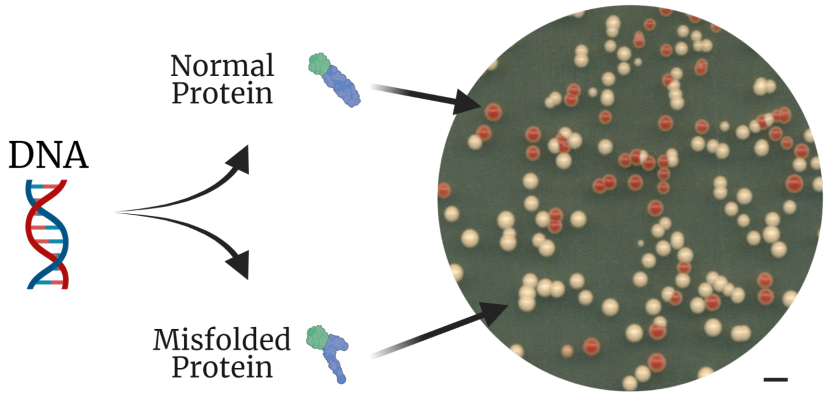


Figure 1.2: **Normal Protein, Protein Misfolding, and Yeast Colony Phenotype.** DNA encodes for the three-dimensional structure of proteins. *Top:* most of the time protein folds into its normal form which results in a red yeast colony. *Bottom:* protein misfolding occurs when a misfolded form of a protein appears and spreads to other cells by acting as a catalyst on normally folded protein to create the misfolded protein form and leads to the formation of a white colony. The scale is 2.75mm.

In order to better understand prions and their aggregation, biologists have turned to the yeast *Saccharomyces cerevisiae* as a biological model for studying prion disease (See Figure 1.2). In yeast, various harmless but heritable phenotypes have been identified that are transmitted by highly stable prion aggregates [135, 136]. Prion aggregates are harmless to the yeast cells, this allows researchers to study the effects of the protein dynamics in the absence of harming the host [74]. Intracellular aggregation and fragmentation dynamics of prion aggregates have been well studied [126, 49, 43, 44], however one of the challenges of working with yeast as a model system is that cells divide during the course of the experiment. Therefore, the number of aggregates in a cell will decrease when the cell divides as any aggregates will be separated between the resulting mother and daughter cells [126]. This means that

for meaningful quantitative comparisons between mathematical models and experimental systems, care must be taken to understand the population averages but also heterogeneity amongst cells in the same population (Chapter 2). Without cell division, we know that the number of aggregates in a cell will increase until a steady-state concentration is reached, where the aggregate number is in balance with the soluble protein level [120, 124, 60, 114] (Chapter 3).

Chapter 2: A Structured Population Model and Likelihood Approach to Estimate Propagon Replication Rates and Their Asymmetric Transmission from Propagon Recovery Experiments in *Saccharomyces Cerevisiae*

We present a structured population model describing the distribution and replication of yeast prion propagons in an actively dividing population of cells. We then develop a likelihood approach for estimating the propagon replication rate and their transmission bias during cell division. We first demonstrate our ability to correctly recover known kinetic parameters from simulated data, then we apply our likelihood approach to estimate the kinetic parameters for six yeast prion variants using propagon recovery data. We find that, under our modeling framework, all variants are best described by a model with an asymmetric transmission bias. This demonstrates the strength of our framework over previous formulations that assume equal partitioning of intracellular constituents during cell division by characterizing the heterogeneity in cell division.

Chapter 3: Numerical Approaches to Division and Label Structured Population Models

In this work, we develop a novel theoretical and numerical framework involving a recursive formulation for a class of division and label structured population models (DLSPMs). We develop this framework for a population of dividing cells with an arbitrary functional form describing the intracellular dynamics. We found that, compared to previously used numerical methods, our recursive framework enables faster and more accurate numerical solutions to DLSPMs. We apply our numerical framework to three common models for intracellular dynamics (exponential, logistic, and constant synthesis growth models) and discuss the potential impact of our findings in the context of combining it with data-driven methods for parameter estimation and uncertainty quantification.

Part 2: Bacterial Antibiotic Resistance

The emergence of bacterial resistance to established antibiotics is a major world health crisis that leads to over 700,000 deaths globally each year, and by the year 2050

it is estimated that antibiotic resistant infections will lead to around 10 million deaths per year [101]. Bacterial resistance has appeared to nearly every antibiotic that has been deployed [33]. Figure 1.3, adapted from information available in [30], presents a timeline that illustrates examples of antibiotic release and resistance identification. Controlling the spread of resistance is complicated by the multiple ways in which bacteria are able to share beneficial genes, including through cell division to resulting daughter cells (vertical transfer of genetic material) and by sharing genetic material with distantly related bacteria (lateral transfer of genetic material) [80]. These forms of gene transfer lead to quick dissemination of resistance phenotypes [41, 131, 67]. It should be noted that the appearance of resistance does not imply that an antibiotic no longer has clinical use, but instead that its use should be targeted to infections that remain susceptible to that antibiotic.

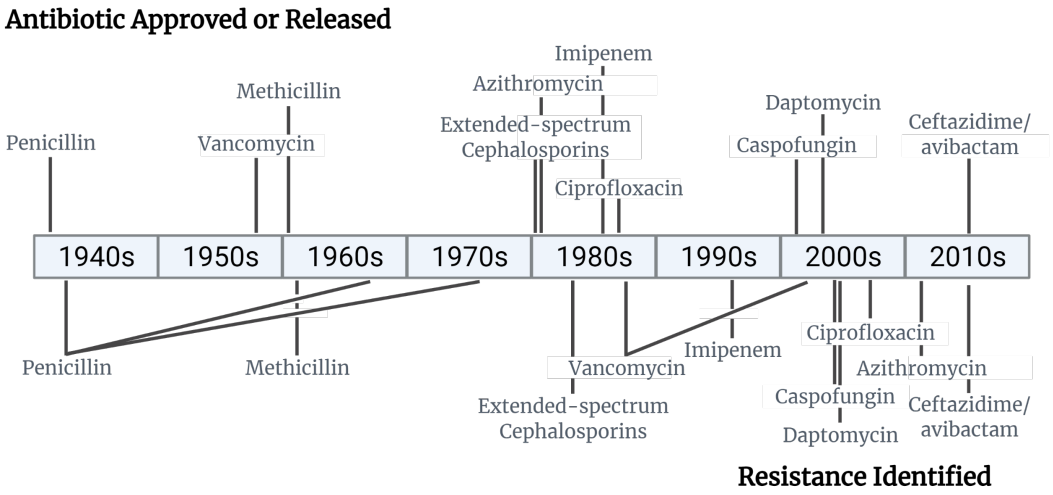


Figure 1.3: **Timeline of Antibiotic Approval (or Release) and First Resistance Detection.** Each line in *resistance identified* indicates detection of resistance in a different bacterial species.

The development of new antibiotics is complicated by both technical difficulties and the lack of economic incentive. Since resistance can emerge quickly, developers know a new antibiotic will have only a limited number of years of clinical effectiveness. Based on these observations, new views in combating antibiotic resistance have emerged: the development of faster and more robust biological and mathematical methods for assessing resistance to particular antibiotics (Chapter 4), and reliable surveillance of antibiotic resistance (Chapter 5), in order to determine effective ways to use the antibiotics currently available.

In this work, antibiotic resistance is investigated through statistical analyses of infectious isolates selected for containing extended spectrum β-lactamases (ESBLs), which are enzymes that confer resistance to β-lactam antibiotics. Since 2013 infectious isolates mostly consisting of *E. coli* from urinary tract infections (UTIs) have

been collected from patients through a partnership with Dignity Health Mercy Medical Center (DHMMC) in Merced, CA. This work focuses on β-lactamase genes due to their prevalence in reducing the effectiveness of the most common antibiotic treatments used to treat UTIs [77, 1, 119, 70].

Chapter 4: Growth rate assays reveal fitness consequences of β-lactamases

An alternative approach for measuring fitness of microbes in the presence of an antibiotic is the use of growth rates. They are a highly sensitive approach for measuring bacterial fitness effects from antibiotics and resistance genes. Growth rates use a single concentration of antibiotic where as in clinical susceptibility testing a bacterial isolate is exposed to a range of concentrations of an antibiotic to determine a minimum inhibitory concentration (MIC) [5]. The range for determining the MIC depends on both the antibiotic and concentration levels deemed safe for the patients.

In this work we analyze isolate growth rates determined during the exponential growth phase of bacterial growth in the presence of a particular antibiotic. We first performed a statistical analysis of the growth rate estimates, under different antibiotic conditions of interest, for deviations from normality using the Shapiro-Wilk test [73], then we compared differences in growth rates by antibiotic conditions using Welch's t-test [118]. We set $\alpha = 0.01$ as the significance level, and performed a post hoc Bonferroni-type multiple testing procedure [21] with a false discovery control level of $q = 0.05$, to control the false positive rate in performing multiple statistical test.

We show that despite the key difference of only using one concentration of antibiotic, the use of growth rates correlate well with clinical determination of resistance by MIC while providing the sensitivity required for direct input as fitness values into mathematical models.

Chapter 5: Distribution of β-Lactamase Genes in Clinical Isolates from California Central Valley Hospital Deviates from the United States Nationwide Trends

In this work, we used two repositories of ESBL-producing isolates collected since 2013 from patients at DHMMC and a nationwide database compiled from clinical isolate genomes reported by the National Center for Biotechnology Information since 2013 [98]. We first show that average resistance gene frequencies over consecutive years are statistically stable using a Z-test [31]. Using this Z-test formulation, we also compare the frequencies of resistance genes in the DHMMC collection with the averages of the nationwide frequencies. We then address the question of statistical evidence for the correlation or pairwise linkage between resistance genes using a χ^2 test [61] and the phi coefficient $\left(\Phi = \pm\sqrt{\frac{\chi^2}{n}}\right)$ as a χ^2 measure of directional deviation

from the null relationship of independent assortment of resistance genes [50]. Controlling for multiple statistical tests was conducted via the FDR-controlling procedure [22], a Bonferroni-type multiple testing procedure, with a false discovery control level of $q* = 0.025$ and a significance level of $\alpha = 0.05$.

We found DHMMC gene frequencies are stable over time and differ significantly from nationwide frequencies throughout the period of time we examined. Our results suggest that local selective pressures are a more important influence on the population structure of resistance genes in bacterial populations than migration. This, in turn, indicates the potential for antibiotic resistance to be controlled at a regional level, making it easier to limit the spread through local stewardship.

Conclusions and Future Directions

Chapter 6 presents concluding remarks regarding the work presented in this dissertation and suggestions for possible extensions to these areas of research.

Chapter 2

A Structured Population Model and Likelihood Approach to Estimate Propagon Replication Rates and Their Asymmetric Transmission from Propagon Recovery Experiments in *Saccharomyces Cerevisiae*

2.1 Introduction

Today in the United States, millions of individuals suffer from Alzheimer's disease and other similar dementias. The cause of these diseases is thought to be the accumulation of misfolded proteins in the brain [6]. Currently, Alzheimer's is the sixth-leading cause of death in the United States and the fifth-leading cause of death for individuals 65 and over [6]. Beyond Alzheimer's, there are a many other disorders caused by protein misfolding. Neurodegenerative diseases such as Parkinson's and Huntington's disease to less well-known diseases like Kuru and Creutzfeldt-Jakob Disease [129, 68] have a similar pathology. In addition, protein misfolding disorders have also been observed in other mammals such as scrapie in sheep, chronic wasting disease in deer and Bovine Spongiform Encephalopathy in cattle [126, 112]. Collectively, these diseases are largely untreatable and nearly universally fatal. While the details differ, all share two key commonalities: (1) a misfolded form of a protein *appears* in an otherwise healthy cell; (2) this misfolded form of the protein *spreads* to other cells.

A promising system allowing for insight into mammalian protein misfolding and prion disease is the yeast *Saccharomyces cerevisiae*. A number of harmless, heritable

phenotypes in yeast are shown to be transmitted vertically to new daughter cells by prion seeds, termed propagons [138, 126]. Because these propagons are harmless to the yeast cells, researchers are able to study the protein dynamics themselves in the absence of harming the host. Indeed, biologists have many experimental tools that have been developed for yeast which allow for a detailed interrogation of the protein aggregation system in ways that are not possible to do in mammalian systems *in vivo* [136]. However, a unique challenge to working with yeast as a model system is that during the experimental time course, the yeast cells themselves continue to divide [49]. Without cell division, we know that the number of propagons in a cell will increase until it reaches a steady-state concentration where the propagon number is in balance with the soluble protein level [49, 124, 60, 114]. However, the number of propagons in a cell will decrease when the cell divides as any propagons (i.e., transmissible aggregates above a critical minimal size, see [138] for more details) will be separated between the resulting mother and daughter cells [126, 49]. This means that for meaningful quantitative comparisons between mathematical models and experimental systems, care must be taken to understand the population averages but also heterogeneity amongst cells in the same population.

In this work, we build upon a previous approach to consider not only the protein dynamics within yeast cells, but the cellular populations themselves. More specifically, in this work we consider a structured population model of yeast cells where the transmission of prion aggregates are tracked as cells divide and as aggregates replicate. While prior models have been developed of this system [8], our work considers a more general framework and has several novel contributions. First, in contrast to prior studies, rather than assuming that the variance in prion aggregate counts is proportional to the system mean, we use the full likelihood of the data. Second, while prior approaches assume symmetric division of propagon between dividing cells, our work allows for non-symmetric transmission of propagons. Finally, we consider prion recovery data from six distinct prion variants.

With this more general framework, we find that all six prion variants are best described by an asymmetric transmission of propagons between actively dividing yeast cells. With our model selection framework we were also able to exclude influential outliers from our prion variant datasets for computing kinetic parameters of prion variants. Moreover, we find differences among the variants for both the propagon replication rate and the transmission bias. Encouragingly, prion variants with similar phenotypic properties are fit with similar kinetic parameters. As such, our framework offers the ability to infer meaningful properties about prion variants even with our simplified model of intracellular aggregate dynamics.

In Section 2.2, we develop the biological background of prion variants and the recovery assays we model. Section 2.3 describes our aggregate and generation structured population model and the likelihood approach we use for fitting kinetic parameters to experimental data. In Section 2.4 we first characterize the ability of our model and inference framework to recover the correct kinetic parameters and then apply our model and inference framework to recovery data from six distinct prion variants. In

Section 2.5 we discuss the implications of our study as well as factors to be considered in future studies on prion aggregate dynamics.

2.2 Biological Background

As mentioned in Section 2.1, yeast prions were not discovered in the context of a disease but in one of mysterious heritable phenotypes [38, 137]. In addition, there is a considerably shorter history of knowledge about yeast prions than their mammalian counterparts [124, 136]. The $[PSI^+]$ phenotype in yeast that we now know to be linked to a prion form of the protein Sup35, was discovered in 1965 by biologist, Brian Cox. The phenotype corresponded to that of a white colored colony and the ability to grow a colony on media lacking adenine [38]. Remarkably, this phenotype appeared to be vertically transmissible (from mother to daughter) but did not obey the laws of Mendelian inheritance as defined by DNA transmission. For example, if two diploid organisms are heterozygous for a recessive trait then, by the rules of Mendelian inheritance, we expect that 1/4 of their offspring would exhibit the recessive trait. With DNA ruled out as the mode of inheritance, the identity for the heritable species associated with the $[PSI^+]$ phenotype began. In 1994, Wickner hypothesized that $[PSI^+]$ and $[URE]$ (another mysterious phenotype) were propagated by a misfolded (prion) form of their respective proteins [144]. In 1996, Paushkin and colleagues [104] demonstrated that the $[PSI^+]$ phenotype was the result of a misfolded form of the protein Sup35 that was self-propagating. Today we know that many proteins in yeast are capable of forming prions and that a given prion protein may have multiple variants - distinct misfolded confirmations - each of which is capable of propagating through this self-propagation process. In particular, as mentioned in the introduction, we are considering six variants of the $[PSI^+]$ prion in yeast [76].

For mammalian prion disease, the disease phenotype is observed at the level of single organism (i.e., a cow, human, mouse, etc). However, in yeast prion biology the prion phenotypes are only observed with a yeast colony consisting of many organisms (cells) and founded by from a single cell [133]. As such, it has been challenging to establish a precise link between the infective species, or propagon (that necessarily) reside in a single cell with the colony level phenotype [126]. While it is clear that the presence of a single propagon in the founding cell is necessary for the appearance of the prion phenotype at the colony level, is it not clear that it is sufficient [138]. In this work, following others [49, 8, 125] we will assume that the presence of a single propagon in a founding cell is both necessary and sufficient for the appearance of the prion phenotype at the colony level. As such, yeast prion dynamics are inherently a multi-scale process [14, 79].

Prion phenotypes occur when a misfolded form of a protein occurs and, rather than be cleared by cellular quality control machinery, the misfolded form persists and associates in aggregates, ordered structures of prion monomers as shown in Figure 2.1(a). More specifically, four steps are essential to the maintenance of prion

phenotypes (see Figure 2.1(b)). First, normal protein is continually produced by the cell. Second, prion aggregates convert the normally folded protein to its misfolded confirmation through a templated conversion process which incorporates the newly misfolded protein. This conversion process increases the size of the aggregate. Third, the total number of aggregates increases when the aggregates are fragmented. This increases the total number of templating units and thus accelerates the conversion process. Fourth, aggregates are transmitted between cells during division. Intriguingly, the cell division process creates an interesting phenomenon from the perspective of a single cell. Between cell divisions, the number of aggregates increases, and it will then decrease when cell division occurs Figure 2.1(c). Due to the low frequency of spontaneous $[PSI^+]$ appearance, $\sim 10^{-8} - 10^{-7}$/generation [75], in this work we assume there is no spontaneous appearance of aggregates.

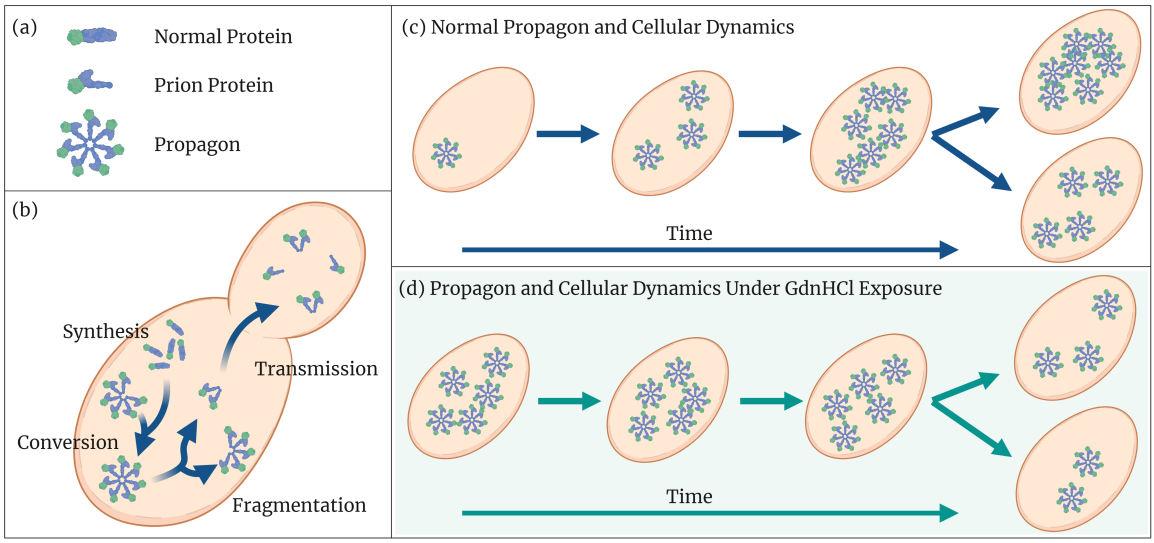


Figure 2.1: **Multiscale Yeast Prion Aggregate (Propagon) Dynamics.** (a) Within each cell in the colony is a mixture of normal protein and prion (misfolded) protein. Prion proteins are contained in aggregates of multiple misfolded monomers. (b) Within each cell normal protein is produced (synthesis) and converted to the prion form and incorporated into existing aggregates (conversion) which increases the size of an aggregate. Aggregates may increase in number by fragmentation and must be spread from mother to daughter cells during division (transmission). (c) Under normal growth conditions, the number of aggregates increases during the lifetime of a cell and is split during cell division. (d) When cells are grown under GdnHCl fragmentation is assumed to stop and the number of aggregates remains unchanged during the lifetime of a cell.

Yeast biologists take advantage of the fact that when cells are exposed to Guanadine Hydrochloride (GdnHCl) cell division is not impacted, but the aggregate fragmentation process is assumed to halt. As such, the number of aggregates within a cell, and indeed in the entire population is kept constant, while the number of cells in

the colony continues to increase, see Figure 2.1(d). This allows for experiments which probe the number of aggregates in a single cell by regrowing colonies and assuming that any cell with at least one aggregate will create a colony with a prion phenotype. A single yeast cell with propagons is introduced to a GdnHCl environment and the population of cells is allowed to grow normally. Because aggregates will be split between mother and daughter cells during division, the expected number of aggregates per cell will continue to decrease until the point where a cell in the population is extremely unlikely to have more than one aggregate [35, 96, 125]. Then each cell in this population is allowed to form their own colony under normal growth conditions.

In propagon recovery experiments, yeast biologists use GdnHCl exposure in two phases to observe the amplification of propagons (see Figure 2.2). In the first phase, yeast cells are treated with GdnHCl until the aggregates have sufficiently diluted. In this first phase, as aggregates present at the beginning of GdnHCl exposure will exist for all time, the expected number of aggregates per cell will continue to decrease as cell division continues as normal. In time, cells in the colony will contain a very low number of propagons per cell, ideally one. In the second phase, the resulting cells from phase one are transferred to a GdnHCl free environment and are allowed to form individual colonies. The number of white colonies is then assumed to correspond exactly to the number of cells with with at least one propagon (see Figure 2.2). This is because the resulting colony whose founding cell had at least one propagon will have the $[PSI^+]$ prion phenotype (white) while those founded by a cell with no propagons will have the $[psi^-]$ non-prion phenotype (red).

2.3 Methods

In this section we begin by presenting our model, a system of partial differential equations (PDEs) for the intracellular process of propagon replication and their transmission through the cellular process of division. We present intermediate quantities that allow the decoupling of the PDE system and present the explicit solutions to the model that we consider in this work. Then we show that the intermediate quantities used to derive explicit solutions facilitate a likelihood formulation for parameter estimation and model selection. We describe how to generate simulated data with the model solutions and that we can recover the true kinetic paramaters with our likelihood formulation and our implementation of the adpative Metrolopolis (AM) algorithm. We conclude this section by detailing how we overcome numerical issues with the implementation of our likelihood formulation.

2.3.1 Asymmetric Transmission of Propagons Model

We seek to model the number of aggregates of propagons or a single prion variant in a population of actively dividing cells. Let $a(t)$ be the number of propagons a cell

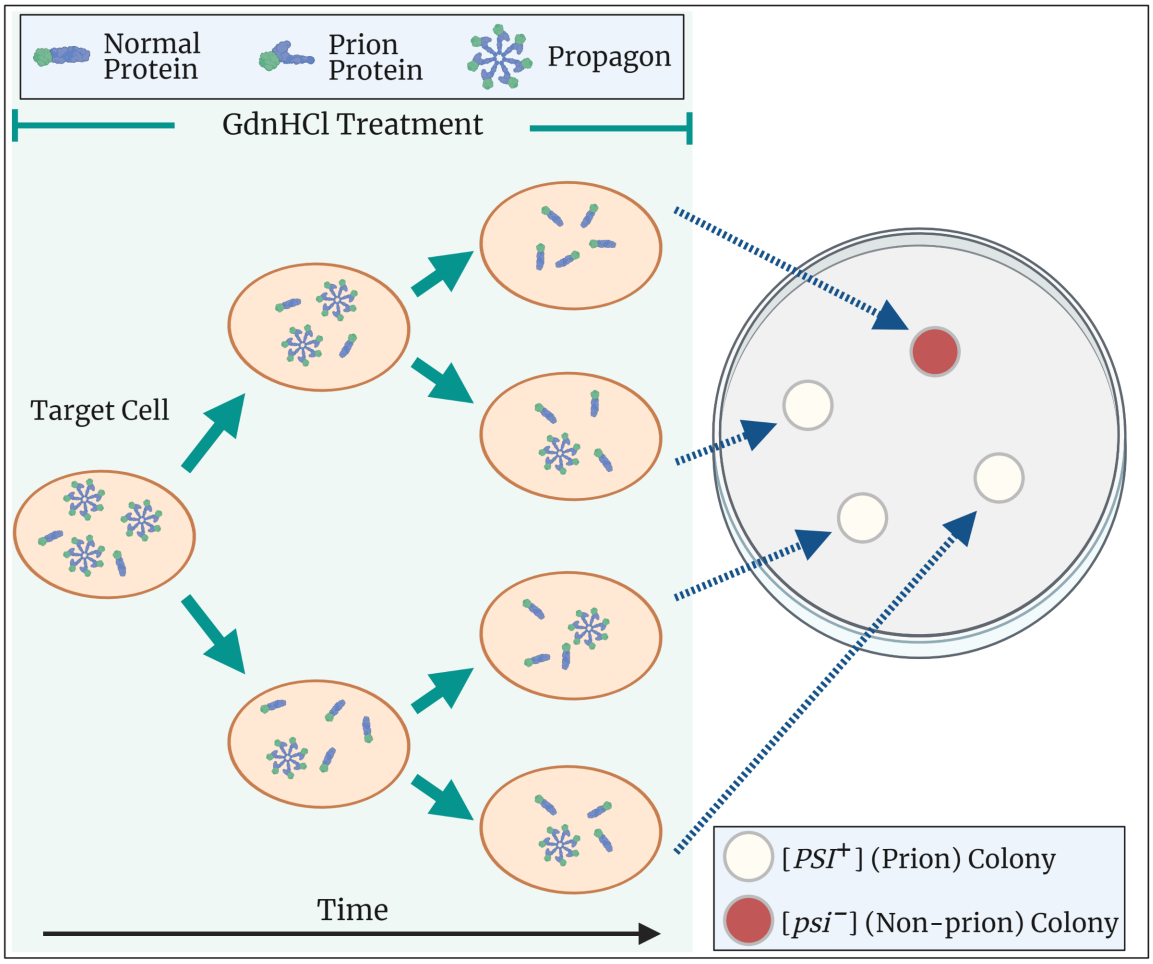


Figure 2.2: **Propagon Amplification Assay**. A two-step process is used to count the number of transmissible prion aggregates (propagons) in a single cell. *Left*: A single target cell is isolated, and aggregate fragmentation is stopped through exposure to GdnHCl. Since aggregates (pinwheels) can not increase in number, they are diluted through cell division (green arrows). *Right*: After sufficient dilution, i.e. each yeast cell is likely to contain at most one aggregate, the colony is replated onto solid media. In the absence of GdnHCl, each single cell serves as a founder of a distinct yeast colony. The number of propagons in the target cell corresponds to the number of white colonies in the plate.

has t hours after dividing, then

$$\frac{da}{dt} = \eta(a; \boldsymbol{\theta}) \tag{2.1}$$

where η is the intracellular propagon amplification model that depends on the current number of propagons a and $\boldsymbol{\theta}$, the kinetic parameter(s) that govern the propagon replication and transmission dynamics (i.e. replication rate and transmission bias). We model the propagon distribution dynamics in the yeast cell population $Y(t, a)$, as evolving in time according to the transport equation:

$$\frac{\partial}{\partial t}Y(t,a) + \frac{\partial}{\partial a}(\eta(a;\boldsymbol{\theta})Y(t,a)) = 0. \tag{2.2}$$

However, we are interested in tracking $Y_i(t, a)$, the distribution of propagons in cells that have undergone i divisions, t hours since the start of the experiment (see Figure 2.3). These propagon dynamics in the population of dividing cells are captured by the following system of $M + 1$ coupled PDEs, which we refer to as the Asymmetric Transmission of Propagons (ATP) model:

$$\begin{aligned}
\frac{\partial}{\partial t}Y_0(t,a) + \frac{\partial}{\partial a}(\eta(a;\boldsymbol{\theta})Y_0(t,a)) &= -(\alpha_0(t) + \beta_0(t))Y_0(t,a), \\
\frac{\partial}{\partial t}Y_1(t,a) + \frac{\partial}{\partial a}(\eta(a;\boldsymbol{\theta})Y_1(t,a)) &= -(\alpha_1(t) + \beta_1(t))Y_1(t,a) + D_1(t,a), \\
&\vdots \\
\frac{\partial}{\partial t}Y_i(t,a) + \frac{\partial}{\partial a}(\eta(a;\boldsymbol{\theta})Y_i(t,a)) &= -(\alpha_i(t) + \beta_i(t))Y_i(t,a) + D_i(t,a), \\
&\vdots \\
\frac{\partial}{\partial t}Y_M(t,a) + \frac{\partial}{\partial a}(\eta(a;\boldsymbol{\theta})Y_M(t,a)) &= -(\alpha_M(t) + \beta_M(t))Y_M(t,a) + D_M(t,a).
\end{aligned} \tag{2.3}$$

The right hand side of each PDE captures the rate of cells lost in that generation from cell division and cell death, and accounts for new cells from cell division in the previous generation. Here $\alpha_i(t)$ is the rate of cell division, and $\beta_i(t)$ the rate of cell death in the i^{th} generation. The rate of increase in the number of cells in the i^{th} generation is represented by the term

$$D_i(t,a) = \rho_1^{-1}\alpha_{i-1}(t)Y_{i-1}(t,\rho_1^{-1}a) + \rho_2^{-1}\alpha_{i-1}(t)Y_{i-1}(t,\rho_2^{-1}a), \tag{2.4}$$

where ρ_1 and ρ_2 control the transmission of propagons between dividing cells. To maintain the conservation of propagons during division we require $\rho_1 + \rho_2 = 1$, showing only one degree of freedom or that if $\rho_1 = \rho \in (0, 1)$, then $\rho_2 = 1 - \rho$. Note that letting $\rho_1 = \rho_2 = 1/\gamma$ reduces our model to that proposed by [66]. Note that this formulation implies that for conservation and symmetric transmission of aggregates during cell division $\rho = 1/\gamma = 1/2$. The work by [83] first generalized the work by

[66] to study T cells traced with carboxyfluorescein diacetate succinimidyl ester to include their resting and cycling phases (delay in division) while considering an asymmetric division with a system of delay hyperbolic PDEs. However, in our modeling framework because of the duration of the experiment (8 hrs), we expect most cell divisions to be the product of mature cells actively producing daughter cells so we do not consider synchronous delays in division.

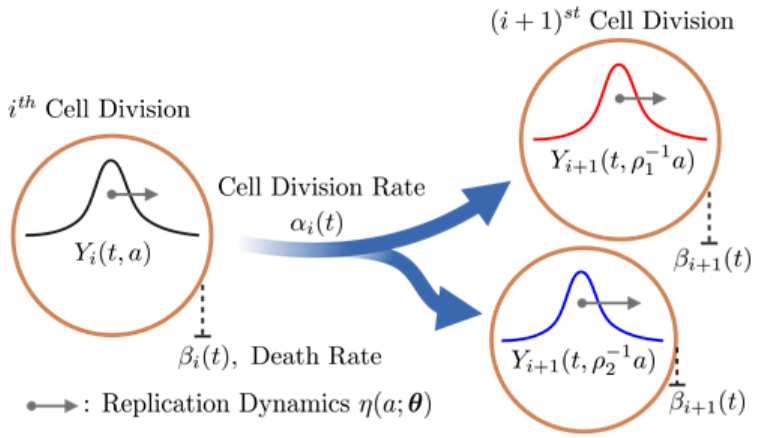


Figure 2.3: **Asymmetric Transmission of Propagons Model Schematic**. The model dynamics of intracellular aggregate replication and cell division from generation i to generation $i+1$. The black arrow ($\longrightarrow$) illustrates the intracellular increase in the number of propagons over time. The remaining parameters are detailed in Section 2.3.1.

In this work, we are interested in determining the effect of ρ, the propagon transmission biased between dividing cells. To this end, we can bound the number of cell divisions (M) because the duration of the experiment of interest is finite. We bound the maximum number of cell divisions up to $M = 6$ generations because yeast cells divide every 1.5 hours and the longest duration of the propagon recovery experiments considered in this work is eight hours. Then to solve the ATP model, Eq. (2.3), we must specify an initial distributions for each generation

$$Y_0(0,a) = \Upsilon(a) \text{ and } Y_i(0,a) = 0 \text{ for all } i > 0. \tag{2.5}$$

Where $\Upsilon(a)$, is the initial intracellular distribution of aggregates at the start of the experiment.

2.3.2 Analytic Solutions and Model Decomposition

The formulation of the ATP model, Eqs.(2.3)-(2.5) presented in Section 2.3.1, allows for computation of intermediate quantities that allow for the decoupling of population and intracellular propagon dynamics in Eq. (2.3). First, the total number

of cells resulting from the i^{th} division, since the beginning of the experiment, at time t is given by

$$n_i(t) = \int_R Y_i(t,a)da, \tag{2.6}$$

where R depends on the intracellular propagon amplification dynamics model in Eq. (2.1), and is taken to be $R = (0, \infty)$ in our work (see Corollary 1). The normalized density of propagons can then be defined by

$$y_i(t,a) = \frac{Y_i(t,a)}{n_i(t)}, \text{ for } n_i(t) > 0, \tag{2.7}$$

and $y_i(t,a) = 0$ otherwise. This quantity represents the intracellular dynamics of the dividing cells in our population of interest. We define the initial number of cells at the start of the experiment by

$$N_0 = \int_R Y_0(0,a)da, \tag{2.8}$$

and the initial normalized propagon density

$$y_0(0,a) = \frac{Y_0(0,a)}{N_0}. \tag{2.9}$$

Using the quantities defined above and Eq. (2.3), the following theorem holds for Eq. (2.1) defined by an exponential growth model $\eta(a;\boldsymbol{\theta}) = \lambda a$, where $a \in R = (0, \infty)$.

Theorem 1. *The solution of the system defined by Eq.* (2.3) *and* (2.4)*, with initial conditions given by Eq.* (2.5) *is given by:*

$$Y_i(t,a) = n_i(t)y_i(t,a), \text{ for } 0 \le i \le M, \tag{2.10}$$

in which:

1. *$n_i(t)$ is the solution of the system of ODEs:*

$$\begin{aligned} i = 0: \quad & \frac{dn_0}{dt} = -(\alpha_0(t) + \beta_0(t))n_0, \\ \text{for } 1 \le i \le M: \quad & \frac{dn_i}{dt} = -(\alpha_i(t) + \beta_i(t))n_i + 2\alpha_{i-1}(t)n_{i-1}, \end{aligned} \tag{2.11}$$

   where $n_0(0) = N_0$, and $n_i(0) = 0$ for $i \ge 1$.

2. *$y_i(t,a)$ is the solution to the PDE*

$$\frac{\partial y_i(t,a)}{\partial t} + \frac{\partial(\eta(a;\boldsymbol{\theta})y_i(t,a))}{\partial a} = 0 \tag{2.12}$$

   with initial conditions $y_i(0,a) := \left(\frac{1}{2}\right)^i \sum_{k=0}^{i} \binom{i}{k} \rho_1^{k-i}\rho_2^{-k} \Upsilon\left(\rho_1^{k-i}\rho_2^{-k}a\right)$, and

3. the solution $y_i(t,a)$ satisfies the recursive property

$$y_i(t,a) = \frac{1}{2}\left(\rho_1^{-1} y_{i-1}(t,\rho_1^{-1}a) + \rho_2^{-1} y_{i-1}(t,\rho_2^{-1}a)\right), \tag{2.13}$$

for all $0 \leq i \leq M$.

Unlike the theorems presented in [66] and [83], in Theorem 1 we highlight the fact that $y_i(t,a)$ must satisfy Eq. (2.12) and Eq. (2.13), in order for Eq. (2.10) to be a solution to the system defined by Eq. (2.3). This is to highlight that for non-linear intracellular dynamics models (Eq. (2.1)), the property given by Eq. (2.13) may no longer be satisfied [120]. The proof of Theorem 1 follows the structure of the proof in [66], therefore we simply outline the proof of Theorem 1 here.

Proof. First, we substitute the $Y_i(t,a)$ terms in Eq. (2.3) with the decomposition given by Eq. (2.10). Then we simplifying the terms of Eq. (2.3), and using Eqs. (2.11) and (2.12) leads to the recursive expression in Eq. (2.13) completing the proof. □

In this work we are modeling propagon replication where at the beginning of the experiment the initial condition is that of a distribution of low propagon counts in the yeast cell population as observed in the propagon recovery data (see Figure 2.5). This initial low number of propagons is followed by a period of exponential growth where a steady state in the number of propagons is not yet observed. Corollary 1 captures this phase of growth, where we assume that the rate of propagon replication is proportional to the current number of propagons present within a yeast cell. This type of propagon proliferation assumes that there is an unlimited amount of normally folded (soluble) protein that can be be misfolded and lead to the continued formation of prion aggregates.

Corollary 1. *The solution of the system defined by Eq.* (2.3) *and* (2.4) *with an intracellular propagon replication model* $\eta(t,a) = \lambda a$*, is*

$$Y_i(t,a) = n_i(t)\sum_{k=0}^{i}\binom{i}{k}{\rho_1}^{k-i}{\rho_2}^{-k}\exp(-\lambda t)\Upsilon\left({\rho_1}^{k-i}{\rho_2}^{-k}a\exp(-\lambda t)\right) \text{ for } 0 \leq i \leq M. \tag{2.14}$$

Where $n_i(t)$ *is the solution to Eq.* (2.11)*.*

Proof. Following [83], we solve Eq. (2.12) via the method of characteristics. This yields

$$y_i(t,a) = \left(\frac{1}{2}\right)^i \sum_{k=0}^{i}\binom{i}{k}{\rho_1}^{k-i}{\rho_2}^{-k}\exp(-\lambda t)\Upsilon\left({\rho_1}^{k-i}{\rho_2}^{-k}a\exp(-\lambda t)\right) \text{ for } 0 \leq i \leq M.$$

To show that $y_i(t,a)$ satisfies the recursive property in Eq. 2.13 we prove the equivalent expression $y_{i+1}(t,a) = \frac{1}{2}\left(\rho_1^{-1} y_i(t,\rho_1^{-1}a) + \rho_2^{-1} y_i(t,\rho_2^{-1}a)\right)$ holds, see Supplementary Section 2.7. Then, replacing $y_i(t,a)$ in Eq. (2.10), and inserting this result in Eq. (2.3) proves Corollary 1. □

The following corollary presented in [66] involves the solution of Eq. (2.11) under a special case of constant cell division rate $\alpha_i(t) = \alpha$ and constant cell death $\beta_i(t) = \beta$. In [66] the authors consider $\beta_i(t) = \beta > 0$, but the solution also holds for $\beta = 0$, which is the assumption in this work.

Corollary 2. *Let $\alpha_i(t) = \alpha \geq 0$ and $\beta_i(t) = \beta \geq 0$, for all $0 \leq i \leq M$, the solution to Eq.* (2.11) *is*

$$n_i(t) = \frac{(2\alpha t)^i}{i!} \exp(-(\alpha + \beta)t) N_0, \quad \textit{for } 0 \leq i \leq M. \tag{2.15}$$

Proof. Straightforward by plugging Eq. (2.15) into Eq. (2.11). □

2.3.3 Likelihood Problem Formulation & Model Selection

In this work we are interested in estimating the kinetic parameters of our model, Eqs.(2.3)-(2.5), using experimental data. We are interested in the probability of the observed data for a given value of the kinetic parameters $\boldsymbol{\theta}$ of the parameter space $\boldsymbol{\Theta}$, denoted $L(\boldsymbol{\theta}|Data)$. Let $N(t) = \sum_{i=0}^{M} n_i(t)$, the total number of cells at time t, $Z(t, a; \boldsymbol{\theta}) = \sum_{i=0}^{M} Y_i(t, a)$ represent the total yeast cell population propagon density after M cell divisions, and $\boldsymbol{\theta}$ represent the kinetic parameters. Further, let the experimental observations a_k observed at time t_k consisting of propagon counts be $\{(a_k, t_k)\}_{k=1}^{m}$ (see Section 2.4). Then, the likelihood of the kinetic parameters $\boldsymbol{\theta}$, given the propagon data is defined as follows

$$\begin{aligned} L(\boldsymbol{\theta}|Data) &= \prod_{k=1}^{m} P(\{t_k, a_k\}; \boldsymbol{\theta}), \\ &= \prod_{k=1}^{m} \left\{ \sum_{i=0}^{M} \left[\left(\frac{n_i(t_k)}{N(t_k)} \right) y_i(t_k, a_k; \boldsymbol{\theta}) \right] \right\}, \\ &= \prod_{k=1}^{m} \sum_{i=0}^{M} \frac{Y_i(t_k, a_k; \boldsymbol{\theta})}{N(t_k)}, \\ &= \prod_{k=1}^{m} \frac{Z(t_k, a_k; \boldsymbol{\theta})}{N(t_k)}. \end{aligned} \tag{2.16}$$

where $P(\{t_k, a_k\}; \boldsymbol{\theta})$ is the probability of observing a_k propagons at time t_k given model parameters $\boldsymbol{\theta}$. This probability is the product of $\frac{n_i(t_k)}{N(t_k)}$, the probability that the observation a_k came from a cell in the i^{th} cell division and $y_i(t_k, a_k; \boldsymbol{\theta})$, the probability that the cell in the i^{th} cell division at time t_k contains a_k propagons. The product of both terms is summed over the number of divisions that can be observed during the experiment. This is because a cell at any point in the experiment results from a finite number of cell divisions since the beginning of the experiment.

In this work we are interested in considering two mathematical modeling scenarios. The first $Z_S(t, a; \lambda, \rho = 0.5)$, symmetric cell division where we estimate the posterior

distribution of the replication rate (λ) and fix the division bias at $\rho = 0.5$, and the second $Z_A(t, a; \lambda, \rho)$, asymmetric cell division where we estimate the posterior distributions of the replication rate λ and the posterior distribution of the propagon division bias ρ. With the likelihood formulation we compute the Akaike Information Criterion (AIC)

$$\text{AIC} = -2 \cdot \log\left(L(\hat{\boldsymbol{\theta}}|Data)\right) + 2K,$$

where $\hat{\boldsymbol{\theta}}$ are the mean parameter estimates that best explain the data (see Section 2.3.4) and K is the number of free parameters [2]. To account for bias due to different number of parameters and data sizes, we use the AIC bias correction

$$\text{AIC}^c = \text{AIC} + \frac{2K(K+1)}{m-K-1},$$

where m is the sample size [69]. These AIC^c values are used to compute the AIC^c difference

$$\Delta_i = \text{AIC}_i^c - \min(\text{AIC}^c),$$

for the i^{th} model and prion variant data. Finally, we can compute the relative model weights

$$\text{W}_i = \frac{\exp(-\frac{1}{2}\Delta_i)}{\sum_{m=1}^{M}\exp(-\frac{1}{2}\Delta_m)},$$

of each model [111] with each prion variant dataset. We interpret this quantity as the probability that a model is the best approximation to the replication and division of propagons during cellular proliferation given the experimental data.

2.3.4 Adaptive Metropolis Algorithm

In this work we follow the procedure described in [128], a Metropolis algorithm with an adaptive Metropolis (AM) step to estimate the target distribution $\pi(\boldsymbol{\theta})$, with p kinetic parameter variables $\boldsymbol{\Theta}$. Using an initial value for each kinetic parameter $\boldsymbol{\theta}_0$, as the starting condition, a random candidate $\boldsymbol{\theta}_{new}$ is drawn from a proposal distribution J of the parameters $\boldsymbol{\theta}$. Thus $\boldsymbol{\theta}_{new} \sim J(\boldsymbol{\theta}_{new}|\boldsymbol{\theta}_{i-1})$, is drawn in every iteration.

The target distribution $\pi(\boldsymbol{\theta})$ is given using previously defined likelihood function $L(\boldsymbol{\theta}|Data)$,

$$\pi(\boldsymbol{\theta}) = \frac{L(\boldsymbol{\theta}|Data)\pi_0(\boldsymbol{\theta})}{\pi(Data)}, \tag{2.17}$$

where the random parameter variables $\boldsymbol{\Theta}$ have a known and possibly uninformative prior density $\pi_0(\boldsymbol{\theta})$, and $\pi(Data) = \int_{\mathbb{R}^p} L(\boldsymbol{\theta}|Data)\pi_0(\boldsymbol{\theta})d\boldsymbol{\theta}$, and p is the dimension of the parameter set. In this work we choose the noninformative prior $\pi_0(\boldsymbol{\theta}) = U(0,1)^p$ and $J(\boldsymbol{\theta}_{new}|\boldsymbol{\theta}_{i-1}) = N(\boldsymbol{\theta}_{i-1}; V)$ to be normally distributed with covariance matrix V. The acceptance probability then follows

$$\alpha(\boldsymbol{\theta}_{new}|\boldsymbol{\theta}_{i-1}) = \min\left\{1, \frac{\pi(\boldsymbol{\theta}_{new})J(\boldsymbol{\theta}_{i-1}|\boldsymbol{\theta}_{new})}{\pi(\boldsymbol{\theta}_{i-1})J(\boldsymbol{\theta}_{new}|\boldsymbol{\theta}_{i-1})}\right\} = \min\left\{1, \frac{L(\boldsymbol{\theta}_{new}|Data)}{L(\boldsymbol{\theta}_{i-1}|Data)}\right\}. \tag{2.18}$$

In Eq. (2.18), the terms involving J cancel because by design J is a symmetric proposal distribution. Now, with probability $\alpha(\boldsymbol{\theta}_{new}|\boldsymbol{\theta}_{i-1})$, we accept $\boldsymbol{\theta}_{new}$, and set $\boldsymbol{\theta}_i = \boldsymbol{\theta}_{new}$. Otherwise, with probability $1 - \alpha(\boldsymbol{\theta}_{new}|\boldsymbol{\theta}_{i-1})$, we set $\boldsymbol{\theta}_i := \boldsymbol{\theta}_{i-1}$.

Following [63], the adaptive step is preceded by a non-adaptive period of length k where $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_k$ are computed using an initial covariant matrix $V_0 = V$. Following the non-adaptive step, the covariance matrix is computed using the previous chain values

$$V_i = s_p \text{cov}\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_i\right) + \varepsilon I_p, \text{ for } i \geq k. \tag{2.19}$$

Here s_p is a design parameter that depends on p. The term εI_p consists of the p-dimensional identity matrix and $\varepsilon \geq 0$ to ensure that V_i remains positive definite. This formulation can quickly incur large computational cost, but this cost can be drastically reduced by use of the recursive update of the covariance [63]

$$V_i = \frac{i-2}{i-1} V_{i-1} + \frac{s_p}{i-1}\left((i-1)\hat{\boldsymbol{\theta}}_{i-2}\hat{\boldsymbol{\theta}}_{i-2}^T - i\hat{\boldsymbol{\theta}}_{(i-1)}\hat{\boldsymbol{\theta}}_{(i-1)}^T + \boldsymbol{\theta}_{(i-1)}\boldsymbol{\theta}_{(i-1)}^T\right). \tag{2.20}$$

In the application of the adaptive Metropolis algorithm to the simulated data and experimental data (Sections 2.4.1 and 2.4.2), we take $s_p = 1.0$, $\varepsilon = 1 \times 10^{-6}$, and

$$V_0 = \begin{bmatrix} \sigma_{\lambda\lambda}^2 & \sigma_{\lambda\rho}^2 \\ \sigma_{\rho\lambda}^2 & \sigma_{\rho\rho}^2 \end{bmatrix},$$

where $\sigma_{\lambda\lambda}^2 = 0.1406$, $\sigma_{\lambda\rho}^2 = \sigma_{\rho\lambda}^2 = 0$, and $\sigma_{\rho\rho}^2 = 0.0156$. A non-adaptive period length of $k = 1500$ for the simulated data and experimental data were established using Geweke's convergence diagnostic as presented in [37], and performed a total of $2{\times}10^6$ iterations. We avoid numerical issues in the direct evaluation of the likelihood formulation presented in Eq. (2.16) by working with natural-logarithm version of the acceptance probability in Eq. (2.18) using the numerically stable procedure outlined in Section 2.3.5.

Algorithm 1 Adaptive Metropolis Algorithm

$D \leftarrow$ load data
$L \leftarrow$ load likelihood formulation
$M \leftarrow$ load Maximum number of iterations
$k \leftarrow$ load non-adaptive period length
$\boldsymbol{\theta}^{(0)} \leftarrow$ initial value for $\boldsymbol{\theta}$
$\boldsymbol{V}^{(0:k)} \leftarrow$ initial covariance matrix values for non-adaptive period V_0
$\texttt{sp} \leftarrow$ set covariance scaling design parameter s_p
$\texttt{ep} \leftarrow$ set identity scaling ε for positive definite V
for $i = 1$ **to** k **do** ▷ Non-adaptive period
    $\boldsymbol{\theta}_{new} \leftarrow N(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{V}^{(i-1)})$
    $u \sim U(0,1)$
    $\alpha \leftarrow \min\left\{1, L(\boldsymbol{\theta}_{new}|D)/L(\boldsymbol{\theta}^{(i-1)}|D)\right\}$
    if $u < \alpha$ **then**
        $\boldsymbol{\theta}^{(i)} \leftarrow \boldsymbol{\theta}_{new}$
    else
        $\boldsymbol{\theta}^{(i)} \leftarrow \boldsymbol{\theta}^{(i-1)}$
for $i = k+1$ **to** M **do** ▷ Adaptive period
    $\boldsymbol{\theta}_{new} \leftarrow N(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{V}^{(i-1)})$
    $u \sim U(0,1)$
    $\alpha \leftarrow \min\left\{1, L(\boldsymbol{\theta}_{new}|D)/L(\boldsymbol{\theta}^{(i-1)}|D)\right\}$
    if $u < \alpha$ **then**
        $\boldsymbol{\theta}^{(i)} \leftarrow \boldsymbol{\theta}_{new}$
    else
        $\boldsymbol{\theta}^{(i)} \leftarrow \boldsymbol{\theta}^{(i-1)}$
    $\boldsymbol{V}^{(i)} = \texttt{sp} * \texttt{cov}\left(\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(i)}\right) + \texttt{ep} * I_p$
return $\boldsymbol{\theta}, V$

2.3.5 Numerical Implementation of Likelihood Formulation

In our initial implementation of the likelihood formulation presented in Section 2.3.3, we encountered significant numerical underflow, so we opted to work with the natural-logarithm form of the likelihood or log-likelihood. Recalling that $N(t) = \sum_{i=0}^{M} n_i(t)$ and that $Z(t) = \sum_{i=0}^{M} Y_i(t,a)$, taking the natural logarithm of Eq. (2.16), we have

$$\begin{aligned} \ln L(\boldsymbol{\theta}|Data) &= \sum_{k=1}^{m} \ln\left(Z(t,a;\boldsymbol{\theta})\right) - \ln\left(N(t)\right), \\ &= \sum_{k=1}^{m} \ln\left(\sum_{i=0}^{M} Y_i(t_k, a_k; \boldsymbol{\theta})\right) - \ln\left(\sum_{i=0}^{M} n_i(t_k)\right). \end{aligned} \tag{2.21}$$

Numerical evaluation of this log-likelihood is performed by generalizing the algebraic and numerically stable property $\ln(a_0+a_1) = \ln(a_0)+\ln(1+\exp(\ln(a_1)-\ln(a_0)))$ where $\ln(a_0) > \ln(a_1)$. That is we evaluate $\ln\left(\sum_{i=0}^{M} Y_i(t_k, a_k; \boldsymbol{\theta})\right)$ by applying the natural log directly to each i^{th} model solution $Y_i(t, a; \boldsymbol{\theta})$, then evaluating the resulting expression. The evaluation of $\ln\left(\sum_{i=0}^{M} Y_i(t_k, a_k; \boldsymbol{\theta})\right)$ is then found recursively using the $\ln\left(Y_i(t_k, a_k; \boldsymbol{\theta})\right)$ terms as follows: let $\ln q_i = \ln Y_i(t, a; \boldsymbol{\theta})\big|_{t=t_k, a=a_k}$, and $\boldsymbol{q} = \{\ln q_j\}_{j=0}^{M}$ be the sorted values $\ln q_i$ in descending order, then $\ln\left(\sum_{i=0}^{M} q_i\right)$ is given by the numerically stable recursive function ξ as follows

$$\xi(\boldsymbol{q}) = \begin{cases} \boldsymbol{q}(1) & |\boldsymbol{q}| = 1 \\ \boldsymbol{q}(1) + \ln(1 + \exp(\xi(\boldsymbol{q}(2:|\boldsymbol{q}|)) - \boldsymbol{q}(1))) & |\boldsymbol{q}| > 1 \end{cases} \tag{2.22}$$

where $|\cdot|$ is the cardinality or the number of elements in $\boldsymbol{q}$ and the notation "$2:|\boldsymbol{q}|$" indicates all elements in $\boldsymbol{q}$ except the first. Implementing this formulation removed numerical underflow when evaluating the likelihood formulation presented in Section 2.3.3.

2.4 Results

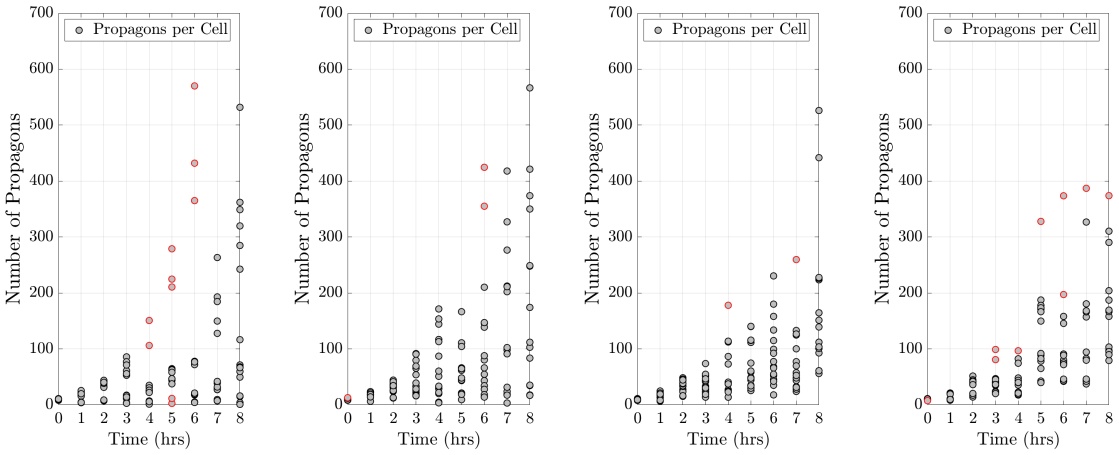
In this section we demonstrate that we can recover known parameters from simulated data using the adaptive Metropolis (AM) algorithm with our likelihood formulation, and study the effects of hourly sampling rate. We then apply these parameter estimation methods to experimental data for six prion variants and perform model selection between symmetric cellular division and asymmetric division of intracellular constituents.

2.4.1 Parameter Inference on Simulated Data

We first verify the capability of our likelihood formulation from Section 2.3.3 and the AM algorithm detailed in Section 2.3.4 to estimate known kinetic parameter values from simulated data before applying our methods to the experimental data. The simulated datasets are created using rejection sampling methods [28] on the ATP model presented in Section 2.3.2. We take the cell division rates $\alpha_i(t)$, to be constant during each period of cell division, with a rate of $\alpha_i(t) = 0.46$ hr^{-1}, and that death is negligible throughout the duration of the experiment by setting $\beta_i(t) = 0$. Furthermore, we assume the initial distribution of aggregates to be a truncated normal distribution defined on the interval $R = (0, \infty)$ with $\mu = 10$ and $\sigma = 1$, that is $\Upsilon(a) = \mathcal{N}^+(a; \mu = 10, \sigma = 1)$, to simulate a colony with low number of aggregates at the beginning of the experiment.

In our investigation we considered the effects of replication rates, division biases, and sampling rates on our ability of recover known kinetic parameter values. Figure

2.4 shows four examples of simulated data produced by rejection sampling in the cases of four intracellular constituents biases $\rho = 0.20, 0.30, 0.40, 0.50$, a replication rate of $\lambda = 0.70$, and a sampling rate of 16 samples per hour.



(a) $\rho = 0.20$ (b) $\rho = 0.30$ (c) $\rho = 0.40$ (d) $\rho = 0.50$

Figure 2.4: **Simulated Data.** The simulated data was generated using a replication rate of $\lambda = 0.70$ and four different division biases (ρ). Samples were generated per experimental hour at a rate of 16 samples per hour. Data points outlined in red were determined to be outliers by the IQR method. Further details are given in Section 2.4.1.

Parameter estimations are made using the likelihood formulation presented in Section 2.3.3 with the AM algorithm outlined in Section 2.3.4. We generate simulated data with three sampling rates: 8, 16, and 32 samples per hour, with replication rates $\lambda = 0.5, 0.7, 0.9$ and division bias $\rho = 0.2, 0.3, 0.4, 0.5$, for 36 total possible combinations. The sampling rates were chosen to reflect the experimental data which contain approximately 16 samples per hour per dataset. To robustly assess the effects of the sampling rates on our ability to recover known parameter choices, we tested our methods on 500 generated datasets for each of the 36 sampling rate and parameter combinations. We compute the mean parameter for each simulated dataset and summarize the distribution of these estimates in the form of 95% credible intervals in Table 2.1.

In all cases we found that we can successfully detect differences in replication rates and asymmetric division biases. As expected, we observed that increasing the sampling rate led to more precise parameter estimates. In the cases where we estimate a symmetric division bias ($\rho = 0.5$), the credible intervals never capture this value. That is due to the fact that we estimate ρ in the interval $[0, 0.5]$ because of the symmetry in our model about the point $\rho = 0.5$ in the interval $[0, 1]$. However, in Table 2.1 we see that as we increase the sampling rate, the estimates become closer to the true value of $\rho = 0.5$. In the experimental data we are interested in removing

outliers using the inner quartile range (IQR) method [31], so we studied the effects of removing possible "outliers" or extreme values using this method from our simulated data to study the effects on our ability to recover the true parameter values. We found that removing such outliers from our simulated data slightly improved our ability to recover the true parameter values (see Supplementary Section 2.7).

$\boldsymbol{\theta}$		8 Samples/Hour		16 Samples/Hour		32 Samples/Hour	
λ	ρ	λ	ρ	λ	ρ	λ	ρ
0.5	0.2	(0.49,0.52)	(0.19,0.21)	(0.50,0.52)	(0.19,0.20)	(0.50,0.51)	(0.20,0.20)
	0.3	(0.49,0.53)	(0.28,0.32)	(0.49,0.51)	(0.29,0.31)	(0.50,0.51)	(0.30,0.31)
	0.4	(0.49,0.51)	(0.38,0.42)	(0.50,0.51)	(0.39,0.41)	(0.50,0.51)	(0.39,0.41)
	0.5	(0.49,0.50)	(0.48,0.49)	(0.50,0.50)	(0.49,0.50)	(0.50,0.50)	(0.49,0.50)
0.7	0.2	(0.69,0.72)	(0.19,0.21)	(0.70,0.72)	(0.19,0.21)	(0.70,0.71)	(0.20,0.20)
	0.3	(0.69,0.73)	(0.28,0.32)	(0.69,0.71)	(0.29,0.31)	(0.69,0.71)	(0.30,0.31)
	0.4	(0.69,0.72)	(0.38,0.42)	(0.70,0.71)	(0.39,0.41)	(0.70,0.71)	(0.39,0.41)
	0.5	(0.69,0.70)	(0.48,0.49)	(0.70,0.70)	(0.48,0.50)	(0.70,0.70)	(0.49,0.50)
0.9	0.2	(0.89,0.92)	(0.19,0.21)	(0.90,0.92)	(0.19,0.21)	(0.90,0.91)	(0.20,0.20)
	0.3	(0.89,0.93)	(0.29,0.32)	(0.89,0.91)	(0.29,0.31)	(0.90,0.91)	(0.30,0.31)
	0.4	(0.89,0.92)	(0.39,0.42)	(0.90,0.91)	(0.39,0.41)	(0.90,0.91)	(0.39,0.41)
	0.5	(0.89,0.91)	(0.48,0.49)	(0.90,0.90)	(0.48,0.50)	(0.90,0.90)	(0.49,0.50)

Table 2.1: **Credible Intervals (95%) for Parameter Estimates on Simulated Data.** The table summarizes the parameter inference results for twelve (λ, ρ) parameter pairs and three sampling rates using data simulated from the ATP model. (see Section 2.3.1 for details).

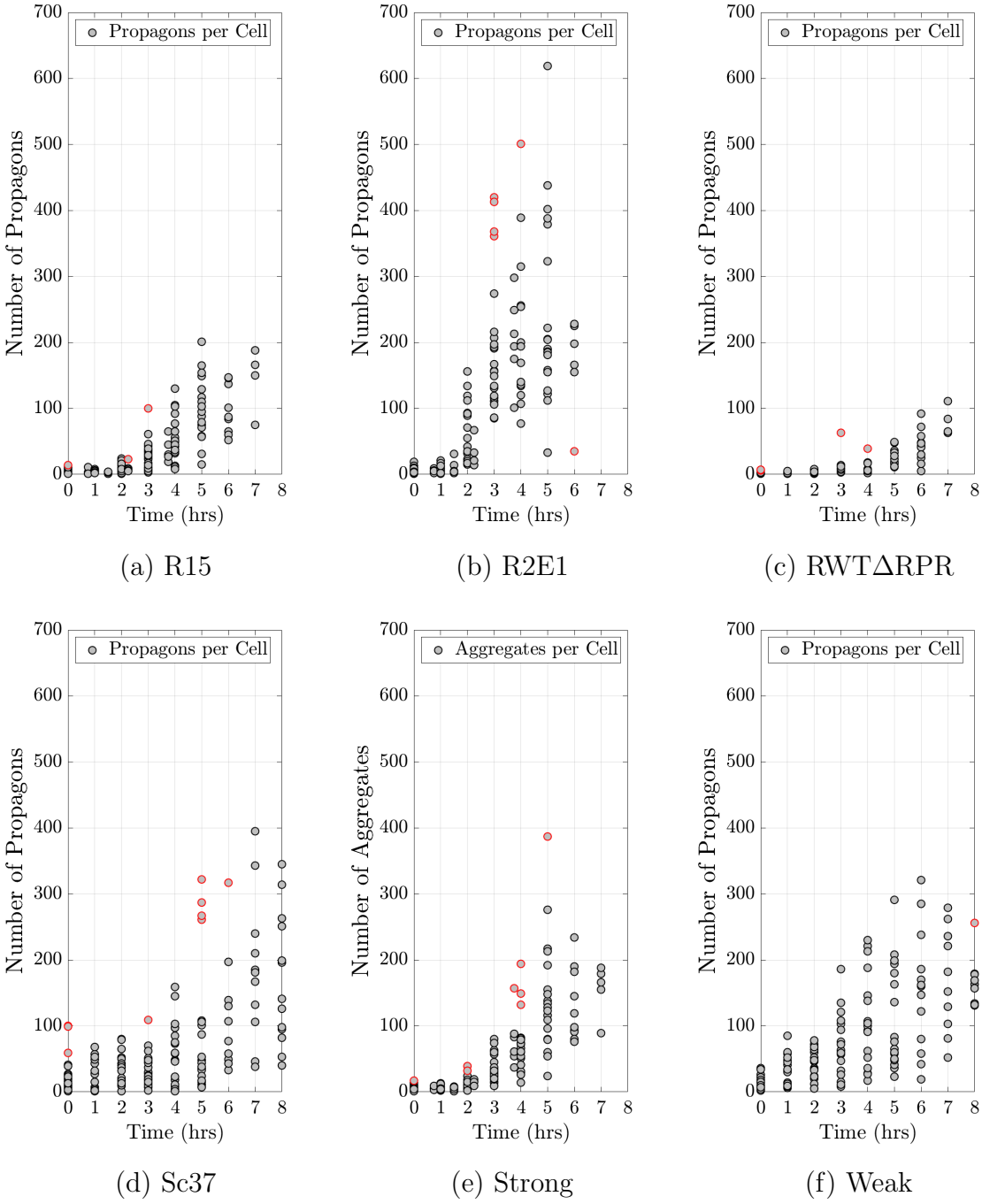
2.4.2 Parameter Inference on Experimental Data

After verifying that our methods allows us to recover known replication rate and division bias parameters from simulated datasets, we apply our methods to experimental data from six prion variants. We consider the aggregate replication experiments for six Sup35 variants that misfold, aggregate, and transmit the $[PSI^+]$ phenotype: Weak, Sc37, Strong [105], RWTΔRPR, R15, and R2E1 [76]. The experimental results from propagon recovery assays for the six prion variants are presented in Figure 2.5. Note that all variants exhibit heteroscedasticity in the number of aggregates in time.

We consider two models for the division of aggregates among dividing yeast cells. First we consider the model $Z_S := Z_S(t, a; \lambda, \rho = 0.5)$, symmetric division of aggregates between dividing cells where our model collapses to the model first proposed by proposed by [66] and we estimate the replication rate λ and fix the division bias at $\rho = 0.5$. Secondly we consider the model $Z_A := Z_A(t, a; \lambda, \rho)$, asymmetric division of prion aggregates where we estimate both the replication rate λ and the aggregate division bias ρ.

The general formulation of our likelihood also allows us to consider the possibility that the start of the exponential phase of prion amplification begins at a time after

the start of the experiment (Figure 2.5). For this, application of the AM algorithm on the experimental data $\{a(t_i)\}_{i=1}^{m}$ is performed by using the data where $t_i \geq T_\Delta$ for $T_\Delta = 0, 1, 2, 3$, and 4 hours into the experiment. Then we apply model selection to determine T_Δ, the point of time into the experiment that best describes the beginning of the exponential phase of aggregate replication.



(a) R15 (b) R2E1 (c) RWTΔRPR

(d) Sc37 (e) Strong (f) Weak

Figure 2.5: **Experimental Propagon Counts for Six Prion Variants.** This experimental data was obtained through propagon recovery experiments (see Section 2.2 for more details). Data points outlined in red were determined to be outliers by the IQR method.

In Table 2.2, we present the replication rate, division bias, the start of exponential growth phase, and the percent model weight for the model, Z_A or Z_S, that best

explains each dataset. Through our analysis we found that a model which includes the cell replication rate and asymmetric division of aggregates during cell division (Z_A), best explains the experimental propagon recovery data for all prion variants (model weights $> 99\%$, Table 2.2). In the case of the Weak prion variant, we found that the dataset where outliers were kept, E_{Weak}, with an asymmetric division model Z_A, best explained the aggregate dynamics for the Weak variant with a model weight of 99.34%. For each variant we detected differences in the replication rates and division bias. However, across prion variants, we detected similar division bias on the order of $\rho \approx 0.30$. For the three prion variants RWTΔRPR, Strong, and R15, we found a delay in the start of the exponential growth phase of aggregate replication. RWTΔRPR was found to have the largest delay ($T_\Delta = 3$ hrs) before the beginning of the exponential growth phase, which was identified from experimental data without the outliers identified through the IQR method. Without the work to detect T_Δ, we underestimate the replication rate for this variant (see Supplemental section 2.7).

Variant	**Rep. Rate (λ, hr^{-1})**	**Trans. Bias (ρ)**	T_Δ	**%W**	**Model**
Weak	0.88(0.87,0.88)	0.27(0.26,0.27)	0*	99.34	Z_A*
Sc37	0.81(0.79,0.82)	0.23(0.22,0.23)	0	100.0	Z_A
RWTΔRPR	0.77(0.72,0.80)	0.31(0.29,0.33)	3	100.0	Z_A
Strong	1.13(1.11,1.16)	0.36(0.34,0.38)	1	100.0	Z_A
R15	1.14(1.10,1.18)	0.32(0.29,0.35)	1	100.0	Z_A
R2E1	1.25(1.22,1.29)	0.29(0.26,0.33)	0	100.0	Z_A

Table 2.2: **Parameter Estimates and Credible Intervals (95%) for Six Prion Variants.** The column labeled T_Δ indicates the point of time into the experiment that best describes the beginning of the exponential phase of aggregate replication. The column labeled $\%W$ presents the percent model weight for Z_A, the model for asymmetric division of aggregates. The asterisk (*) indicates the dataset not filtered for outliers (raw data) was selected.

The parameter estimates, corrected AIC, percent model weights, and the postprocessing of the Metropolis chain iterations for all the cases considered are summarized in Supplemental Section 2.7.

2.5 Discussion and Conclusions

In this work we presented a structured population model that generalizes the work presented in [8, 66], to study propagon replication dynamics and the bias in their transmission among proliferating yeast cells. We developed an inverse problem formulation that consists of an interpretable likelihood formulation descriptive of the propagon recovery experiment. We first verified that we could recover known propagon replication parameters from simulated data, then used our inverse problem formulation to study propagon replication in six prion variants. Additionally, the

likelihood formulation allowed us to consider the presence of influential outliers and if there was a delay in observing the exponential phase of propagon replication in the experimental data.

We were able to detect differences in propagon replication rates and bias in their transmission during cell division among the six prion variants that we studied. We found that the Weak prion variant had a lower replication rate and transmission bias, $\lambda = 0.88$ and $\rho = 0.27$ respectively, when compared to the Strong prion variant which had a higher replication rate and transmission bias, $\lambda = 1.13$ and $\rho = 0.36$ respectively. These findings are consistent with previous studies of propagon size [138, 74], and fragmentation rates [49, 133] of Weak and Strong aggregates, where Weak aggregates are larger and replicate at a lower rate than Strong aggregates, thus it would be reasonable to observe a higher rate of propagon replication and transmission bias in Strong than in Weak.

We would expect the transmission bias to be proportional to the relative volume of the daughter cell to that of the mother and daughter volumes or $\rho \approx V_D/(V_M+V_D)$. From Table 2.3 adapted from [148], these ratios are generation dependent. However, our estimates for the transmission biases remain close to these generational dependent values. Therefore, it is possible that the transmission bias values we are capturing are a combination of these generation volume proportions and diffusion effects due to propagon aggregate size. In future models, we will need to incorporate differences in daughter volumes to discern transmission biases from daughter cell volume size based bias.

Generation	$\boldsymbol{G_1}$	$\boldsymbol{G_2}$	$\boldsymbol{G_3}$	$\boldsymbol{G_4}$
$V_D/(V_D + V_M)$	0.40	0.32	0.29	0.25

Table 2.3: **Relative Volume of Daughter Cells by Generation.** Where V_D is the daughter volume, V_M is the mother volume, and G_i indicate a cell in the i^{th} generation. This table was adapted from Table 2 in [148].

To our knowledge our work is the first to use propagon amplification assays and a structured population model (ATP model) to recover both the propagon replication rate and the transmission bias during cell division. The work in [8] uses propagon amplification assays and a structured population model that assumes symmetric transmission of propagon between dividing cells to determine propagon replication rates for two yeast variants but did not directly address the asymmetric bias in propagon transmission during cell division. Separate work presented in [26] has used curing experiments, where the application of GdnHCl causes inhibition of propagon replication and leads to elimination of the $[PSI^+]$ phenotype through dilution by cell division and a model that captures asymmetric cell division through unequal transmission of propagons through cell division [34], are able to recover the propagon transmission bias, considered the probability of propagon transmission to a budding daughter cell in their work. The curing experiment eliminates $[PSI^+]$ phenotype and as such, a propagon replication rate cannot be computed. We note that our model can also be

used to calculate the asymmetric transmission bias using data from curing experiments by setting the replication rate equal to zero ($\lambda = 0$). The number of propagon will decrease through asymmetric transmission of aggregates between dividing cells.

In this work we detail the importance of the propagon replication rate λ, the asymmetric propagon transmission bias ρ and detection of a delay T_Δ, before observing the exponential growth phase of prion replication, to explain prion replication in a proliferating yeast cell colony. In future work we intend to build on this framework by incorporating cell maturation, a state during which cells can grow but cannot divide until they are fully mature. We have made an effort to be clear about the difference between *asymmetric cell division* and *asymmetric transmission of intracellular constituents*, but in future models we plan to incorporate cell volume into our models, so that we can distinguish between effects due to volume and those due to transmission bias. Also, in this work our intracellular aggregate replication model assumed exponential growth because a clear steady state, or carrying capacity, was not observed in the experimental data, but perhaps by running the propagon amplification assays over a longer period of time we would begin to observe such a steady state. This would require the application of a more complex intracellular model to capture such a carrying capacity and would require methods such as those presented in [120] to numerically solve such a propagon replication model. As previously mentioned, another possible application with our modeling framework that we have not yet considered is the possibility of its application to study curing experiments.

2.6 Acknowledgments

Funding

This research was funded in part by the National Science Foundation (Award No. DMS-1840265), and the Joint Initiative to Support Research at the Interface of the Biological and Mathematical Sciences between the National Science Foundation Division of Mathematical Sciences (NSF-DMS) and National Institutes of Health National Institute of General Medical Sciences (NIH-NIGMS) (Grant No. R01-GM126548).

Data

We would like to acknowledge Dr. Tricia Serio for providing us with the experimental data for this investigation and for the insightful discussions regarding this work.

Computing

The Multi-Environment Computer for Exploration and Discovery (MERCED) cluster at UC Merced, funded by National Science Foundation (Grant No. ACI-1429783).

Graphics

Figures 2.1, 2.2 and 2.3 created with `BioRender.com`.

2.7 Supplemental Materials

2.7.1 Recursive Property to Corollary 1

We show that the recursive property of Eq. 2.13, in Theorem 1 re-stated below in an equivalent form

$$y_{i+1}(t,a) = \frac{1}{2}\left(\rho_1^{-1} y_i(t,\rho_1^{-1}a) + \rho_2^{-1} y_i(t,\rho_2^{-1}a)\right), \tag{2.23}$$

is satisfied by the solution

$$y_i(t,a) = \left(\frac{1}{2}\right)^i \sum_{k=0}^{i} \binom{i}{k} {\rho_1}^{k-i} {\rho_2}^{-k} \exp(-\lambda t) \Upsilon\left({\rho_1}^{k-i} {\rho_2}^{-k} a \exp(-\lambda t)\right),$$

in Corollary 1. That is, we show that the left hand side (LHS) of Eq. (2.23) is equal to the right hand side (RHS) of Eq. (2.23) with our solution. First we replace $y_i(t,a)$ in the RHS of the recursive equation with the solution from Corollary 1, resulting in $(s1)$. We write out the first term $(k=0)$ from the first sum and the last term $(k=i)$ in the second sum. The second sum is re-indexed with $k := k-1$ and we combine the two sums, resulting in $(s2)$. Using Pascal's rule

$$\binom{i}{k} + \binom{i}{k-1} = \binom{i+1}{k},$$

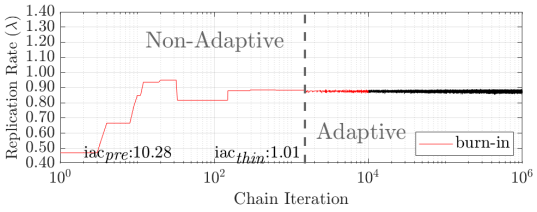
the equivalent expressions $\binom{i}{0} = \binom{i+1}{0}$ and $\binom{i}{i} = \binom{i+1}{i+1}$, results in $(s3)$. The three terms can now be written as one sum $(s4)$ which is the LHS of Eq. (2.23), or $y_{i+1}(t, a)$.

$$
\begin{aligned}
\text{RHS} = \frac{1}{2}\Bigg(& \left(\frac{1}{2}\right)^i \sum_{k=0}^{i} \binom{i}{k} {\rho_1}^{k-i+1} {\rho_2}^{-k} \exp(-\lambda t) \Upsilon\left({\rho_1}^{k-i+1} {\rho_2}^{-k} a \exp(-\lambda t)\right) \\
& + \left(\frac{1}{2}\right)^i \sum_{k=0}^{i} \binom{i}{k} {\rho_1}^{k-i} {\rho_2}^{-k+1} \exp(-\lambda t) \Upsilon\left({\rho_1}^{k-i} {\rho_2}^{-k+1} a \exp(-\lambda t)\right) \Bigg), \qquad (s1) \\
= & \left(\frac{1}{2}\right)^{i+1} \exp(-\lambda t) \Bigg(\binom{i}{0} \rho_1^{-i+1} \Upsilon\left({\rho_1}^{-i+1} a \exp(-\lambda t)\right) \\
& + \sum_{k=1}^{i} \left[\binom{i}{k} + \binom{i}{k-1} \right] \rho_1^{k-i+1} \rho_2^{-k} \Upsilon\left({\rho_1}^{k-i+1} {\rho_2}^{-k} a \exp(-\lambda t)\right) \\
& + \binom{i}{i} \rho_2^{i+1} \Upsilon\left({\rho_2}^{-i+1} a \exp(-\lambda t)\right) \Bigg), \qquad (s2) \\
= & \left(\frac{1}{2}\right)^{i+1} \exp(-\lambda t) \Bigg(\binom{-i+1}{0} \rho_1^{-i+1} \Upsilon\left({\rho_1}^{-i+1} a \exp(-\lambda t)\right) \\
& + \sum_{k=1}^{i} \binom{i+1}{k} \rho_1^{k-i+1} \rho_2^{-k} \Upsilon\left({\rho_1}^{k-i+1} {\rho_2}^{-k} a \exp(-\lambda t)\right) \\
& + \binom{i+1}{i+1} \rho_2^{i+1} \Upsilon\left({\rho_2}^{-i+1} a \exp(-\lambda t)\right) \Bigg), \qquad (s3) \\
= & \left(\frac{1}{2}\right)^{i+1} \sum_{k=0}^{i+1} \binom{i+1}{k} \rho_1^{k-i+1} \rho_2^{-k} \exp(-\lambda t) \Upsilon\left({\rho_1}^{k-i+1} {\rho_2}^{-k} a \exp(-\lambda t)\right), \qquad (s4) \\
= & \, y_{i+1}(t, a), \\
= & \, \text{LHS}.
\end{aligned}
$$

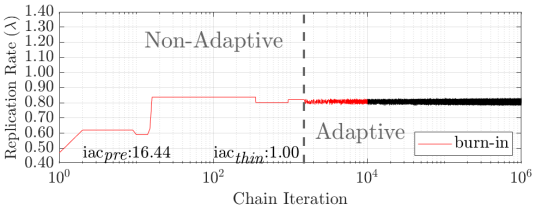
2.7.2 Adaptive Metropolis

In our application of Adaptive Metropolis (AM) detailed in Section 2.3.4, on experimental data discussed in Section 2.4, we performed 2×10^6 iterations with $k = 1500$ non-adaptive steps with the asymmetric cell division model $Z_A(t, a; \lambda, \rho)$ and the symmetric cell division model $Z_S(t, a; \lambda, \rho = 0.5)$. Figures 2.6 and 2.7 illustrate the chain iterations for the models and datasets selected after our model selection procedure. The appearance of increasing variance in the posterior estimates, as the number of iterations increases, is an artifact from plotting the AM algorithm chain iterations on a logscale. To reduce the correlation among the AM estimates, estimates after the burn-in period and after thinning at regular intervals of 50 iterations are used in our work. In this work we use the integrated autocorrelation time (iac) as a measure of autocorrelation across the autocorrelation function (ACF) [117]. Note

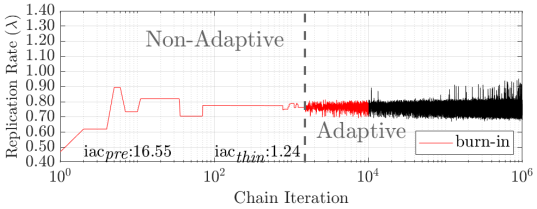
that there are no chain iteration and autocorrelation function figures for the division bias (ρ) under the model $Z_S(t, a; \lambda, \rho = 0.5)$ because in these cases the division bias parameter is fixed at $\rho = 0.5$ (symmetric cell division).



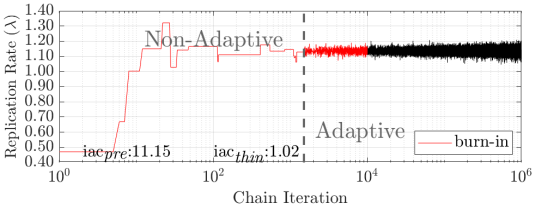
(a) Weak Replication Rate



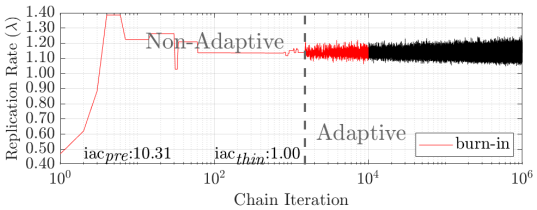
(b) Sc37 Replication Rate



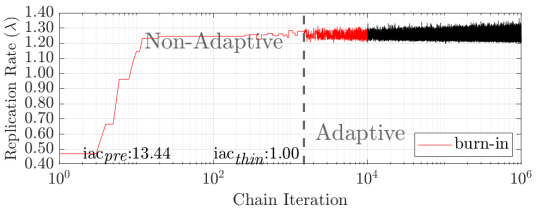
(c) RWTΔRPR Replication Rate



(d) Strong

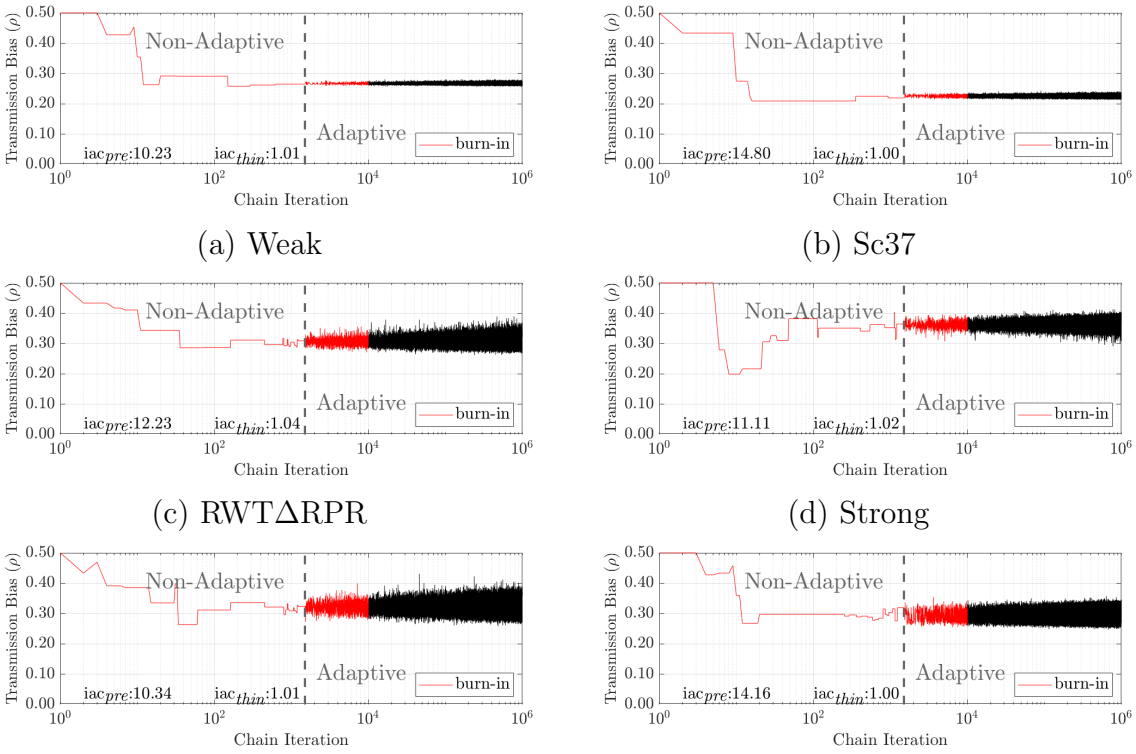


(e) R15



(f) R2E1

Figure 2.6: **Adaptive Metropolis Chain Iterations: Replication Rate.** The non-adaptive and adaptive steps are displayed and the burn-in period is highlighted in red. The iac for each chain is presented for both before thinning and dropping the burn-in period (iac$_{pre}$) and after dropping the burn-in period and thinning (iac$_{thin}$).



(a) Weak (b) Sc37

(c) RWTΔRPR (d) Strong

(e) R15 (f) R2E1

Figure 2.7: **Adaptive Metropolis Chain Iterations: Transmission Bias.** The non-adaptive and adaptive steps are displayed and the burn-in period is highlighted in red. The iac for each chain is presented for both before thinning and dropping the burn-in period (iac$_{pre}$) and after dropping the burn-in period and thinning (iac$_{thin}$).

2.7.2.1 Parameter Estimates: Raw Data

Prion Variant	$T_\Delta = 0$		$T_\Delta = 1$		$T_\Delta = 2$		$T_\Delta = 3$		$T_\Delta = 4$	
	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A
Weak	0.71	0.88	0.56	0.62	0.68	0.81	0.58	0.72	0.44	0.51
Sc37	0.48	0.51	0.55	0.59	0.62	0.63	0.81	0.91	0.81	1.11
RWTΔRPR	0.77	0.80	0.96	0.99	1.08	1.43	0.58	0.62	0.96	0.92
Strong	0.78	0.86	1.09	1.17	1.19	1.33	0.93	1.06	0.85	1.06
R15	0.79	0.84	1.13	1.20	1.17	1.39	0.99	1.12	0.83	0.85
R2E1	1.23	1.26	1.79	2.01	0.92	1.17	0.38	0.53	0.62	0.54

Table 2.4: **Replication Rate (λ) by T_Δ: Raw Data.**

Prion Variant	$T_\Delta = 0$		$T_\Delta = 1$		$T_\Delta = 2$		$T_\Delta = 3$		$T_\Delta = 4$	
	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A
Weak	0.50	0.27	0.50	0.33	0.50	0.32	0.50	0.27	0.50	0.29
Sc37	0.50	0.28	0.50	0.29	0.50	0.26	0.50	0.25	0.50	0.25
RWTΔRPR	0.50	0.36	0.50	0.37	0.50	0.23	0.50	0.28	0.50	0.30
Strong	0.50	0.31	0.50	0.35	0.50	0.32	0.50	0.32	0.50	0.33
R15	0.50	0.30	0.50	0.31	0.50	0.29	0.50	0.33	0.50	0.31
R2E1	0.50	0.27	0.50	0.19	0.50	0.29	0.50	0.33	0.50	0.37

Table 2.5: **Division Bias (ρ) by T_Δ: Raw Data.**

2.7.2.2 Parameter Estimates: Data Filtered for Outliers

Prion Variant	$T_\Delta = 0$		$T_\Delta = 1$		$T_\Delta = 2$		$T_\Delta = 3$		$T_\Delta = 4$	
	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A
Weak	0.71	0.89	0.56	0.62	0.68	0.81	0.58	0.72	0.45	0.51
Sc37	0.65	0.81	0.46	0.55	0.51	0.59	0.67	0.90	0.51	0.63
RWTΔRPR	0.82	0.82	0.89	0.87	0.91	0.86	0.80	0.77	1.06	1.07
Strong	0.78	0.86	1.08	1.13	1.29	1.43	0.73	0.80	0.94	1.12
R15	0.81	0.84	1.09	1.14	1.10	1.19	1.04	1.18	0.83	0.85
R2E1	1.21	1.25	1.73	2.02	0.79	0.87	0.52	0.66	0.71	0.77

Table 2.6: **Replication Rate (λ) by T_Δ: Data Filtered for Outliers.**

Prion Variant	$T_\Delta = 0$		$T_\Delta = 1$		$T_\Delta = 2$		$T_\Delta = 3$		$T_\Delta = 4$	
	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A
Weak	0.50	0.28	0.50	0.33	0.50	0.32	0.50	0.27	0.50	0.29
Sc37	0.50	0.23	0.50	0.30	0.50	0.26	0.50	0.26	0.50	0.33
RWTΔRPR	0.50	0.43	0.50	0.40	0.50	0.32	0.50	0.31	0.50	0.32
Strong	0.50	0.32	0.50	0.36	0.50	0.30	0.50	0.33	0.50	0.31
R15	0.50	0.31	0.50	0.32	0.50	0.31	0.50	0.32	0.50	0.31
R2E1	0.50	0.29	0.50	0.19	0.50	0.33	0.50	0.36	0.50	0.42

Table 2.7: **Division Bias (ρ) by T_Δ: Data Filtered for Outliers.**

2.7.2.3 Parameter Estimates of Simulated Data Filtered for Outliers

θ		8 Samples/Hour		16 Samples/Hour		32 Samples/Hour	
λ	ρ	λ	ρ	λ	ρ	λ	ρ
0.5	0.2	(0.49,0.52)	(0.19,0.21)	(0.49,0.52)	(0.19,0.21)	(0.50,0.51)	(0.19,0.20)
	0.3	(0.49,0.52)	(0.29,0.32)	(0.49,0.51)	(0.29,0.31)	(0.49,0.51)	(0.30,0.31)
	0.4	(0.49,0.52)	(0.38,0.42)	(0.49,0.51)	(0.39,0.41)	(0.50,0.51)	(0.39,0.41)
	0.5	(0.49,0.50)	(0.48,0.49)	(0.50,0.50)	(0.49,0.50)	(0.50,0.50)	(0.49,0.50)
0.7	0.2	(0.69,0.72)	(0.19,0.21)	(0.69,0.72)	(0.19,0.21)	(0.70,0.71)	(0.19,0.20)
	0.3	(0.69,0.72)	(0.28,0.32)	(0.69,0.71)	(0.29,0.31)	(0.69,0.70)	(0.30,0.31)
	0.4	(0.69,0.71)	(0.38,0.42)	(0.69,0.71)	(0.39,0.41)	(0.70,0.71)	(0.39,0.41)
	0.5	(0.69,0.70)	(0.48,0.49)	(0.70,0.70)	(0.48,0.50)	(0.70,0.70)	(0.49,0.50)
0.9	0.2	(0.89,0.92)	(0.19,0.21)	(0.89,0.92)	(0.19,0.21)	(0.90,0.91)	(0.19,0.20)
	0.3	(0.88,0.92)	(0.29,0.32)	(0.89,0.91)	(0.29,0.31)	(0.89,0.90)	(0.30,0.31)
	0.4	(0.89,0.91)	(0.39,0.42)	(0.89,0.91)	(0.39,0.41)	(0.90,0.91)	(0.39,0.41)
	0.5	(0.89,0.90)	(0.48,0.49)	(0.90,0.90)	(0.49,0.50)	(0.90,0.90)	(0.49,0.50)

Table 2.8: **Credible Intervals (95%) for Parameter Estimates of Simulated Data: Filtered for Outliers.** The table summarizes the parameter inference results for twelve (λ, ρ) parameter pairs and three sampling rates using data simulated from the ATP model (see Section 2.3.1 for details).

2.7.3 Model Weights

Tables 2.9 and 2.10 present the model weights for every model and dataset combination using the formulation presented in Section 2.3.3. Note that the model weights for each row and across both tables sum to 100%.

Prion Variant	$T_\Delta = 0$		$T_\Delta = 1$		$T_\Delta = 2$		$T_\Delta = 3$		$T_\Delta = 4$	
	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A
Weak (1/2)	0.00	**99.3**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sc37 (1/2)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RWTΔRPR (1/2)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Strong (1/2)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R15 (1/2)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R2E1 (1/2)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 2.9: **Percent Model Weights by T_Δ: Raw Data.**

Prion Variant	$T_\Delta = 0$		$T_\Delta = 1$		$T_\Delta = 2$		$T_\Delta = 3$		$T_\Delta = 4$	
	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A	Z_S	Z_A
Weak (2/2)	0.00	**0.70**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sc37 (2/2)	0.00	**100**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RWTΔRPR (2/2)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**100**	0.00	0.00
Strong (2/2)	0.00	0.00	0.00	**100**	0.00	0.00	0.00	0.00	0.00	0.00
R15 (2/2)	0.00	0.00	0.00	**100**	0.00	0.00	0.00	0.00	0.00	0.00
R2E1 (2/2)	0.00	**100**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 2.10: **Percent Model Weights by T_Δ:Data Filtered for Outliers.**

Chapter 3

Numerical Approaches to Division and Label Structured Population Models

This chapter originally published as, "Santiago, F., Sindi, S., Flores, K. (2020). Numerical Approaches to Division and Label Structured Population Models. Letters in Biomathematics, 7(1), 153-170." Reprinted in accordance with the Creative Commons Attribution 4.0 International License (`http://creativecommons.org/licenses/by/4.0/`) with minor changes.

3.1 Introduction

Division and label structured population models have recently been investigated in the context of cell proliferation assays. These assays yield insight into the role of cell cycle dynamics [127] and intercellular heterogeneity [57] to the overall growth of a population of cells. A widely used proliferation assay, developed by Lyons and Parish [87], uses the intracellular dye carboxyfluorescein succinimidyl ester (CFSE). CFSE is nonradioactive and enables a durable and sufficiently uniform labeling of a population of cells without adverse effects on proliferation and death. Label structured population models are an ideal tool for understanding these type of data because the CFSE dye is modeled as a label that is approximately partitioned into two equal halves when cells divide into two daughter cells. Flow cytometry can be used to rapidly measure CFSE fluorescence intensity for each cell in a population [87, 86, 103, 116, 141, 146], resulting in a population histogram. Peaks in the histogram are typically interpreted to represent a separate generation of cells, i.e., the number of times a subpopulation of cells has divided.

Prior to the application of division and structured population models to CFSE data [11, 9, 84, 85], the number of cells in each generation was approximated using simple binning or deconvolution techniques [45, 46, 47, 55, 56]. CFSE labeling has been used to study immune system dynamics through monitoring lymphocytes pro-

liferation [87], and is also able to track the proliferation behavior of specific types of lymphocytes through the use of fluorescent antibodies with specificity to markers on the lymphocyte cell surface [86]. The typical response of various cell populations within the immune system is to clonally expand when presented with foreign antigens. Thus, being able to more accurately quantify these responses, e.g., by estimating rates of activation, proliferation, and death through coupling structured population models with data, will lead to a more complete characterization of how various diseases and stimuli affect the immune system. With increasing advances in the ability to quantify the contents in single-cells [91], such structured population models will become more applicable.

Previous efforts to fit CFSE label decay experimental data with division and label structured population models used a least squares inverse problem framework [84, 85, 11, 9]. The inverse problem formulation typically involves estimating parameters by minimizing a least squares cost function that measures the discrepancy between a solution of the PDE system at a set of parameter values and the histogram data from flow cytometry. Optimization routines that have been used to solve the inverse problem with flow cytometry data include derivative free methods [85], such as the Nelder-Mead simplex based method, and gradient based methods relying on the computation of a local gradient of the cost function [12]. Both of these optimization methods use an iterative scheme requiring thousands of simulations of the forward solution of the label structured PDE system at different parameter values. In the case where division structure is modeled with separate compartments, i.e., one per cell division cycle, the computational requirements are exacerbated since one needs to solve a large system (>10) of coupled PDEs. Under simplifying assumptions on the form of the CFSE label decay rate, such as decay that is directly proportional to the amount of intracellular label, analytical solutions to the PDEs can be obtained through the method of characteristics. The main drawbacks to these assumptions are that the resulting decay rate functional forms may not accurately explain the data and that they do not extend to other scenarios involving the measurement of an intracellular label such as protein production [24, 54] or prion aggregate amplification [8]. Thus, more efficient numerical methods are needed to enable validation of division and labeled structured population models from data across a broader range of assumptions describing the rate of change of the intracellular label.

In this work, we develop a novel numerical and theoretical framework for a class of division and label structured population models. We develop our framework in the more general context of population of dividing cells that is structured by the concentration of a single intracellular species evolving under its own dynamics (rather than a decaying label). In Section 3.2, we derive a recursive relation allowing the density of the $(i+1)$-generation to be solved in terms of the i-th generation. In Section 3.3 we apply our recursive formulation to three common models for intracellular dynamics. In Section 3.3.3.2 we demonstrate that our recursive framework facilitates rapid and accurate numerical solutions through recursive numerical integration. In Section 3.4 we summarize our findings and discuss their implications in the context of parameter

inference and uncertainty quantification.

3.2 Division and Label Structured Population

We seek to model the concentration of a single intercellular constituent in a population of actively dividing cells. Let $a(t)$ be the concentration of an intracellular species in a cell t hours after the start of an experiment. In the absence of division, the intracellular dynamics are governed by the following ordinary differential equation

$$\frac{da}{dt} = \eta(a; \boldsymbol{\theta}), \tag{3.1}$$

where η is assumed to be a known smooth function of possibly unknown parameters $\boldsymbol{\theta}$. When cells divide, the intracellular constituent is divided between daughter cells. Let $Y_i(t, a)$ represents the constituent density of cells that have undergone i cell divisions since the beginning of the experiment. The dynamics of each generation in the dividing cellular population evolves according to the following system of coupled partial differential equations:

$$\begin{aligned}
\frac{\partial}{\partial t}Y_0(t,a) + \frac{\partial}{\partial a}(\eta(a;\boldsymbol{\theta})Y_0(t,a)) &= -(\alpha_0(t)+\beta_0(t))Y_0(t,a),\\
\frac{\partial}{\partial t}Y_1(t,a) + \frac{\partial}{\partial a}(\eta(a;\boldsymbol{\theta})Y_1(t,a)) &= -(\alpha_1(t)+\beta_1(t))Y_1(t,a) + 2\gamma\alpha_0(t)Y_0(t,\gamma a),\\
&\vdots\\
\frac{\partial}{\partial t}Y_M(t,a) + \frac{\partial}{\partial a}(\eta(a;\boldsymbol{\theta})Y_M(t,a)) &= -(\alpha_M(t)+\beta_M(t))Y_M(t,a) + 2\gamma\alpha_{i-1}(t)Y_{i-1}(t,\gamma a),
\end{aligned} \tag{3.2}$$

where $\alpha_i(t)$ and $\beta_i(t)$ represent the division and death rates respectively. As mentioned above i represents the number of divisions since the beginning of the experiment (not the generational age) and so our problem formulation includes the following initial conditions:

$$Y_0(0,a) = \Upsilon(a), \text{ and } Y_i(0,a) = 0 \text{ for } i \geq 1. \tag{3.3}$$

Where $\Upsilon(a)$ is the initial constituent density of cells at the start of an experiment. Following previous convention we refer to this system as the Division and Label Structured Population Model (DLSPM).

In this study, we derive solutions to the DLSPM (Eqs. (3.2) & (3.3)) through recursive integration of the previous generation. The decomposition approach for finding solutions to $Y_i(t, a)$ in terms of $\Upsilon(a)$, used in Theorem 2.1 of [10] and first presented in [66], is valid only for functions of the form $\eta(a; \boldsymbol{\theta}) = v(t; \boldsymbol{\theta})a$. Theorem 2 stated below is valid for a broader set of flux functions, namely those which allow for solution via the method of characteristics, and whose characteristics are invertible. This includes the flux functions considered by [66].

Theorem 2. *The solution of the system defined by Eq.* (3.2) *with initial conditions in Eq.* (3.3)*, is given by an analytic solution and a recursive integral equation:*

$$\begin{aligned} Y_0(t,a) &= \Upsilon(s(t,a))\mu_0(t,s(t,a))^{-1}, \text{ and} \\ Y_i(t,a) &= \frac{2\gamma\alpha_{i-1}(t)}{\mu_i(t,s(t,a))}\int_0^t \mu_i(\tau,s(t,a))Y_{i-1}(\tau,\gamma a(\tau,s(t,a)))d\tau, \text{ for } 1\le i\le M, \end{aligned} \tag{3.4}$$

in which:

1. *$s(t,a)$ comes from solving the characteristic equation:*

$$\frac{da}{dt} = \eta(a;\boldsymbol{\theta}), \tag{3.5}$$

   with the parametric initial condition $a(0,s)=s$, and

2. *$\mu_i(t,s)$ is an integration factor of the form:*

$$\mu_i(t,s) = \exp\left(\int_0^t \alpha_i(\tau)+\beta_i(\tau)+\frac{d\eta(a;\boldsymbol{\theta})}{da}\Big|_{a=a(\tau,s)}d\tau\right). \tag{3.6}$$

Before presenting the proof of Theorem 2 we offer two comments on its use. First, we note it is possible to write $Y_i(t,a)$ as an explicit high dimensional integral equation of Υ; however, because the notation is cumbersome we omit it from the presentation. Second, the integral form of the solution requires being able to explicitly solve for and invert the characteristic Eq. (3.1).

Proof. To determine the solution of the linear system defined by Eq. (3.2) with initial conditions, Eq. (3.3), we solve these using the method of characteristics following [108]. For the initial generation $Y_0(t,a)$, we solve the characteristic equations

$$\frac{da}{dt} = \eta(a;\boldsymbol{\theta}), \text{ and } \frac{dY_0}{dt} + \left(\alpha_0(t)+\beta_0(t)+\frac{d\eta(a;\boldsymbol{\theta})}{da}\right)Y_0 = 0, \tag{3.7}$$

with parametric initial conditions $a(0,s)=s$, and $Y_0(0,s)=\Upsilon(s)$ respectively. Note that solution to the second characteristic equation can be written using an integrating factor. Then combining the solutions to the ODEs (Eqs. (3.7)) and the initial conditions, the solution to the initial generation $Y_0(t,a)$, is given by

$$Y_0(t,a) = \mu_0(t,s(t,a))^{-1}\Upsilon(s(t,a)),$$

where

$$\mu_0(t,s) = \exp\left(\int_0^t \alpha_0(\tau)+\beta_0(\tau)+\frac{d\eta(a;\boldsymbol{\theta})}{da}\Big|_{a=a(\tau,s)}d\tau\right).$$

For generations $1\le i\le M$, we solve the characteristic equations

$$\frac{da}{dt} = \eta(a;\boldsymbol{\theta}), \text{ and } \frac{dY_i}{dt} + \left(\alpha_i(t)+\beta_i(t)+\frac{d\eta(a;\boldsymbol{\theta})}{da}\right)Y_i = 2\gamma\alpha_{i-1}(t)Y_{i-1}(t,\gamma a(t,s)), \tag{3.8}$$

with parametric initial conditions $a(0,s) = s$, and $Y_i(0,s) = 0$. Solving the second equation using an integrating factor, and using the solution on the first equation in the system of ODEs (Eqs. (3.8)), the solution to the i^{th} generation $Y_i(t,a)$, is given by

$$Y_i(t,a) = \frac{2\gamma\alpha_{i-1}(t)}{\mu_i(t,s(t,a))}\int_0^t \mu_i(\tau,s(t,a))Y_{i-1}(\tau,\gamma a(\tau,s(t,a)))d\tau, \tag{3.9}$$

where

$$\mu_i(t,s) = \exp\left(\int_0^t \alpha_i(\tau)+\beta_i(\tau)+\frac{d\eta(a;\boldsymbol{\theta})}{da}\Big|_{a=a(\tau,s)}d\tau\right).$$

□

3.3 Analytic and Numerical Solutions

In this section we apply Theorem 2 to three biologically relevant intracellular models: (1) linear growth, (2) constant growth and linear decay, and (3) logistic growth. In each case, we first derive the problem specific recursive integral formulation and then compare the stability of numerical solutions using recursive numerical integration (RNI) [4] to the explicit Lax-Wendroff (LxW) method [123], a standard numerical approach for hyperbolic PDEs.

Theorem 2 gives an analytic solution for the initial generation, however iterative substitution of previous generations to determine the following generation in the recursive formulation leads to high dimensional integral solutions that need to be numerically estimated, which requires high computational overhead. The application of RNI takes advantage of the recursive formulation in Theorem 2 and allows for numerical estimations of the structured population density by recursively iterating through previous generations until the initial generation (analytic solution) is reached. In what follows, we compare our numerical solutions for both each generation $Y_i(t,a)$ and, when informative, the "total population" which we define as the sum of densities up to some generation M:

$$Z_M(t,a) = \sum_{i=0}^{M} Y_i(t,a). \tag{3.10}$$

Note that Theorem 2 combined with RNI leads to a meshfree method that allows for the evaluation of each generation at a single point in terms of intracellular constituent level and time (t,a), which is not possible with methods, such as LxW, that require a spatial grid.

In this implementation of RNI, Guass-Legendre quadrature [139] is applied for solving the integral equations. In order to compare RNI to LxW numerical solutions, the number of Gauss-Legendre quadrature points were determined systematically as we detail in the Supplemental Information (Section 3.5). Briefly, the number of quadrature points were increased until the addition of a new quadrature point did not

lead to an appreciable change in the solution. Also, the same number of quadrature points were used for integration at every level of recursion.

Intriguingly, for our third example the number of quadrature points needed per generation decreased while for our second example the number of points increased. Therefore, in this work, RNI outperformed LxW with intracellular dynamic models with a linear and logistic growth flux terms, but not in the case of a flux that consisted of constant synthesis and linear degradation (Table 3.1). Because of this, we conjecture that using Theorem 2 and RNI will outperform LxW in terms of accuracy and speed when the flux term, Eq. (3.1), does not have a constant non-zero additive or degradation term.

Flux Term	**Average Runtime (s)**			
	LxW_1	LxW_2	RNI	Exact
Linear Growth	11.1620	870.7137	0.3205	0.0559
Constant Growth and Linear Decay	0.09089	7.0570	180.6480	-
Logistic	0.4466	31.9893	0.2792	-

Table 3.1: **Average runtimes.** Average runtimes for numerical solutions, and evaluation time of the analytic solution with a linear growth, using two numerical methods Lax-Wendroff (LxW) and Recursive Numerical Integration (RNI), for three different flux terms: Linear Growth (Section 3.3.1), Constant Growth with Linear Decay (Section 3.3.2), and Logistic Growth (Section 3.3.3). These averages were computed over four different model evaluation times $T = 1, 2, 3$, and 4 hours.

3.3.1 Linear Growth

We first consider linear growth of an intracellular constituent:

$$\eta(a; \boldsymbol{\theta}) = v(t; \boldsymbol{\theta})a. \tag{3.11}$$

In this case, an analytic solution is known from the decomposition method first presented in [66]. We recover the same solution in our recursive integral formulation and use the analytic solution to compare to our numerical solutions with RNI and LxW.

3.3.1.1 Recursive Integral Formulation

We present the general solution to Eq. (3.2) with initial conditions stated in Eq. (3.3) and $\eta(a; \boldsymbol{\theta})$ as in Eq. (3.11). To apply Theorem 2 we solve the following equation

$$\frac{da}{dt} = v(t; \boldsymbol{\theta})a, \tag{3.12}$$

with the parametric initial condition $a(0, s) = s$.

This yields $a(t,s) = s\exp\left(\int_0^t v(t;\boldsymbol{\theta})dt\right)$, and $s(t,a) = a\exp\left(-\int_0^t v(t;\boldsymbol{\theta})dt\right)$. We then solve

$$\mu_i(t,s) = \exp\left(\int_0^t \alpha_i(\tau) + \beta_i(\tau) + \frac{d\eta(a;\boldsymbol{\theta})}{da}\Big|_{a=a(\tau,s)} d\tau\right).$$

This is simply $\mu_i(t,s) = \exp\left(\int_0^t \alpha_i(\tau) + \beta_i(\tau) + v(\tau;\boldsymbol{\theta})d\tau\right)$. Then the analytic solution to the initial generation, $Y_0(t,a)$, is

$$Y_0(t,a) = \mu_0(t,s(t,a))^{-1}\Upsilon\left(a\exp\left(-\int_0^t v(\tau;\boldsymbol{\theta})d\tau\right)\right). \tag{3.13}$$

Then from Theorem 2, we have that $Y_i(t,a)$ is given by

$$Y_i(t,a) = \frac{2\gamma\alpha_{i-1}(t)}{\mu_i(t,s(t,a))}\int_0^t \mu_i(\tau,s(t,a))Y_{i-1}(\tau,\gamma a(\tau,s(t,a)))d\tau. \tag{3.14}$$

For simplicity we consider constant cell division and constant cell death rates, $\alpha_i(t) = \alpha$ and $\beta_i(t) = \beta$, where $\alpha,\beta \geq 0$, for $0 \leq i \leq M$, and we have $\mu_i(t,s) = \mu(t,s) = \exp((\alpha+\beta)t)\exp\left(\int_0^t v(\tau;\boldsymbol{\theta})d\tau\right)$. We drop the subscript notation because there is no longer a dependence on the generation i. We can then write the initial solution $Y_0(t,a) = \mu(t,s(t,a))^{-1}\Upsilon\left(a\exp\left(-\int_0^t v(\tau;\boldsymbol{\theta})d\tau\right)\right)$. Now we can write the first generation as

$$Y_1(t,a) = \frac{2\gamma\alpha}{\mu(t,s(t,a))}\int_0^t \mu(\tau,s(t,a))Y_0(\tau,\gamma a(\tau,s(t,a)))d\tau.$$

When we substitute $Y_0(t,a)$ we see that

$$Y_1(t,a) = \frac{2\gamma\alpha}{\mu(t,s(t,a))}\int_0^t \Upsilon(\gamma s(t,a))d\tau = \frac{2\gamma\alpha t}{\mu(t,s(t,a))}\Upsilon(\gamma s(t,a)),$$

where the integrand term $\Upsilon(\gamma s(t,a))$ does not depend on τ so we can perform the integration. Further replacement of the terms $s(t,a)$, and $\mu(t,a)$, we have an analytic solution for the first generation

$$Y_1(t,a) = 2\gamma\alpha t\exp(-(\alpha+\beta)t)\exp\left(-\int_0^t v(\tau;\boldsymbol{\theta})d\tau\right)\Upsilon\left(\gamma a\exp\left(-\int_0^t v(\tau;\boldsymbol{\theta})d\tau\right)\right). \tag{3.15}$$

Continuing iteratively by substitution of the previous solution in the recursive integral form of Eq. (3.14) to determine the following generation we find that

$$Y_i(t,a) = \frac{(2\gamma\alpha t)^i}{i!\exp((\alpha+\beta)t)}\exp\left(-\int_0^t v(\tau;\boldsymbol{\theta})d\tau\right)\Upsilon\left(\gamma^i a\exp\left(-\int_0^t v(\tau;\boldsymbol{\theta})d\tau\right)\right), \tag{3.16}$$

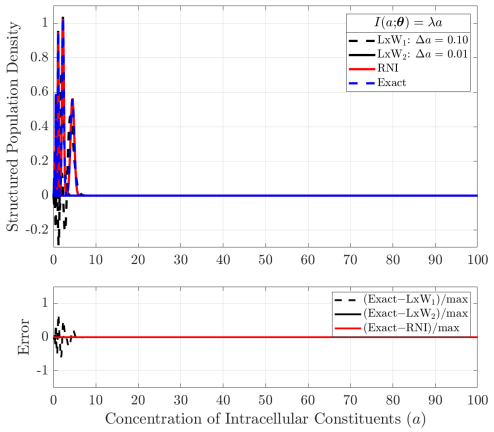
for $1 \leq i \leq M$. As expected we have reproduced the solution to this problem presented in [66]. Note that in this case, iterative substitutions of previous generations to compute the following generations leads to an explicit analytic solution and not high dimensional integral equations as we will see in Sections 3.3.2 and 3.3.3 with different intracellular dynamics models.

3.3.1.2 Numerical Solution

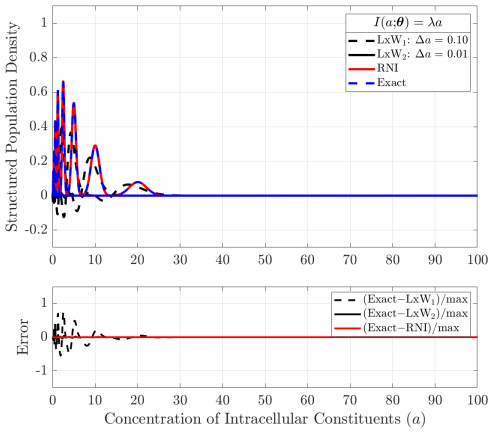
Numerically solving the DLSPM with a linear flux (Eq. (3.11)) requires specifying several parameters in our model. For these numerical experiments we assumed that there is no cell death ($\beta_i(t) = 0$) and a time-independent rate of cell division ($\alpha_i(t) = \log(2)/1.5$). We solved the DLSPM up to generation $M = 7$ on spatial grids of intracellular constituents that range from zero to one-hundred ($a \in [0, 100]$). The spatial discretizations for Lax-Wendroff for comparison with RNI are $\Delta a = 0.1$ (LxW_1) and $\Delta a = 0.01$ (LxW_2). Recursive numerical integration is applied on the same intracellular constituent values as LxW_1. For simplicity, we consider a flux term that is linear in intracellular concentration $\eta(a; \lambda) = \lambda a$ with $\lambda = 1.5$. The number of initial cells was set equal to 1 and the initial intracellular density, $\Upsilon(a)$, was set to a normal distribution with mean 1 and variance 0.1.

For the LxW method, the CFL condition for stability that requires that $\Delta t < \frac{\Delta a}{(M+1)\cdot \max_a\{\lambda a\}} = \frac{\Delta a}{(M+1)\lambda a_{\max}}$. Here $a_{\max}$ is the largest concentration of intracellular constituent considered in the spatial discretization. The number of quadrature points used in our implementation of RNI and our systematic way of determining them are presented in the Supplemental Information (Section 3.5). In this case, at most four quadrature points were applied in this implementation of RNI.

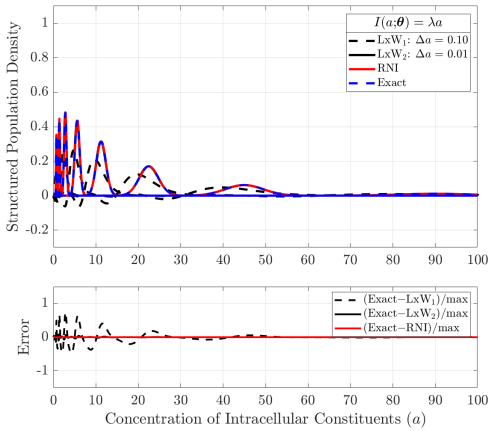
We now compare numerical solutions with RNI using Theorem 2 and the application of LxW on the original DLSPM, Eq. (3.2), to the known analytic solution to this system. As shown in Figures 3.1 and 3.2, using Theorem 2 and RNI leads greater accuracy than a LxW method at equal mesh size with a significantly faster runtime. Figure 3.2 and Table 3.2 show that RNI and the exact model evaluation significantly outperform both applications of LxW in terms of runtime.



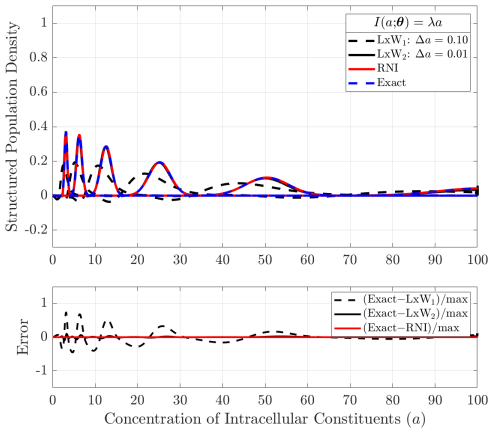
(a) $T = 1$ hour of experimental time.



(b) $T = 2$ hour of experimental time.



(c) $T = 3$ hours of experimental time.



(d) $T = 4$ hours of experimental time.

Figure 3.1: **Numerical Solutions Linear Growth: Separate Generations.** We compare numerical solutions (RNI, LxW) of the DLSPM, (Eqs. (3.2)-(3.3)) with Linear Flux Term (Eq. (3.12)) to the exact solution. (Top) We separately plot each generation at different points in time $T = 1, 2, 3$, and 4 hours. For the LxW method we used a uniform mesh with $\Delta a = 0.1$ (LxW$_1$) and $a = 0.01$ (LxW$_2$), and $\Delta a = 0.01$ for the RNI method. (Bottom): The error between the exact solution and each of the two numerical methods LxW and RNI. (See Section 3.3.1 for further details).

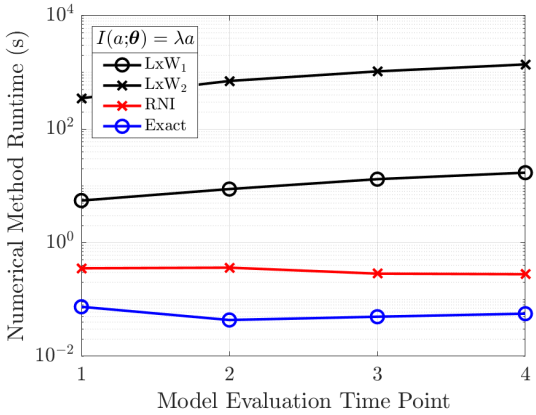


Figure 3.2: **Numerical Runtimes Linear Growth.** Comparison of runtimes for numerical solutions (LxW and RNI) and the exact solution [66] for the DLSPM (Eqs. (3.2)-(3.3)) with a linear flux term (Eq. (3.11)) at four different model evaluation times $T = 1, 2, 3$, and 4 hours.

Method	**Runtime (s)**			
	$T = 1$	$T = 2$	$T = 3$	$T = 4$
LxW$_1$	5.528	8.812	13.17	17.14
LxW$_2$	350.6	706.6	1043.0	1383.0
RNI	0.3545	0.3636	0.2853	0.2784
Exact	0.07435	0.0435	0.04969	0.05625

Table 3.2: **Numerical Runtimes Linear Growth.** Runtimes for numerical solutions, and evaluation time of the analytic solution for the DLSPM, (Eqs. (3.2)-(3.3)) with a linear flux term (Eq. (3.11)) at four different model evaluation times $T = 1, 2, 3$, and 4 hours.

3.3.2 Constant Synthesis and Linear Degradation

Now we consider solutions to Eq. (3.2) with initial conditions from Eq. (3.3) and a constant synthesis and linear degradation model

$$\eta(a;\boldsymbol{\theta}) = \lambda - \delta a. \tag{3.17}$$

Flux functions of this form have been previously used in modeling cellular populations [54]. In this case the DLSPM, Eq. (3.2), no longer admits analytic solutions.

3.3.2.1 Recursive Integral Formulation

Applying Theorem 2, we first solve the characteristic equation

$$\frac{da}{dt} = \lambda - \delta a, \tag{3.18}$$

with the parametric initial condition $a(0,s) = s$. Using an integrating factor, Eq. (3.18) has solutions

$$a(t,s) = \frac{\lambda}{\delta} + \exp(-\delta t)\left(s - \frac{\lambda}{\delta}\right), \text{ and } s(t,a) = \frac{\lambda}{\delta} + \exp(\delta t)\left(a - \frac{\lambda}{\delta}\right).$$

Then we solve

$$\mu_i(t,s) = \exp\left(\int_0^t \alpha_i(\tau) + \beta_i(\tau) + \frac{d\eta(a;\boldsymbol{\theta})}{da}\Big|_{a=a(\tau,s)} d\tau\right).$$

This is simply $\mu_i(t,s) = \exp\left(\int_0^t \alpha_i(\tau) + \beta_i(\tau) - \delta d\tau\right)$. The dynamics of the initial generation are then given by the equation

$$Y_0(t,a) = \mu_0(t, s(t,a))^{-1}\Upsilon\left(\frac{\lambda}{\delta} + \exp(\delta t)\left(a - \frac{\lambda}{\delta}\right)\right). \tag{3.19}$$

Continuing with Theorem 2, $Y_i(t,a)$ is then given by the recursive equation

$$Y_i(t,a) = \frac{2\gamma\alpha_{i-1}(t)}{\mu_i(t,s(t,a))}\int_0^t \mu_i(\tau, s(t,a)) Y_{i-1}(\tau, \gamma a(\tau, s(t,a)))d\tau, \tag{3.20}$$

for $i > 0$. Thus, the first generation dynamics are given by the equation

$$\begin{aligned} Y_1(t,a) = &\frac{2\gamma\alpha_1(t)}{\exp\left(\int_0^t \alpha_1(\tau) + \beta_1(\tau) - \delta d\tau\right)} \times \\ &\int_0^t \exp\left(\int_0^t \alpha_1(\tau) + \beta_1(\tau) - \delta d\tau\right) Y_0(\tau, \gamma a(\tau, s(t,a)))d\tau. \end{aligned} \tag{3.21}$$

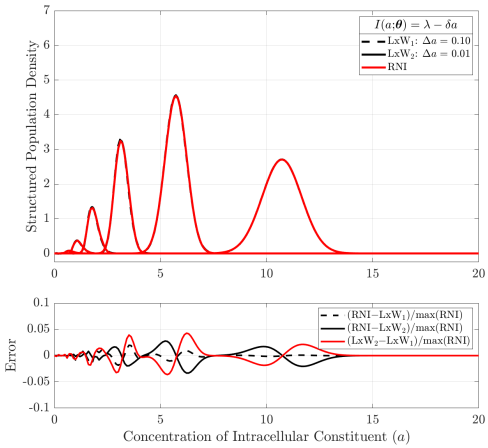
Here, substitution of the term $\gamma a(\tau, s(t,a))$ in $Y_0(t,a)$ for a leads to an integral form where the existence of an analytic solution to the integral in τ depends on the form of the initial distribution Υ. Therefore, continuing inductively to determine solutions for $Y_i(t,a)$ leads to high dimensional integral equations whose dimensionality (number of nested integrals) scales with the generation (i). The formulation in Eq. (3.20) allows for the estimate of $Y_i(t,a)$ for a particular value of intracellular constituents (a) at any time (t), without explicit knowledge of $Y_{i-1}(t,a)$.

3.3.2.2 Numerical Solutions

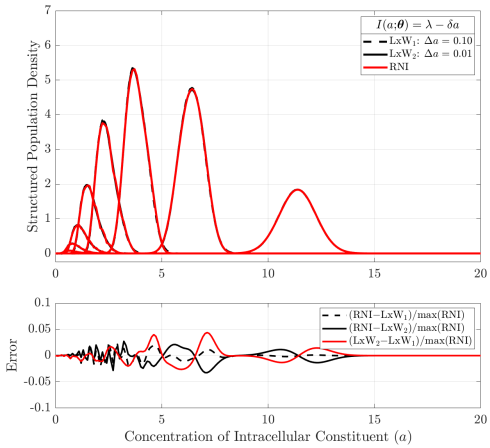
As in the linear flux case, we specify parameter values used to numerically solve our model. It is assumed that there is no cell death ($\beta_i(t) = 0$) and a time-independent rate of cell division ($\alpha_i(t) = \log(2)/1.5$). The flux term parameter values used are $\lambda = 1.5$, and $\delta = \frac{20}{1.5}$. We solved our system up to the seventh generation, $M = 7$. We set initial intracellular density $\Upsilon(a)$ to follow a normal distribution with mean 10, a variance 1, and we begin with 10 cells in the initial population.

We compare solutions with RNI using Theorem 2 to application of LxW on the original DLSPM, Equation (3.2). The CFL condition for stability of the LxW method requires that our timestep be $\Delta t < \frac{\Delta a}{(M+1)\cdot \max_a\{\lambda-\delta a\}} = \frac{\Delta a}{(M+1)\lambda}$. The spatial discretizations for Lax-Wendroff for comparison with RNI are $\Delta a = 0.1$ (LxW_1) and $\Delta a = 0.01$ (LxW_2). The number of quadrature points used in our implementation of RNI and our systematic way of determining them are presented in the Supplemental Information (Section 3.5). Recursive numerical integration is applied on the same intracellular constituent values as LxW_1.

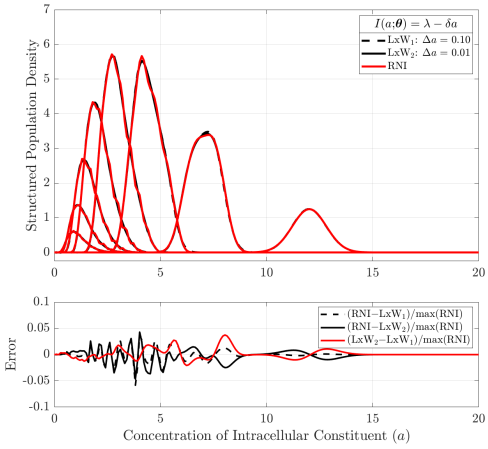
We show our numerical solutions in two contexts, first as each generation separately plotted (Figure 3.3) and then as the sum over all generations (Figure 3.4), which is consistent with what would be obtained with an experimental assay. While we can no longer compare with an exact solution, we see that RNI achieves results comparable to a higher order LxW method at the cost of longer runtimes than both LxW implementations. These longer runtimes for RNI are due to an increasing trend in necessary quadrature points as we estimate later generations, see Supplemental Information. We hypothesize that this increase in quadrature points is necessary due to the constant synthesis term in Eq. (3.17), which adds to the recursive computational complexity of the problem.



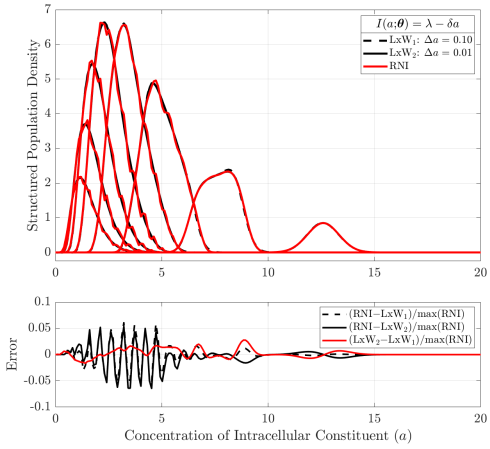
(a) $T = 1$ hour of experimental time.



(b) $T = 2$ hours of experimental time.

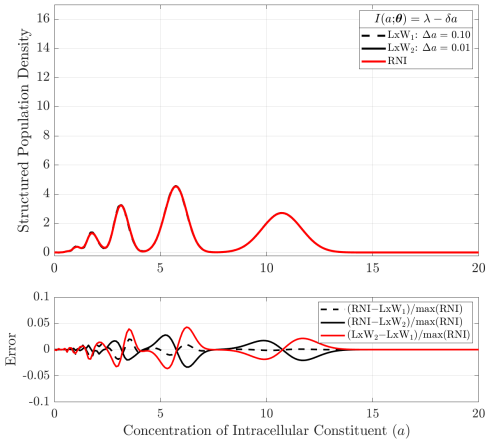


(c) $T = 3$ hours of experimental time.

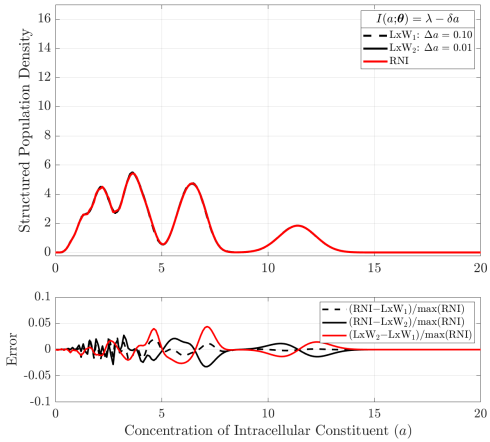


(d) $T = 4$ hour of experimental time.

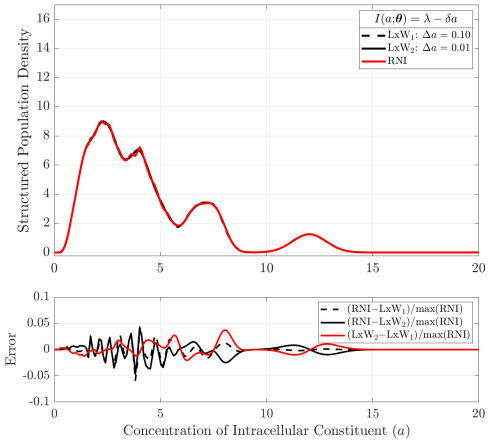
Figure 3.3: **Numerical Solutions Constant Growth and Linear Decay: Separate Generations.** We compare numerical solutions (RNI, LxW) of the DLSPM, (Eqs. (3.2)-(3.3)) with Flux Term (Eq. (3.17)). (Top) We separately plot each generation at different points in time $T = 1, 2, 3$, and 4 hours. For the LxW method we used a uniform mesh with $\Delta a = 0.1$ (LxW$_1$) and $a = 0.01$ (LxW$_2$), and $\Delta a = 0.01$ for the RNI method. (Bottom) The error between the RNI and LxW solutions. (See Section 3.3.2 for further details).



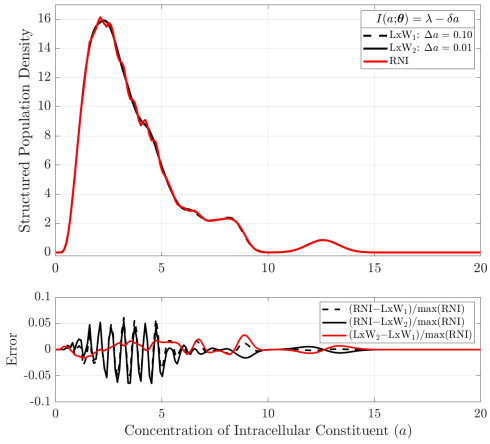
(a) $T = 1$ hour of experimental time.



(b) $T = 2$ hours of experimental time.



(c) $T = 3$ hours of experimental time.



(d) $T = 4$ hour of experimental time.

Figure 3.4: **Numerical Solutions Constant Growth and Linear Decay: Total Population.** We compare numerical solutions (RNI, LxW) of the DLSPM, (Eqs. (3.2)-(3.3)) with Flux Term (Eq. (3.17)). (Top) We plot the sum of all generations at different points in time $T = 1, 2, 3$, and 4 hours. For the LxW method we used a uniform mesh with $\Delta a = 0.1$ (LxW$_1$) and $a = 0.01$ (LxW$_2$), and $\Delta a = 0.01$ for the RNI method. (Bottom) The error between the RNI and LxW solutions. (See Section 3.3.2 for further details).

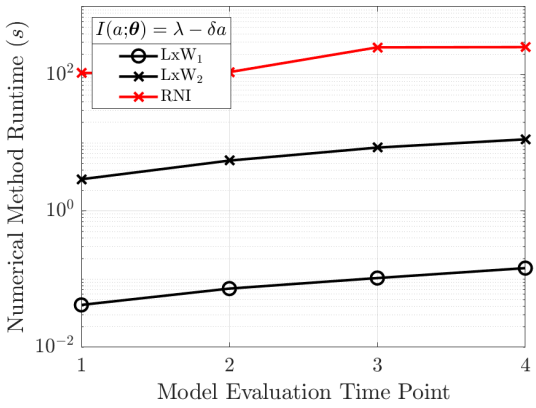


Figure 3.5: **Numerical Runtimes Constant Growth and Linear Decay.** Comparison of runtimes for numerical solutions (LxW and RNI) for the DLSPM (Eqs. (3.2)-(3.3)) with constant growth and linear decay (Eq. (3.17)) at four different model evaluation times $T = 1, 2, 3$, and 4 hours.

Method	**Runtime (s)**			
	$T = 1$	$T = 2$	$T = 3$	$T = 4$
LxW$_1$	0.04184	0.07266	0.1037	0.1451
LxW$_2$	2.926	5.497	8.533	11.27
RNI	106.1	109.3	252.6	254.6

Table 3.3: **Numerical Runtimes Constant Growth and Linear Decay.** Comparison of runtimes for numerical solutions (LxW and RNI) for the DLSPM (Eqs. (3.2)-(3.3)) with constant growth and linear decay (Eq. (3.17)) at four different model evaluation times $T = 1, 2, 3$, and 4 hours.

3.3.3 Logistic Growth

Finally, we consider solutions to Eq. (3.2) with initial conditions from Eq. (3.3) with intracellular dynamics driven by logistic growth

$$\eta(a; \lambda, K) = \lambda a \left(1 - \frac{a}{K}\right). \tag{3.22}$$

The logistic equation is a standard model in mathematical biology and considers growth within a resource limited environment [51, 121]. It is characterized by two parameters λ, the maximum growth rate, and K, the maximum possible concentration. As in our previous example, Eq. (3.2) does not admit analytic solutions. Therefore, we begin by solving for the recursive formulation and then compare numerical solutions with RNI to LxW.

3.3.3.1 Recursive Integral Formulation

We consider solutions to Eq. (3.2) with initial conditions given in Eq. (3.3) under logistic growth (Eq. (3.22)). As before, following Theorem 2, we first solve the characteristic equation

$$\frac{da}{dt} = \lambda a\left(1 - \frac{a}{K}\right),$$

with the parametric initial condition $a(0, s) = s$. This has solutions

$$a(t, s) = \frac{Ks\exp(\lambda t)}{K - s + s\exp(\lambda t)}, \text{ and } s(t, a) = \frac{Ka}{a - a\exp(\lambda t) + K\exp(\lambda t)}.$$

Secondly, we solve for $\mu_i(t, s(t, a))$ by

$$\mu_i(t, s(t, a)) = \exp\left(\int_0^t \alpha_i(\tau) + \beta_i(\tau) + \frac{d\eta(a;\boldsymbol{\theta})}{da}\Big|_{a=a(\tau,s)} d\tau\right),$$

so that

$$\mu_i(t, s(t, a)) = \left(\frac{K}{K + s(t, a)(\exp(\lambda t) - 1)}\right)^2 \exp\left(\lambda t + \int_0^t \alpha_i(\tau) + \beta_i(\tau)d\tau\right).$$

We have an analytic solution for the initial generation,

$$Y_0(t, a) = \mu_0(t, s(t, a))^{-1}\Upsilon\left(\frac{Ka}{a - a\exp(\lambda t) + K\exp(\lambda t)}\right).$$

Then, continuing with Theorem 2, $Y_i(t, a)$ is given by the recursive equation

$$Y_i(t, a) = \frac{2\gamma\alpha_{i-1}(t)}{\mu_i(t, s(t, a))}\int_0^t \mu_i(\tau, s(t, a))Y_{i-1}(\tau, \gamma a(\tau, s(t, a)))d\tau, \tag{3.23}$$

for $i > 0$. However, we no longer obtain non-integral solutions for $Y_i(t, a)$ for any initial distribution Υ by inductively using the previous solution $Y_{i-1}(t, a)$ by substituting the term $Y_{i-1}(t, \gamma a(t, s(t, a)))$ in the integral form of $Y_i(t, a)$. In this case, substitution of $Y_{i-1}(t, \gamma a(t, s(t, a)))$ leads to high dimensional integral solutions for $Y_i(t, a)$, whose dimension depends on the generation (i). The advantage to numerically solving the formulation in Eq. (3.23), over numerically solving the system in Eq. (3.2), is that for later generations, we can estimate this solution for a particular value of intracellular constituents (a) at any time (t).

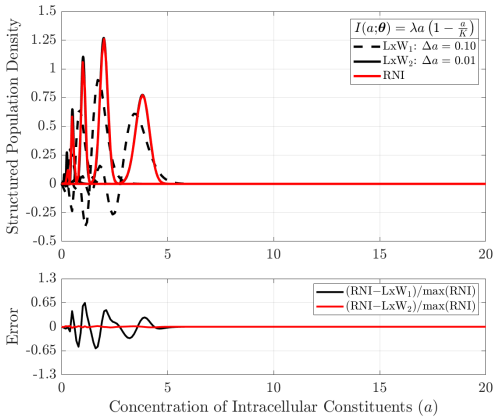
3.3.3.2 Numerical Solutions

In our numerical solutions, we assume there is no cell death ($\beta_i(t) = 0$) and a time-independent rate of cell division ($\alpha_i(t) = \log(2)/1.5$). We fix the parameters in the logistic model: $\lambda = 1.5$ and $K = 20$. The number of initial cells was set equal to

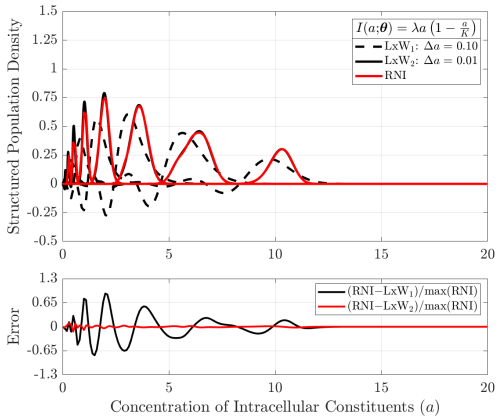
1 and the initial population density, $\Upsilon(a)$ was set to a normal distribution with mean and variance 1 and 0.1, respectively. We solved our system up to generation $M = 7$.

As before, we compare solutions with RNI to LxW. The CFL condition for stability of the explicit LxW method requires that our timestep be $\Delta t < \frac{\Delta a}{(M+1)\cdot \max_a\{\lambda a\left(1-\frac{a}{K}\right)\}} = \frac{4\Delta a}{(M+1)\lambda K}$. The spatial discretizations for Lax-Wendroff for comparison with RNI are $\Delta a = 0.1$ (LxW $_1$) and $\Delta a = 0.01$ (LxW$_2$). The number of quadrature points used in our implementation of RNI and our systematic way of determining them are presented in the Supplemental Information (Section 3.5). Recursive numerical integration is applied on the same intracellular constituent values as LxW$_1$.

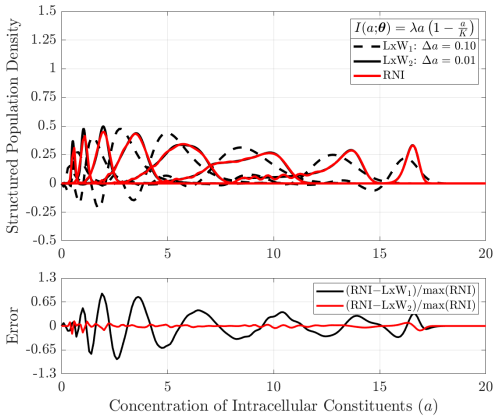
As in the previous example, we show our solutions in two ways: each generation separately (Figure 3.6) and the sum of all generations to be more consistent with experimental results (Figure 3.7). Notice that our implementation of Theorem 2 with RNI achieves an accuracy that is comparable to LxW at a higher refinement ($\Delta a = 0.01$) with a much faster runtime than both LxW implementations (Figure 3.8 and Table 3.4). These also show that while LxW with $\Delta a = 0.1$ and RNI take a fraction of a second of computational time, LxW with $\Delta a = 0.01$ takes more than 100 times longer to run. We also observe that unlike in the case of constant synthesis and linear degradation (Section 3.3.2), application of Theorem 2 with RNI required a decreasing number of quadrature points to estimate later generations, see Supplemental Information.



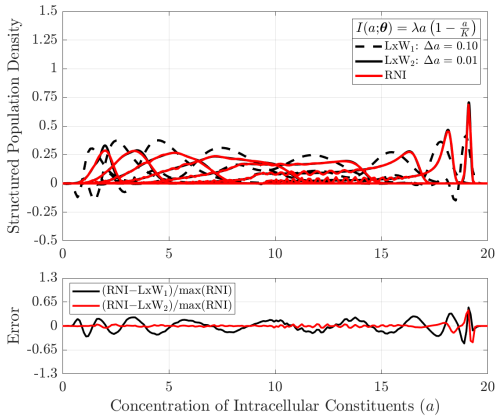
(a) $T = 1$ hour of experimental time.



(b) $T = 2$ hour of experimental time.

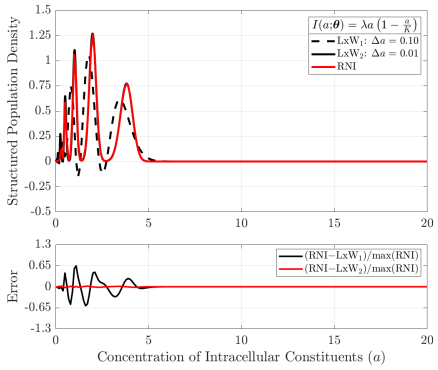


(c) $T = 3$ hours of experimental time.

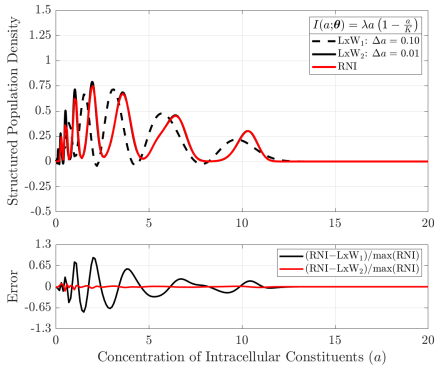


(d) $T = 4$ hours of experimental time.

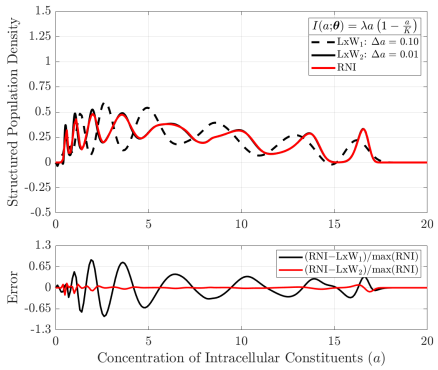
Figure 3.6: **Numerical Solutions Logistic Growth: Separate Generations.** We compare numerical solutions from (RNI and LxW) of the DLSPM (Eqs. (3.2)-(3.3)) with Flux Term (Eq. (3.22)). (Top) We separately plot each generation at different points in time $T = 1, 2, 3$, and 4 hours. For the LxW method we used a uniform mesh with $\Delta a = 0.1$ (LxW$_1$) and $a = 0.01$ (LxW$_2$), and $\Delta a = 0.01$ for the RNI method. (Bottom) The error between the RNI and LxW solutions. (See Section 3.3.3 for further details).



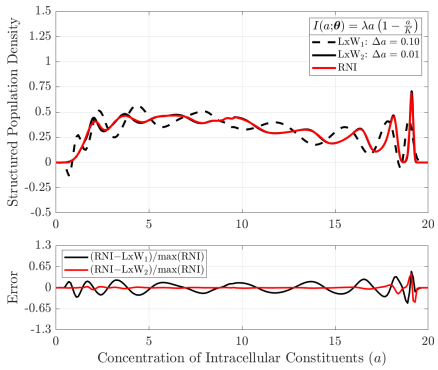
(a) $T = 1$ hour of experimental time.



(b) $T = 2$ hour of experimental time.



(c) $T = 3$ hours of experimental time.



(d) $T = 4$ hours of experimental time.

Figure 3.7: **Numerical Solutions Logistic Growth: Total Population.** We compare numerical solutions from (RNI and LxW) of the DLSPM (Eqs. (3.2)-(3.3)) with Flux Term (Eq. (3.22)). (Top) We plot the sum of generation at different points in time $T = 1, 2, 3$, and 4 hours. different points in time. For the LxW method we used a uniform mesh with $\Delta a = 0.1$ (LxW$_1$) and $a = 0.01$ (LxW$_2$), and $\Delta a = 0.01$ for the RNI method. (Bottom) The error between the RNI and LxW solutions. (See Section 3.3.3 for further details).

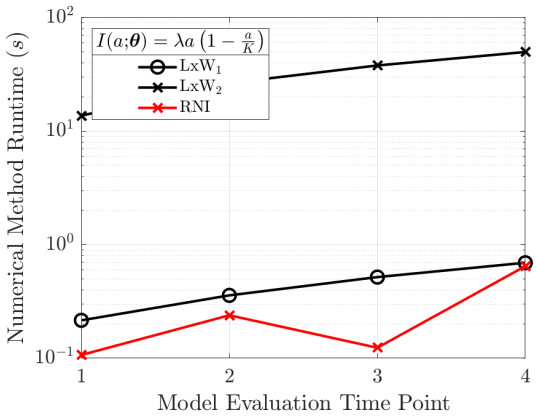


Figure 3.8: **Numerical Runtimes Logistic Growth.** Comparison of runtimes for LxW and RNI for numerical solutions (LxW and RNI) for for the DLSPM (Eqs. (3.2)-(3.3)) with logistic growth dynamics (Eq. (3.22)) at four different model evaluation times $T = 1, 2, 3$, and 4 hours.

Method	**Runtime (s)**			
	$T = 1$	$T = 2$	$T = 3$	$T = 4$
LxW$_1$	0.2158	0.358	0.5186	0.6939
LxW$_2$	13.71	26.52	37.91	49.82
RNI	0.1069	0.2391	0.1237	0.6472

Table 3.4: **Numerical Runtimes Logistic Growth.** Comparison of runtimes for LxW and RNI for numerical solutions (LxW and RNI) for for the DLSPM (Eqs. (3.2)-(3.3)) with logistic growth dynamics (Eq. (3.22)) at four different model evaluation times $T = 1, 2, 3$, and 4 hours.

3.4 Discussion & Conclusions

In this work, we presented a new theorem that gives a recursive solution to the label structured population models. This theorem applies to more general functional forms of the flux term for DLSPMs than addressed by previous work in [10, 66] whose aims were to speed up computation of the forward solution to the PDE system with decomposition techniques. Our recursive solution structure allows us to define $Y_1(t, a)$, by using the solution $Y_0(t, a)$, then $Y_2(t, a)$ can be written by assuming that $Y_1(t, a)$ has already been defined, and so on for $i \geq 2$. For some examples, an analytic solution for each generation can be determined, as in the case of Section 3.3.1. In other cases our formulation allows for novel numerical solutions.

We provided examples for applying this theorem with RNI to functional forms of the flux term which have been previously used to describe the production or decay of intracellular constituents, including linear growth, constant growth with linear

degradation, and logistic growth. Numerical results for a DLSPM with linear flux, in which we can also compute an exact solution for a baseline comparison, showed that the RNI method was significantly more efficient than the Lax-Wendroff (LxW) finite difference method, a standard numerical technique for approximating the solution of first order hyperbolic PDE systems. We found that application of Theorem 2 with RNI was more accurate than finite differences even when a very small step size was employed with the LxW method. We observed that RNI was between 900 to 5,000 times faster than the LxW method for this base case. For flux terms with constant growth and linear degradation, we found that RNI was 36 times slower than the LxW method. However, with the more complex flux term of logistic growth, RNI was between 77 to 300 times faster than the LxW. We believe that in the case of flux terms with constant synthesis, the additive term adds to the computational complexity when applying RNI methods to evaluate the DLSPM. However, the LxW method can efficiently handle first order hyperbolic PDE systems with a constant flux term.

An advantage of using Theorem 1 and RNI is in that this leads to a meshfree method for evaluating the DLSPM for an individual point in time and an intracellular constituent amount for a particular generation. The runtimes presented in Figures 3.2, 3.5, and 3.8 are the result of applying RNI on the coarse spatial grid for LxW, so evaluation of the DLSPM with Theorem 2 using RNI at one point takes less time than was reported. We also explored a forward method (iterative approach) of using Theorem 2, by saving the solution to the previous generation (i) to approximate the next generation $(i+1)$ in hopes that this would lead to a speedup in runtime; however, we found that this is not the case because in saving the solution to the previous generation we incur two costs. First, we incur an overhead cost from storing the previous solution by discretizing in time and space. Second, we incur additional overhead in having to use an interpolation method on previous generation in order to produce the next generation. The need for an interpolation method can be seen in the form of Equation (3.4) in Theorem 2 because of the $\gamma a(\tau, s(t,a))$ in the integrand:

$$Y_i(t,a) = \frac{2\gamma\alpha_{i-1}(t)}{\mu_i(t,s(t,a))}\int_0^t \mu_i(\tau,s(t,a))Y_{i-1}(\tau,\gamma a(\tau,s(t,a)))d\tau.$$

The RNI method on the other hand takes advantage of the recursive formulation in Theorem 2.

There are ways in which application of the RNI method can be sped up that were not considered in this work. We used Gaussian quadrature to approximate the integrals observed in the recursion formulation in Theorem 2, but there are other methods and problem specific quadrature rules that lead to greater stability and faster integration times [142]. In our work, we found that generating the quadrature points during each recursive integration using the scripts from [139] was another limiting factor. While we do not address these problems in our work, we believe that these are problem specific details that open up other research directions.

Our results are significant because the increased efficiency will drastically speed up

the ability to perform inverse problems, e.g., for parameter estimation, in which thousands of forward solutions of the entire PDE system are typically required. Thereby, Theorem 2 combined with the RNI method will enable model comparisons [13, 8], i.e., the testing of many different model structures including more detailed molecular processes, which may have previously been unexplored due to computational burden. Moreover, RNI will enable researchers to perform uncertainty quantification for DLSPMs, since methods such as Markov Chain Monte Carlo Bayesian estimation or frequentist bootstrapping techniques can require several orders of magnitude more forward solves than parameter estimation with gradient based optimization [72, 128].

Code and Data

Open-source code for both RNI and LxW methods are available at `https://github.com/biomathlab/DLSPM_solver`.

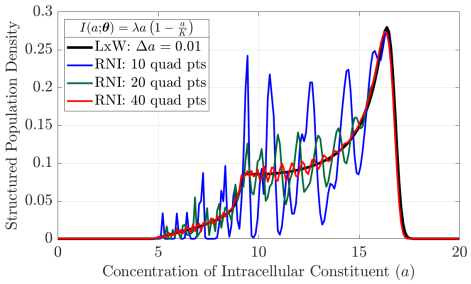
Acknowledgements

This research was funded in part by the Joint Initiative to Support Research at the Interface of the Biological and Mathematical Sciences between the National Science Foundation Division of Mathematical Sciences (NSF-DMS) and National Institutes of Health National Institute of General Medical Sciences (NIH-NIGMS) (Grant No. R01-GM126548) and used the Multi-Environment Computer for Exploration and Discovery (MERCED) cluster at UC Merced, funded by National Science Foundation (Grant No. ACI-1429783); in part by the National Science Foundation (Grant No.s DMS-1514929 and IOS-1838314), and in part by the National Institute of Aging (Grant No. R21AG059099).

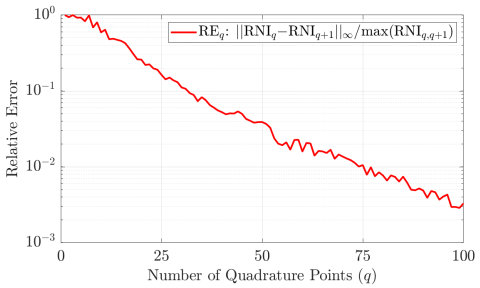
3.5 Supplemental Information

3.5.1 Determining Number of Quadrature Points

In the analysis presented in the main text, in the implementation of Recursive Numerical Integration, the number of quadrature points used (RNI_q) were chosen such that the addition of a quadrature point (RNI_{q+1}) no longer led to an improvement in the solution with a tolerance of $tol = 10^{-1}$, using an adjusted ∞-norm (see Figure 3.9). That is the number of quadrature points (q) used in the text satisfy $||\mathrm{RNI}_q - \mathrm{RNI}_{q+1}||_\infty / \max(\mathrm{RNI}_{q,q+1}) < tol$, for each generation and point in time presented. Here the norm adjustment term $\max(\mathrm{RNI}_{q,q+1})$ allows dynamically adapting the norm to each generation.



(a) Numerical solutions using Lax-Wendroff and three examples of RNI using a different number of quadrature points



(b) The convergence of RNI as we increase the number of quadrature points

Figure 3.9: **Number of Quadrature Points and Convergence.** Using a logistic flux term, the concentration of intracellular constituents is determined for the second generation at four hours into the experiment. The parameter values used here are the same as in the text.

Tables 3.5-3.7 present the number of quadrature points used for each flux term at each point in time. Note that we have an analytic solution for the initial generation (Y_0) and so we do not use RNI to approximate this solution.

Time	**Generation**						
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
$T = 1$	3	3	3	3	3	4	4
$T = 2$	3	3	3	3	3	4	4
$T = 3$	3	3	3	3	3	4	4
$T = 4$	3	3	3	3	3	4	4

Table 3.5: **Linear Growth Flux Term.** Number of quadrature points by generation for each model evaluation time.

Time	**Generation**						
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
$T = 1$	3	4	6	8	10	12	14
$T = 2$	3	6	9	11	13	14	14
$T = 3$	4	7	11	13	14	14	16
$T = 4$	6	9	13	13	14	14	16

Table 3.6: **Constant Synthesis Linear Degradation Flux Term.** Number of quadrature points by generation for each model evaluation time.

Time	**Generation**						
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
$T = 1$	3	3	3	3	4	3	4
$T = 2$	5	4	4	4	4	3	5
$T = 3$	18	13	7	5	5	5	4
$T = 4$	54	33	22	13	7	7	5

Table 3.7: **Logistic Flux Term.** Number of quadrature points by generation for each model evaluation time.

Chapter 4

Growth Rate Assays Reveal Fitness Consequences of β-Lactamases

This chapter originally published as, "Santiago, F., Doscher, E., Kim, J., Camps, M., Meza, J., Sindi, S., & Barlow, M. (2020). Growth rate assays reveal fitness consequences of β-lactamases. PloS one, 15(1), e0228240." Reprinted in accordance with the Creative Commons Attribution 1.0 International License (`https://creativecommons.org/publicdomain/zero/1.0/`) with minor changes.

4.1 Introduction

Antibiotic resistance has become a powerful model system for studying evolutionary biology. The emergence, mutation and selection of antibiotic resistance genes has created a nearly unique opportunity to study fitness [143], adaptation [64],[18], pleiotropy [59], epistasis [48], adaptive landscapes [95], and evolutionary potential [16]. Studying these specific aspects of antibiotic resistance requires the measurement of fitness of bacterial isolates. Although antibiotic clinical classifications were not designed to measure fitness, antibiotic susceptibility testing has been heavily used as an approximation of fitness [143]. The justification in the field is that as resistance to an antibiotic increases, the fitness of organisms exposed to that antibiotic is likewise increasing. More recently, growth rates [65] have been implemented as a direct measurement of fitness [94] but without direct experimental connection to susceptibility testing results. While this method has been rapidly catching on, and its results are being used to answer important questions about evolution [107], there are few studies investigating the effects of this change in methodology upon the data and results to which we now have access. In this study, we investigate the correlation of growth rate assays to susceptibility testing and find that growth rates are a much more sensitive method for assessing fitness in bacteria.

Clinical testing of antibiotic resistance is critical for the development of effective

treatment options. By determining the susceptibility of an isolate to an antibiotic, health care providers can properly administer antibiotics [5]. Three common methods used to detect resistance among bacterial isolates are: disc diffusion, E-testing, and broth dilution minimum inhibitory concentrations (MICs) [88]. In a clinical setting, bacterial isolates are classified as resistant, intermediate, or susceptible (RIS) to individual antibiotics. While these methods are useful for detecting antibiotic resistance, they lack the sensitivity for precise comparison of fitness in different antibiotics. Moreover, these clinical classifications are not amenable to quantitative analyses and were not designed to measure fitness.

A more sensitive measurement that reflects fitness is bacterial growth rate, which is a measure of the rate at which bacteria go through binary fission. Bacterial growth is characterized by four phases: lag phase, exponential phase, stationary phase, and death phase [17]. Typically, bacterial growth rates measure the exponential phase since this is the period in which the most growth occurs in the bacterial population. Growth rates have not yet been compared to clinical assays and it is unknown how growth rates correlate with clinical susceptibility assays. One of the goals of this study is to investigate if growth rates correlate with clinical resistance classification.

We hypothesized that growth rates can provide evidence about the relative contributions of resistance genes to the susceptibility of clinical isolates [94]. We reasoned that a low bacterial growth rate in the presence of an antibiotic would indicate susceptibility to that antibiotic, whereas a high bacterial growth rate would indicate elevated resistance to that antibiotic [94]. In either case, growth rates are a more quantitative measure of susceptibility to antibiotics and are more amenable to mathematical modeling and analysis. We additionally hypothesized that the quantitative sensitivity of growth rates would make it possible to generate reliable estimates about the predicted effect that the presence of resistance genes has on fitness.

To investigate these relationships, we partnered with Dignity Health Mercy Medical Center. Since 2013 we have collected patient isolates consisting primarily of E. coli from urinary tract infections (UTIs) that are resistant to extended spectrum beta-lactamases (ESBLs). While antibiotics are usually an effective treatment against UTIs, the rise of antibiotic resistant bacteria is starting to limit their effectiveness. Moreover, many isolates of bacteria are resistant to multiple antibiotics so identifying appropriate treatments remains a challenge [149]. We have previously used these patient isolates to study broad trends in resistance [122]; in this work we study the correlation between isolate growth rates and clinical resistance classifications as well as the presence of known resistance genes.

4.1.1 β-lactamases and antibiotic resistance

β-lactamase genes are among the most prevalent resistance genes in UTIs. This study considers the effects of three β-lactamases: $bla_{\text{TEM-1}}$, $bla_{\text{CTX-M-15}}$, and $bla_{\text{OXA-1}}$. We consider these three specifically because they are the most common resistance genes in our isolates (Table 4.5, Fig 4.3, and accompanying text). Although the

resistance gene ampC occurred with high frequency in our isolates we did not include it in our analysis because it is almost always found within the chromosome of E. coli and usually not expressed.

Historically the most common β-lactamase has been bla_{TEM}, which has accounted for approximately 90% of ampicillin resistance in E. coli [27]. $bla_{\text{TEM-1}}$, the most common bla_{TEM} causes penicillin resistance, while most of the 200+ variants that have evolved from it confer cephalosporin resistance as well. Based on structure, bla_{TEM}s have been categorized as Class A β-lactamases. Conformational changes in the active site of the enzyme, caused by amino acid substitutions result in the different resistance phenotypes (cephalosporin resistance) conferred by extended spectrum β-lactamases (ESBLs). More recently, another family of Class A resistance genes, the $bla_{\text{CTX-M}}$s, have been replacing bla_{TEM} β-lactamases in clinical isolates [27], [78], [82]. These confer cephalosporin resistance, including cefepime resistance at levels equal to or surpassing bla_{TEM}s. This trend has caused a reduction of available treatment options and primary treatment has shifted to carbapenems. We note, bla_{TEM} and $bla_{\text{CTX-M}}$ are structurally similar as they are both Class A β-lactamases [62].

bla_{OXA} β-lactamases are Class D serine β-lactamases that were named for their ability to hydrolyze oxacillin [52]. They are typically located on large plasmids and have been present in this location before the antibiotic era [15]. While they are not the most efficient at cephalosporin hydrolysis, these resistance genes eventually evolved the ability to confer resistance to cephalosporins and carbapenems [52].

4.2 Methods

4.2.1 Ethics statement

This project was evaluated by the IRB committees of both UC Merced and Dignity Health Mercy Medical Center and was classified as an exempt study by both. The isolates were isolated in the course of normal patient care, and provided to us after their clinical usefulness was fulfilled. Furthermore, the samples were de-identified, and for these reasons patient consent was not required.

4.2.2 Hospital isolates

The isolates used in this study were collected from patients admitted to Dignity Health Mercy Medical Center in Merced, California. They are representatives of the different resistance phenotypes and species that were collected from the hospital. Our goal in selecting these were to make sure each phenotype was represented at least once. For each sample we recorded: the date of sample isolation, the age and gender of the patient, the species of the bacteria, the source of the isolate, and its phenotype: resistant, intermediate, or susceptible for 16 different antibiotics (ampicillin, ampicillin/sulbactam, piperacillin/tazobactam, cefazolin, ceftazidime, ceftriaxone, cefepime, ertapenem, imipenem, amikacin, gentamicin, to-

bramycin, ciprofloxacin, levofloxacin, nitrofurantoin and sulfamethoxazole/trimethoprim) using standard breakpoints CLSI M100-S26 (2015).

4.2.3 Growth rate assays

Growth rate inocula were taken from standing overnight cultures and diluted to a final working concentration of 10^5 cells per mL in Mueller Hinton (MH) broth. The growth rate assay was performed in a BIOTEK (Model# 267638) spectrophotometer for 22 hours on stationary cultures at a temperature of 37° C at a wavelength of 600nm. Reads were collected every 20 minutes and after the 22-hour incubation period the optical density (OD) readings were then converted into growth rates, estimated during the exponential growth phase, using the freely available GrowthRates software package [65] (see Fig 4.1). The growth rate for each isolate was measured with six replicates for each antibiotic and control (no antibiotic). The cephalosporin antibiotics used were ceftazidime (CAZ), ceftriaxone (CRO), cefepime (FEP), all at a concentration of 64 μg/mL. Isolates were also tested against ampicillin (AMP) at 32 μg/mL. We chose these concentrations because after measuring the growth rates of 213 separate clinical isolates, we found that these concentrations provided a broad range of growth rates that indicated phenotypic differences between isolates (Figures 4.8-4.10). It was necessary to use concentrations that in some cases exceeded CLSI breakpoints for resistance because the variance in growth rates was insufficient at other concentrations to detect fitness differences.

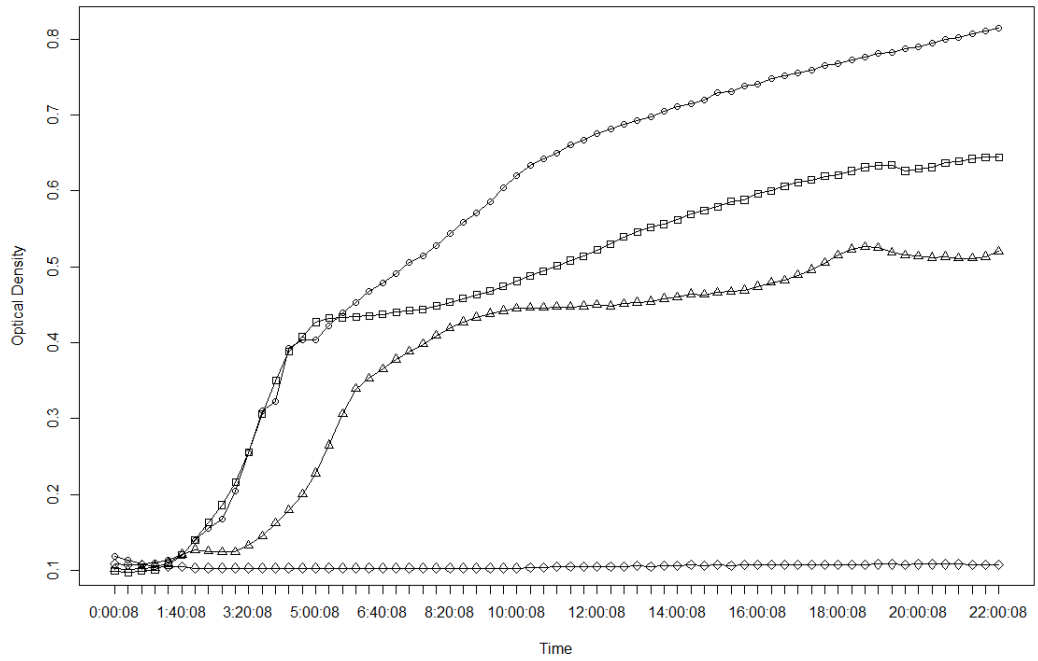


Figure 4.1: **Optical density (OD) measurements of E.coli isolates in the presence of ceftriaxone at a concentration of 64 μg/mL and control over a period of 22 hours.** OD measurements were made at 600nm every 20 minutes. The circles ($\circ$) represent OD measurements for isolate99; the squares ($\square$) represent the OD measurements for isolate155; the triangles ($\triangle$) represent the OD measurements for isolate109; and the diamonds ($\diamond$) represent the OD measurements for isolate105.

4.2.4 Statistical analysis of growth rates

We analyzed isolate growth rates in the presence of an antibiotic as a function of antibiotic susceptibility (S/R) and genetic differences (presence or absence of resistance alleles). We performed all analyses with R version 3.3.2 [134] and used $\alpha = 0.01$. In order to control for multiple statistical tests, we used The False Discovery Rate (FDR) Controlling Procedure [21, 22], which is a Bonferroni-type multiple testing procedure for both Welch's t-tests and the Shapiro-Wilk tests. Applying this procedure[1] with a false discovery control level of $q* = 0.05$, we report as significant only those results that remained significant after using FDR.

We tested the growth rates in each condition using the Shapiro-Wilk test, which rejects the null hypothesis of normality based on the skew and kurtosis of the sample [73]. Most of our data did not deviate from normality (Tables 1 and 2), which allowed

[1]The FDR Controlling Procedure follows from Theorem 1 in [22], which is restated with minor changes for clarity. If we consider n null hypotheses $h_1, h_2, \ldots, h_n$, and their corresponding p-values $p_1, p_2, \ldots, p_n$. Let $P_1 \leq P_2 \leq \cdots \leq P_n$ be the ordered p-value, and H_i be the corresponding null hypothesis to P_i. The Bonferroni-type multple-testing procedure is as follows: let k be the larges i for which $P_i \leq \frac{i}{n} q*$. Then reject all H_i for $i = 1, 2, \ldots, k$. This procedure controls the FDR at the $q*$ control level.

us to compare the means between conditions using Welch's t-test [118]. The skew we observed in growth rates for cefepime (Fig 4.2), may be due to greater responsiveness to slight concentration differences in cefepime than in other antibiotics. Though the growth rates data may have different variances, equal variance is not a requirement of Welch's t-test.

4.3 Results and discussion

The growth rates of these isolates were found using the optical density (OD) measurements of their growth in MH medium with a single concentration of the antibiotics CAZ, CRO, FEP, and AMP. We began by determining how well these growth rates corresponded to clinical determinations of susceptibility (Table 4.3). For simplicity, and because of sample size, we included only resistant and susceptible classifications. All isolates were either resistant or susceptible to FEP and CRO, and all isolates were found to be resistant to AMP. Three of the isolates displayed intermediate susceptibility to CAZ and were excluded from this analysis.

4.3.1 Isolate growth rates and clinical antibiotic susceptibility tests

In the case of all three cephalosporin antibiotics (CAZ, CRO, and FEP), growth rates for isolates clinically determined to be resistant were higher than growth rates of susceptible isolates (Table 4.4 and Fig 4.2, $p < 2{\times}10^{-3}$). The entries marked with an asterisk in Table 4.4 indicate that one (or both) of the growth rate groups under analysis did not pass a test for normality (Table 4.1). The cephalosporin CAZ had the greatest inhibitory effect on the growth rates of both resistant and susceptible isolates (Fig 4.2). We believe this is due to the high frequency of the $bla_{\text{CTX-M-15}}$ gene in our isolate population. The immediate ancestor of this gene, $bla_{\text{CTX-M-3}}$ has innate activity against ceftriaxone and cefepime, and through mutation, $bla_{\text{CTX-M-15}}$ more recently evolved activity against ceftazidime (see also Fig 4.4) [25]. All isolates were clinically determined to be resistant to ampicillin, and so we could not compare susceptible and resistant isolate growth rates.

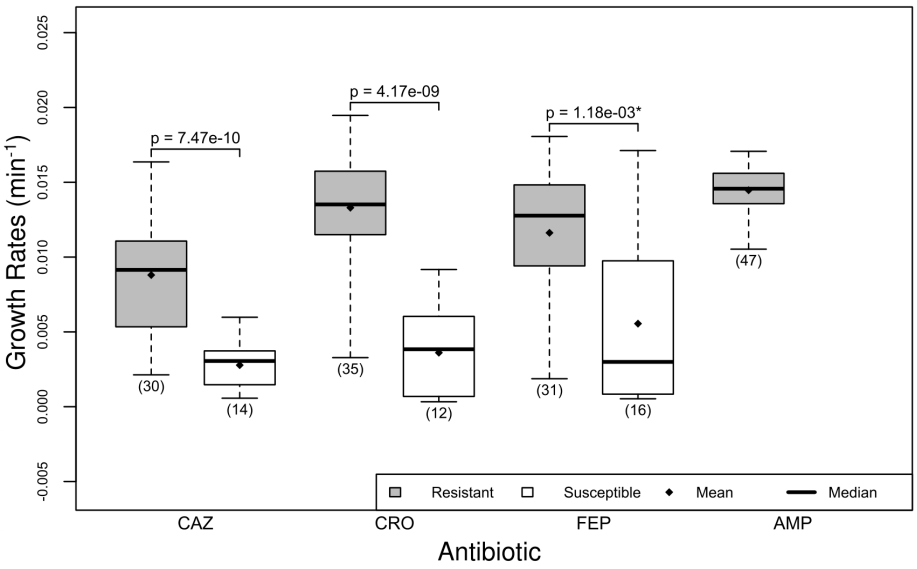


Figure 4.2: **Resistant isolates grow faster than susceptible isolates.** There are two boxplots per antibiotic, the first (gray) shows the distribution of growth rates for isolates that are resistant to the antibiotic, and the second (white) shows the distribution of growth rates for isolates that are susceptible to the antibiotic. The boundaries on the boxes indicate the 25th (Q1) and the 75th (Q3) percentiles, the line in the box represents the median, the diamond represents the arithmetic mean, and the whiskers indicate the minimum (below) and maximum (above) growth rate. The number of isolates used to create each boxplot is given in parenthesis. The asterisk indicates that one (or both) of the groups of growth rates under consideration did not pass a test for normality (Table 4.1).

	Resistant/(+)	**Susceptible/(−)**
	Ceftazidime	
Sensitivity	0.67	0.47
$bla_{\text{CTX-M-15}}$	0.73	0.62
$bla_{\text{TEM-1}}$	0.11	0.13
$bla_{\text{OXA-1}}$	0.62	0.14
	Ceftriaxone	
Sensitivity	0.32	0.12
$bla_{\text{CTX-M-15}}$	0.04	0.12
$bla_{\text{TEM-1}}$	0.27	0.08
$bla_{\text{OXA-1}}$	0.24	0.10
	Cefepime	
Sensitivity	0.03	2.93×10^{-3}*
$bla_{\text{CTX-M-15}}$	0.04	4.29×10^{-4}*
$bla_{\text{TEM-1}}$	0.02	7.00×10^{-3}*
$bla_{\text{OXA-1}}$	0.09	3.58×10^{-3}*

Table 4.1: ***p*-values for Shapiro-Wilk normality test in each condition.** For the row labeled "Sensitivity", the first column corresponds to resistance to the antibiotic, and the second column corresponds to susceptibility to the antibiotic. For the rows labeled with a resistance gene ($bla_{\text{CTX-M-15}}$, $bla_{\text{TEM-1}}$, $bla_{\text{OXA-1}}$) the first column indicates the presence of a resistance gene (+), while the second column indicates the gene is not present (−). All cases except growth rates in the presence of cefepime in the second column (Susceptible/(−)) indicate normality (with $p < 0.01$).

$\boldsymbol{bla_{\text{CTX-M-15}}}$**:**$\boldsymbol{bla_{\text{TEM-1}}}$	(+) : (+)	(+) : (−)	(−) : (+)	(−) : (−)
Ceftazidime	0.56	0.53	0.65	0.07
Ceftriaxone	0.82	0.16	0.11	0.32
Cefepime	0.28	0.20	1.70×10^{-3}*	0.13

Table 4.2: ***p*-values for Shapiro-Wilk normality test with combined resistance genes.** For each condition we test for normality with the Shapiro-Wilk Normality Test. All cases except one (growth rates for isolates with no $bla_{\text{CTX-M-15}}$ and $bla_{\text{TEM-1}}$ in presence of cefepime) indicate normality (with $p < 0.01$).

	CAZ	**CRO**	**FEP**	**AMP**
Resistant	30	35	31	47
Susceptible	14	12	16	0

Table 4.3: **Susceptibility to ceftazidime (CAZ), ceftriaxone (CRO), cefepime (FEP) and ampicillin (AMP).** Clinical susceptibility classifications of isolates in the presence of four antibacterial agents. Three intermediate samples excluded.

	μ_1	μ_2	99% CI
Sensitivity	**Resistant**	**Susceptible**	$\mu_1 - \mu_2$
CAZ	8.80×10^{-3}	2.78×10^{-3}	(3.97×10^{-3},8.08×10^{-3})
CRO	1.33×10^{-2}	3.61×10^{-3}	(6.74×10^{-3},1.26×10^{-2})
FEP	1.16×10^{-2}	5.55×10^{-3}	(1.45×10^{-3},1.07×10^{-2})
CTX-M-15	(+)	(−)	–
CAZ	9.19×10^{-3}	3.25×10^{-3}	(4.03×10^{-3},7.87×10^{-3})
CRO	1.34×10^{-2}	6.66×10^{-3}	(3.04×10^{-3},1.04×10^{-2})
FEP	1.26×10^{-2}	4.62×10^{-3}	(4.39×10^{-3},1.16×10^{-2})
TEM-1	(+)	(−)	–
CAZ	5.17×10^{-3}	8.59×10^{-3}	(-6.29×10^{-3},-5.51×10^{-4})
CRO	7.73×10^{-3}	1.38×10^{-2}	(-9.71×10^{-3},-2.40×10^{-3})
FEP	7.01×10^{-3}	1.20×10^{-2}	(-9.06×10^{-3},-8.96×10^{-4})
CTX-M-15:TEM-1	(+) : (+)	(+) : (−)	–
CAZ	8.08×10^{-3}	9.78×10^{-3}	(-5.59×10^{-3},2.20×10^{-3})
CRO	1.15×10^{-2}	1.44×10^{-2}	(-8.47×10^{-3},2.57×10^{-3})
FEP	1.16×10^{-2}	1.31×10^{-2}	(-6.72×10^{-3},3.66×10^{-3})
CTX-M-15:TEM-1	(+) : (−)	(−) : (+)	–
CAZ	9.78×10^{-3}	2.93×10^{-3}	(-4.53×10^{-3},9.17×10^{-3})
CRO	1.44×10^{-2}	4.85×10^{-3}	(-5.93×10^{-3},1.32×10^{-2})
FEP	1.31×10^{-2}	3.47×10^{-3}	(-6.21×10^{-3},1.31×10^{-2})
CTX-M-15:TEM-1	(+) : (+)	(−) : (+)	–
CAZ	8.08×10^{-3}	2.93×10^{-3}	(-1.42×10^{-3},8.90×10^{-3})
CRO	1.15×10^{-2}	4.85×10^{-3}	(8.12×10^{-4},1.24×10^{-2})
FEP	1.16×10^{-2}	3.47×10^{-3}	(2.68×10^{-3},1.36×10^{-2})

Table 4.4: **Average growth rates ($\mathbf{min^{-1}}$) for all conditions.** Mean isolate growth rates and 99% confidence intervals (CI) for the difference in mean growth rates based on a t-statistic.

	$\boldsymbol{bla}_{\textbf{CTX-M-15}}$	$\boldsymbol{bla}_{\textbf{TEM-1}}$	$\boldsymbol{bla}_{\textbf{OXA-1}}$
Present (+)	29	23	26
Absent (−)	18	24	21

Table 4.5: **Presence of $\boldsymbol{bla}_{\textbf{CTX-M-15}}$, $\boldsymbol{bla}_{\textbf{TEM-1}}$, and $\boldsymbol{bla}_{\textbf{OXA-1}}$ resistance genes.** The frequency of resistance genes *bla*CTX-M-15, $bla_{\text{TEM-1}}$, and $bla_{\text{OXA-1}}$ in our 47 isolates.

4.3.2 Resistance genes and isolate growth rates

After confirming that growth rates are associated with the broad clinical categorization of resistance to cephalosporins, we wanted to analyze the effect of individual

resistance genes on isolate growth rates. We hypothesize that isolates with genes known to confer resistance to cephalosporins would have higher growth rates than isolates without those resistance genes. We first identified the β-lactamases in our isolates (Table 4.5). Only three β-lactamases occurred at high frequency (more than 20 isolates) in our data set: $bla_{\text{CTX-M-15}}$, $bla_{\text{TEM-1}}$ and $bla_{\text{OXA-1}}$ (Table 4.5, Fig 4.3). Although we observed four other β-lactamases: $bla_{\text{TEM-1}}9$ (1 isolate), $bla_{\text{CTX-M-14}}$ (4 isolates), $bla_{\text{CTX-M-27}}$ (3 isolates), $bla_{\text{CTX-M-65}}$ (2 isolates), we focused on $bla_{\text{CTX-M-15}}$, $bla_{\text{TEM-1}}$ and $bla_{\text{OXA-1}}$ because of their frequency.

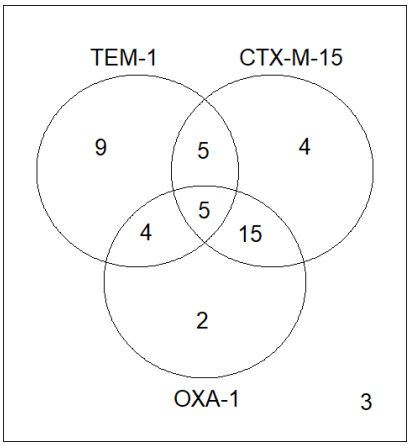


Figure 4.3: **Resistance gene combinations for our 47-isolate population.**

4.3.3 Observed effect of $bla_{\text{CTX-M-15}}$ on isolate growth rates

We found that isolates with the $bla_{\text{CTX-M-15}}$ gene had a higher growth rate in the presence of all three cephalosporin antibiotics than isolates that did not have the resistance marker (Table 4.4, Fig 4.4, $p < 2.05\times10^{-5}$). We did not observe this difference in either of the two controls (i.e., isolates grown with no antibiotic) or isolates grown in the presence of ampicillin.

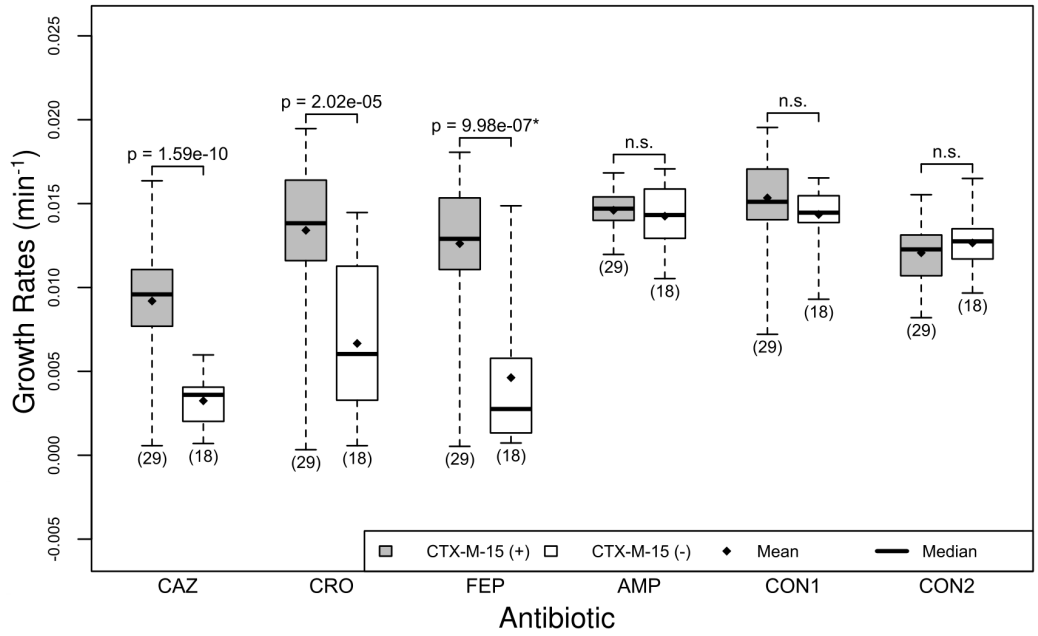


Figure 4.4: **Isolates with $bla_{\text{CTX-M-15}}$ exhibit higher growth rates in the presence of cephalosporins.** There are two boxplots per antibiotic, the first (gray) shows the distribution of growth rates for isolates that have the $bla_{\text{CTX-M-15}}$ gene, labeled $bla_{\text{CTX-M-15}}(+)$ and the second (white) shows the distribution of growth rates for isolates that do not have the $bla_{\text{CTX-M-15}}$ gene, labeled $bla_{\text{CTX-M-15}}(-)$. The n.s. indicates the difference in growth rates is not significant. Abbreviations for controls: CON1, the control growth rates for the experiment CAZ, CRO and FEP; CON2, the control growth rates for the experiment AMP. See Fig 4.2 for interpretation of boxplots.

4.3.4 Observed effect of $bla_{\text{TEM-1}}$ on isolate growth rates

When comparing the growth rates of isolates with the $bla_{\text{TEM-1}}$ gene and those without the gene, we observed that isolates without the $bla_{\text{TEM-1}}$ gene had a higher growth rate than those that had the marker (Table 4.4, Fig 4.5, $p < 2.5\times10^{-3}$). This difference in growth rates was not observed in the ampicillin experiments or the controls.

There are multiple potential causes for this effect that we have not yet explored which may not directly result from $bla_{\text{TEM-1}}$ expression. As $bla_{\text{CTX-M}}$ is a powerful cephalosporinase, the fact that our isolates without $bla_{\text{TEM-1}}$ were likely to have $bla_{\text{CTX-M-15}}$ may be the primary reason for these differences (Fig 4.3). However, the difference in frequency has been observed but not addressed as $bla_{\text{CTX-M}}$ genes have replaced bla_{TEM} genes in bacterial populations [27], [78], [82]. This unexplained shift in populations suggests that something more may be at work. Another possibility is that the two genes result in negative sign epistasis when they co-occur. Since the $bla_{\text{TEM-1}}$ enzyme appears to hydrolyze cephalosporins more slowly than $bla_{\text{CTX-M}}$s, its

presence could interfere with $bla_{\text{CTX-M}}$ hydrolysis of cephalosporins and decrease their overall efficiency. While we do not fully understand the biochemical basis for such a fitness cost, it does provide a mechanistic reason for the observation that $bla_{\text{CTX-M-15}}$ and $bla_{\text{OXA-1}}$ are replacing $bla_{\text{TEM-1}}$ in bacterial populations [27].

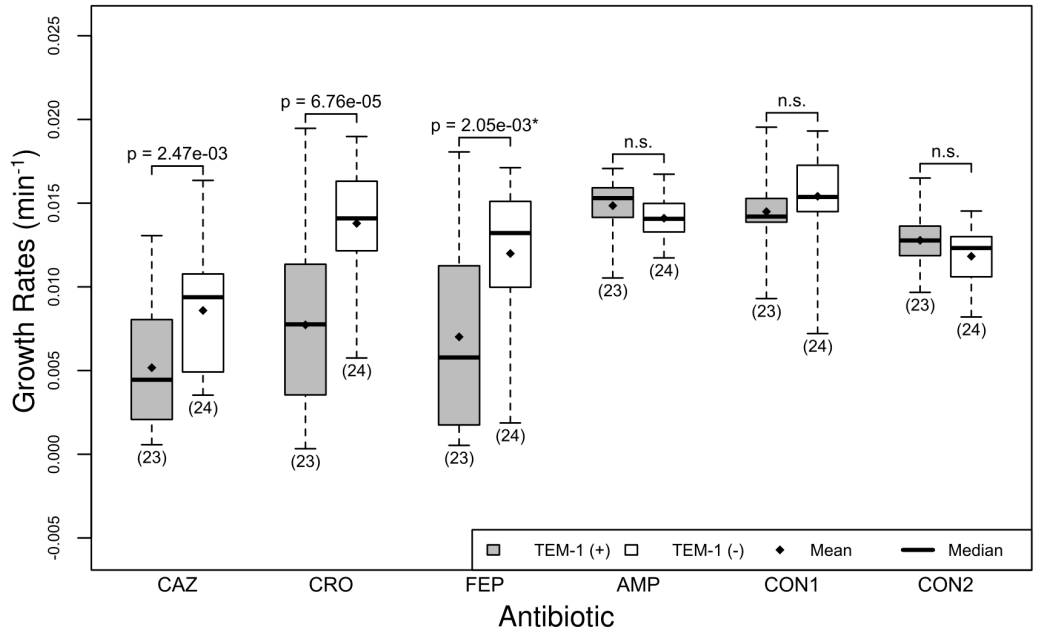


Figure 4.5: **Isolates with $bla_{\text{TEM-1}}$ exhibit lower growth rates in the presence of cephalosporins.** There are two boxplots per antibiotic, the first (gray) shows the distribution of growth rates for isolates that have the $bla_{\text{TEM-1}}$ gene, labeled $bla_{\text{TEM-1}}$ gene (+), and the second (white) shows the distribution of growth rates for isolates that do not have the $bla_{\text{TEM-1}}$ gene, labeled $bla_{\text{TEM-1}}(-)$. The n.s. indicates the difference in growth rates is not significant. See Fig 4.2 for interpretation of boxplots.

4.3.5 Observed effect of $bla_{\text{OXA-1}}$ on isolate growth rates

We did not observe a difference in growth rates based on the presence or absence of the $bla_{\text{OXA-1}}$ gene (Fig 4.6). This may be due to the high frequency of the co-occurrence of $bla_{\text{CTX-M-15}}$ and $bla_{\text{OXA-1}}$ (Fig 4.3). Also, $bla_{\text{OXA-1}}$ is penicillinase whose emergence was thought to result from the clinical introduction of oxacillin and methicillin. The different antibiotic specificities of $bla_{\text{OXA-1}}$ and $bla_{\text{CTX-M-15}}$ may confer an advantage when bacteria expressing both are exposed to a wide variety of antibiotics [97].

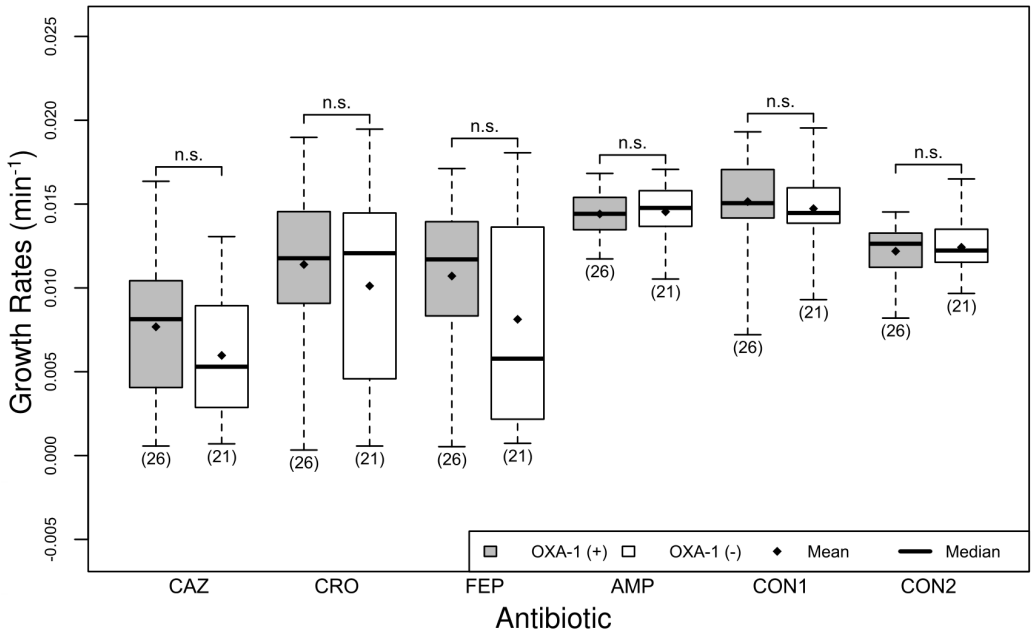


Figure 4.6: **The presence of $bla_{\text{OXA-1}}$ does not affect isolate growth rates.** There are two boxplots per antibiotic, the first (gray) shows the distribution of growth rates for isolates that have the $bla_{\text{OXA-1}}$ gene, labeled $bla_{\text{OXA-1}}$ gene (+), and the second (white) shows the distribution of growth rates for isolates that do not have the $bla_{\text{OXA-1}}$ gene, labeled $bla_{\text{OXA-1}}(-)$. See Fig 4.2 for interpretation of boxplots.

4.3.6 Separate and combined effects of $bla_{\text{CTX-M-15}}$ and $bla_{\text{TEM-1}}$ on isolate growth rates

Based on our previous observations about $bla_{\text{CTX-M-15}}$ and $bla_{\text{TEM-1}}$ in our isolates, we considered their combined effects. We found that the presence of $bla_{\text{TEM-1}}$ is not significantly associated with changes in growth rates for isolates that carry $bla_{\text{CTX-M-15}}$ even though there is a slight increase in mean growth rate for isolates lacking $bla_{\text{TEM-1}}$ (Table 4.4, Fig 4.7). It is possible that this result may become significant in future studies as the sample size increases since there is a tendency for isolates without $bla_{\text{TEM-1}}$ to have a faster growth rate than those that have it (Figs 4 and 6). Additionally, in the presence of ampicillin, isolates carrying $bla_{\text{TEM-1}}$ had a statistically significant faster growth rate (Table 4.4, Fig 4.7, $p < 9.64\times10^{-3}$). This result is expected because $bla_{\text{TEM-1}}$ is an efficient penicillinase whereas $bla_{\text{CTX-M-15}}$ is primarily a cephalosporinase.

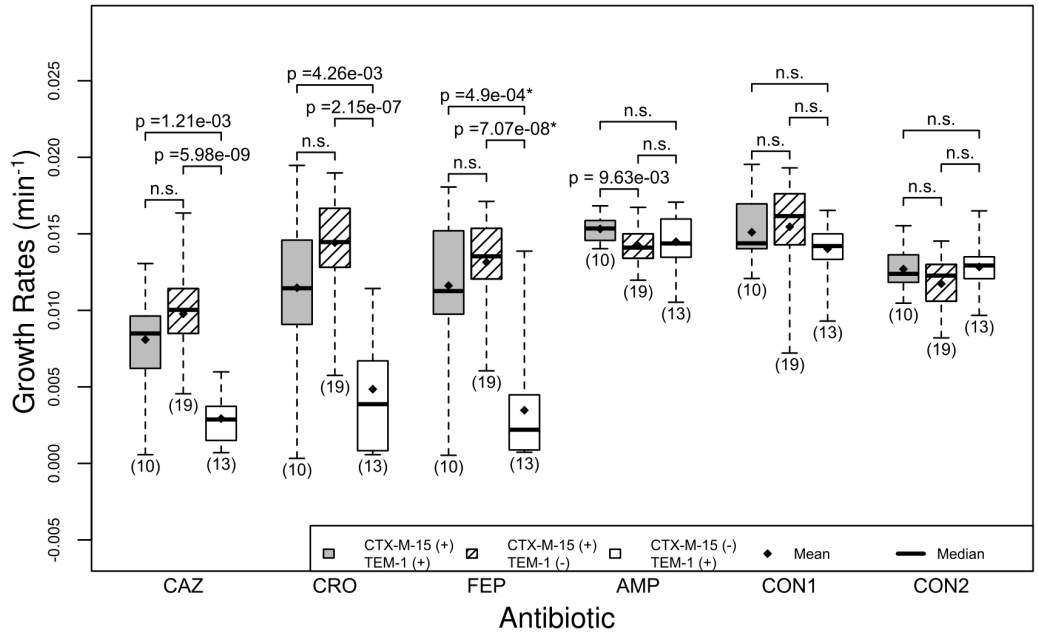


Figure 4.7: **Isolates with $bla_{\text{CTX-M-15}}$ exhibit higher growth rates irrespective of $bla_{\text{TEM-1}}$ presence.** There are three boxplots per antibiotic, the first boxplot (gray) shows the distribution of growth rates for isolates that have the $bla_{\text{CTX-M-15}}$ gene and the $bla_{\text{TEM-1}}$ gene (labeled $bla_{\text{CTX-M-15}}(+)$ and $bla_{\text{TEM-1}}(+)$), the second boxplot (hash marked) shows the distribution of growth rates for isolates that have the $bla_{\text{CTX-M-15}}$ gene but not the $bla_{\text{TEM-1}}$ gene (labeled $bla_{\text{CTX-M-15}}(+)$ and $bla_{\text{TEM-1}}(-)$), and the third boxplot (white) shows the distribution of growth rates for isolates that have the $bla_{\text{TEM-1}}$ gene but not the $bla_{\text{CTX-M-15}}$ gene (labeled $bla_{\text{CTX-M-15}}(-)$ and $bla_{\text{TEM-1}}(+)$). See Fig 4.2 for interpretation of boxplots.

4.4 Conclusion

The continued evolution of bacterial resistance to clinically useful antibiotics is a major world health crisis. The ability of bacteria to evolve and transfer resistance genes rapidly throughout a bacterial population presents an ongoing challenge for healthcare providers and will require the development of more effective treatment plans. Coping with this challenging problem requires not only effort of clinicians, but the expertise of researchers in evolutionary biology, statistics and mathematics. Methods in these latter fields require sensitive techniques for quantifying the fitness of a bacterial isolate to an antibiotic. We propose that bacterial growth rates have the potential to be a useful technique to the study of antibiotic resistance in these fields because growth rates can provide more information about the fitness effects of resistance genes than current clinical classifications of resistance.

In this study, we used E. coli isolates from Dignity Health, Mercy Medical Center to measure bacterial growth rates in the presence of different antibiotics. In particular,

we determined growth rates for each of 47 bacterial isolates in the presence of three cephalosporins and one penicillin. Through growth rate assays, we were able to quantify an isolate's fitness in the presence of antibiotics at single concentrations and found these results to be highly consistent with the information provided by clinical antibiotic susceptibility testing. Additionally, the sensitivity of growth rates enabled us to observe the effects that resistance genes had on isolate growth rates, both individually and in various combinations. We also confirmed that $bla_{\text{CTX-M-15}}$ is an excellent cephalosporinase.

Overall, we found that growth rates have greater sensitivity for assaying interactions between resistance genes and may be useful in helping to develop predictive models for assessing antibiotic susceptibility based on presence or absence of resistance genes. While growth rates are not a good candidate for susceptibility testing, their added sensitivity may help reveal the interactions between resistance genes and may facilitate improvements in the reliability of molecular diagnostic methods. Our findings suggest that there may be inhibitory interactions between resistance genes, and we intend to explore the relationships in future studies with isogenic strains.

4.5 Supporting information

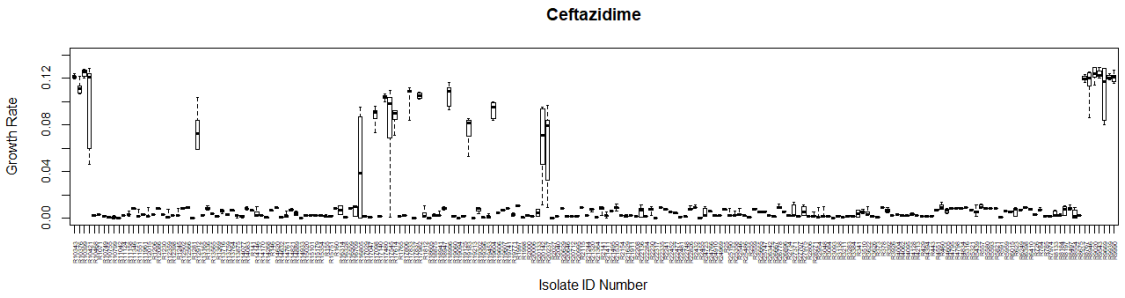


Figure 4.8: **Boxplot of E.coli isolates growth rates in the presence of ceftazidime at a concentration of 64 μg/mL.** The boundaries on the boxes indicate the 25th (Q1) and the 75th (Q3) percentiles (quartiles), the line in the box represents the median, and the whiskers indicate the minimum (below) and maximum (above) growth rate. There are 214 boxplots, one box-plot per isolate from six technical replicates. This figure shows that at a concentration of 64 μg/mL, we observe phenotypic differences between isolate growth rates.

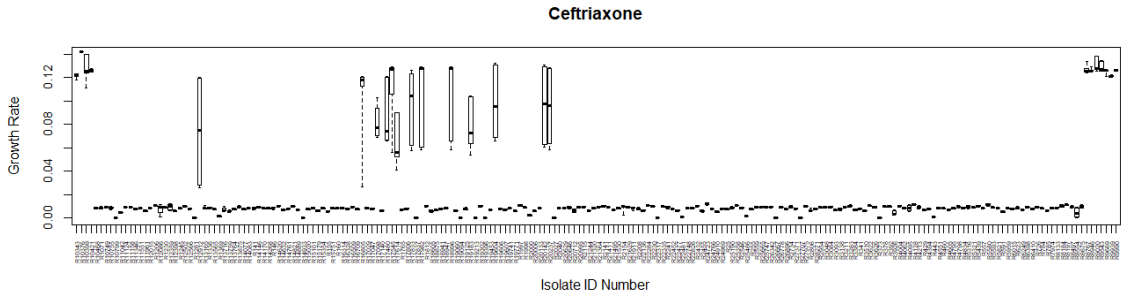


Figure 4.9: **Boxplot of E.coli isolates growth rates in the presence of ceftriaxone at a concentration of 64 μg/mL.** The boundaries on the boxes indicate the 25th (Q1) and the 75th (Q3) percentiles (quartiles), the line in the box represents the median, and the whiskers indicate the minimum (below) and maximum (above) growth rate. There are 214 boxplots, one box-plot per isolate from six technical replicates. This figure shows that at a concentration of 64 μg/mL, we observe phenotypic differences between isolate growth rates.

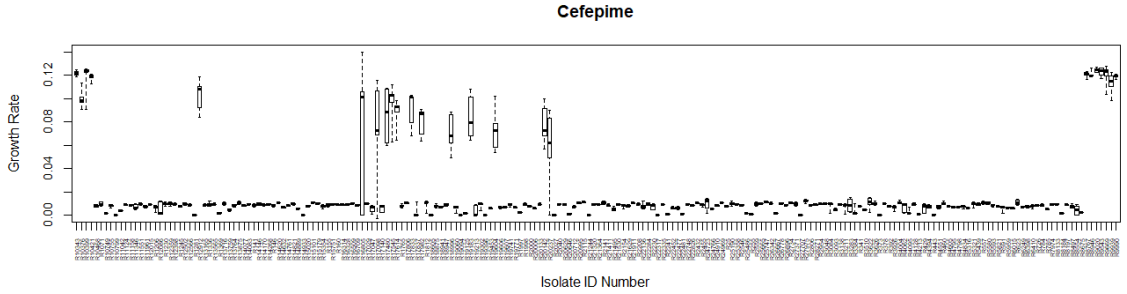


Figure 4.10: **Boxplot of E.coli isolates growth rates in the presence of cefepime at a concentration of 64 μg/mL.** The boundaries on the boxes indicate the 25th (Q1) and the 75th (Q3) percentiles (quartiles), the line in the box represents the median, and the whiskers indicate the minimum (below) and maximum (above) growth rate. There are 214 boxplots, one box-plot per isolate from six technical replicates. This figure shows that at a concentration of 64 μg/mL, we observe phenotypic differences between isolate growth rates.

Chapter 5

Distribution of β-Lactamase Genes in Clinical Isolates from California Central Valley Hospital Deviates from the United States Nationwide Trends

This chapter originally published as, "Guzman-Cole, C., Santiago, F., Garsevanyan, S., Sindi, S., & Barlow, M. (2021). Distribution of β-Lactamase Genes in Clinical Isolates from California Central Valley Hospital Deviates from the United States Nationwide Trends. Antibiotics, 10(5), 498." Reprinted in accordance with the Creative Commons Attribution 4.0 International License (`http://creativecommons.org/licenses/by/4.0/`) with minor changes.

5.1 Introduction

β-lactam antibiotics have been in use since the discovery of penicillin in the 1940s, and they continue to be the most widely used antibiotics due to their high effectiveness, ease of delivery, and low toxicity [145, 102]. The longstanding use of β-lactam antibiotics has led to the emergence of resistant strains in clinical care settings [65]. The continuous selection and evolution of β-lactamase genes by β-lactam antibiotic use has led to the diversification of successful β-lactamase genes: bla_{TEM}, bla_{SHV}, $bla_{\text{CTX-M}}$, and bla_{OXA} [20]. β-lactamase genes produce extended spectrum β-lactamase (ESBL) enzymes that work by hydrolyzing β-lactam antibiotics, rendering them ineffective. bla_{TEM} and bla_{SHV} were the first β-lactamase enzymes identified in 1963 and 1972, respectively, and were implicit in outbreaks in the 1990s [42, 81, 109]. Today, bla_{SHV} composes 10% of ESBLs, and bla_{TEM} has become somewhat less common in the U.S. [8]. $bla_{\text{CTX-M}}$ was first identified in 1989, and was identified with increasing frequency throughout the 1990s [9]. By the 2000s, the frequency of $bla_{\text{CTX-M}}$ enzymes surpassed

those of bla_{TEM} and bla_{SHV}.

Although first discovered in 1976, bla_{OXA} enzymes have been increasing in prevalence due to the frequent association of bla_{OXA}-1 with $bla_{\text{CTX-M}}$-15 [10,11]. Today, $bla_{\text{CTX-M}}$ enzymes are the most identified ESBLs, and have displaced bla_{TEM} and bla_{SHV} in many individual hospitals [81, 27, 39, 132, 93]. However, this trend is not uniform across publications originating from different surveillance locations [110, 90, 130]. Bajpai et al. (2017) found bla_{TEM} to be the most abundant ESBL enzyme in a single hospital, although other reports detail different ESBL gene frequencies [7]. In the United States, few recent nationwide surveillance studies have specifically examined the frequencies of specific ESBL genes. One recent survey of 26 hospitals identified bla_{TEM} as the most abundant ESBL enzyme in clinical isolates (47%), followed by $bla_{\text{CTX-M}}$ (36%), bla_{SHV} (35%), and bla_{OXA} (20%) [29].

Regional variance in the frequencies of ESBLs enables the assessment of which factors are contributing the most to ESBL frequencies. Due to the strong selection that bacteria experience from antibiotics and the rapid migration of bacteria that occurs in human populations, selection and migration were the two factors we chose to investigate. To understand the relative contributions of selection and migration, it was important to obtain and compare updated ESBL gene frequencies. We chose to compare the frequencies of ESBLs in a local repository of ESBL positive isolates collected from a single hospital, with average frequencies nationwide across the U.S. obtained from ESBL-positive clinical isolates whose genomic sequences have been deposited in the NCBI Isolates Browser Database. When comparing genetic variations over two populations, there are four possible outcomes depending on the stability of gene frequencies within a site and comparisons of those frequencies between sites. First, gene frequencies that are stable over time within a population and non-uniform across populations indicate low migration between bacterial populations and that selection for resistance within a given population is strong and constant. However, if the gene frequencies are unstable over time within a population and non-uniform across populations, this would indicate alternating local selective pressures and rapid migration as "immigrants would increase the mutation supply rate" [89] and would compete with "better-adapted residents maintaining the population away from the local fitness optimum" [106]. Stability over time and uniformity between populations suggest rapid continuous migration between populations and strong consistent selection resulting in a highly resistant and optimized strain [106]. Finally, unstable (alternating) frequencies over time and uniformity between populations indicate strong alternating selective pressures in large areas (or populations). Moreover, this signal also indicates rapid migration because variation between populations averages out as immigration leads to a decrease in genetic differentiation between populations [99]. We compared ESBL gene frequencies from Dignity Health Mercy Medical Center (DHMMC) and the rest of the U.S. over a period of six years as follows.

5.2 Materials and Methods

5.2.1 Hospital Isolates

Clinical patient isolates ($n = 872$) were collected from patients at DHMMC in Merced, CA, USA, from 2013 to 2018. These isolates were flagged as ESBL using Vitek 2 (bioMérieux, Inc. Hazelwood, MO, USA). These patient samples were collected from urine, blood, sputum, and wounds.

5.2.2 Molecular Methods

Genomic DNA was isolated using the boil preparation method by adding a single colony to 100 μL sterile water and boiling at 100 °C for 15 min. From this 100 μL solution, 1 μL was used in the PCR reaction for the respective genes with the primers listed in Table 5.1. Multiplex PCR was used to determine the presence of $bla_{\text{CTX-M}}$, bla_{TEM}, and bla_{OXA}. Detection of bla_{SHV} was run in a separate reaction. Each PCR reaction consisted of 1 μL of template DNA, 10 μM of each primer, and Taq 2X master mix (NEB) at a final concentration of 1X, and the reactions were run under the following conditions: initial denaturation at 94 °C for 10 min, 30 cycles of 94 °C for 40 s, 60 °C for 40 s, 72 °C for 1 min, and a final elongation at 72 °C for 7 min [40]. PCR amplicons were run out on 2% agarose gel at 100 V for 30 min and visualized using a ChemiDoc$^{\text{TM}}$ Touch Imaging System. (Figures S1 and S2).

5.2.3 US Database

We obtained clinical isolate genomes from the NCBI RefSeq database [100], using the NIH Isolate Browser [98] to identify clinical isolates of *E. coli* and *K. pneumoniae* from the United States from 2013 to 2018. Using the Comprehensive Antibiotic Resistance Database (CARD) [3], we identified isolate genomes containing ESBL genes to compile an ESBL clinical database ($n = 1060$) using the BLAST+ program. In combination with a 98% identity cut-off to positively identify the frequency of bla_{TEM}, bla_{OXA}, $bla_{\text{CTX-M}}$, and bla_{SHV}, we applied an additional base pair match cutoff for each gene to limit partial gene matches. For bla_{TEM}, we required a base pair (bp) match at or above 753 bp; for bla_{OXA}, we required 831 bp; for $bla_{\text{CTX-M}}$, we required 876 bp; and for bla_{SHV}, we required 861 bp. The metadata for nationwide clinical isolate genomes were downloaded from the NIH Isolate Browser and included date, species, and location. The list of genome assemblies used to perform this analysis can be found in the Supplementary Materials.

Gene	Primer Sequence (5' to 3')	Product Size (bp)
bla_{SHV}	Forward GCCTGTGTATTATCTCCCTGTTAG Reverse TCCCGGCGATTTGCTGATTCC	813
bla_{TEM}	Forward TGACGCCGGGCAAGAGCA Reverse AAGGGCCGAGCGCAGAAGTG	424
bla_{OXA}	Forward AGCGCCAGTGCATCAACAG Reverse GCAAAACCCAAACAACAGAAA	300
$bla_{\text{CTX-M}}$	Forward CGGCCGCGGTGCTGAAGAA Reverse GCTGCCGGTTTTATCCCCCACAA	482

Table 5.1: **List of primers pairs used to identify bla_{SHV}, bla_{TEM}, bla_{OXA}, and $bla_{\text{CTX-M}}$ and their expected product size.**

5.2.4 Statistical Analysis

We used one-way analysis of variance (ANOVA) to compare the means of the resistance gene frequencies to identify significant differences between frequencies across months. We compared the same months across years, different months within the same year, different months across all years, and bins of 2, 3, 4, and 6 months. We tested for significant differences between the means of the resistance gene frequencies across years at DHMMC and the nationwide database using a Z-test[1]. We tested for significant differences in the proportions of a resistance marker between isolates from DHMMC and the nationwide database using a Z-test. Pairwise linkage among the resistance alleles in each of the two clinical isolate populations, DHMMC and the nationwide database, was assessed using a χ^2 test [61]. The phi coefficient $\left(\Phi = \pm\sqrt{\frac{\chi^2}{n}}\right)$ was used as a χ^2 measure of directional deviation from the null relationship of independent assortment [50]. The PC has the desired property of accounting for sample size n (often >500 in this work) and because it has a known sampling distribution, it allows us to compute significance and to form confidence intervals. Controlling for multiple statistical tests was conducted via the FDR-controlling procedure [21, 22], a Bonferroni-type multiple testing procedure[2], with a false discovery control level of $q* = 0.025$. Only those results that remained significant after using FDR are reported as significant. All analyses were performed using the Statistics and Machine Learning Toolbox of MATLAB R2020a [92].

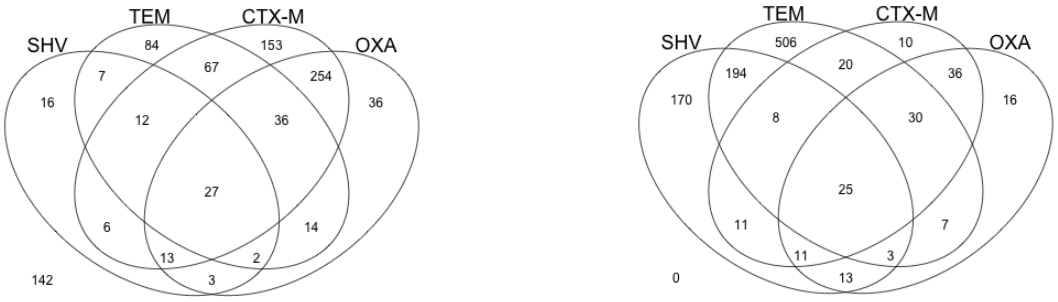
[1]Following [31], let $X_1 \sim \mathcal{B}(n_1, p_1)$ and $X_2 \sim \mathcal{B}(n_2, p_2)$, where $\mathcal{B}$ is the binomial distribution with n trials and success probability p, be independent. To test $h_0 : p_1 = p_2$ against $h_a : p_1 \neq p_2$, we form the Z test statistic $\text{Z} = (\hat{p}_1 - \hat{p}_2)/\sqrt{\hat{p}_*(1-\hat{p}_*)(1/n_1 + 1/n_2)}$, where $\hat{p}_i$ is the sample proportion and $\hat{p}_* = \frac{X_1+X_2}{n_1+n_2}$, and $\text{Z} \sim \mathcal{N}(0,1)$.

[2]The FDR Controlling Procedure follows from Theorem 1 in [22], which is restated with minor changes for clarity. If we consider n null hypotheses $h_1, h_2, \ldots, h_n$, and their corresponding p-values $p_1, p_2, \ldots, p_n$. Let $P_1 \leq P_2 \leq \cdots \leq P_n$ be the ordered p-value, and H_i be the corresponding null hypothesis to P_i. The Bonferroni-type multple-testing procedure is as follows: let k be the larges i for which $P_i \leq \frac{i}{n} q*$. Then reject all H_i for $i = 1, 2, \ldots, k$. This procedure controls the FDR at the $q*$ control level.

5.3 Results

5.3.1 Regional gene frequencies

We performed a molecular surveillance study of common β-lactamases among isolates. At DHMMC, the most common ESBL gene we identified was $bla_{\text{CTX-M}}$, followed by bla_{OXA}, bla_{TEM}, and bla_{SHV} (Figure 5.1a). Their yearly frequencies are provided in Table 5.2. Mathematical analysis of those frequencies over time revealed no significant differences over months or agricultural seasons. However, there were some significant differences (p-value < 0.05) in yearly frequencies (Table 5.2). bla_{SHV} and bla_{TEM} frequencies were stable over time. $bla_{\text{CTX-M}}$ frequencies increased after the first year in 2014 and remained stable over time (Table S3). bla_{OXA} frequencies significantly decreased in 2016 from previous years but returned to stable in 2017 (Table S4).



(a) DHMMC (b) National DB

Figure 5.1: **Venn diagrams of bla_{SHV}, bla_{TEM}, $bla_{\text{CTX-M}}$, and bla_{OXA} combinations from both repositories.** (a) Venn diagram of the resistance genes found in the clinical isolates from DHMMC. There were 142 isolates without any of these resistance genes (inconclusive data). (b) Venn diagram of the resistance genes found in the nationwide database of ESBL clinical isolates.

DHMMC	bla_{SHV}	bla_{TEM}	$bla_{\text{CTX-M}}$	bla_{OXA}
2013($n = 106$)	9.4(4.6,16.7)	37.7(28.5,47.7)	52.8(42.9,62.6)	52.8(42.9,62.6)
2014($n = 88$)	8.0(3.3,15.7)	29.5(20.3,40.2)	75.0(64.6,83.6)	54.5(43.6,65.2)
2015($n = 255$)	7.5(4.5,11.4)	29.8(24.3,35.8)	71.4(65.4,76.8)	49.8(43.5,56.1)
2016($n = 207$)	12.6(8.4,17.9)	24.6(18.9,31.1)	62.8(55.8,69.4)	35.7(29.2,42.7)
2017($n = 126$)	9.5(5.0,16.0)	23.8(16.7,32.2)	61.1(52.0,69.7)	36.5(28.1,45.6)
2018($n = 90$)	13.3(7.1,22.1)	28.9(19.8,39.4)	63.3(52.5,73.2)	37.8(27.8,48.6)

Table 5.2: **The yearly frequency of bla_{SHV}, bla_{TEM}, $bla_{\text{CTX-M}}$, and bla_{OXA} from DHMMC.** Each frequency is presented with a 95% confidence interval. The number of isolates is given in the first column in parenthesis.

We then measured the frequencies at which these genes co-occurred in each isolate population. Statistical analysis suggests a genetic linkage (or correlation) between resistance genes in isolates from DHMMC (Table 5.3). Two statistical methods (Pearson's χ^2 test and the Φ) revealed a significant positive correlation between bla_{TEM} and bla_{SHV} (p-value < 0.05) and between $bla_{\text{CTX-M}}$ and bla_{OXA} (p-value < 0.05). There are significant negative correlations between bla_{TEM} and $bla_{\text{CTX-M}}$ (p-value < 0.05) and bla_{TEM} and bla_{OXA} (p-value < 0.05).

Species	**Markers**	$\boldsymbol{\chi^2}$ ***p*-value**	$\boldsymbol{\Phi}$	$\boldsymbol{\Phi}$ ***p*-value**
All ($n = 872$)	bla_{SHV}:bla_{TEM}	3.75×10^{-9}*	0.20	2.74×10^{-9}*
	bla_{SHV}:$bla_{\text{CTX-M}}$	6.37×10^{-1}	0.02	6.37×10^{-1}
	bla_{SHV}:bla_{OXA}	1.08×10^{-1}	0.05	1.08×10^{-1}
	bla_{TEM}:$bla_{\text{CTX-M}}$	1.49×10^{-3}*	-0.11	1.46×10^{-3}*
	bla_{TEM}:bla_{OXA}	3.00×10^{-6}*	-0.16	2.68×10^{-6}*
	$bla_{\text{CTX-M}}$:bla_{OXA}	8.59×10^{-30}*	0.38	5.24×10^{-32}*
Ec ($n = 787$)	bla_{SHV}:bla_{TEM}	5.22×10^{-1}	0.02	5.23×10^{-1}
	bla_{SHV}:$bla_{\text{CTX-M}}$	2.58×10^{-1}	-0.04	2.59×10^{-1}
	bla_{SHV}:bla_{OXA}	4.89×10^{-1}	-0.02	4.89×10^{-1}
	bla_{TEM}:$bla_{\text{CTX-M}}$	1.24×10^{-6}*	-0.17	1.07×10^{-6}*
	bla_{TEM}:bla_{OXA}	1.40×10^{-9}*	-0.22	9.50×10^{-10}*
	$bla_{\text{CTX-M}}$:bla_{OXA}	9.22×10^{-26}*	0.37	1.53×10^{-27}*
Kp ($n = 85$)	bla_{SHV}:bla_{TEM}	3.36×10^{-4}*	0.39	2.34×10^{-4}*
	bla_{SHV}:$bla_{\text{CTX-M}}$	1.97×10^{-5}*	0.46	8.15×10^{-6}*
	bla_{SHV}:bla_{OXA}	6.84×10^{-3}*	0.29	6.44×10^{-3}*
	bla_{TEM}:$bla_{\text{CTX-M}}$	2.58×10^{-5}*	0.46	1.13×10^{-5}*
	bla_{TEM}:bla_{OXA}	3.84×10^{-2}	0.22	3.88×10^{-2}
	$bla_{\text{CTX-M}}$:bla_{OXA}	5.69×10^{-6}*	0.49	1.72×10^{-6}*

Table 5.3: **Linkage analysis summary for DHMMC isolates.** Here *Ec* abbreviates *E. coli*, and *Kp* abbreviates *K. pneumoniae*. The p-value for a χ^2 test for linkage, the phi coefficient (Φ), and the associated p-value are presented for each resistance marker pair comparison. The number of isolates for each species is given in parenthesis. An asterisk (*) indicates a statistically significant comparison after the FDR-controlling procedure ($q* = 0.025$) for both the χ^2 test and the Φ.

5.3.2 U.S. Database Gene Frequencies

We conducted an analogous surveillance study using a nationwide database of ESBL clinical isolates from the NIH Pathogen Detection Isolates Browser (Figure 5.1b). Nationwide, the resistance gene frequencies were different from the DHMMC repository. The most common ESBL gene was bla_{TEM}, followed by bla_{SHV}, $bla_{\text{CTX-M}}$, and bla_{OXA}. Those frequencies are provided in Table 5.4. Analysis of these gene frequencies over time also showed no significant differences in the frequencies of common ESBL genes in the U.S. over a period of months. However, there were significant

differences (p-value < 0.05) in the yearly frequencies for bla_{SHV}, bla_{TEM}, and $bla_{\text{CTX-M}}$ (Table 5.4). bla_{SHV} significantly increased in 2015, followed by a significant decrease in 2016, but returned to stable in 2017 (Table S5). bla_{TEM} frequencies significantly decreased in 2017 and 2018 from previous years (Table S6). There was a non-significant decrease in $bla_{\text{CTX-M}}$ frequency in 2015 from previous years. The 2015 $bla_{\text{CTX-M}}$ frequency was significantly different than that of 2018 due to an increase that year (Table S7).

Nationwide	$\boldsymbol{bla_{\text{SHV}}}$	$\boldsymbol{bla_{\text{TEM}}}$	$\boldsymbol{bla_{\text{CTX-M}}}$	$\boldsymbol{bla_{\text{OXA}}}$
2013($n = 6$)	16.7(0.4,64.1)	83.3(35.9,99.6)	0.0(0.0,45.9)	0.0(0.0,45.9)
2014($n = 179$)	4.5(1.9,8.6)	82.7(76.3,87.9)	14.0(9.2,19.9)	16.8(11.6,23.1)
2015($n = 268$)	66.8(60.8,72.4)	74.6(69.0,79.7)	8.2(5.2,12.2)	9.7(6.4,13.9)
2016($n = 190$)	19.5(14.1,25.8)	84.7(78.8,89.5)	16.8(11.8,22.9)	13.2(8.7,18.8)
2017($n = 251$)	50.6(44.2,56.9)	66.9(60.7,72.7)	13.9(9.9,18.9)	11.6(7.9,16.2)
2018($n = 166$)	50.0(42.2,57.8)	66.9(59.2,74.0)	22.3(16.2,29.4)	18.7(13.1,25.4)

Table 5.4: **The yearly frequency of $\boldsymbol{bla_{\text{SHV}}}$, $\boldsymbol{bla_{\text{TEM}}}$, $\boldsymbol{bla_{\text{CTX-M}}}$, and $\boldsymbol{bla_{\text{OXA}}}$ from the U.S. Nationwide Database.** Each frequency is presented with a 95% confidence interval. The number of isolates is listed in the first column in parenthesis.

In the nationwide database, the co-occurrence of resistance genes differs from that of our local samples from DHMMC (Table 5.5). We observed a negative association between bla_{TEM} and the other resistance markers (p-value < 0.05), and a positive association between $bla_{\text{CTX-M}}$ and bla_{OXA} (p-value < 0.05). The negative association of bla_{TEM} with both $bla_{\text{CTX-M}}$ and bla_{OXA} was almost double that found at DHMMC for all species (Table 5.3 and Table 5.5). When stratified by species, *E. coli* upholds these trends, but in *K. pneumoniae*, bla_{SHV} is negatively correlated with the other resistance genes (p-value < 0.05) and $bla_{\text{CTX-M}}$ is positively correlated bla_{OXA} (p-value < 0.05).

Species	Markers	χ^2 p-value	Φ	Φ p-value
All ($n = 1060$)	bla_{SHV}:bla_{TEM}	7.03×10^{-43}*	-0.42	6.29×10^{-47}*
	bla_{SHV}:$bla_{\text{CTX-M}}$	2.13×10^{-1}	-0.04	2.14×10^{-1}
	bla_{SHV}:bla_{OXA}	2.81×10^{-1}	-0.03	2.81×10^{-1}
	bla_{TEM}:$bla_{\text{CTX-M}}$	1.31×10^{-9}*	-0.19	9.80×10^{-10}*
	bla_{TEM}:bla_{OXA}	3.31×10^{-17}*	-0.26	1.02×10^{-17}*
	$bla_{\text{CTX-M}}$:bla_{OXA}	1.01×10^{-99}*	0.65	7.63×10^{-129}*
Ec ($n = 559$)	bla_{SHV}:bla_{TEM}	2.03×10^{-6}*	-0.20	1.68×10^{-6}*
	bla_{SHV}:$bla_{\text{CTX-M}}$	1.48×10^{-1}	-0.06	1.49×10^{-1}
	bla_{SHV}:bla_{OXA}	1.52×10^{-1}	-0.06	1.52×10^{-1}
	bla_{TEM}:$bla_{\text{CTX-M}}$	6.72×10^{-35}*	-0.52	2.90×10^{-40}*
	bla_{TEM}:bla_{OXA}	1.85×10^{-64}*	-0.72	2.38×10^{-89}*
	$bla_{\text{CTX-M}}$:bla_{OXA}	1.83×10^{-53}*	0.65	1.05×10^{-68}*
Kp ($n = 501$)	bla_{SHV}:bla_{TEM}	6.81×10^{-10}*	-0.28	3.45×10^{-10}*
	bla_{SHV}:$bla_{\text{CTX-M}}$	4.69×10^{-5}*	-0.18	4.23×10^{-5}*
	bla_{SHV}:bla_{OXA}	7.37×10^{-3}*	-0.12	7.30×10^{-3}*
	bla_{TEM}:$bla_{\text{CTX-M}}$	4.73×10^{-1}	0.03	4.74×10^{-1}
	bla_{TEM}:bla_{OXA}	5.68×10^{-1}	0.03	5.69×10^{-1}
	$bla_{\text{CTX-M}}$:bla_{OXA}	4.37×10^{-48}*	0.65	1.04×10^{-61}*

Table 5.5: **Linkage analysis summary for the Nationwide Database Isolates.** Here *Ec* abbreviates *E. coli*, and *Kp* abbreviates *K. pneumoniae*. The p-value for a χ^2 test for linkage, the phi coefficient (Φ), and the associated p-value are presented for each resistance marker pair comparison. The number of isolates for each species is given in parenthesis. An asterisk (*) indicates a statistically significant comparison after the FDR-controlling procedure ($q* = 0.025$) for both the χ^2 test and the Φ.

5.3.3 Comparison of DHMMC and US populations

We then performed a formal statistical comparison between the frequencies of resistance genes in the DHMMC repository and the U.S. database to determine significant differences (Table 5.6). In the nationwide database, bla_{SHV} and bla_{TEM} occur more frequently (all p-values < 0.05) than at DHMMC. At DHMMC, $bla_{\text{CTX-M}}$ and bla_{OXA} occur at higher frequencies than they do nationwide (all p-values < 0.05). A further breakdown of the gene frequencies by species and FDR controlling revealed no significant frequency difference in bla_{SHV} within species between datasets. bla_{TEM} occurs in 57-68% more *E. coli* isolates in the U.S. database than *E. coli* isolates at DHMMC (p-values < 0.05). There is no significant difference in bla_{TEM} frequency in *K. pneumoniae* isolates. $bla_{\text{CTX-M}}$ occurs much more frequently in *E. coli* and *K. pneumoniae* isolates from DHMMC than in the U.S. database (all p-values < 0.05). $bla_{\text{CTX-M}}$ occurs in 47-58% more *E. coli* isolates and 33-52% more *K. pneumoniae* isolates at DHMMC than isolates from the nationwide U.S. database. This trend is similar but less drastic for bla_{OXA}, which occurs in 26-36% more *E. coli* isolates and

26-44% more *K. pneumoniae* isolates than isolates nationwide.

Species	Marker	F_M	F_N	$F_M - F_N$	p-value	95% CI
All	bla_{SHV}	9.9	41.0	-31.2	2.79×10^{-53}*	(-35.2,-27.2)
	bla_{TEM}	28.6	74.8	-46.3	1.34×10^{-91}*	(-50.7,-41.8)
	$bla_{\text{CTX-M}}$	65.1	14.2	50.9	2.40×10^{-117}*	(46.6,55.2)
	bla_{OXA}	44.2	13.3	30.8	6.58×10^{-52}*	(26.9,34.8)
Ec	bla_{SHV}	2.8	2.5	0.3	7.44×10^{-1}	(-1.5,2.0)
	bla_{TEM}	25.9	89.1	-63.2	1.14×10^{-115}	(-68.6,-57.8)
	$bla_{\text{CTX-M}}$	65.8	12.7	53.1	1.75×10^{-83}	(47.7,58.5)
	bla_{OXA}	43.6	12.5	31.1	4.16×10^{-34}	(26.1,36.1)
Kp	bla_{SHV}	75.3	84.0	-8.7	4.86×10^{-2}	(-17.4,-0.1)
	bla_{TEM}	52.9	58.9	-5.9	3.05×10^{-1}	(-17.3,5.4)
	$bla_{\text{CTX-M}}$	58.8	16.0	42.9	1.46×10^{-18}	(33.3,52.4)
	bla_{OXA}	49.4	14.2	35.2	2.65×10^{-14}	(26.2,44.3)

Table 5.6: **Percent frequency differences between resistance markers at DHMMC and the Nationwide Database.** Here *Ec* abbreviates *E. coli*, and *Kp* abbreviates *K. pneumoniae*. The frequency of a resistance marker from DHMMC is denoted FM, and the frequency of a resistance marker from the National Database is denoted by FN. An asterisk (*) indicates a statistically significant comparison after the FDR-controlling procedure ($q* = 0.025$). The last column provides the 95% confidence interval for the percent difference for a particular resistance marker between the two datasets.

5.4 Discussion

We answered our initial question about the relative importance of selection and migration in small and large regions. The stability of resistance genes over time in a distinct community that differ from the nationwide frequencies strongly suggests that local selective pressures have a larger impact on frequencies than migration. DHMMC is unique with regards to the presence of $bla_{\text{CTX-M}}$ and bla_{TEM} genes. At DHMMC, $bla_{\text{CTX-M}}$ occurs more frequently than in the nationwide database, while bla_{TEM} occurs less frequently at DHMMC. The negative correlation at DHMMC between bla_{TEM} and $bla_{\text{CTX-M}}$ and between bla_{TEM} and bla_{OXA} implies incompatibilities between bla_{TEM} and at least one of the other genes. Since $bla_{\text{CTX-M}}$ and bla_{OXA} are commonly linked with each other, it is not surprising that bla_{TEM} is negatively associated with both of them, and genetic incompatibilities may exist for only one of those pairings. As those genetic compatibilities were not observed throughout the U.S., it is likely that they are the product of local selective pressures. This negative relationship results mainly from *E. coli* isolates; this relationship is not observed in *K. pneumoniae* isolates. This result indicates that either the genetic background of *K. pneumoniae* eliminates the incompatibilities of these genes, or that the antibiotic exposures of these pathogens is

different from *E. coli*. Additionally, at DHMMC, there appears to be strong selection for $bla_{\text{CTX-M}}$, which may be displacing bla_{TEM} likely due to antagonism between these genes [119].

In terms of antimicrobial stewardship, our results suggest resistance may be modulated at a regional level, facilitating the implementation of effective strategies to limit and control selection of the antibiotic resistance genes. However, we also found evidence that resistance must be modulated differently for separate species, which may be difficult to conduct because of the environmental presence of antibiotics and the use of empiric therapies without species identification.

High bla_{SHV} frequencies are unique to our nationwide dataset. However, when we considered *E. coli* and *K. pneumoniae* isolates separately, DHMMC and nationwide frequencies of bla_{SHV} are similar. The greater number of *K. pneumoniae* isolates in the nationwide database likely accounts for the overall higher nationwide frequencies of bla_{SHV} (Simpson's paradox [140]). A high bla_{TEM} frequency is also unique to our nationwide dataset. There are 57-68% more *E. coli* isolates with bla_{TEM} nationwide than isolates from DHMMC. *E. coli* isolates nationwide have a negative correlation between bla_{TEM} and the other resistance genes, meaning these isolates are likely to have bla_{TEM} and no other resistance genes. In *K. pneumoniae* isolates nationwide, there is a negative correlation between bla_{SHV} and the other resistance genes, meaning these isolates are likely to have bla_{SHV} and no other resistance genes, which is the opposite of the *K. pneumoniae* isolates from DHMMC. ESBLs were initially derived from bla_{SHV} and bla_{TEM}, explaining the relatively high bla_{TEM} frequencies nationwide, as bla_{TEM} has proceeded to fixation [147]. bla_{SHV} and bla_{TEM} have been responsible for most ESBL infections since at least the 1980s, so it is reasonable for them to be widely distributed in a large-scale dataset [71].

We observed stable gene frequencies over time within the DHMMC population which differ from nationwide frequencies. This implies there is low migration or low survival of immigrants between populations, and that the selective forces within the DHMMC population are strong and constant. Overall, there is strong evidence that local selective pressures have a much stronger effect on the frequencies of ESBLs in local populations. This suggests that communities and specific regions have the potential to effectively manage ESBL frequencies through intentional antibiotic stewardship practices.

5.5 Supplementary Materials

The following are available online at `https://www.mdpi.com/article/10.3390/antibiotics10050498/s1`. Table S1: Pairwise yearly frequency comparison for bla_{SHV} in the DHMMC database, Table S2: Pairwise yearly frequency comparison for bla_{TEM} in the DHMMC database, Table S3: Pairwise yearly frequency comparison for $bla_{\text{CTX-M}}$ in the DHMMC database, Table S4: Pairwise yearly frequency comparison for bla_{OXA} in the DHMMC database, Table S5: Pairwise yearly frequency comparison for bla_{SHV} in the Nationwide database, Table S6: Pairwise yearly frequency comparison for

bla_{TEM} in the Nationwide database, Table S7: Pairwise yearly frequency comparison for $bla_{\text{CTX-M}}$ in the Nationwide database, Table S8: Pairwise yearly frequency comparison for bla_{OXA} in the Nationwide database, Figure S1: 2% agarose gel of $bla_{\text{CTX-M}}$, bla_{TEM}, and bla_{OXA} PCR products, Figure S2: 2% agarose gel of bla_{SHV} PCR products, Table S9: Number of clinical isolates annually by state.

Chapter 6

Conclusions and Future Directions

6.1 Conclusions

In this dissertation, mathematical models, numerical methods, and data driven investigations of inheritable cellular phenotypes were presented. In the first part, Chapters 2 and 3, novel mathematical models and numerical approaches for the dynamics of protein misfolding in yeast cells were presented. In the second part, Chapters 4 and 5, data driven mathematical approaches to evaluate new biological techniques used to study the evolution of bacterial antibiotic resistance were presented. The following summarizes my direct contributions towards the study of prion dynamics in yeast and in the area of bacterial antibiotic resistance by dissertation chapter.

Prion Dynamics

Chapter 2

- I proposed a structured population model describing the intracellular distribution and replication dynamics of yeast prion propagons in an actively dividing population of cells.
- I developed a likelihood formulation that is interpretable in terms of the propagon recovery experiment we are modeling, and detail how it can be applied to estimate the propagon replication rate and their transmission bias during cell division of a prion variant.
- I showed that under our modeling framework, prion dynamics in yeast are best described by a model with an asymmetric transmission of propagons.

Chapter 3

- We develop a novel numerical and theoretical framework involving a recursive formulation for a class of Division and Label Structured Population Models (DLSPMs). I developed this framework for a population of dividing cells with an arbitrary functional form describing the intracellular dynamics.

- I demonstrate that recursive numerical integration can be used to efficiently evaluate the integral form of the DLSPM models.
- I created an open-source code base that can be used to implement the numerical methods using the proposed theoretical framework.

Antibiotic Resistance

Chapters 4 & 5: My contribution to these chapters involved mathematical formalization of questions addressed. I performed, interpreted the results, and wrote the sections on all statistical analyses in these chapters.

6.2 Future Directions

6.2.1 Prion Dynamics

The work presented in Chapters 2 and 3 highlight the importance of descriptive structured population models to understand intracellular processes while capturing colony level dynamics in yeast. However, there are more colony level dynamics that can be incorporated into our modeling framework. Thus, in future investigations we plan to adapt our modeling framework to account for the following processes.

- *Cell maturation.* In yeast cell division, while a mature yeast cell divides about every 90 minutes, a new cell bud requires about 120 minutes to mature before it can begin dividing. In future iteration of the model presented in Chapter 2, we plan to incorporate a maturation process to capture its effects on the transmission bias and replication rate.
- *Cell volume.* In our modeling framework of Chapter 2 we have made an effort to be clear about the difference between *asymmetric cell division* and *asymmetric division of intracellular constituents*, and that our model captures the asymmetric division of intracellular constituents (propagons), but in future models we plan to incorporate cell volume to distinguish between effects due to volume and those due to division bias.
- *More complex intracellular dynamics.* In future investigations we plan to include a more complex intracellular propagon replication than linear replication as presented in Chapter 2. This will be possible by applying the methods presented in Chapter 3.
- *Removing high frequencies in numerical solutions.* Some of our numerical results in Chapter 3 contain high frequencies due to numerically solving integral equations using Gaussian quadrature. In future work, we plan to include an additional post-processing step to filter numerical solutions for high frequencies. This will allow us to use less quadrature points in our recursion while improving the accuracy of our numerical solutions.

6.2.2 Antibiotic Resistance

In this work we highlight the need for better mathematical approaches to quantify bacterial resistance to particular antibiotics and to assess local trends in antibiotic resistance. The goal of this research is to apply statistical and numerical approaches to aid in model building and interpretation to combat antibiotic resistance. In future investigations the following ideas will be explored.

- *A mathematical model to assess the efficacy of treatment strategies to combat antibiotic resistance.* Current mathematical models used to assess the efficacy of treatment strategies for combating antibiotic resistance [58, 32, 19, 23, 53], do not take into account the fact that an individuals' infection can be tested for resistance to particular antibiotics. Therefore, these models fail to consider treatment strategies where an appropriate treatment can eventually be prescribed at the patient level. Development of such a model would allow us to study antibiotic resistance in the context of test turnaround time, laboratory tests for resistance, as well as allow us to study the efficacy of different treatment strategies prior to receiving test results and administration of the appropriate antibiotic.

- *A model to predict antibiotic resistance.* The goal is to create a machine learning model to predict the probability of resistance to particular antibiotics using genetic information. While methods for detecting antibiotic resistance are available, they are time intensive (results typically take multiple days) and during that time the infection can progress within the patient. These predictive methods in combination with rapid molecular polymerase chain reaction (PCR) tests (results can typically take less than an hour) could lead to faster methods for detecting antibiotic resistance and application of the correct antibiotic for treatment.

References

Bibliography

[1] Ahmed, M., W. N. W. Long, S. Javed, and T. Reynolds (2021). Rising resistance of urinary tract pathogens in children: a cause for concern. *Irish Journal of Medical Science (1971-)*, 1–4.

[2] Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control 19*(6), 716–723.

[3] Alcock, B. P., A. R. Raphenya, T. T. Lau, K. K. Tsang, M. Bouchard, A. Edalatmand, W. Huynh, A.-L. V. Nguyen, A. A. Cheng, S. Liu, et al. (2020). Card 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic acids research 48*(D1), D517–D525.

[4] Ammon, A., A. Genz, T. Hartung, K. Jansen, H. Leövey, and J. Volmer (2016). Applying recursive numerical integration techniques for solving high dimensional integrals. *PoS LATTICE2016*, 335.

[5] Andrews, J. M. (2001, 7). Determination of minimum inhibitory concentrations. *Journal of Antimicrobial Chemotherapy 48*, 5–16.

[6] Association, A. (2019). 2019 alzheimer's disease facts and figures. *Alzheimer's & dementia 15*(3), 321–387.

[7] Bajpai, T., M. Pandey, M. Varma, and G. Bhatambare (2017). Prevalence of tem, shv, and ctx-m beta-lactamase genes in the urinary isolates of a tertiary care hospital. *Avicenna journal of medicine 7*(1), 12.

[8] Banks, H., K. B. Flores, C. R. Langlois, T. R. Serio, and S. S. Sindi (2018). Estimating the rate of prion aggregate amplification in yeast with a generation and structured population model. *Inverse Problems in Science and Engineering 26*(2), 257–279.

[9] Banks, H., K. L. Sutton, W. C. Thompson, G. Bocharov, D. Roose, T. Schenkel, and A. Meyerhans (2011). Estimation of cell proliferation dynamics using cfse data. *Bulletin of mathematical biology 73*(1), 116–150.

[10] Banks, H. T., K. B. Flores, and S. Sindi (2016). On analytical and numerical approaches to division and label structured population models. *Applied Mathematics Letters 60*, 81–88.

[11] Banks, H. T., K. L. Sutton, W. C. Thompson, G. Bocharov, M. Doumic, T. Schenkel, J. Argilaguet, S. Giest, C. Peligero, and A. Meyerhans (2011). A new model for the estimation of cell proliferation dynamics using cfse data. *Journal of immunological methods 373*(1), 143–160.

[12] Banks, H. T., W. C. Thompson, C. Peligero, S. Giest, J. Argilaguet, and A. Meyerhans (2012, October). A division-dependent compartmental model for computing cell numbers in CFSE-based lymphocyte proliferation assays. *Mathematical biosciences and engineering: MBE 9*(4), 699–736.

[13] Banks, H. T. and H. T. Tran (2009, January). *Mathematical and Experimental Modeling of Physical and Biological Processes*. CRC Press. Google-Books-ID: SSRapIe8p3QC.

[14] Banwarth-Kuhn, M., J. Collignon, and S. Sindi (2020). Quantifying the biophysical impact of budding cell division on the spatial organization of growing yeast colonies. *Applied Sciences 10*(17), 5780.

[15] Barlow, M. and B. G. Hall (2002, 9). Phylogenetic analysis shows that the oxa b-lactamase genes have been on plasmids for millions of years. *Journal of Molecular Evolution 55*, 314–321.

[16] Barlow, M. and B. G. Hall (2003, 4). Experimental prediction of the natural evolution of antibiotic resistance. *Genetics 163*, 1237–1241.

[17] Bartholomew, J. W. and T. Mittwer (1952). The gram stain. *Bacteriological reviews 16*, 1.

[18] Baym, M., T. D. Lieberman, E. D. Kelsic, R. Chait, R. Gross, I. Yelin, and R. Kishony (2016, 9). Spatiotemporal microbial evolution on antibiotic landscapes. *Science 353*, 1147–1151.

[19] Beardmore, R. E. (2010). Antibiotic cycling versus mixing: the difficulty of using mathematical models to definitively quantify their relative merits. *Mathematical Biosciences & Engineering 7*(4), 923–933.

[20] Ben-Ami, R., M. Schwaber, S. Navon-Venezia, D. Schwartz, M. Giladi, I. Chmelnitsky, A. Leavitt, and Y. Carmeli (2006). Influx of extended-spectrum β-lactamase—producing enterobacteriaceae into the hospital. *Clinical infectious diseases 42*(7), 925–934.

[21] Benjamini, Y., D. Drai, G. Elmer, N. Kafkafi, and I. Golani (2001, 11). Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research 125*, 279–284.

[22] Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological) 57*(1), 289–300.

[23] Bergstrom, C. T., M. Lo, and M. Lipsitch (2004). Ecological theory suggests that antimicrobial cycling will not reduce antimicrobial resistance in hospitals. *Proceedings of the National Academy of Sciences of the United States of America 101*(36), 13285–13290.

[24] Blake, W. J., M. Kærn, C. R. Cantor, and J. J. Collins (2003). Noise in eukaryotic gene expression. *Nature 422*(6932), 633–637.

[25] Bonnet, R. (2004, 1). Growing group of extended-spectrum β-lactamases: the ctx-m enzymes. *Antimicrobial Agents and Chemotherapy 48*, 1.

[26] Byrne, L. J., D. J. Cole, B. S. Cox, M. S. Ridout, B. Morgan, and M. F. Tuite (2009). The number and transmission of $[PSI^+]$ prion seeds (propagons) in the yeast saccharomyces cerevisiae. *PLoS One 4*(3), e4670–e4670.

[27] Cantón, R., J. M. González-Alba, and J. C. Galán (2012). Ctx-m enzymes: Origin and diffusion. *Frontiers in Microbiology 3*, 110.

[28] Casella, G., C. P. Robert, M. T. Wells, et al. (2004). Generalized accept-reject sampling schemes. In *A festschrift for herman rubin*, pp. 342–347. Institute of Mathematical Statistics.

[29] Castanheira, M., S. E. Farrell, L. M. Deshpande, R. E. Mendes, and R. N. Jones (2013). Prevalence of β-lactamase-encoding genes among enterobacteriaceae bacteremia isolates collected in 26 us hospitals: report from the sentry antimicrobial surveillance program (2010). *Antimicrobial agents and chemotherapy 57*(7), 3012–3020.

[30] CDC (2020, Mar). *About Antibiotic Resistance*. `https://www.cdc.gov/drugresistance/about.html`.

[31] Chihara, L. M. and T. C. Hesterberg (2018). *Mathematical statistics with resampling and R*. John Wiley & Sons.

[32] Chow, K., X. Wang, R. Curtiss, and C. Castillo-Chavez (2011). Evaluating the efficacy of antimicrobial cycling programmes and patient isolation on dual resistance in hospitals. *Journal of Biological Dynamics 5*(1), 27–43.

[33] Clatworthy, A. E., E. Pierson, and D. T. Hung (2007). Targeting virulence : a new paradigm for antimicrobial therapy. *Nature chemical biology 3*(9), 541–548.

[34] Cole, D. J., B. J. Morgan, M. S. Ridout, L. J. Byrne, and M. F. Tuite (2004). Estimating the number of prions in yeast cells. *Mathematical Medicine and Biology 21*(4), 369–395.

[35] Cole, D. J., M. S. Ridout, B. J. T. Morgan, L. J. Byrne, and M. F. Tuite (2007). Approximations for Expected Generation Number. *Biometrics 63*(4), 1023–1030.

[36] Collinge, J. (2001). Prion diseases of humans and animals: their causes and molecular basis. *Annual review of neuroscience 24*(1), 519–550.

[37] Cowles, M. K. and B. P. Carlin (1996). Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association 91*(434), 883–904.

[38] Cox, B. (1965). *psi*, a cytoplasmic suppressor of super-suppressor in yeast. *Heredity 20*(4), 505.

[39] Critchley, I. A., N. Cotroneo, M. J. Pucci, and R. Mendes (2019). The burden of antimicrobial resistance among urinary tract isolates of escherichia coli in the united states in 2017. *PloS one 14*(12), e0220265.

[40] Dallenne, C., A. Da Costa, D. Decré, C. Favier, and G. Arlet (2010). Development of a set of multiplex pcr assays for the detection of genes encoding important β-lactamases in enterobacteriaceae. *Journal of Antimicrobial Chemotherapy 65*(3), 490–495.

[41] Dantas, G. and M. O. Sommer (2014). How to fight back against antibiotic resistance. *Am Sci 102*(1), 42–51.

[42] Datta, N., P. Kontomichalou, et al. (1965). Penicillinase synthesis controlled by infectious r factors in enterobacteriaceae. *Nature 208*, 239–41.

[43] Davis, J. K. and S. S. Sindi (2015). A study in nucleated polymerization models of protein aggregation. *Applied Mathematics Letters 40*, 97–101.

[44] Davis, J. K. and S. S. Sindi (2016). A mathematical model of the dynamics of prion aggregates with chaperone-mediated fragmentation. *Journal of mathematical biology 72*(6), 1555–1578.

[45] De Boer, R. J., V. V. Ganusov, D. Milutinović, P. D. Hodgkin, and A. S. Perelson (2006). Estimating lymphocyte division and death rates from cfse data. *Bulletin of mathematical biology 68*(5), 1011–1031.

[46] De Boer, R. J. and A. S. Perelson (2005). Estimating division and death rates from cfse data. *Journal of computational and applied mathematics 184*(1), 140–164.

[47] Deenick, E. K., A. V. Gett, and P. D. Hodgkin (2003). Stochastic model of t cell proliferation: a calculus revealing il-2 regulation of precursor frequencies, cell cycle time, and survival. *The Journal of Immunology 170*(10), 4963–4972.

[48] Dellus-Gur, E., M. Elias, E. Caselli, F. Prati, M. L. Salverda, J. A. G. de Visser, J. S. Fraser, and D. S. Tawfik (2015, 7). Negative epistasis and evolvability in tem-1 β-lactamase—the thin line between an enzyme's conformational freedom and disorder. *Journal of Molecular Biology 427*, 2396–2409.

[49] Derdowski, A., S. S. Sindi, C. L. Klaips, S. DiSalvo, and T. R. Serio (2010). A size threshold limits prion transmission and establishes phenotypic diversity. *Science 330*(6004), 680–683.

[50] Dunlap, W. P., C. J. Brody, and T. Greer (2000). Canonical correlation and chi-square: Relationships and interpretation. *The Journal of General Psychology 127*(4), 341–353. PMID: 11109997.

[51] Edelstein-Keshet, L. (2005). *Mathematical models in biology*. SIAM.

[52] Evans, B. A. and S. G. B. Amyes (2014, 4). Oxa -lactamases. *Clinical Microbiology Reviews 27*, 241.

[53] Evin, B. R. R. L. (1997). Evaluating treatment protocols to prevent antibiotic resistance. *Proceedings of the National Academy of Sciences 94*(October), 12106–12111.

[54] Flores, K. B. (2013, July). A structured population modeling framework for quantifying and predicting gene expression noise in flow cytometry data. *Applied Mathematics Letters 26*(7), 794–798.

[55] Ganusov, V. V., S. S. Pilyugin, R. J. de Boer, K. Murali-Krishna, R. Ahmed, and R. Antia (2005). Quantifying cell turnover using cfse data. *Journal of immunological methods 298*(1-2), 183–200.

[56] Gett, A. V. and P. D. Hodgkin (2000, September). A cellular calculus for signal integration by T cells. *Nature Immunology 1*(3), 239–244.

[57] Glauche, I., K. Moore, L. Thielecke, K. Horn, M. Loeffler, and I. Roeder (2009). Stem cell proliferation and quiescence-two sides of the same coin. *PLoS Comput Biol 5*(7), e1000447.

[58] Gori, F., J. Iredell, R. E. Beardmore, and R. Pe (2017). Antibiotic Cycling and Antibiotic Mixing : Which One Best Mitigates Antibiotic Resistance? Article Fast Track. *Molecular biology and evolution 34*(4), 802–817.

[59] Goulart, C. P., M. Mahmudi, K. A. Crona, S. D. Jacobs, M. Kallmann, B. G. Hall, D. C. Greene, and M. Barlow (2013, 2). Designing antibiotic cycling strategies by determining and understanding local adaptive landscapes. *PLoS ONE 8*, e56040.

[60] Greer, M. L., L. Pujo-Menjouet, and G. F. Webb (2006). A mathematical analysis of the dynamics of prion proliferation. *Journal of Theoretical Biology 242*(3), 598–606.

[61] Griffiths, J. F., A. J. Griffiths, S. R. Wessler, R. C. Lewontin, W. M. Gelbart, D. T. Suzuki, J. H. Miller, et al. (2005). *An introduction to genetic analysis*. Macmillan.

[62] Gupta, N., B. M. Limbago, J. B. Patel, and A. J. Kallen (2011, 7). Carbapenem-resistant enterobacteriaceae: Epidemiology and prevention. *Clinical Infectious Diseases 53*(1), 60–67.

[63] Haario, H., E. Saksman, J. Tamminen, et al. (2001). An adaptive metropolis algorithm. *Bernoulli 7*(2), 223–242.

[64] Hall, B. G. (2004, 5). Predicting the evolution of antibiotic resistance genes. *Nature Reviews Microbiology 2*, 430–435.

[65] Hall, B. G., H. Acar, A. Nandipati, and M. Barlow (2014). Growth rates made easy. *Molecular biology and evolution 31*(1), 232–238.

[66] Hasenauer, J., D. Schittler, and F. Allgöwer (2012). Analysis and simulation of division-and label-structured population models. *Bulletin of mathematical biology 74*(11), 2692–2732.

[67] Hastings, P., S. M. Rosenberg, and A. Slack (2004). Antibiotic-induced lateral transfer of antibiotic resistance. *Trends in microbiology 12*(9), 401–404.

[68] Head, M. W. (2014). *Human prion diseases*. Cambridge University Press.

[69] Hurvich, C. M. and C.-L. Tsai (1989, 06). Regression and time series model selection in small samples. *Biometrika 76*(2), 297–307.

[70] Kahlmeter, G. and H. O. Poulsen (2012). Antimicrobial susceptibility of escherichia coli from community-acquired urinary tract infections in europe: the eco·sens study revisited. *International journal of antimicrobial agents 39*(1), 45–51.

[71] Kalp, M., C. R. Bethel, R. A. Bonomo, and P. R. Carey (2009). Why the extended-spectrum β-lactamases shv-2 and shv-5 are "hypersusceptible" to mechanism-based inhibitors. *Biochemistry 48*(41), 9912–9920.

[72] Kenz, Z. R., H. T. Banks, and R. C. Smith (2013, January). Comparison of Frequentist and Bayesian Confidence Analysis Methods on a Viscoelastic Stenosis Model. *SIAM/ASA Journal on Uncertainty Quantification 1*(1), 348–369. Publisher: Society for Industrial and Applied Mathematics.

[73] Komsta, L. and F. Novomestky (2015). Moments, cumulants, skewness, kurtosis and related tests. *R package version 14*, 1.

[74] Kryndushki, D., I. Alexandrov, M. Ter-Avanesyan, and V. Kushnirov (2003). Yeast [PSI^+] Prion Aggregates Are Formed by Small Sup35 Polymers Fragmented by Hsp104. *Journal of Biological Chemistry 278*(49), 49636–43.

[75] Lancaster, A. K., J. P. Bardill, H. L. True, and J. Masel (2010). The spontaneous appearance rate of the yeast prion [psi+] and its implications for the evolution of the evolvability properties of the [psi+] system. *Genetics 184*(2), 393–400.

[76] Langlois, C. R., F. Pei, S. S. Sindi, and T. R. Serio (2016). Distinct prion domain sequences ensure efficient amyloid propagation by promoting chaperone binding or processing in vivo. *PLoS genetics 12*(11), e1006417.

[77] Larramendy, S., A. Gaultier, J.-P. Fournier, J. Caillon, L. Moret, and F. Beaudeau (2021). Local characteristics associated with higher prevalence of esbl-producing escherichia coli in community-acquired urinary tract infections: an observational, cross-sectional study. *Journal of Antimicrobial Chemotherapy 76*(3), 789–795.

[78] Leflon-Guibout, V., C. Jurand, S. Bonacorsi, F. Espinasse, M. C. Guelfi, F. Duportail, B. Heym, E. Bingen, and M.-H. Nicolas-Chanoine (2004, 10). Emergence and spread of three clonally related virulent isolates of ctx-m-15-producing escherichia coli with variable resistance to aminoglycosides and tetracycline in a french geriatric hospital. *Antimicrobial Agents and Chemotherapy 48*, 3736.

[79] Lemarre, P., L. Pujo-Menjouet, and S. S. Sindi (2019). Generalizing a mathematical model of prion aggregation allows strain coexistence and co-stability by including a novel misfolded species. *Journal of mathematical biology 78*(1), 465–495.

[80] Lerminiaux, N. A. and A. D. Cameron (2019). Horizontal transfer of antibiotic resistance genes in clinical environments. *Canadian journal of microbiology 65*(1), 34–44.

[81] Liakopoulos, A., D. Mevius, and D. Ceccarelli (2016). A review of shv extended-spectrum β-lactamases: neglected yet ubiquitous. *Frontiers in microbiology 7*, 1374.

[82] Livermore, D. M., R. Canton, M. Gniadkowski, P. Nordmann, G. M. Rossolini, G. Arlet, J. Ayala, T. M. Coque, I. Kern-Zdanowicz, F. Luzzaro, L. Poirel, and N. Woodford (2006, 11). Ctx-m: changing the face of esbls in europe. *Journal of Antimicrobial Chemotherapy 59*, 165–174.

[83] Luzyanina, T., J. Cupovic, B. Ludewig, and G. Bocharov (2014). Mathematical models for cfse labelled lymphocyte dynamics: asymmetry and time-lag in division. *Journal of mathematical biology 69*(6), 1547–1583.

[84] Luzyanina, T., D. Roose, and G. Bocharov (2009). Distributed parameter identification for a label-structured cell population dynamics model using cfse histogram time-series data. *Journal of mathematical biology 59*(5), 581–603.

[85] Luzyanina, T., D. Roose, T. Schenkel, M. Sester, S. Ehl, A. Meyerhans, and G. Bocharov (2007). Numerical modelling of label-structured cell population growth using cfse distribution data. *Theoretical Biology and Medical Modelling 4*(1), 26.

[86] Lyons, A. B., J. Hasbold, and P. D. Hodgkin (2001). Flow cytometric analysis of cell division history using dilution of carboxyfluorescein diacetate succinimidyl ester, a stably integrated fluorescent probe. In *Essential Cytometry Methods*, pp. 375–398. Wiley New York.

[87] Lyons, A. B. and C. R. Parish (1994, May). Determination of lymphocyte division by flow cytometry. *Journal of Immunological Methods 171*(1), 131–137.

[88] Maalej, S., M. Meziou, F. Rhimi, and A. Hammami (2011, 11). Comparison of disc diffusion, etest and agar dilution for susceptibility testing of colistin against enterobacteriaceae. *Letters in Applied Microbiology 53*, 546–551.

[89] MacLean, R. C., A. R. Hall, G. G. Perron, and A. Buckling (2010). The population genetics of antibiotic resistance: integrating molecular mechanisms and treatment contexts. *Nature Reviews Genetics 11*(6), 405–414.

[90] Markovska, R., I. Schneider, E. Keuleyan, and A. Bauernfeind (2004). Extended-spectrum β-lactamase (esbl) ctx-m-15-producing escherichia coli and klebsiella pneumoniae in sofia, bulgaria. *Clinical microbiology and infection 10*(8), 752–755.

[91] Marx, V. (2019). A dream of single-cell proteomics. *Nature methods 16*(9), 809–812.

[92] MathWorks (2020). *Statistics and Machine Learning Toolbox^{TM} R2020a*. `https://www.mathworks.com/help/stats/`.

[93] Mendes, R. E., R. N. Jones, L. N. Woosley, V. Cattoir, and M. Castanheira (2019). Application of next-generation sequencing for characterization of surveillance and clinical trial isolates: Analysis of the distribution of β-lactamase resistance genes and lineage background in the united states. In *Open forum infectious diseases*, Volume 6, pp. S69–S78. Oxford University Press US.

[94] Mira, P. M., K. Crona, D. Greene, J. C. Meza, B. Sturmfels, and M. Barlow (2015, 5). Rational design of antibiotic treatment plans: A treatment strategy for managing evolution and reversing resistance. *PLOS ONE 10*, e0122283.

[95] Mira, P. M., J. C. Meza, A. Nandipati, and M. Barlow (2015, 10). Adaptive landscapes of resistance genes change as antibiotic concentrations change. *Molecular Biology and Evolution 32*, 2707–2715.

[96] Morgan, B. J. T., M. S. Ridout, and L. W. Ruddock (2003). Models for Yeast Prions. *Biometrics 59*(3), 562–569.

[97] Naas, T. and P. Nordmann (1999). Oxa-type -lactamases. *Current pharmaceutical design 5*, 865–880.

[98] NCBI (2013). *Isolates Browser–Pathogen Detection*. `https://www.ncbi.nlm.nih.gov/pathogens/isolates/`.

[99] Norén, K., L. Carmichael, E. Fuglei, N. E. Eide, P. Hersteinsson, and A. Angerbjörn (2011). Pulses of movement across the sea ice: population connectivity and temporal genetic structure in the arctic fox. *Oecologia 166*(4), 973–984.

[100] O'Leary, N. A., M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, and K. D. Pruitt (2015, 11). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research 44*(D1), D733–D745.

[101] O'Neill, J. (2014). *Antimicrobial resistance: tackling a crisis for the health and wealth of nations*. Review on Antimicrobial Resistance.

[102] Pandey, N. and M. Cascella (2020). Beta lactam antibiotics. `https://www.ncbi.nlm.nih.gov/books/NBK545311/`.

[103] Parish, C. R. (1999). Fluorescent dyes for lymphocyte migration and proliferation studies. *Immunology and cell biology 77*(6), 499–508.

[104] Paushkin, S., V. Kushnirov, V. Smirnov, and M. Ter-Avanesyan (1996). Propagation of the yeast prion-like [PSI^+] determinant is mediated by oligomerization of the sup35-encoded polypeptide chain release factor. *The EMBO journal 15*(12), 3127–3134.

[105] Pei, F., S. DiSalvo, S. S. Sindi, and T. R. Serio (2017). A dominant-negative mutant inhibits multiple prion variants through a common mechanism. *PLoS genetics 13*(10), e1007085.

[106] Perron, G. G., A. Gonzalez, and A. Buckling (2007). Source–sink dynamics shape the evolution of antibiotic resistance and its pleiotropic fitness cost. *Proceedings of the Royal Society B: Biological Sciences 274*(1623), 2351–2356.

[107] Peña-Miller, R., A. Fuentes-Hernandez, C. Reding, I. Gudelj, and R. Beardmore (2014, 7). Testing the optimality properties of a dual antibiotic treatment in a two-locus, two-allele model. *Journal of The Royal Society Interface 11*, 20131035.

[108] Pinchover, Y. and J. Rubinstein (2005). *An introduction to partial differential equations*. Cambridge university press.

[109] Pitton, J.-S. (1972). Mechanisms of bacterial resistance to antibiotics. *Ergebnisse der Physiologie Reviews of Physiology, Volume 65 65*, 15–93.

[110] Poirel, L., T. Naas, and P. Nordmann (2010). Diversity, epidemiology, and genetics of class d β-lactamases. *Antimicrobial agents and chemotherapy 54*(1), 24–38.

[111] Posada, D. (2009). *Bioinformatics for DNA sequence analysis*. Springer.

[112] Prusiner, S. (1982). Novel proteinaceous infectious particles cause scrapie. *Science 216*(4542), 136–144.

[113] Prusiner, S. B. (1991). Molecular biology of prion diseases. *Science 252*(5012), 1515–1522.

[114] Pruss, J., L. Pujo-Menjouet, G. Webb, and R. Zacher (2006, January). Analysis of a model for the dynamics of prions. *Discrete and Continuous Dynamical Systems - Series B 6*(1), 225–235.

[115] Qian, G. and A. Mahdi (2020). Sensitivity analysis methods in the biomedical sciences. *Mathematical biosciences 323*, 108306.

[116] Quah, B. J., H. S. Warren, and C. R. Parish (2007). Monitoring lymphocyte proliferation in vitro and in vivo with the intracellular fluorescent dye carboxyfluorescein diacetate succinimidyl ester. *Nature protocols 2*(9), 2049.

[117] Roberts, G. O. and J. S. Rosenthal (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science 16*(4), 351 – 367.

[118] Ruxton, G. D. (2006, 7). The unequal variance t-test is an underused alternative to student's t-test and the mann–whitney u test. *Behavioral Ecology 17*, 688–690.

[119] Santiago, F., E. Doscher, J. Kim, M. Camps, J. Meza, S. Sindi, and M. Barlow (2020). Growth rate assays reveal fitness consequences of β-lactamases. *PloS one 15*(1), e0228240.

[120] Santiago, F., S. Sindi, and K. Flores (2020). Numerical approaches to division and label structured population models. *Letters in Biomathematics 7*(1), 153–170.

[121] Segel, L. A. and L. Edelstein-Keshet (2013). *A Primer in Mathematical Models in Biology*, Volume 129. Siam.

[122] Seigal, A., P. Mira, B. Sturmfels, and M. Barlow (2017, 1). Does antibiotic resistance evolve in hospitals? *Bulletin of Mathematical Biology 79*, 191–208.

[123] Shampine, L. F. (2005). Solving hyperbolic pdes in matlab. *Applied Numerical Analysis & Computational Mathematics 2*(3), 346–358.

[124] Sindi, S. S. (2017). Mathematical Modeling of Prion Disease. In *Prion-An Overview*, pp. 207–227. InTech.

[125] Sindi, S. S. and P. Olofsson (2013). A discrete-time branching process model of yeast prion curing curves. *Mathematical Population Studies 20*(1), 1–13.

[126] Sindi, S. S. and T. R. Serio (2009). Prion dynamics and the quest for the genetic determinant in protein-only inheritance. *Current opinion in microbiology 12*(6), 623–630.

[127] Smith, J. and L. Martin (1973). Do cells cycle? *Proceedings of the National Academy of Sciences 70*(4), 1263–1267.

[128] Smith, R. C. (2014). *Uncertainty quantification: theory, implementation, and applications*. Philadelphia: SIAM. OCLC: 875327904.

[129] Soto, C. (2003). Unfolding the role of protein misfolding in neurodegenerative diseases. *Nature Reviews Neuroscience 4*(1), 49–60.

[130] Sugumar, M., K. Kumar, A. Manoharan, A. Anbarasu, and S. Ramaiah (2014). Detection of oxa-1 β-lactamase gene of klebsiella pneumoniae from blood stream infections (bsi) by conventional pcr and in-silico analysis to understand the mechanism of oxa mediated resistance. *PLoS One 9*(3), e91800.

[131] Summers, A. O. (2006). Genetic linkage and horizontal gene transfer, the roots of the antibiotic multi-resistance problem. *Animal biotechnology 17*(2), 125–135.

[132] Tamma, P. D., S. L. Sharara, Z. D. Pana, J. Amoah, S. L. Fisher, T. Tekle, Y. Doi, and P. J. Simner (2019). Molecular epidemiology of ceftriaxone-nonsusceptible enterobacterales isolates in an academic medical center in the united states. In *Open forum infectious diseases*, Volume 6, pp. ofz353. Oxford University Press US.

[133] Tanaka, M., S. R. Collins, B. H. Toyama, and J. S. Weissman (2006). The physical basis of how prion conformations determine strain phenotypes. *Nature 442*(7102), 585–589.

[134] Team, R. C. et al. (2013). R: A language and environment for statistical computing.

[135] Tuite, M. and B. Cox (2003, November). Propagation of yeast prions. *Nature Reviews Molecular Cell Biology 4*, 878–889.

[136] Tuite, M. and T. Serio (2010, December). The prion hypothesis : from biological anomaly to regulatory mechanism. *Nature Reviews Molecular Cell Biology 11*(12), 823–833.

[137] Tuite, M. F. and B. S. Cox (2006). The [PSI+] prion of yeast: a problem of inheritance. *Methods 39*(1), 9–22.

[138] Villali, J., J. Dark, T. M. Brechtel, F. Pei, S. S. Sindi, and T. R. Serio (2020). Nucleation seed size determines amyloid clearance and establishes a barrier to prion appearance in yeast. *Nature structural & molecular biology 27*(6), 540–549.

[139] von Winckel, G. (2004). Legendre-gauss quadrature weights and nodes. `http://www.mathworks.com/matlabcentral/fileexchange/4540`.

[140] Wagner, C. H. (1982). Simpson's paradox in real life. *The American Statistician 36*(1), 46–48.

[141] Wallace, P. K., J. D. Tario Jr, J. L. Fisher, S. S. Wallace, M. S. Ernstoff, and K. A. Muirhead (2008). Tracking antigen-driven responses by flow cytometry: monitoring proliferation by dye dilution. *Cytometry Part A 73*(11), 1019–1034.

[142] Weideman, J. and D. Laurie (2000). Quadrature rules based on partial fraction expansions. *Numerical Algorithms 24*(1-2), 159.

[143] Weinreich, D. M. (2006, 4). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science 312*, 111–114.

[144] Wickner, R. B. (1994). [*URE*3] as an altered *URE*2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science 264*(5158), 566–569.

[145] Wilke, M. S., A. L. Lovering, and N. C. Strynadka (2005). β-lactam antibiotic resistance: a current structural perspective. *Current opinion in microbiology 8*(5), 525–533.

[146] Witkowski, J. M. (2008). Advanced application of cfse for cellular tracking. *Current Protocols in Cytometry 44*(1), 9–25.

[147] Woerther, P.-L., C. Angebault, M. Lescat, E. Ruppé, D. Skurnik, A. El Mniai, O. Clermont, H. Jacquier, A. Da Costa, M. Renard, et al. (2010). Emergence and dissemination of extended-spectrum β-lactamase-producing escherichia coli in the community: lessons from the study of a remote and controlled population. *The Journal of infectious diseases 202*(4), 515–523.

[148] Woldringh, C., P. Huls, and N. Vischer (1993). Volume growth of daughter and parent cells during the cell cycle of saccharomyces cerevisiae a/alpha as determined by image cytometry. *Journal of bacteriology 175*(10), 3174–3181.

[149] Zowawi, H. M., P. N. A. Harris, M. J. Roberts, P. A. Tambyah, M. A. Schembri, M. D. Pezzani, D. A. Williamson, and D. L. Paterson (2015, 10). The emerging threat of multidrug-resistant gram-negative bacteria in urology. *Nature Reviews Urology 12*, 570.