Mathematical Modelling of Bacteria and Phage: Coevolution, Ecology and Stochastic Decision Making

A thesis presented for the degree of
Doctor of Philosophy of Imperial College London

by

Matthew Lee Robb

Department of Mathematics
Imperial College
180 Queen’s Gate, London SW7 2BZ

FEBRUARY 9, 2014
I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Matthew Lee Robb
Copyright

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.
To my family.
Abstract

The aim of this thesis is to use mathematical models to develop understanding of bacteria and phage interactions. The work focused on both population scale interactions and analysis at the cellular level. Models developed in this thesis reveal the importance of connecting work at the single cell level and population scale level and also the significance of global cell effects of noise in cellular systems.

At the population level, the effects of gene flow on diversity in coevolving bacteria and phage were analysed. It was found experimentally that the effects of gene flow on diversity depend on the direction of gene flow. Through a deterministic model it was found that this conclusion is dependent on both the rate of gene flow and the genetic interaction. Recent experimental work giving additional information on the factors affecting the rate of choosing lysogeny in lambda phage enabled us to revisit an ecological model to understand the reasons for being temperate. This was carried out using bifurcation analysis and numerically solving differential equations. At the single cell level, stochastic modelling of genetic networks was used to develop a mechanistic understanding of decision making in lambda phage. It was found using a simple representation of the genetic switch that the effects of intrinsic noise can largely explain experimental observations on the dependence of rate of lysogeny on the number of infecting phage. Further analysis revealed that there are also possible contributions from spatial and cell cycle effects. Stochastic models were also used to investigate the effects of random partitioning at cell partitioning, generation time and cell size on protein noise. Finally, the effects of growth rate on global cell parameters was investigated for simple genetic circuits including the phage genetic switch.
Acknowledgements

I would like to thank my supervisors Vahid Shahrezaei and Ivana Gudelj for their support throughout my studies. Thank you also to collaborators Samantha Forde, Michael Sieber, Lanying Zeng and Ido Golding. Thank you to the organisers of the weekly Biomath seminars for regularly hosting interesting talks. While I would also like to thank John Gibbons and Emma McCoy for their support. Funding from the Natural Environment Research Council (NERC) has made this research possible. Finally, I would like to thank my family for their love and immeasurable support.

Matthew Lee Robb
# Table of contents

Abstract ................................. 5

List of Publications .................... 14

1 Introduction ................................ 15
   1.1 General Introduction .................. 15
   1.2 Bacteria-Phage Interaction ............ 17
       1.2.1 Introduction to Bacteria and Phage .. 17
   1.3 Ecology and Coevolution ............... 20
   1.4 Stochasticity in Cellular Systems .... 23

2 Methods ................................ 27
   2.1 Introduction .......................... 27
   2.2 ODE Models of Ecology and Coevolution . 27
   2.3 Bifurcation Analysis ................... 28
   2.4 Linear Noise Approximation .......... 28
   2.5 Modelling Extrinsic Noise ............ 34
   2.6 Cellular Image Processing ............. 37
   2.7 Stochastic Simulations ............... 39
       2.7.1 Non-Spatial Simulations .......... 39
       2.7.2 Spatial Simulations ............... 40

3 Models of Bacteria-Phage Coevolution .... 41
   3.1 Overview ................................ 41
   3.2 Introduction .......................... 41
   3.3 Background Model ...................... 44
   3.4 Gene Flow Problem ..................... 48
       3.4.1 Experimental Methods .............. 48
       3.4.2 Model ............................ 50
   3.5 Results ................................ 51
       3.5.1 Experimental Results .............. 51
       3.5.2 Model Results ..................... 54
       3.5.3 Direction ........................ 54
       3.5.4 Infection Mechanism ............... 58
       3.5.5 No Resource Transfer .............. 60
   3.6 Discussion ........................... 61
# 4 Population Scale Model of Lambda Phage

4.1 Overview ................................................. 64
4.2 Introduction ............................................ 64
4.3 Model .................................................. 66
4.4 Results ................................................ 68
   4.4.1 Lytic Phage and Bacteria Dynamics ...................... 68
   4.4.2 Bifurcation Analysis - Temperate Phage with Constant Rate of Lysogeny ....................................... 69
   4.4.3 Bifurcation Analysis - Temperate Phage with Responsive Rate of Lysogeny ........................................ 75
   4.4.4 Fluctuating Conditions ................................. 79
4.5 Conclusion ............................................. 85

# 5 Stochastic Cellular Fate Decision Making by Individual Lambda Bacteriophage

5.1 Overview ................................................. 89
5.2 Introduction ............................................ 90
5.3 Model .................................................. 91
5.4 Results ................................................ 95
   5.4.1 Motivation .......................................... 95
   5.4.2 Threshold and Noise .................................. 98
   5.4.3 Extrinsic Noise ....................................... 102
   5.4.4 Spatial Effects ....................................... 103
   5.4.5 Cell Growth Effects .................................. 106
   5.4.6 Model Robustness .................................... 109
5.5 Discussion ............................................. 110

# 6 The Effects of Cell Partitioning on Noise in Genetic Regulatory Networks

6.1 Overview ................................................. 114
6.2 Introduction ............................................ 114
6.3 Methods ................................................ 116
6.4 Results ................................................ 118
   6.4.1 Motivation .......................................... 118
   6.4.2 Constitutive Gene Expression .......................... 122
   6.4.3 Simple Genetic Networks .............................. 126
   6.4.4 Decision Making in Lambda Phage ....................... 131
6.5 Discussion ............................................. 132

# 7 Conclusion and Outlook

References ................................................. 153

A .......................................................... 154
List of Figures

1.1 Phage infection and replication mechanism ............................................. 18
1.2 Lysis/lysogeny decision in Temperate Phage ........................................... 19
1.3 Simple Genetic Circuit ............................................................................. 23
2.1 Image Processing Segmentation ............................................................... 38
2.2 Image Processing Tracking ....................................................................... 38
2.3 Asymmetric Growth in Mycobacterium smegmatis ...................................... 39
3.1 Description of how phenotypes evolve in both Bacteria and Phage ............... 46
3.2 Illustration of transfer method for different treatments .............................. 49
3.3 The diversity (± standard error se) for the experimental results averaged across time .............................................................................................................. 52
3.4 The proportion of phenotypes observed under each direction of gene flow 53
3.5 The number of unique phenotypes at each resource level for each direction of gene flow ........................................................................................................... 54
3.6 Diversity over time for the host from day 0 to day 13, at high and low resource, with gene flow rate $0.025 h^{-1}$ ................................................................. 55
3.7 Host diversity plotted as a function of gene flow rate for high and low resource .................................................................................................................. 56
3.8 Host diversity plotted as a function of gene flow rate for high and low resource with 3 levels of resource in model ......................................................... 57
3.9 Host diversity plotted as a function of gene flow rate for high and low resource with discrete transfers ................................................................................ 58
3.10 Diversity as a function of gene flow rate and point on the matching allele - gene-for-gene continuum .................................................................................. 59
3.11 Host diversity plotted as a function of gene flow rate for high and low resource with no transfer of resource ................................................................. 60
4.1 Illustration of dynamics of Bacteria (top) and Phage dynamics (bottom) in a well mixed environment based on parameters in Table 4.1. This observation represents lytic phage, however a similar figure can be obtained for temperate phage with appropriate parameters. ................................... 68
4.2 Bifurcation Plot for system 4.1 changing parameter $\mu_{Max}$ ......................... 71
4.3 Location of branch points in system 4.1 when changing bacteria growth parameter $\mu_{Max}$ and cost to growth of being a lysogen $q$ ............................... 72
4.4 Location of branch points in system 4.1 when changing bacteria growth parameter $\mu_{Max}$ and burst rate of virulent phage $\beta_P$ ........................................... 73
4.5 Bifurcation Plot for system 4.1 changing parameter $\beta_P$ ................. 74
4.6 Location of branch points in system 4.1 when changing the burst rate of virulent phage $\beta_P$ and rate of lysogeny $\text{lys}$ ......................... 75
4.7 Bifurcation plot for system of equations 4.1 changing parameter $\mu_{\text{Max}}$, where temperate phage has responsive rate of lysogeny $f = f_r$ ............ 76
4.8 Location of branch points in the system with responsive temperate phage $T$ when changing bacteria growth parameter $\mu_{\text{Max}}$ and population MOI parameter $\kappa$ .................................................. 77
4.9 Location of branch points in the system with responsive temperate phage $T$ when changing bacteria growth parameter $\mu_{\text{Max}}$ and cost to growth parameter $q$ .................................................. 77
4.10 ................................................................. 78
4.11 Location of branch points in the system with responsive temperate phage $T$ when changing burst rate parameter $\beta_P$ and population MOI parameter $\kappa$ 79
4.12 Dynamics of System Under Fluctuating Resource Conditions ............ 80
4.13 Parameters at Which All Species Survive Under Fluctuating Conditions 82
4.14 Parameters at Which Lytic Phage is Eliminated Under Fluctuating Conditions ............................................................ 83
4.15 Parameters at Which Temperate Phage is Eliminated Under Fluctuating Conditions ............................................................ 84
4.16 Parameters at Which Lytic Phage Eliminates Bacteria Under Fluctuating Conditions ............................................................ 85
5.1 Lysis-lysogeny Genetic Switch .................................................. 93
5.2 Rate of lysogeny as a function of MOI and Viral Concentration ............ 96
5.3 Decision Criteria for Lysis-Lysogeny Decision .................................. 97
5.4 Average Stochastic Trajectories of CII Concentration Over Time .......... 98
5.5 How the Relationship Between Noise and Threshold Level Affect the Decision Outcome .......................................................... 99
5.6 How changing the level of noise further affects the decision outcome .... 101
5.7 The Effect of Extrinsic Noise on Decision Outcome ............................ 102
5.8 The Role of Spatial Effects on CII Average .................................... 104
5.9 The Role of Spatial Effects on Rate of Lysogeny ............................... 105
5.10 The Role of Spatial Effects on Rate of Lysogeny for Slower Diffusion Rate. ........................................................................... 106
5.11 The Role of Cell Cycle Effects - Experiments .................................... 107
5.12 The Role of Cell Cycle Effects - Modelling Results ............................ 108
5.13 Model Robustness ......................................................................... 110
6.1 Outline of positive and negative feedback regulatory networks .......... 117
6.2 Reported data relating volume to doubling time ............................... 118
6.3 Approximation of Effects of Growth Rate on Noise ........................ 120
6.4 Comparison of LNA for growing cells to our simulations .................. 121
6.5 Convergence of LNA for growing cells to our simulations ................. 122
6.6 Effect of growth rate on mean and copy number of constitutively expressed proteins .......................................................... 124
6.7 Effect of growth rate on coefficient of variation for constitutively expressed proteins ........................................ 125
6.8 Effect of growth rate on system with negative feedback ................................................................. 127
6.9 Effect of growth rate on system with positive feedback ................................................................. 127
6.10 Effect of growth rate on a bound protein ...................................................................................... 128
6.11 Bistable Growth Rate Due to a Constitutively Expressed Toxin ..................................................... 130
6.12 Effect of Growth Rate on Switching Rate .................................................................................... 130
6.13 Effect of MOI on Division Time in Lambda Phage ........................................................................ 131
6.14 Effect of Growth Rate on Rate of Lysogeny in Lambda Phage ..................................................... 132

A.1 Population densities of equations 4.1 along different solution branches for temperate phage $T$ with constant rate of lysogeny when changing bacteria growth parameter $\mu_{Max}$ .................................................. 154
A.2 Population densities of equations 4.1 along different solution branches for responsive temperate phage $T$ when changing bacteria growth parameter $\mu_{Max}$ .................................................. 155
List of Tables

3.1 Model Parameters for Simulating Coevolution of Bacteria and Phage . . . 48
3.2 Explanation of resource transfers in model . . . . . . . . . . . . . . . . 51
3.3 Analysis of Variance of Shannon-Wiener Index of Diversity for experi-
mental results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52
4.1 Model Parameters for Simulating the Ecology of Lambda Phage . . . . 81
5.1 Model Parameters for Stochastic Simulation of Lambda Phage Genetic
Switch . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 94
5.2 Diffusion Model Parameters for Stochastic Simulation of Lambda Phage
Genetic Switch . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 94
6.1 Model Parameters - Constitutive Gene Expression in Dividing Cells . . . 125
List of Abbreviations

CV - coefficient of variation
LNA - linear noise approximation
LPS - lipopolysaccharides
MOI - multiplicity of infection
ODE - ordinary differential equation
OMP - outer membrane protein
se - standard error
TUD - time until death
V - cell volume
VC - viral concentration
List of Publications

Most of the author’s own research can be found in chapters 3-6 of this thesis. Some of the research presented in this thesis can also be found in the following publications:


The following papers are in preparation:


Matthew L. Robb, and Vahid Shahrezaei. The Effects of Cell Partitioning on Noise in Genetic Regulatory Networks.

Additional publications involving the author:


Chapter 1

Introduction

1.1 General Introduction

Biological systems are complex. At every level of life, from DNA through to large scale food webs, there are many features that introduce complexity, which make systems as a whole difficult to untangle. Advances in experimental systems and cellular image processing allow us to gather more data on systems at many biological levels and make inferences. The increasing level of detail with which we can observe systems gives us more information so that we can gain a more quantitative understanding. However, observing systems alone is not always sufficient to gain a true picture of the underlying mechanisms present. Mathematics gives us a language with which to translate this complexity. Models have long been used to study biological systems and it is increasingly being understood that these can in turn inform further experimental work in what becomes a cycle of new information. In this thesis, we use mathematical models to develop understanding of bacteria and its interaction with bacteriophage at both the single cell and population level. It is hoped that this work can then lead to further experimental studies and continue this cycle of information for these topics.

In studies over evolutionary timescales, microbial systems are useful due to their short replication time and their application in medicine and agriculture (Anderson and May, 1982). A common system for studying coevolution in host-parasite interactions is the bacteria-phage interaction. Typically this pairs *Escherichia coli* with a lytic phage to understand for example what factors increase diversity and adaptation (Buckling and Rainey, 2002; Forde et al., 2004, 2008). While analysis of ecological systems can be ben-
1.1 General Introduction

In revealing how species interact on shorter timescales and how this can be influenced by for example environmental factors before analysis over evolutionary timescales (Stewart and Levin, 1984; Jessup et al., 2004; Jessup and Bohannan, 2008). In systems biology, it is increasingly being understood that stochasticity has a major influence on cellular systems. For example, it can give rise to different phenotypes (Arkin et al., 1998; Elowitz et al., 2002; Chang et al., 2008) in isogenic populations, which can be useful for cell survival (Balaban et al., 2004). The level of stochasticity or noise in cellular systems can be influenced by copy number, extracellular fluctuations and the cell cycle (Shahrezaei et al., 2008; Huh and Paulsson, 2011). It is therefore of interest to investigate how the amount of noise changes in different networks (Alon, 2007) and whether this level of noise can be maintained in different environments. Research in different biological systems including lambda phage can lead to new insight at both the ecological and single cell level (Arkin et al., 1998; Avlund et al., 2009).

This thesis attempts to describe the importance of not only using mathematical models at different levels in biology, but also connecting them with experimental data. We begin in this chapter with an introduction outlining the Bacteria-Phage interaction which features in much of the work and then giving a general background on ecology and co-evolution, and the effect of noise in biochemical networks. Chapter 2 outlines some of the methods used in the thesis. Over the 4 chapters that follow we describe work from 4 studies in detail. Chapter 3 relates to the effect of gene flow on diversity in a coevolving system. The problem starts with analysis of experiments using Escherichia coli and T3 bacteriophage, and uses a mathematical model to understand the generalizability of the results and make predictions in other systems. The model is an extension of that used in Forde et al. (2008). This work has been accepted for publication and was carried out in collaboration with Ivana Gudelj, Samantha Forde and Michael Sieber. In Chapter 4 we utilise information recently derived on the probability of choosing the lysogenic decision (Zeng et al., 2010) and use this to gain a deeper understanding of the reason for evolving this latent function. This was investigated at the population scale, using an ecological model of bacteria and lambda phage. Following on from work in Stewart and Levin (1984) we perform a bifurcation analysis on the system and follow this with exploratory simulations under fluctuating resource conditions. This was carried out in collaboration with Vahid Shahrezaei and Ivana Gudelj and is in preparation for submission. Chapter 5 also focuses on lambda phage, specifically lysis-lysogeny decision making. In this work stochastic simulations of the genetic regulatory network are used to understand mechanistically how the rate of choosing the lysogenic fate is affected by the viral con-
centration. The decision is mediated by a well studied genetic switch. We investigate the effects of intrinsic noise, extrinsic noise, spatial effects and cell cycle effects on the decision outcome. This work is under revision for submission and was conducted with Vahid Shahrezaei. Experimental data was provided by Ido Golding and Lanying Zeng. In the final study of the thesis, Chapter 6, we analyse the effects of cell division and cell growth rate on noise in simple genetic networks. We know from previous studies that cell division increases noise in a cell (Huh and Paulsson, 2011) and also that changes in growth rate can affect transcription rates, gene dosage and cell volume (Klumpp et al., 2009). We investigate the effect of the combination of these factors under different network motifs. This work was carried out in collaboration with Vahid Shahrezaei and is in preparation for submission. Chapter 7 is the concluding chapter and gives an outlook on further work.

1.2 Bacteria-Phage Interaction

1.2.1 Introduction to Bacteria and Phage

*Escherichia coli* is a Gram-negative bacteria commonly used as a subject in microbial experiments. It is found most commonly in the gut of mammals, and has a large variety of different strains. The cells typically have a length of $2\mu m$ and replicate by mitosis. They are useful in studying evolution because of their short lifespan and fast replication time, typically of the order of 1 hour. They are also relatively easy to manipulate experimentally.

The bacteriophage virus is a natural parasite of bacteria. It is found in most places where bacteria is found. They are typically around $24 - 200 nm$ in length, and therefore much smaller than the host cells. The phage decay in short periods outside bacteria. This is typically at rates between $0.2 h^{-1}$ and $1.4 h^{-1}$ in water (Heldal and Bratbak, 1991). Phage infect bacteria by adsorption to the cell. This step is regulated by the phage tail fibres, which attach to the cell after detecting proteins such as lipopolysaccharides (LPS), and pili on the host cells surface. The rate at which they attach depends on the specific tail fibres and host cell receptors. Once attached, the phage inject their viral DNA into the cell. This allows the phage to use the replication mechanisms of the host cell to reproduce through the lytic cycle. This is illustrated in Figure 1.1. Once the number of new virions (individual phage particles) reach a limit known as burst size the host cell dies and the new phage are released.
Some phages are able to lay dormant in the host cell once they adsorb, rather than kill the cell, and replicate by cell division through integrating their genome into the host cell. This is called the lysogenic cycle. These phages are called temperate phages (Lederberg and Lederberg, 1953). It was observed that these phage do not always kill the host cells shortly after infection as would have been expected with other phage types. In fact, temperate phages infect the cell and make a decision as to whether to kill the cell immediately by the lytic cycle, or to lie dormant until replicating and destroying the cell later, this is called lysogeny (see Figure 1.2). At a later time the phage can enter the lytic cycle. The lysogenic state is described as being stable, and once the state is reached cells only switch back at the rate of about 1 × 10⁻⁴ cells per generation (Zong et al., 2010), this is called the rate of induction. Induction can be triggered more quickly if the cells are exposed to some form of stress such as UV radiation which leads to all cells bursting rapidly (Zong et al., 2010). The phage also replicate along with the host cells. The growth rate of lysogens is thought to be slower than uninfected host cells (Stewart and Levin, 1984). Some cells replicate with the daughter cells not containing any of the phage and this is a form of vegetative segregation. This also happens at a relatively low rate. When host cells are in the lysogenic state, then phage of the same type are not able to infect, this is called superinfection inhibition. While different strains are able to infect and therefore treat lysogens as susceptible cells. The mechanism that enables this is now fairly well understood (Arkin et al., 1998; Ptashne, 2004; Zeng et al., 2010). However, the reasons for
Figure 1.2: Lysis/lysogeny decision in Temperate Phage. Upon infection, the cell can undergo lysis or lysogeny. Under lysis, the phage replicates in the cell and destroys the cell. Under lysogeny, the virus lies dormant and integrates its DNA into the host genome. At a later stage it undergoes induction, in which it enters the lytic cycle.

this behaviour in temperate phages has been investigated in the past (Stewart and Levin, 1984) but it is still not completely clear.

Typically phage can only infect specific types of bacteria and this specificity depends on the infection mechanism. There are a number of infection mechanisms evident in the bacteria-phage, for example lock-key (Weitz et al., 2005) and CRISPR mediated resistance (Young, 2008; Vale and Little, 2010) that store information of previously infecting phage to acquire resistance. How species interact with each other depends on their genetic makeup and is very important in determining the outcome of their interaction. As species evolve, the way they interact changes (Adamson and Caira, 1994), and how they change depends on their infection (or resistance) mechanism. A description of interactions lying on the Matching Alleles - Gene-for-Gene continuum is provided by Agrawal and Lively (2002). The Matching Alleles model suggests that the parasite has unique infection properties such that once a host has evolved resistance to the parasite then the parasite subsequently has no effect on host fitness. They are thus described as specialists. In turn when the parasite evolves to be able to infect this more resistant host, it cannot infect any other host species. This model induces negative-frequency dependent selection dynamics where the least abundant phenotype at a particular time is selected for (Howard and Lively, 1994; Gandon, 2002). In contrast, the Gene-for-Gene model allows virulent parasites to be able to infect less-resistant hosts giving them a wider host range. The parasites are thus described as generalists. Therefore, where an organism lies
1.3 Ecology and Coevolution

Since Darwin postulated the theory of evolution (Darwin, 1859) we have been interested in what factors influence this process. For instance, what affects the rate of evolution, species diversity and speciation? These processes are not only interesting in view of the worlds decreasing diversity levels but also in relation to other applications such as epidemiology and phage therapy. Innovations in biology have also led us to be able to manipulate different species genetically and in turn observe the effect this has on the course of evolution (Datsenko and Wanner, 2000). This field has been developed over the past 30 years since the work of Anderson and May (1982). Current themes of research focus on investigating biologically relevant questions as well as developing the methodology in the area. For example, what leads to increased diversity and the high levels we observe in nature? Why do novel types arise and what allows them to coexist with existing types? Are the tools we currently use to investigate such questions adequate?

Microbial experiments provide an ideal system for studying various aspects of evolution, for example to analyse costs of antibiotic resistance (Lenski, 1998) and gene flow on processes such as adaptation (Morgan et al., 2005). Their short generation time makes them ideal to analyse how manipulating different aspects of their environment and their interactions affect outcomes of interest over the course of evolution. This has been extensively shown by Richard Lenski with a long term study of the evolution of Escherichia coli bacteria (Lenski et al., 1991; Lenski and Travisano, 1994). Furthermore, advance-
ments in the field have led to vastly increased knowledge of the makeup and underlying systems of microbial organisms. Genetic sequencing of subjects (Blattner et al., 1997) and greater understanding of regulatory networks (Gardner et al., 2000) give rise to a more complete picture of the makeup and function of the organisms constituent parts.

It is now also clear that environmental factors alone do not shape the course of evolution, but that antagonistic processes between interacting species play a major role (Dawkins and Krebs, 1979; Buckling and Rainey, 2002). Evolutionary problems are also intertwined with ecological factors which, together lead to the complex patterns of diversity observed in the natural world. This is described in detail in the Geographic Mosaic Theory of Coevolution (Thompson, 2005, 2009). To study coevolution, bacteriophage is often observed with *E. coli* (Forde et al., 2008). Due to the way that problems are related to real world problems it is no surprise that this research area contains many observational field studies. These are often very useful in learning what ecological factors are influential, directly from nature (Koskella and Lively, 2009). In addition, they can highlight features that have not been unaccounted for in models, or alternatively enable confirmation of theoretical findings.

Theoretical studies can provide useful insights into how varying the levels of factors of interest for different interactions, which are difficult to manipulate in experiments, affect the evolutionary outcome. There are a number of ways to approach and model the problems in this research area; two of the most frequently used methods are population genetics and adaptive dynamics. Population Genetics is a method that describes the change in frequencies of genes of interest over generations of a population. This enables one to deduce the strongest genotypes to particular species over a large number of generations. The models typically use many assumptions including large population size, same frequency of alleles in subsequent generations, and also principles such as Hardy-Weinberg (Stern, 1943) to achieve this. Another method of investigating problems is Adaptive Dynamics (Geritz et al., 1998). This is a more abstract method that looks at how strategies of particular traits develop over the course of evolution. The analysis focuses on invasion conditions of mutant species into a system in equilibrium and through stability analysis conditions for stable strategies can be derived. It is a useful method in determining how shapes of trade offs influence evolutionary strategies. Meanwhile, network models can be used to make inferences in complex systems (Rosvall et al., 2006).

The most insightful approach uses both theoretical models and experiments to inform each other. This gives rise more accurate models and more valuable experiments and ultimately more informed predictions. A recent example is given in Forde et al. (2008),
where the model uses deterministic differential equations and includes a basic level of genetic detail in an attempt to relate the theory to the observed phenomena and as a result the findings are directly applicable to the problem at hand. The model can then be adjusted to make predictions by changing factors of the interaction that are difficult to manipulate experimentally. This gives an idea of how generalisable these predictions are.

Research into biodiversity and related areas has many applications including explaining the decline of species richness in ecosystems (Helm et al., 2009), and epidemiology of infectious disease (Anderson and May, 1991). It is therefore necessary to investigate which factors affect diversity on both ecological and evolutionary timescales (Cadotte, 2006; Venail et al., 2008). This helps form a deeper understanding of which factors have positive effect on diversity, which factors have a negative effect, and how these effects change across time and across different types of interaction. For example Buckling and Hodgson (2007) demonstrate how rates of parasite evolution drive host diversity, while Forde et al. (2008) looked at how resource gradients and infection mechanism can affect host diversity. Diversity can be measured both in terms of the number of observed phenotypes and also in terms of phenotypic evenness. While typically diversity between populations can be thought of by measuring the number of unique phenotypes, evenness of the observed types is also an important measure which can increase population productivity (Wittebolle et al., 2009).

Another important factor in ecological and evolutionary research is adaptation. In particular why are some species better adapted to their environment than others? What factors influence levels of adaptation? Local adaptation more formally is a measure of fitness in an organisms local environment compared with its fitness in a foreign environment. When considering this in terms of hosts and parasites, it measures the hosts performance on sympatric (those from the same habitat) parasites compared to performance on allopatric (those from a different habitat) parasites. Alternatively, maladapted infers that they perform better on allopatric individuals. There are many studies investigating the role of ecological and evolutionary factors on adaptation (Hamilton et al., 1990; Gandon, 2002; Gandon and Michalakis, 2002; Gandon et al., 2008; Hoeksema and Forde, 2008). Experimental studies with *Pseudomonas Fluorescens* (Morgan et al., 2005) and *E. coli* (Forde et al., 2004) have been used to investigate the effects of gene flow on local adaptation. However, we find little work connecting models and experiments that use data on the specific genetic interaction over the course of coevolution.
1.4 Stochasticity in Cellular Systems

Biological cells are noisy, exhibiting great phenotypic variability (Elowitz et al., 2002; Shahrezaei and Swain, 2008b). By this it is meant that even though a group of cells may be genetically identical (isogenic), expression of their genes and their phenotypes can vary. This has been observed in many cellular systems including eukaryotes (Raser and O’Shea, 2004) and prokaryotes (Ozbudak et al., 2002; Swain et al., 2002) although the sources of noise may differ. There are different types of noise, intrinsic and extrinsic noise (Swain et al., 2002). Intrinsic noise in gene expression is caused by the probabilistic nature of the birth death process of mRNA and protein. This effect is larger when there are low molecule numbers in the cell. A temporal change of 1 molecule in a population of 1000 will not affect the cell, whereas if there are less than 10 molecules in a cell then this small fluctuation could have large repercussions for the cell. It is also known that mRNA transcription is a ‘bursty’ process (Raj et al., 2006; So et al., 2011). This is due to the transition between the inactive and active states of a gene (Figure 1.3). Extrinsic noise is caused by the interaction of the network with the rest of the cell and the surrounding environment. The sources of extrinsic noise are not definitively known but contribute to the total noise observed in a network. Noise can also be caused by cell division (Huh and Paulsson, 2011). If biomolecules are partitioned binomially during mitosis then this is also a source of variation.

![Simple Genetic Circuit](image)

**Figure 1.3:** Simple Genetic Circuit. When a gene is turned on at a particular rate $k_{on}$, it can transcribe mRNA at a rate $k_m$. It can also return to the off state with rate $k_{off}$. mRNA degrades at rate $d_m$ and can translate protein at rate $k_p$. Protein degrades at rate $d_p$.

The way in which the noise affects the cell depends on the genetic network motif
1.4 Stochasticity in Cellular Systems

(Alon, 2007). If there is a negative feedback loop in the network then transcriptional noise can be reduced (Thattai and van Oudenaarden, 2001; Swain, 2004). In the case of a positive feedback loop, the level of noise can be amplified (Delbrück, 1940). This can lead to bimodal populations (Becskei et al., 2001). Why would a cell want to increase or decrease the amount of noise? In some cases where noise affects a response signal, it may be of benefit to reduce the level of noise to give a clear signal (Andrews et al., 2006). In other cases, noise can lead to population heterogeneity, which may act as a bet-hedging mechanism and be beneficial in stress response (Balaban et al., 2004; Kussell et al., 2005). To emphasise the importance of noise and the control of noise in the genetic regulatory network it has been demonstrated that noise can be tuned and exploited (Barkai and Leibler, 2000; Raser et al., 2002; Süssel et al., 2007). Moreover, it has also been suggested that noise is an evolvable trait (Fraser et al., 2004; Mehta et al., 2008; Eldar and Elowitz, 2010).

There is a great deal of evidence demonstrating the stochastic nature of cell processes and how their effects can be utilised by the cells (Arkin et al., 1998; Kuchina et al., 2011). In eukaryotes, it has been observed that gene expression noise is in large part due to the slow transition rates of promoters (Blake et al., 2003). Therefore transcription happens in larger bursts over shorter periods of time. This has been observed in *Saccharomyces cerevisiae* (Blake et al., 2003). In prokaryotes, the transition times are much shorter, and noise is more often generated from inefficient translation (Rosenfeld et al., 2005). Whereas intrinsic noise is caused by the probabilistic nature of the birth death process and affects specific genes, extrinsic noise has unknown source and affects multiple genes equally, affecting both eukaryotes (Raser and O’Shea, 2004) and prokaryotes (Elowitz et al., 2002). This additional noise can affect the overall level of noise in a network and also mean concentrations (Lei, 2009). Possible sources of extrinsic noise are cell cycle effects, upstream sources and fluctuations in the number of ribosomes (Shahrezaei et al., 2008). The lifetime of intrinsic noise is short such that bursts occur and can be corrected on timescales much shorter than the cell cycle (Swain et al., 2002) and is described as ‘white noise’. However extrinsic noise has a similar lifetime to the cell cycle (Shahrezaei et al., 2008) and fluctuations are described as ‘coloured fluctuations’. Methods have also been developed to identify how much noise in experiments is due to intrinsic and extrinsic noise respectively (Elowitz et al., 2002; Swain et al., 2002). Noise observed in a system can be identified as being intrinsic or extrinsic dependent on how correlated the noise is (Swain et al., 2002). Comparing two genes, noise that is not correlated can be attributed to intrinsic noise, whereas noise that is correlated between genes can be attributed to
extrinsic noise. The source of each can be important in determining how much of an observed phenomenon is down to the genetic regulatory network, and how much is due to apparent background noise. This is possible due to advancements in cellular image processing (Saxton and Jacobson, 1997; Schwille et al., 1999; Thompson et al., 2002; Jaqaman et al., 2008; Huang et al., 2009; Young et al., 2012). While cell partitioning can affect noise (Huh and Paulsson, 2011), the rate of division can also affect cell volume (Chien et al., 2012; Turner et al., 2012) and global cell parameters (Klumpp et al., 2009). It is possible that these phenomena confound each other in their effect on protein noise. To our knowledge there is not much data available relating protein noise to rate of cell growth although advances in experimental procedures make it possible to investigate.

In contrast to the methods used in tools above, single cell analysis usually requires a probabilistic model. While deterministic models can be used to infer the average behaviour, under certain conditions noise can have a large influence on the cellular behaviour. This influence would therefore be missed in a deterministic setting. There are many examples of the use of stochastic modelling in biological systems (Arkin et al., 1998; Friedman et al., 2006; Zong et al., 2010). Stochastic processes in cellular systems are commonly modelled by some form of the Gillespie algorithm (Gillespie, 1977), allowing one to produce statistically accurate trajectories of individual biomolecules and reactions. Using the Gillespie algorithm for a given gene expression model (for example Figure 1.3) is sufficient to simulate the intrinsic noise present in the cellular process. To simulate extrinsic noise potentially affecting all parameters of the system, more complex methodology is required (Shahrezaei et al., 2008). To simulate this, parameter values are varied in time about the given value (Paulsson, 2004). Extrinsic noise can interact with intrinsic noise to complicate the effect of the additional variation (Paulsson, 2004; Lei, 2009). It can therefore be useful to untangle which source of noise is dominant to understand what drives the observed behaviour in any system. It is increasingly being realised that this noise must be taken into account in parameter estimation (Komorowski et al., 2010). It is shown in Komorowski et al. (2010) that when using fluorescence reporters in estimating transcription, translation and degradation rates that taking sources of noise such as extrinsic noise and protein maturation into account is crucial to getting an accurate measure. It is difficult to use analytical methods to understand the effects of extrinsic noise, as a result most investigations are carried out computationally using the Gillespie algorithm (Gillespie, 1977). However some attempts have been made to understand its effect on steady state behaviours (Mackey et al., 2011). It is also possible to model the effects of cell partitioning using the linear noise approximation (Huh and
Paulsson, 2011), as outlined in Chapter 2. It can be seen that even for a constitutively expressed protein, the effects can be difficult to untangle from those caused by the birth death process.

Cellular noise can also affect the outcome of cell fate decisions. It has been observed that noise can have an effect on decision making in Lambda phage (Arkin et al., 1998; Joh and Weitz, 2011), and the importance of this is explored further in Chapter 5. It has also been hypothesised that noise can be of importance in decision making in HIV (Weinberger et al., 2008). Furthermore, the importance of noise in persister cells has been explored (Kussell et al., 2005), which has major relevance in antibiotic resistance (Balaban et al., 2004).
Chapter 2

Methods

2.1 Introduction

In this chapter, we describe methods that are based in previous literature which are used in this work. We describe the general form of the linear noise approximation, following work from Elf and Ehrenberg (2003). We then outline a specific use for it in approximating the mean and variance of biomolecules in a dividing cell as described in Huh and Paulsson (2011). This tool is used in the work described in Chapter 6. In the section that follows we outline the Gillespie algorithm (Gillespie, 1977) and the work in Shahrezaei et al. (2008), in order to simulate extrinsic noise in a genetic regulatory network. This is used in the work on lambda phage described in Chapter 5. Finally we describe methods used in cellular image processing. This was used to extract data from original experimental movies obtained from the Golding lab, and was shared with us by the authors kind permission. The work used Schnitzcell software courtesy of Michael Elowitz (Elowitz et al., 2002).

2.2 ODE Models of Ecology and Coevolution

To study models in ecology and coevolution, we used deterministic ordinary differential equations (ODEs). To solve the systems of equations numerically, Matlab was used. Variables describe densities of bacteria and phage species and phenotypes, and also resource levels. Infection outcomes are determined by an infectivity matrix. Analysis of the systems was performed by looking at either the population densities at steady states or at transient time points.
2.3 Bifurcation Analysis

In the work on ecology of lambda phage (Chapter 4) we use bifurcation analysis to analyse stability of steady states and how these are affected by different parameters. This was performed using AUTO software. This allows one to input a system of equations and a non-trivial steady state and find all steady state solutions across different parameters by continuing solutions. Initial steady states were found by solving in Matlab.

2.4 Linear Noise Approximation

The linear noise approximation (Van Kampen, 2007) is a method of approximating the distribution of the solution to stochastic processes. While very simple systems may be solved analytically, more than 2 variables and any nonlinear rate constants makes the system difficult to solve. This methodology gives a general method of simplifying the Master equation (Elf and Ehrenberg, 2003). It is obtained by using van Kampen’s $\Omega$ expansion method. It is assumed that the system is spatially homogeneous.

For a set of reactants $X_i$ for $i = 1, \ldots, N$, and set of reactions $R_j$ for $j = 1, \ldots R$, the stoichiometry matrix $S_{ij}$ defines the reaction network. The system has volume $\Omega$. The macroscopic concentration is defined as $\phi = (\phi_1, \ldots, \phi_N)$ while the mesoscopic concentration is defined as $x = (x_1, \ldots, x_N)$. We also have $X = \Omega x$. The probability that reaction $j$ occurs in the small time interval $\delta t$ is $\Omega \tilde{f}_j(x, \Omega)\delta t$. The macroscopic reactions are defined

$$\frac{d\phi_i}{dt} = \sum_{j=1}^{R} S_{ij} f_j(\phi) = [1, 2, \ldots, N]. \quad (2.1)$$

To start the approximation we take the limit as $\Omega$ gets large. Approaching this limit, fluctuations in the mesoscopic concentrations become insignificant and thus the system becomes close to deterministic

$$\phi = \lim_{\Omega \to \infty} \Omega^{-1} X = \lim_{\Omega \to \infty} x$$

$$\lim_{\Omega \to \infty} \tilde{f}_j(x, \Omega) = f_j(\phi).$$

In reduced form, the Master equation for this system can be written as follows
\[
\frac{dP(X,t)}{dt} = \Omega \sum_{j=1}^{R} \left( \prod_{i=1}^{N} E^{-S_{ij}} - 1 \right) \tilde{f}_j(x, \Omega) P(X, t) \quad (2.2)
\]

where \(E^{-S_{ij}}\) is an operator such that the object upon which it acts has \(S_{ij}\) molecules removed from it. As \(\Omega \to \infty\) this approaches the macroscopic relation shown in equation 2.1. To understand this, we first define a random variable \(\xi\) as follows:

\[
X_i = \Omega \phi_i + \Omega^{1/2} \xi_i. \quad (2.3)
\]

This therefore splits the copy number into its macroscopic component and a difference factor, which is our random variable \(\xi\). The probability distribution for the copy number \(X, P(X, t)\), is related to the equivalent probability distribution for \(\xi\) via

\[
P(X, t) = P(\Omega \psi + \Omega^{1/2} \xi, t) = \Pi(\xi, t).
\]

If we differentiate this with respect to time for constant copy number we find that

\[
\frac{\partial P(X, t)}{\partial t} = \frac{\partial \Pi(\xi, t)}{\partial t} + \sum_{i=1}^{N} \frac{\partial \xi_i}{\partial t} \frac{\partial \Pi(\xi, t)}{\partial \xi_i} = \frac{\partial \Pi(\xi, t)}{\partial t} - \Omega^{1/2} \sum_{i=1}^{N} \frac{\partial \phi_i}{\partial t} \frac{\partial \Pi(\xi, t)}{\partial \xi_i}
\]

when taking into account

\[
\frac{\partial X_i}{\partial t} = 0
\]

due to constant copy number and

\[
\frac{\partial \xi_i}{\partial t} = -\Omega^{1/2} \frac{\partial \phi_i}{\partial t}
\]

which follows from equation 2.3. A Taylor expansion of the mesoscopic transition rates \(\tilde{f}_j\) about the macroscopic values for the first 2 terms is

\[
\tilde{f}_j(x) = \tilde{f}_j(\psi + \Omega^{-1/2} \xi) = f_j(\psi) + \Omega^{-1/2} N \sum_{i=1}^{N} \frac{\partial f_j(\psi)}{\partial \phi_i} \xi_i + O(\Omega^{-1})
\]

where terms of order \(\Omega^{-1}\) are single molecules. The step operator can be approximated (Van Kampen, 2007) by

\[
E^k f(x) = f(\phi + \Omega^{-1/2}(\xi + \Omega^{-1/2}k)) = \left[ 1 + \Omega^{-1/2}k \frac{\partial}{\partial \xi} + \frac{\Omega^{-1}k^2}{2} \frac{\partial^2}{\partial \xi^2} + O(\Omega^{-3/2}) \right] f(x).
\]
2.4 Linear Noise Approximation

Expanding this to a multidimensional system gives us

\[
E^{-S_{ij}} = \prod_{i=1}^{N} E^{-S_{ij}} \approx 1 - \Omega^{-1/2} \sum_{i} S_{ij} \frac{\partial}{\partial \xi_i} + \frac{\Omega^{-1}}{2} \sum_{i} \sum_{k} S_{ij} S_{kj} \frac{\partial^2}{\partial \xi_i \partial \xi_k} + O(\Omega^{-3/2}).
\]

where \(-S_{ij}\) is the stoichiometry matrix across all biomolecules. Using the above in the Master equation (2.2), we have

\[
\frac{\partial \Pi(\xi,t)}{\partial t} - \Omega^{-1/2} \sum_{i=1}^{N} \frac{\partial \phi_i}{\partial t} \frac{\partial \Pi(\xi,t)}{\partial \xi_i} = \Omega \sum_{j=1}^{R} \left( -\Omega^{-1/2} \sum_{i} S_{ij} \frac{\partial}{\partial \xi_i} + \frac{\Omega^{-1}}{2} \sum_{i} \sum_{k} S_{ij} S_{kj} \frac{\partial^2}{\partial \xi_i \partial \xi_k} + O(\Omega^{-3/2}) \right)
\]

\[
\left( f_j(\psi) + \Omega^{-1/2} \sum_{i} \frac{\partial f_j(\psi)}{\partial \phi_i} \xi_i + O(\Omega^{-1}) \right) \Pi(\xi,t).
\]

We then compare terms of the same order. Starting with those of order \(\Omega^{1/2}\) we have

\[
\Omega^{1/2} : \sum_{i=1}^{N} \frac{\partial \phi_i}{\partial t} \frac{\partial \Pi(\xi,t)}{\partial \xi_i} = \sum_{i=1}^{N} \sum_{j=1}^{R} S_{ij} f_j(\psi) \frac{\partial \Pi(\xi,t)}{\partial \xi_j}
\]

which is the same as the macroscopic equation. For those of order \(\Omega^0\)

\[
\Omega^0 : \frac{\partial \Pi(\xi,t)}{\partial t} = \sum_{j} \left[ \sum_{i,k} -S_{ij} \frac{\partial f_j}{\partial \phi_k} \frac{\partial (\xi_k \Pi(\xi,t))}{\partial \xi_i} + \frac{1}{2} f_j \sum_{i,k} S_{ij} S_{kj} \frac{\partial^2 \Pi(\xi,t)}{\partial \xi_i \partial \xi_k} \right] = -\sum_{i,k} A_{ik} \frac{\partial (\xi_k \Pi)}{\partial \xi_i} + \frac{1}{2} \sum_{i,k} S_{ij} S_{kj} \frac{\partial^2 \Pi}{\partial \xi_i \partial \xi_k}
\]

where

\[
f_i = f_i(\psi)
\]

\[
A_{ik} = \sum_{j=1}^{R} S_{ij} \frac{\partial f_j}{\partial \phi_k} = \frac{\partial (S_{ij} f)}{\partial \phi_k}
\]

\[
[BB^T]_{ik} = \sum_{j=1}^{R} S_{ij} S_{kj} f_j(\psi) = [S_{\text{diag}}(f(\psi))S^T]_{ik}.
\]
This is the Fokker-Plank equation (Van Kampen, 2007) which we can solve to give us the distribution of the molecules. Higher orders are not considered in the linear noise approximation. Following the standard method of solving this equation

\[ 0 = - \sum_{ik} A_{ik} \frac{\partial}{\partial \xi_i} \xi_k \Pi(\xi) + \frac{1}{2} \sum_{ik} [BB^T]_{ik} \frac{\partial^2 \Pi(\xi)}{\partial \xi_i \partial \xi_k} \]

gives us the stationary solution. It is known that the solution is a multidimensional Gaussian distribution as follows

\[ P(\xi) = \left( \frac{(2\pi)^{N/2}}{\sqrt{\det \Xi}} \right)^{-1} \exp\left(-\xi \Xi \xi / 2\right). \]

This distribution has zero mean vector and covariance matrix \( \Xi = \langle \xi \xi^T \rangle \) that follows from the Lyapunov equation

\[ \frac{d\Xi}{dt} = A\Xi + \Xi A^T + BB^T. \quad (2.4) \]

As an example of how to use this approximation we derive the equations in Huh and Paulsson (2011). This is the linear noise approximation for a dividing cell. For a protein \( x \) and mRNA \( y \) in a simple circuit we have

\[
\begin{align*}
y & \xrightarrow{\lambda_1} y + 1 \\
x & \xrightarrow{\lambda_2} x + 1 \\
y & \xrightarrow{\beta_1} y - 1 \\
x & \xrightarrow{\beta_2} x - 1
\end{align*}
\]

for transcription rate \( \lambda_1 \), translation rate \( \lambda_2 \) and degradation rates \( \beta_1 \) and \( \beta_2 \). We first show the derivation of the differential equations for the mean and variance. For the mean this is simply

\[
\begin{align*}
\frac{dx}{dt} &= \lambda_2 \langle y \rangle - \beta_2 \langle x \rangle \\
\frac{dy}{dt} &= \lambda_1 - \beta_1 \langle y \rangle.
\end{align*}
\]

The derivation of the variance relations requires more work. First we define

\[
X = \begin{bmatrix} x \\ y \end{bmatrix}
\]
while the mesoscopic rates are given in

\[ f(\tilde{x}) = \begin{bmatrix} \lambda_1 \\ \lambda_2 y \\ \beta_1 y \\ \beta_2 x \end{bmatrix}. \]

The stoichiometry matrix \( S \) is given by

\[ S = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \end{bmatrix}. \]

Matrix \( A_{ik} \) then follows

\[ A = \begin{bmatrix} -\beta_2 & \lambda_2 \\ 0 & \beta_1 \end{bmatrix}, \]

while, \( BB^T \) is given by

\[ BB^T = \begin{bmatrix} \lambda_2 y + \beta_2 x & 0 \\ 0 & \lambda_1 + \beta_1 y \end{bmatrix}. \]

Using these in the non-stationary version of equation 2.4, it can be seen that

\[ \frac{d\sigma_{xx}}{dt} = -2\beta_2 \sigma_{xx} + 2\lambda_2 \sigma_{xy} + \lambda_2 \langle y \rangle + \beta_2 \langle x \rangle \] (2.5)

\[ \frac{d\sigma_{xy}}{dt} = -(\beta_2 + \beta_1) \sigma_{xy} + \lambda_2 \sigma_{yy} \] (2.6)

\[ \frac{d\sigma_{yy}}{dt} = -2\beta_1 \sigma_{yy} + \lambda_1 + \beta_1 \langle y \rangle. \] (2.7)

The next step is to solve these equations for the appropriate initial conditions. For the mean, we have

\[ \langle X_i \rangle_{t=0}^{g+1} = \frac{\langle X_i \rangle_{t=T}^g}{2} \]

since at cell division the average number of each molecular species is halved, where \( g \) is the current generation, \( g + 1 \) is the next generation and \( t = T \) is the doubling time. For the variance we have
Chapter 2. Methods

\[ \frac{\sigma_{ij}^{g+1}}{\langle X_i \rangle \langle X_j \rangle} \bigg|_{t=0} = \frac{\sigma_{ij}^g}{\langle X_i \rangle \langle X_j \rangle} \bigg|_{t=T} + Q_{ij}^g \]

where

\[ Q_{ij} = \frac{\left(\langle L_i - R_i \rangle (L_j - R_j)\right)}{\langle X_i \rangle \langle X_j \rangle} \bigg|_{t=T} \]

for \( i, j = 1, 2 \) where \( X_1 = x \) and \( X_2 = y \) with \( L_{i/j} \) and \( R_{i/j} \) the numbers of molecule \( i/j \) in each daughter cell. For binomial partitioning of molecules at cell division, as we focus on in this work, the numerator of \( Q_{ii} \) is 4 times the binomial variance

\[ Q_{ij} = 4X_i \times 0.5 \times 0.5 \]

while for the \( i \neq j \), \( Q_{ij} = 0 \). Putting this together we have the initial conditions for the means to be

\[ x_T = 2x_0 \]
\[ y_T = 2y_0 \]

and for the variance

\[ 4\sigma_{xx,0} = \sigma_{xx,T} + A_x \langle x \rangle_T \]
\[ 4\sigma_{yy,0} = \sigma_{yy,T} + A_y \langle y \rangle_T \]
\[ 4\sigma_{xy,0} = \sigma_{xy,T} \]

If we apply these to equations 2.5-2.7 then using the integrating factor method we can solve for \( \sigma_{xx}, \sigma_{xy} \) and \( \sigma_{yy} \). The general solutions for different forms of cell division are given by

\[ \langle y \rangle_t = \lambda_1 T c_{1,t} \]
\[ \langle x \rangle_t = \lambda_1 \lambda_2 T c_{1,t} \frac{c_{2,t}}{\beta_2 - \beta_1} \]
\[ CV_{y,t}^2 = \left[ 1 + k_{11,t} \frac{c_{1,T}}{c_{1,t}} (A_y - 1) \right] \frac{1}{\langle y \rangle_t} \quad (2.8) \]
\[ CV_{x,t}^2 = (S_{x,t} + U_{x,t} A_x) \frac{1}{\langle x \rangle_t} + (S_{y,t} + U_{y,t} A_y) \frac{1}{\langle y \rangle_t} \quad (2.9) \]
where

\[ S_{x,t} = 1 - U_{x,t} \]
\[ U_{x,t} = k_{22,t} \frac{c_{1,T} - c_{2,T}}{c_{1,t} - c_{2,t}} \]

\[ S_{y,t} = \frac{\left[ r + c_{1,T} \frac{k_{22,t}}{1 - 2R} + 2 \left( \frac{r}{R + 1} + \frac{c_{1,T}}{R} \right) k_{12,t} - c_{1,T} k_{11,t} + \frac{R - 1}{2R - 1} \left( 2 \frac{R - 1}{R} c_{1,t} - \frac{r}{R + 1} \right) \right]}{(c_{1,t} - c_{2,t})^2} \]

\[ U_{y,t} = [k_{22,t} - 2k_{12,t} + k_{11,t}] \frac{c_{1,T} c_{1,t}}{(c_{1,t} - c_{2,t})^2} \]

\[ R = \frac{\beta_2}{\beta_1} \]
\[ r = \frac{3}{T} \left( \frac{\beta_2 - \beta_1}{\beta_1 \beta_2} \right) \]
\[ c_{i,t} = \frac{1}{\beta_i T} \left( 1 - \frac{e^{-\beta_i t}}{2 - e^{-\beta_i T}} \right) \]
\[ k_{ij,t} = \frac{e^{(\beta_i + \beta_j) t}}{4 - e^{(\beta_i + \beta_j) T}} \]

and where the parameters \( A_x \) and \( A_y \) define the type of division and if division of biomolecules is binomial then \( A_x = A_y = 1 \).

### 2.5 Modelling Extrinsic Noise

The Gillespie algorithm (Gillespie, 1977) is a tool that uses Monte Carlo methods to simulate discrete Markov stochastic processes. It works by generating random times for each successive reaction and probabilities of reactions are determined by the rate of reaction and the concentrations of the biomolecules. A computer algorithm is then used to simulate these probabilistic reactions. This is both efficient and sufficient for replicating intrinsic noise in genetic regulatory networks. To include extrinsic noise, one wishes to vary the reaction rates about the given values over time. This means that the reaction rates become fluctuating. Here Shahrezaei et al. (2008) describe a method of extending the Gillespie algorithm to deal with extrinsic noise by simulating models with time-varying parameters. It works by firstly calculating the potential reaction times for
the simulation. The variations in the reaction rates are then calculated at each of these
time points. The simulations can then be run with these reaction rates implemented at the
reaction times. In the usual way, we first want to calculate the putative reaction time $\tau$.
We have the probability of a time dependent reaction with propensity $a(t)$ at time $t = \tau$
as follows

$$
P(\tau) = a(\tau) \exp \left( - \int_0^\tau a(t) dt \right).$$

(2.10)

In the algorithm, a sample is taken from the uniform distribution in solving the equation
2.10

$$
\int_0^\tau P(t) dt = r
$$

where $r$ is a uniform random number. Using this in equation 2.10 we have

$$
\int_0^\tau dt a(t) \exp \left( - \int_0^t a(t') dt' \right) = r,
$$

which is integrated to give

$$
1 - \exp \left( - \int_0^\tau a(t) dt \right) = r.
$$

Rearranging this we have the exponential term equal to $1 - r$. For $r$ belonging to a uniform
distribution, we also have $1 - r$ belonging to a uniform distribution. This therefore allows
us to write

$$
\int_0^\tau a(t) dt = \log(1/r).
$$

As explained above, we now want to change the reaction rates at these generated
reaction times, which will make $a(t)$ discontinuous. Therefore to integrate the above
equation, we have to take into account that $a(t)$ is a series of step functions or piecewise
linear functions. We define

$$
a(t) = \begin{cases} 
a_<(t) & \text{for } t < t_0 \\
a_>(t) & \text{for } t > t_0 
\end{cases}
$$

where $t_0$ is the point at which the reaction rate changes discontinuously. If $r = r_1$ is large
then for $t < t_0$ we have
2.5 Modelling Extrinsic Noise

\[ \int_0^\tau dta_<(t) = \log(1/r_1). \]

Conversely for \( r = r_2 \) with small \( r_2 \) we have

\[ \int_0^{t_0} dta_<(t) + \int_0^{\tau-t_0} dta_>(t) = \log(1/r_2) \]

for \( t > t_0 \). If we define \( c \) such that

\[ \int_0^{t_0} dta_<(t) = \log(1/c) \]

where \( c \geq r_2 \), then we can write

\[ \int_0^{\tau-t_0} dta_>(t) = \log(1/r_3) \]

with \( r_3 = r_2/c \). This gives us \( r_3 \) uniformly distributed. We now have a new form of propensity which can be used in the same way as the Gillespie algorithm (Gillespie, 1977). This allows us to calculate the putative next reaction times. The changes in reaction rates at these reaction times can then be calculated as follows. First an Ornstein-Uhlenbeck process is initiated using

\[ \frac{d\epsilon}{dt} = -\frac{\epsilon}{\tau} + \frac{\xi_0}{\tau} \]

where \( \xi \) is a white noise source such that

\[ \langle \xi_0(t_1)\xi_0(t_2) \rangle = 2\tau \eta_\epsilon^2 \delta(t_1 - t_2) \]

where \( \epsilon \) is normally distributed with mean 0 and variance \( \eta_\epsilon^2 \). It also has an exponentially decaying autocorrelation function

\[ C_\epsilon(t) = \eta_\epsilon^2 e^{-t/\tau}. \]

The reaction rate \( k \) is then normalised such that

\[ C_e(t) = k^2(\eta_\epsilon^2 e^{-t/\tau} - 1) \]

\[ = k^2 \sum_{r=1}^{\infty} \frac{\eta_\epsilon^{2r}}{r!} e^{-rt/\tau}. \]

This process allows us to expand the Gillespie algorithm to model extrinsic fluctuations.
2.6 **Cellular Image Processing**

The intricately detailed images produced by time lapse microscopy cannot be interpreted by the naked eye alone. Cell lineages can be hard to trace and fluorescence levels need to be measured accurately. There is now software available so that such images can be interpreted for analysis. Schnitzcell was developed by Michael Elowitz (Elowitz et al., 2002), and given to the author by kind permission. I have used this in the work in Chapter 5 and also in a collaborative work (Joyce et al., 2012). Here we demonstrate how it can be used. Given a set of microscopy images, the first task is the segmentation. This process identifies the cells in the images. This is carried out by a program written in C++ that detects contrast from a background. This needs to be carried out for each frame in a movie. An example of this is shown in Figure 2.1. In some cases, microscopy images can be noisy. In which case, segmentation can be impossible for the software to elucidate. To overcome this the images need to be processed first for example by increasing the contrast level but without losing too much information of the image.

The process can encounter mistakes in the segmentation such as mistaking 1 cell as 2, or missing a cell completely. These errors can be corrected manually. The next task is the cell tracking. This attempts to connect the cells from successive frames so that information can be tracked on cell generations over time. Processes such as cell death and cells moving outside of the frame can also lead to errors, which again need to be corrected. An example of this process is shown in Figure 2.2.

Once this has been completed, information can be extracted from cell lineages from generated arrays. These arrays contain information on fluorescence levels, cell length and position, and how many frames the cell was present for. The information has to be interpreted carefully. For example, cell length is calculated by the Euclidean distance between the cell poles. For a small *E. coli* cell this can be accurate, however for longer, more curved cells such as Mycobacteria then this measurement does not accurately represent the true length of the cell. In such a case, it is currently better to make such calculations manually.

This methodology was used by the author in Joyce et al. (2012). In this study features of asymmetric growth in *Mycobacterium smegmatis* were investigated. Mycobacteria although rod-shaped lack the proteins that are present in other such bacteria (for example *E. coli*), that are necessary for symmetrical cell division. It was demonstrated by cellular image processing of time lapse movies that the septum placement of the centre was accurate but that the difference in growth rates between septum placement and cell division is the reason for the apparent asymmetric growth. This is demonstrated in Figure 2.3.
Figure 2.1: Image Processing Segmentation. The left panels show the original images and the right panels show the software’s interpretation of the image. A) Segmentation algorithms can sometimes not give completely accurate results. It can be seen in this example that one cell is completely missed and others are interpreted as being two cells where they are actually one. B) These errors are easily corrected manually.

Figure 2.2: Image Processing Tracking. Cell lineages are tracked through the software so that information from individual cells can be analysed for the course of the movie.
Figure 2.3: Asymmetric Growth in *Mycobacterium smegmatis*. Cell centre represented by vertical axis. The yellow represents the cell at septum placement, and the blue represents the cell at cell division. This figure was originally published in Joyce et al. (2012).

2.7 Stochastic Simulations

Here we briefly describe the processes used in producing stochastic results in the work in Chapters 5 and 6.

2.7.1 Non-Spatial Simulations

For simulations where we do not believe spatial effects to be important we use the Gillespie algorithm (Gillespie, 1977) to simulate genetic regulatory networks. To do this we use the free software Facile (Siso-Nadal et al., 2007) and Easystoch (Shahrezaei et al., 2008). Facile enables one to input reactions as they would be displayed in a report and outputs a file that can be read by Easystoch. The software can then perform the stochastic simulations based on the Gibson-Bruck version (Gibson and Bruck, 2000) of the Gillespie algorithm.
2.7 Stochastic Simulations

2.7.2 Spatial Simulations

Where we believe there is a possibility that spatial effects can alter the behaviour of the cell we have conducted spatial simulations. Reasons that diffusion may have an effect could be slow diffusion or spatial structure within a cell. To test whether they are important we have used Smoldyn software (Andrews and Bray, 2004). Smoldyn implements a full particle based Monte Carlo simulation of diffusion reactions. By simply entering the reactions in a similar way to Facile, we are able to simulate these spatial simulations.
Chapter 3

Models of Bacteria-Phage Coevolution

3.1 Overview

In this Chapter, we describe models and experiments outlining research in evolution and ecology. The work builds on previous investigations in Forde et al. (2008). The previous experiments tracked the evolution of *Escherichia coli* and T3 bacteriophage in different resource environments. The system was described by an ODE model, this is outlined in the methods section. The focus of this work was to look at the effect of adding complexity to the system by the way of gene flow, which is thought to have a large effect on ecosystems. We extended the previous ODE model for this work by adding extra terms and equations to the system. The analysis focussed on the effect of gene flow on diversity and resistance. The experiments were performed by Samantha Forde’s group at the University of Santa Cruz. The model was able to capture the observations from the experiments. This gave support to our model and allowed us to make further predictions that would be more difficult to make through experiments. This included making predictions for different genetic interaction models and rates of gene flow. Together the use of both models and experiments improves our understanding of the relationship between diversity, environmental factors and genetic interactions. The work was conducted in collaboration with the Forde group and Ivana Gudelj and continued by her post doctoral student Michael Sieber who carried out the work on resistance. This work has been accepted for publication Sieber et al. (2013).

3.2 Introduction

Factors affecting evolution and diversity are of major interest. Research in this area has developed such that we now realise the importance of coevolutionary interactions
3.2 Introduction

(Dawkins and Krebs, 1979; Bull, 1994; Thompson, 1998; Brockhurst et al., 2007). Host-parasite interactions can have a dramatic effect on the course of evolution (Anderson and May, 1982; Lenski, 1988; Turelli, 1994; Kawecki, 1998; Best et al., 2009). While there is a large volume of literature in this area from field studies (Thompson, 1999; Koskella and Lively, 2009; Laine, 2009), experiments with microbial systems are increasingly being used to develop insight and make inferences (Lenski and Levin, 1985; Bohannan and Lenski, 2000; Buckling and Rainey, 2002; Forde et al., 2008). In particular, *E. coli* and bacteriophage is a model system in which much is known about the genetic interaction between the virus and the pathogen. Coevolution occurs in this system over a short period of time and therefore, it is ideal for analysis over evolutionary timescales. The relationship between genetic interactions and the path of evolution is outlined in the The Geographic Mosaic Theory of Coevolution (Thompson, 1999, 2005). It states that the path of evolution and selection is not only determined by the ecological and evolutionary interaction between players, but also the environmental landscape in which the interaction takes place. While the basis of this hypothesis was mainly formed using observational studies, there is also a growing body of evidence from experimental studies (Buckling and Rainey, 2002; Forde et al., 2004, 2008). These different paths of evolution can have a large bearing on the level of diversity of a given population and its ability to adapt.

Aside from environmental factors such as resource availability, and genetic interactions, migration of subpopulations can also have an effect on the selection outcomes of evolution. Dispersal between communities is well studied across ecological timescales (Cadotte, 2006), but less so across evolutionary timescales (Venail et al., 2008). A meta-analysis by Cadotte (2006) demonstrated the difficulty in coming to a universal conclusion of the role of gene flow in driving diversity over ecological timescales when analysing multiple studies across different organisms. The meta-analysis showed that while most studies are able to surmise a positive effect of dispersal on diversity, this is not always the case. This indicates that the relationship between gene flow and diversity may be influenced by the way in which organisms interact. These interactions often occur across environmental gradients such as temperature or resource gradients, which can also influence ecological and evolutionary interactions.

Increasingly, *in vitro* model systems are widely used to infer outcomes and effects of ecological and evolutionary processes by manipulating different features of the experiment. For example, experiments have demonstrated the effects of the environment on competitive ability (Jessup and Bohannan, 2008) and the effect of dispersal on adapta-
tion and rate of evolution (Lopez-Pascua et al., 2010). Furthermore, several studies have demonstrated that relatively low levels of gene flow among communities of coevolving species can lead to increased adaptation, relative to conditions with either high or no gene flow (Forde et al., 2004; Morgan et al., 2005, 2007). When gene flow is relatively low, it can act primarily as a source of variation on which selection can act (Gomulkiewicz et al., 1999; Lenormand, 2002; Forde et al., 2004; Garant et al., 2007). In contrast, high levels of gene flow can dominate patterns of spatially varying adaptation that would be driven by heterogeneity among local environments (Morgan et al., 2005) and can impede the rate of evolution in parasite populations (Vogwill et al., 2008).

Mathematical models have explored the role of gene flow across resource gradients. The effect of dispersal on communities with the same environment has been explored using models (Gandon and Michalakis, 2002) and has predicted that this can have the effect of homogenising the communities. A study by Hochberg and van Baalen (1998) showed that coevolution combined with gene flow along a gradient of prey productivity can result in higher investments in prey defences and predator (or parasite) counter-defences in high productivity populations than in low productivity populations. Their model predicted that the highest level of predator diversity should occur in patches with low to intermediate levels of prey productivity, whereas the highest prey diversity should be found in intermediate to high productivity patches. Furthermore, gene flow across spatial gradients in productivity was predicted to lead to a decrease in overall diversity summed across the gradient (Hochberg and van Baalen, 1998). However, these predictions were not explored experimentally. Loeuille and Leibold (2008) modelled the adaptive evolutionary dynamics of plant defences in a metacommunity food web along a productivity gradient. Their model predicted that dispersal increased diversity in resistance related traits and that the movement of nutrients among local communities altered trophic structure and diversity along the productivity gradient. Related models have explored symmetric and asymmetric movement of species between coevolutionary hot spots and cold spots and have indicated that relatively high levels of gene flow from hot spots can strongly influence evolutionary processes in neighbouring cold spots, and vice versa (Gomulkiewicz et al., 2000). More recent models have reinforced and expanded these results, demonstrating that gene flow across heterogeneous environments can fundamentally alter the outcome of coevolution (Gavrilets and Michalakis, 2008; Gandon and Nuismer, 2009). While theoretical studies discussed above have made important predictions it is rare that the models are tested and refined through being fully challenged by data.

If possible, it is useful to use models and experiments together so that models can in-
form experiments and vice versa. One such study that uses experiments to inform models is Forde et al. (2008). This followed the coevolution of *Escherichia coli* and T3 bacteriophage in different resource environments. A deterministic model was used to understand the genetic basis of interaction between host and the parasite. The model was able to capture the outcome of the experiments in terms of the host phenotypes observed. This was then used to make predictions based on other genetic interactions, this would be difficult to replicate experimentally. Here, we expand on Forde et al. (2008) by analysing the effects of gene flow on this system across resource gradients over evolutionary timescales. The productivity gradient adds more complexity to the system. Experiments were carried out on this system and using data on the host phenotypes observed, the diversity and resistance was analysed for different directions of gene flow. The expanded model was used to make predictions at different rates of gene flow and for other genetic interactions. We investigated the effects of gene flow for matching alleles (Gandon and Michalakis, 2002; Agrawal and Lively, 2002) and gene-for-gene interactions (Gandon and Nuismer, 2009; Agrawal and Lively, 2002). This demonstrates the benefit of having models that inform experiments that in turn inform models. Specifically, it was found experimentally that bidirectional gene flow had a negative effect on diversity. It was also observed that unidirectional gene flow can have a positive effect on diversity, although this was not found to be a significant increase. The model was used to illustrate that we do not expect to observe these effects at all rates of gene flow, and also that the predictions are dependent on the genetic interaction.

### 3.3 Background Model

The system that is used in the experiments is a continuous culture chemostat with glucose, *E. coli* and T3 phage. Here, we reproduce the work from Forde et al. (2008) which used ODEs to describe the population levels of the interacting players across time until a steady state was reached. Data from the experiments in this study gave data relating the growth rates of different phenotypes together with their relative ability to resist infection from phage phenotypes. The basis of the interaction was then incorporated in the model. Instead of simply applying for example, an existing gene-for-gene model, a model based on how the coevolving players interact and mutate was developed. Parameters in the infectivity matrix were directly related to genetic properties of the coevolving species as observed in the experiments. Wild type bacteria have long lipopolysaccharides (LPS) receptors, which are relatively easily detected by phage tail fibres. The bacteria then evolve by shortening their LPS to be more evasive to the phage. There are also
pleiotropic effects associated with these mutations (Sen and Nikaido, 1991). In particular, this mutation causes outer membrane proteins (OMPs) to reduce in size and this in turn lowers the ability of the bacteria to absorb glucose and thus reduces growth rate. For resistance to T2 phage, OMP F undergoes mutation. For resistance to K3 phage, OMP A undergoes mutation. To counteract the bacterial mutations, the phage mutate by shortening their tail fibres. This also has the effect of lowering their growth rate (burst rate) in infected bacteria. These coevolving properties are outlined in Figure 3.1. The wild type bacteria is denoted by $B_0$, and subsequent evolved types are defined by mutations in their lipopolysaccharides length (Laird et al. 1994), $L$, and OMP measure (Sen and Nikaido, 1991), $O$ (wild type $= 0$, or mutated $= 1$). For $B_0$, these characteristics are $L = 0, O = 0$. The bacteria can then evolve to become resistant to the different phage types by attaining mutations in these characteristics (Qimron et al., 2006). Type $B_1$ has $L = 0, O = 1$, type $B_2$ has $L = 1, O = 0$ and finally the most resistant type $B_3$ has $L = 1, O = 1$. Due to pleiotropic effects between the loci, each of these phenotypes has a unique LPS length $l$ where $l = 4 - (2 \times L + O)$. This gives the wild type bacteria $B_0$ the longest LPS length of 4, and the most resistant type $B_3$ the shortest length of 1. The wild type phage is denoted by $P_0$, with $P_1-3$ denoting evolved phage of increasing virulence and shortening tail fibre length. The infectivity matrix $\Phi$, which determines the outcomes of interactions between the bacteria and phage is shown below

$$
\Phi = \begin{pmatrix}
P_0 & P_1 & P_2 & P_3 \\
1  & \lambda  & \lambda^2  & \lambda^3 \\
B_0 & 0 & \lambda \nu & \lambda^2 \nu & \lambda^3 \nu \\
B_1 & 0 & 0 & \lambda^2 \nu^2 & \lambda^3 \nu^2 \\
B_2 & 0 & 0 & 0 & \lambda^3 \nu^3 \\
B_3 & 0 & 0 & 0 & 0 \\
\end{pmatrix}.
$$

This is referred to as the modified gene-for-gene model. The matrix is also multiplied by a small constant $\alpha = 2 \times 10^{-8} ml/(virions \cdot h)$ which is the adsorption rate. Here $\lambda$ represents the change of adsorption rate caused by alterations in the structure of phage tail-fibre protein (Paterson et al. 2010, Scanlan et al. 2011) and $\nu$ represents the change in adsorption rate caused by the loss of a single sugar from bacterial LPS complex (Sen and Nikaido, 1991; Qimron et al., 2006). These have constraints $\lambda < 1, \nu < 2$ respectively.

An alternative description of the infection mechanism is given by the matching alleles - gene-for-gene continuum Agrawal and Lively (2002). The corresponding infectivity matrix $\Phi_{MA-GFG}$ for the case with 4 host and 4 parasite phenotypes is as follows
3.3 Background Model

Figure 3.1: Description of how phenotypes evolve in both bacteria and phage. As wild type bacteria \( B_0 \) gain resistance to wild type phage \( P_0 \) it mutates by truncating its LPS. Bacteria can also gain resistance by a mutation to its OMP. This mutation also has pleiotropic effects such that it also truncates the LPS. In turn, phage mutate to be able to infect the bacteria \( B_1 \) by shortening their tail fibre length. Further infective phage \( P_1 \), \( P_2 \) and \( P_3 \) and resistant bacteria \( B_2 \) and \( B_3 \) evolve as shown.

\[
\Phi_{MA-GFG} = \begin{pmatrix}
P_0 & P_1 & P_2 & P_3 \\
B_0 & 1 & a(1 - ak) & a(1 - ak) & a^2(1 - ak)^2 \\
B_1 & 0 & 1 - ak & 0 & a(1 - ak)^2 \\
B_2 & 0 & 0 & 1 - ak & a(1 - ak)^2 \\
B_3 & 0 & 0 & 0 & (1 - ak)^2
\end{pmatrix}
\]

This is also multiplied by \( \alpha \) in the model. Agrawal and Lively suggested that both the matching alleles and gene-for-gene infection mechanisms are extremes lying on the same continuum. The \( a \) parameter \((0 < a < 1)\) describes the point along the continuum that the interaction of focus lies on. When \( a \) is 0, the interaction is matching alleles, and when \( a \) is 1, the interaction is gene-for-gene, while other values suggest an intermediate between the two. Finally, \( k \) is a cost of becoming an increasingly virulent parasite.

Mutational matrices \( M_b \) and \( M_p \) were used to enable the evolution of bacteria and phage respectively. For phage, the model requires only one mutation event to go from
any one phenotype to another, thus $M_p = I_4 + \epsilon(M_{p1} - I_4)$, where $M_{p1}$ is given by

$$M_{p1} = \frac{1}{3} \times \begin{pmatrix}
0 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 \\
1 & 1 & 1 & 0
\end{pmatrix},$$

where $\epsilon$ is the rate of mutations and $I_4$ is the $4 \times 4$ identity matrix.

For the bacteria, we base the mutations on two genetic loci, one representing an LPS mutation and the other representing an OMP mutation. Therefore to mutate from one phenotype to another, sometimes two mutations are required (for example from $B_0$ to $B_3$). The mutations that require a change at one locus are described by $M_{b1}$

$$M_{b1} = \begin{pmatrix}
0 & 1 & 1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 \\
0 & 1 & 1 & 0
\end{pmatrix}.$$

While those that involve changes at two loci are described by the matrix $M_{b2}$

$$M_{b2} = \begin{pmatrix}
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0
\end{pmatrix}.$$

To combine the information in these matrices we use the following formula

$$M_b = I_4 + \epsilon(M_{b1} - 2 \times I_4) + \epsilon^2(M_{b2} - I_4),$$

where we multiply $M_{b2}$ by $\epsilon^2$ to account for mutations at two loci.

These details were incorporated into the model in the following way

$$\frac{dS}{dt} = D(S_0 - S) - c\mu(S)B^T$$

$$\frac{dB}{dt} = M_b(\mu(S) \cdot B) - (\Phi P) \cdot B - DB$$

$$\frac{dP}{dt} = M_p(\beta \cdot (\Phi^T B) \cdot P) - DP.$$

(3.1)
The constant $D$ represents the dilution rate, $S$ represents the chemostat resource level and $S_0$ is the input vessel resource concentration. The bacterial density vector is given by $B = (B_0, B_1, B_2, B_3)$ while $P = (P_0, P_1, P_2, P_3)$ represents the phage densities vector. Resource consumption is modelled using Michaelis-Menten bacterial growth function $\mu(S) = \frac{\mu_{\text{Max}} S}{K + S}$ where $\mu_{\text{Max}}$ is the maximal growth rate of a bacterial phenotype $i$, for $i = 0, 1, 2, 3$ and $K$ is the half-saturation constant, and resource conversion rate $c$. The phage production is illustrated with a vector of burst sizes $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$ for the respective phage $j$, for $j = 0, 1, 2, 3$. Latent period was not explicitly modelled. All parameters were derived from the experimental data, apart from $\lambda$ and $\nu$, which were fitted numerically. The parameter values used are shown in Table 3.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{\text{Max}}^i$</td>
<td>Maximal growth rate of bacterial type $i$</td>
<td>$\mu_{\text{Max}}^0 = 1.18h^{-1}$, $\mu_{\text{Max}}^1 = 1.009h^{-1}$, $\mu_{\text{Max}}^2 = 0.89h^{-1}$, $\mu_{\text{Max}}^3 = 0.66h^{-1}$.</td>
</tr>
<tr>
<td>$K$</td>
<td>Bacterial half-saturation constant</td>
<td>$0.06\mu g/ml$</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>Burst size of phage type $i$</td>
<td>$\beta_0 = 306$, $\beta_1 = 153$, $\beta_2 = 99$, $\beta_3 = 72$ virions/cell</td>
</tr>
<tr>
<td>$D$</td>
<td>Chemostat dilution rate</td>
<td>$0.2h^{-1}$</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Rate of point mutations</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>$c$</td>
<td>Resource conversion rate</td>
<td>$2.3 \times 10^{-5} \mu g/cell$</td>
</tr>
<tr>
<td>$S_0$</td>
<td>Resource input parameter</td>
<td>high = $10^3 \mu g/ml$, low = $10^1 \mu g/ml$</td>
</tr>
<tr>
<td>$\nu, \lambda$</td>
<td>Fitted infectivity parameters</td>
<td>$\nu = 0.677$, $\lambda = 0.94$</td>
</tr>
</tbody>
</table>

### 3.4 Gene Flow Problem

#### 3.4.1 Experimental Methods

The experiment was used to observe the effects of direction of gene flow on diversity. The experiment also allowed us to check that the expanded model was fit for purpose. The chemostats were set at 3 resource levels; high resource (1000$\mu g/ml$ glucose), intermediate (100$\mu g/ml$) and low resource (10$\mu g/ml$). Each chemostat was inoculated with *E. coli* and T3 bacteriophage with the total solution maintained at 30ml. These were all
left to evolve for 24 hours and at this point they were treated differently depending on the
direction of the gene flow treatment. For the ‘no gene flow’ treatment, there was no dis-
placement of communities. For the ‘unidirectional’ treatment each chemostat was sam-
ped by withdrawal of 7.5 ml of the communities. In addition to the sampling, 3 ml from
the high resource sample were dispersed to the intermediate environment, and 3 ml of the
intermediate sample were dispersed to the low resource environment. Saline solution was
also added to the chemostats to keep the volumes constant. For the ‘bidirectional’ treat-
ment, samples from low resource were also added to the intermediate environment, and
samples from the intermediate environment were dispersed to the high resource chem-
ostat. The setup is illustrated in Figure 3.2. The chemostats were then allowed to run for
a further 48 hours, when sampling and dispersals were performed again. This occurred
at every subsequent 48 hours until day 13 when the experiment ended. To have some
measure of the effect of gene flow on diversity, the proportion of bacterial phenotypes
were measured at two time points at both high and low resource environments. The phe-
notypes were identified by their resistance to reference phages T2, Tu1a and K3 on agar
plates, as previously identified in Forde et al. (2008).

Figure 3.2: Illustration of transfer method for different directions of gene flow
3.4 Gene Flow Problem

3.4.2 Model

We expanded the deterministic model based on that described in Forde et al. (2008). The wild type bacteria and phage were identical to those used in that experiment, and the no gene flow treatment experiment was identical to the setup previously. Therefore the modified gene-for-gene infection mechanism used in the model and all of the parameter values (Table 3.1) could be used from that study. Terms to represent dispersal in the model were added. The experiment was only modelled at high and low resource concentrations to keep the model as simple as possible, but could have been expanded to also include intermediate resource equations should the model not have shown agreement with the experimental results. The equations are displayed below

\[
\frac{dS_L}{dt} = D(S_{0L} - S_L) - c\mu(S_L)B_L^T - F_1S_L + F_2S_H \\
\frac{dS_H}{dt} = D(S_{0H} - S_H) - c\mu(S_H)B_H^T - F_1S_H + F_3S_L \\
\frac{dB_L}{dt} = M_b(\mu(S_L) \cdot B_L) - (\Phi P_L) \cdot B_L - DB_L - F_1B_L + F_2B_H \\
\frac{dB_H}{dt} = M_b(\mu(S_H) \cdot B_H) - (\Phi P_H) \cdot B_H - DB_H - F_1B_H + F_3B_L \\
\frac{dP_L}{dt} = M_p(\beta \cdot (\Phi^T B_L) \cdot P_L) - DP_L - F_1P_L + F_2P_H \\
\frac{dP_H}{dt} = M_p(\beta \cdot (\Phi^T B_H) \cdot P_H) - DP_H - F_1P_H + F_3P_L.
\]

Here \( S_L \) and \( S_H \) represent the resource levels at low and high resource environments respectively. The resource concentrations in the low and high resource input vessel are given by \( S_{0L} \) and \( S_{0H} \). While similarly \( B_L = (B_L^0, B_L^1, B_L^2, B_L^3) \) and \( B_H = (B_H^0, B_H^1, B_H^2, B_H^3) \) represent the bacteria densities at low and high resource and \( P_L = (P_L^0, P_L^1, P_L^2, P_L^3) \) and \( P_H = (P_H^0, P_H^1, P_H^2, P_H^3) \) represent the phage densities at low and high resource. Resource consumption is again modelled using Michaelis-Menten bacterial growth function \( \mu(S) = \frac{\mu_{\text{Max}}^i S}{K + S} \) where \( \mu_{\text{Max}}^i \) is the maximal growth rate of a bacterial phenotype \( i \) for \( i = 0, 1, 2, 3 \) and \( K \) is the half-saturation constant, and resource conversion rate \( c \). The terms that describe gene flow (modelled continuously in this model) are given by \( F_1, F_2 \) and \( F_3 \) and are explained in more detail in Table 3.2 and Figure 3.2. In the experiment, sampling occurs in the unidirectional and bidirectional treatments. This is expressed by \( F_1 \), although setting this term to zero does not affect the
results qualitatively. In the unidirectional and bidirectional treatments there is dispersal from the high resource vessel to the low resource vessel, expressed by $F^2$. In addition to this, in the bidirectional treatment dispersal occurs from the low resource to the high resource vessel, expressed by $F^3$. The constant $A$, the rate of gene flow, varies between 0 and 1 in the model. The remaining constants, mutation matrices ($M_b$, $M_p$) and interaction matrix ($\Phi$) are the same as those used in Forde et al. (2008) as described in Section 3.3.

Table 3.2: Explanation of resource transfers in model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>No Gene Flow</th>
<th>Unidirectional</th>
<th>Bidirectional</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>0</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>F2</td>
<td>0</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>F3</td>
<td>0</td>
<td>0</td>
<td>A</td>
</tr>
</tbody>
</table>

3.5 Results

3.5.1 Experimental Results

The diversity was calculated using the Shannon-Wiener index. This is given by $-\sum_{i=1}^{N} p_i \ln p_i$, where $p_i$ is the proportion of phenotype $i$ for $i = 1, ..., N$. The results are displayed in Figure 3.3. It was observed that unidirectional gene flow increased diversity at high resource levels, while bidirectional gene flow decreased diversity in comparison to no gene flow. The data was also analysed in an analysis of variance, this is shown in Table 3.3. It was found that the direction of gene flow had a significant effect on diversity. There was no effect of resource level or interaction between the factors. A significant difference in diversity was found between unidirectional and bidirectional gene flow (p-value $< 0.0001$, 1 d.f.), and also between no gene flow and bidirectional (p-value $= 0.01$, 1 d.f.), and while there was a difference between the unidirectional and no gene flow treatments there was no evidence to suggest this was a significant difference. The finding for bidirectional gene flow is in contrast with previous findings which show gene flow to have a positive effect on diversity (Venail et al., 2008).
Table 3.3: Analysis of Variance of Shannon-Wiener Index of Diversity for experimental results

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum Sq.</th>
<th>d.f.</th>
<th>Mean Sq.</th>
<th>F</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direction</td>
<td>0.6345</td>
<td>2</td>
<td>0.31724</td>
<td>10.68</td>
<td>0.0009</td>
</tr>
<tr>
<td>Resource</td>
<td>0.02055</td>
<td>1</td>
<td>0.02055</td>
<td>0.69</td>
<td>0.4164</td>
</tr>
<tr>
<td>Direction*Resource</td>
<td>0.02749</td>
<td>2</td>
<td>0.01375</td>
<td>0.46</td>
<td>0.6368</td>
</tr>
<tr>
<td>Error</td>
<td>0.5347</td>
<td>18</td>
<td>0.02971</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.3: The diversity (± se) for the experimental results averaged across time for low resource (blue) and high resource (yellow). Dashed blue line is a reference for the level of no gene flow diversity, and the yellow dashed line is the reference diversity level for high resource.

Another measure of diversity is the total number of observed phenotypes. This data is displayed in Figure 3.4 where phenotypes are displayed in order of increasing resistance to reference phenotypes in accordance with Sieber et al. (2013). Across all of the treatments 17 phenotypes were observed. Of these, 11 were observed under no gene flow, 13 were observed under unidirectional and 10 were observed under bidirectional gene flow. This observation that the diversity is highest in the unidirectional case and lowest in the bidirectional case also agrees with the other measure of diversity in that bidirectional gene flow resulted in the lowest observed diversity. The number of phenotypes unique to
a particular resource environment was also calculated as a measure of diversity (Sieber et al., 2013) and this is displayed in Figure 3.5. Here it was observed that the bidirectional treatment had the highest percentage of shared phenotypes (60.0%). The treatment with the lowest percentage of shared phenotypes was the no gene flow (27.3%). This was to be expected given that in the no gene flow case there is no interaction between the environments, whereas in the bidirectional case, phenotypes in each environment are introduced into the other.

![Figure 3.4](image-url)

**Figure 3.4:** The proportion of phenotypes observed under each direction of gene flow averaged over time and resource level (± se). No gene flow - yellow, unidirectional - blue and bidirectional - green.
3.5 Results

Figure 3.5: The number of unique phenotypes at each resource level for each direction of gene flow. Number of phenotypes observed in both high and low resource - Blue, observed in high resource only - Yellow, and observed in low resource only - Brown.

3.5.2 Model Results

3.5.3 Direction

The model as outlined in Section 3.4.2 was simulated deterministically for gene flow rates ($A$) between 0 and $0.4h^{-1}$ for each direction of gene flow. Above this rate of sampling, the bacteria was not supported. This value of diversity is calculated as the average over time between days 3 and 13. It can be seen in Figure 3.6 that the diversity is dynamic over this period and has not reached a steady state. Moreover, gene flow in both the unidirectional and bidirectional cases causes the level of diversity to fluctuate. The Shannon-Wiener index as a function of rate of gene flow for each treatment is shown in Figure 3.7. The model results agree qualitatively with the experimental results for high rates of gene flow (see Figure 3.3) between $0.01h^{-1}$ and $0.05h^{-1}$. Specifically it shows that unidirectional gene flow maintains high diversity while bidirectional gene flow has a negative effect. The model also predicted for lower rates of dispersal that bidirectional gene flow increased levels of host diversity. It further suggests that very high rates of dispersal have a negative effect on diversity when compared to the no gene flow treatment.
The model is therefore able to capture the experimental results at high rates of gene flow, but also the expected results from previous literature at low and intermediate rates (Vénail et al., 2008). Crucially, the comparison with the experimental results show that the model is fit for purpose. It is interesting to note the observation that direction and rate of gene flow interact with respect to their effect on diversity. At low rates, bidirectional gene flow has a positive effect on diversity, but at higher rates it has a negative effect. In contrast, unidirectional gene flow had its highest diversity at high gene flow rates. The benefit of having such a model that reflects the experimental system well is that aspects of the model can be changed to infer information that would be difficult to carry out experimentally. We also note that this form of the model cannot be used to compare the number of phenotypes observed, as carried out for the experimental data.

Figure 3.6: Diversity over time for the host from day 0 to day 13 as measured using the Shannon-Wiener index, at high and low resource, with gene flow rate $0.025h^{-1}$. Blue - no gene flow, green - unidirectional, red - bidirectional gene flow.
3.5 Results

Figure 3.7: Host diversity plotted as a function of gene flow rate for high and low resource. This plot is based on the modified gene-for-gene infectivity matrix. Blue - no gene flow, green - unidirectional, red - bidirectional gene flow.

We also simplified the model by reducing the model from three levels of resource to two. We note that the observed proportion of phenotypes have the same steady states for both the intermediate and high resource levels under no gene flow. We do not therefore expect that this has any significant effect on the outcome of our predictions. We checked this hypothesis for diversity, this is shown in Figure 3.8. It can be seen bidirectional gene flow had a negative effect on diversity at high rates of gene flow as for the two chemostat model. We therefore conclude that it is sufficient to use two levels of resource, high and low only.
Figure 3.8: Host diversity plotted as a function of gene flow rate for high and low resource with 3 levels of resource in model. This plot is based on the modified gene-for-gene infectivity matrix. Blue no gene flow, green unidirectional, red bidirectional gene flow.

Finally, we want to look at the effect of using a discrete transfer in the model as carried out experimentally. Due to the highly dynamic nature of the interacting players, picking a single time point to conduct the transfer in a deterministic model may not be sufficient to replicate the experimental situation. We also note the added complexity that adding a step function to the equations brings to the simulations. Therefore, to try and replicate this we make change the constant transfer rate to a cosine function that peaks at the time points where the transfers were made experimentally. In this case the transfer rate becomes \( A(0.5 + 0.5 \cos((2\pi(t + 24))/48 - \pi)) \). The results from the model using this form of transfer are shown in Figure 3.9. We observe the same qualitative results as for the constant rate. Therefore, for simplicity we continue to use the constant rate of transfers for the remainder of the work.
3.5 Results

Figure 3.9: Host diversity plotted as a function of gene flow rate for high and low resource. In this case we use a cosine function to approximate the discrete transfers performed experimentally. This plot is based on the modified gene-for-gene infectivity matrix. Blue - no gene flow, green - unidirectional, red - bidirectional gene flow.

3.5.4 Infection Mechanism

It was also investigated whether the results were unique to the modified gene-for-gene interaction. To do this we looked at interactions lying on the matching alleles - gene-for-gene continuum (Agrawal and Lively, 2002) by changing the $\Phi$ matrix to the matching alleles - gene-for-gene form ($\Phi_{MA-GFG}$, Section 3.3). The $a$ parameter was varied between 0 and 1 at intervals of 0.1, and the model was run at each of these parameter values; this would be impossible to study experimentally. The ratio of diversity of unidirectional and bidirectional gene flow in comparison to no gene flow is shown in Figure 3.10. While we note that the level of diversity observed varies along the continuum (Forde et al., 2008), direction of gene flow can have less of an effect on diversity in comparison to the modified gene-for-gene model. We observe for matching alleles model ($a = 0$), where we observe high diversity under no gene flow, that there is less difference between the three types of gene flow for most rates. At very high rates we see a negative effect on diversity for both unidirectional and bidirectional gene flow. We also observe a larger peak in diversity at high resource for high rates of bidirectional gene flow. As the infectivity moves towards gene-for-gene, we observe more differences between the directions of gene flow. We observe both peaks and troughs in diversity at high rates.
of gene flow for both unidirectional and bidirectional gene flow, dependent on $a$. Looking vertically at lines of constant gene flow rate it can be seen how this effect changes across the continuum. However, what is most striking is the lack of difference between the unidirectional case and the bidirectional case at low resource. For the full gene-for-gene model ($a = 1$) we observe more similarity to the modified gene-for-gene results, while the effects of gene flow in the matching alleles model are smaller. This demonstrates the importance of the infection mechanism in determining the effect of gene flow on diversity. Further examples are shown in Sieber et al. (2013).

Figure 3.10: Diversity as a function of gene flow rate and point on the matching allele - gene-for-gene continuum. A) Ratio of Shannon-Wiener index of unidirectional to no gene flow at high resource. B) Ratio of Shannon-Wiener index of unidirectional to no gene flow at low resource. C) Ratio of Shannon-Wiener index of bidirectional to no gene flow at high resource. D) Ratio of Shannon-Wiener index of bidirectional to no gene flow at low resource. When $a = 0$, the infection mechanism is of the matching alleles form, and when $a = 1$ it is of the gene-for-gene form.
3.5 Results

3.5.5 No Resource Transfer

The experiments looked at the effect of gene flow where resource is also transferred in addition to the coevolving players. This is likely to be of most interest for aquatic environments. It is also possible that coevolving species interact without transfer of resource. While this would be difficult to study experimentally due to the transfer methods, this can be easily analysed using the model. The results are shown in Figure 3.11. In this case we find that bidirectional gene flow has a positive effect on diversity for most rates of gene flow, and doesn’t have the large negative effect that is observed when resource is also transferred. It is thought that this is because when the most resistant phenotypes are transferred to the low resource environment, they are unable to take a stronghold on the system as there is not enough resource available to sustain them. The gene flow therefore has the effect of introducing new phenotypes to the different environments for all rates of gene flow. In addition, unidirectional gene flow gives a positive effect on diversity at low resource across a wide range of gene flow rates. This increase is likely due to the same reasoning as for bidirectional gene flow.

![Figure 3.11: Host diversity plotted as a function of gene flow rate for high and low resource with no transfer of resource. This plot is based on the modified gene-for-gene infectivity matrix. Blue - no gene flow, green - unidirectional, red - bidirectional gene flow.](image-url)
3.6 Discussion

This study was set out to develop understanding of the relationship between gene flow and diversity in a coevolving system. The conclusion reached using *E. coli* and T3 phage chemostat experiments was that the direction of gene flow is crucial in determining the effect of gene flow. In particular, it was found that unidirectional gene flow can maintain the level of diversity, while bidirectional gene flow had a negative effect in comparison to no gene flow (Figure 3.3). Predictions of how this observation is affected by other factors not testable by the experiment were then made by developing an ODE model. It was found that this relationship is quite complex. Results from the model showed that there were different outcomes of the effect of gene flow for different rates of gene flow and different genetic interactions.

Experimentally it was observed that the direction of gene flow has a key effect on diversity. Namely it was found that diversity increased at high resource when the gene flow was unidirectional and maintained at low resource, and conversely it decreased across all resource levels when the gene flow was bidirectional. These observations for diversity were seen by analysing phenotypic evenness (Figure 3.3). In terms of number of observed phenotypes (Figure 3.4) it was found that most phenotypes were seen in the unidirectional case, with the fewest phenotypes observed in the bidirectional case. A model was introduced to explain these results. The model had good agreement with the experimental results for diversity properties at a particular set of gene flow rates (Figure 3.7). This demonstrated that the model was fit for purpose. Aspects of the model were then changed to look at some potential driving factors in more detail. This enabled us to analyse the robustness of the observations and whether they were due to particular conditions of the experiment. Factors investigated were the rate of gene flow and the genetic interaction. To provide insight into the experimental and model results, we start by noting that in our system at high resource environments, the more resistant phenotypes dominate while at low resource environments all phenotypes are present under no gene flow. This results in a higher level of diversity at low resource. The dominance of more resistant types at higher resource levels occurs as the cost to growth rate of being resistant to virulent phage is offset by the amount of resource. This can be seen in Forde et al. (2008). This holds for the no gene flow treatment in results here. For unidirectional gene flow at high resource, increases in gene flow rate can increase levels of diversity. This is due to the dilution of resource, which is more likely to be accommodating to less resistant types and reduce dominance of high resistance phenotypes. When the phenotypes from the high resource environment are introduced into the low resource environment the di-
versity is maintained, and even increased over no gene flow at particular gene flow rates. For bidirectional gene flow at low rates of gene flow, we find that a small introduction of allopatric phenotypes can increase the diversity. This is in agreement with previous studies (Venail et al., 2008). However, at higher rates of gene flow we observe the dominant types at high resource beginning to dominate in all resource environments as the introduction of sufficient resource into low resource environments allow them to survive there. This regime also shows agreement with previous work (Hochberg and van Baalen, 1998). The system therefore becomes homogeneous and this results in lower diversity in comparison to no gene flow. This therefore tells us that we would not expect to see the same observations experimentally if we were to change the amount transferred between the environments.

Further work on the study by Michael Sieber found that the unidirectional treatment led to the highest level of host population resistance against reference phages experimentally (Sieber et al., 2013). From the contributions of the respective phenotypes to the total resistance we infer that this is due to the number of phenotypes. Bidirectional gene flow was also found to have a beneficial effect on resistance in comparison to no gene flow. In this case it is believed that this was due to the dominance of the most resistant phenotypes. This also agrees with our model findings for diversity. The model found that the bidirectional gene flow gave the most resistance against the allopatric phages. The model did not find that unidirectional gave the most resistance. This is possibly due to the structure of the model, which had with a constrained number of phenotypes. It was therefore unable to capture the diversity in terms of the larger number of phenotypes observed under unidirectional gene flow. As a result the resistance seen experimentally could not be replicated with the model.

It was found that for other genetic interactions, the direction of gene flow had varying degrees of importance (Figure 3.10). Specifically, for matching alleles it was found that the direction of gene flow has less effect on diversity. In this case the form of the interaction has less selection pressure. This form of coevolution doesn’t allow phenotypes to acquire multiple resistance to phage phenotypes and doesn’t allow the most resistant phenotypes to dominate. Therefore, gene flow only had the effect of mixing the system. In contrast, for the gene-for-gene case we observed similar differences between bidirectional gene flow and no gene flow to those seen experimentally in the modified gene-for-gene case. Namely that there was a peak of diversity at low to intermediate rates of gene flow and a negative effect at high rates. However, in this case we also observed that this also held for unidirectional gene flow. Therefore direction of gene flow had little effect
under this infection mechanism. Using the matching alleles - gene-for-gene continuum (Agrawal and Lively, 2002) we were able to observe how this effect changes across the continuum. Again, this demonstrates that the experimental observations are specific to the experimental system.

It would be beneficial to test some of the hypotheses made by the model. Experiments could be used to see whether our predictions at lower rates of gene flow held in this system. It would also be useful to have more replicates of the experiments to give us a higher level of certainty of the difference between replicates. In this experiment there were a maximum of 4 replicates per gene flow direction. Moreover, it may be of interest to have information on the experiment at more of the time points that were sampled and this would not require any additional disturbance of the system. In this study, information was used on the samples at days 3 and 13 for continuity with Forde et al. (2008). Information at more time points would allow us to check whether we observe the oscillatory dynamics expected from the model. We focus on phenotypes in terms of the model that were obtained without gene flow. It is possible that the phenotypes observed in Forde et al. (2008) are not identical to those that were observed when gene flow was introduced. Here we argue that the phenotypes are general in terms of their resistance to reference phages and thus our predictions for the modified gene-for-gene model are likely to hold. We only investigated the results using the parameters derived in Forde et al. (2008). These gave strong agreement with the data in Forde et al. (2008) and this study. The system may not be robust to changes in parameters used. However we note that in nature, differences in physical parameters such as growth rates are could also have a large effect on ecosystems (Gudelj et al., 2004). We do not believe that this infringes on the validity of the results. Therefore, one would expect that small changes in parameters in the gene flow would alter what we observe, but this could potentially be the case in nature also.

In summary, it has been shown that relationship between gene flow and diversity is complex. The effect of gene flow depends on the amount of gene flow, the direction and the genetic basis of interaction. This finding is in line with the Geographic Mosaic Theory of Coevolution (Thompson, 1999, 2005). In making predictions it is therefore crucial to have information on such factors as genetic interaction, resource consumption, rate and direction of gene flow as lack of knowledge could easily lead to a false prediction.
Chapter 4

Population Scale Model of Lambda Phage

4.1 Overview

A paper by Stewart and Levin (1984) presents a number of hypotheses for reasons to be temperate. It concludes that the reason is to preserve the phage when the density of bacteria oscillates below a level at which phage can grow. This conclusion was reached without information that we now have, on how the decision to undergo lysogeny is made in lambda phage. Furthermore, it was not demonstrated conclusively that this hypotheses was plausible using the model. We now have more information on how the decision is made; specifically we have a functional form to describe how the number of infecting phage (multiplicity of infection) and host cell length affect the rate of lysogeny (Zeng et al., 2010). It has been demonstrated that a higher viral concentration (the ratio of infecting phage to cell length) increases the rate of lysogeny. Does this new information fit in with the conclusion in the Stewart and Levin (1984) paper or is there an alternative hypothesis that can be drawn? We use bifurcation analysis and numerical solutions of ODEs to investigate this question.

4.2 Introduction

There are many different types of phage. Many replicate by the lytic cycle (Section 1.2), but others such as lambda phage (Lederberg and Lederberg, 1953) can also choose to undergo a dormant stage and replicate by bacterial division. These types are called
temperate phages. We are interested in why certain phages act in this way, what is the 
benefit in being temperate over being virulent? This has been asked before in Stewart 
and Levin (1984), however the analysis does not give conclusive evidence to justify the 
resulting hypothesis. It is therefore interesting to further investigate and compare the 
claims from this study and develop them further using the more detailed information we 
now have on the rate of lysogeny in lambda phage from Zeng et al. (2010).

In Stewart and Levin (1984) reasons for being temperate are discussed. It is suggesed 
that being temperate is of benefit in hard times, that is in low resource conditions. How-
ever the analysis does not demonstrate this conclusively, that is, it is not demonstrated 
that being temperate gives a survival advantage over lytic phage under such conditions. A 
study by Avlund et al. (2009) uses a game theoretic approach to suggest that when there 
is one infecting phage, a deterministic decision of either lysis of lysogeny gives the phage 
a better chance of survival. If there are more than one infecting phage then a stochastic 
choice is advantageous to minimise chance of extinction. This work was formulated with 
experimental data from classic experiments by Kourilsky (1973) as a basis. This work 
suggested that one infecting phage always chooses lysis, whereas more than one gives a 
stochastic decision which is biased towards choosing lysogeny. The experimental work 
in Zeng et al. (2010) does not agree with this simple relationship between multiplicity of 
infection (MOI) and rate of lysogeny. Recent studies have suggested that different factors 
affect the rates of lysogeny in lambda phage. In particular, it has been demonstrated that 
MOI (Arkin et al., 1998), and cell volume (St-Pierre and Endy, 2008), are largely respon-
sible for the decision to become lysogenic. This work was then combined in Zeng et al. 
(2010) to demonstrate that the phage concentration (MOI divided by cell length) largely 
determines the outcome. However, the decision is still stochastic. It is not clear whether 
rate of lysogeny in all types of temperate phage are influenced by the same factors. Work 
in other systems, for example persisters cells, has shown the benefit of constant rates of 
switching over responsive rates (Kussell and Leibler, 2005).

A review of the ecology of phage by Weinbauer (2004) gives information on where 
lysogens are more prevalent. It is said that in general, the highest percentage of lysogens 
is found in offshore seawaters. These are low productivity environments in comparison 
to coastal waters. The percentage of lysogens in these environments has been recorded 
between 10% and 50%. This gives an indication that temperate phage have evolved to be 
of benefit in hard times. However, a study in coastal waters in a Canadian fjord recorded 
lysogeny rates as high as 80%. This gives indication that their advantage may be in 
highly productive environments with fluctuating resource conditions (Mittler, 1996). In
an epidemiological context it has been observed that higher latency is an advantage in later stages of an epidemic when the number of susceptible hosts is low (Berngruber et al., 2013). Therefore, we find examples that possibly support different hypotheses, but not conclusive evidence for championing one over the other. We suggest that while they may have evolved to be of use in one of these situations, they are potentially advantageous under both sets of conditions.

We now attempt to include this new information from Zeng et al. (2010) at the single cell level and revisit the work carried out in Stewart and Levin (1984) and further develop the understanding of the ecological reasons for being temperate using population scale models. It was suggested in Stewart and Levin (1984) that lysogenic types persist when the host population fall on hard times. We investigate whether this is captured with a more detailed model that includes more information on the behaviour of the rate of lysogeny for lambda phage using bifurcation analysis. We also investigate whether there is an alternative conclusion for the advantage of being temperate. We then briefly explore the effects of fluctuating resource conditions on the system. Using these methods we look to demonstrate why the lysogenic switch responds to the environment in this way.

### 4.3 Model

The system is represented by the ODE model below, following Stewart and Levin (1984)

\[
\begin{align*}
\frac{dS}{dt} &= D(S_0 - S) - c\mu(S)B - \frac{c\mu(S)L}{q} \\
\frac{dB}{dt} &= \mu(S)B - \Phi_TTB - \Phi_PPB + \tau L - DB \\
\frac{dL}{dt} &= \frac{\mu(S)L}{q} + \Phi_TfTB - \text{ind}L - \Phi_PPL - \tau L - DL \\
\frac{dT}{dt} &= \beta_T\Phi_T(1 - f)BT + \beta_T\text{ind}L - DT \\
\frac{dP}{dt} &= \beta_P\Phi_P(B + L)P - DP 
\end{align*}
\] (4.1a-4.1e)

where \( S \) is resource concentration, \( B \) is density of susceptible bacteria, \( L \) is the density of lysogens, \( T \) is the density of free temperate lambda phage and \( P \) is the density of free virulent phage. The model is based on a well mixed environment such as a chemostat where input vessel resource concentration is given by \( S_0 \), while \( D \) is the dilution rate. The adsorption rate of virulent phage is given by \( \Phi_P \), and \( \Phi_T \) is the adsorption rate of
temperate phage. Burst rate of virulent phage is $\beta_P$ and $\beta_T$ is the burst rate of temperate phage undergoing lysis. The bacterial growth function $\mu(S) = \mu_{\text{Max}} \frac{S}{K+S}$, where $\mu_{\text{Max}}$ is the maximal growth rate of bacteria, $K$ is the half-saturation constant, and $c$ is resource conversion rate. The cost on growth rate for being a lysogen is given by $q$. The rate of induction, that is the rate at which cells switch from lytic to lysogenic, is given by $\text{ind}$. While $\tau$ is the rate of vegetative segregation, which is the rate that lysogens divide with no phage and hence become susceptible bacteria. Here $f$ is the rate of lysogeny. This can be a constant such that

$$f = f_c = \text{lys},$$

where $0 < \text{lys} < 1$ (Stewart and Levin, 1984). Alternatively it can be a responsive rate such as the one obtained in Zeng et al. (2010) as

$$f = f_r(M, h, J, l) = \left( \frac{M/l}{J^h + (M/l)^h} \right),$$

where $M$ is the multiplicity of infection, $l$ is the cell length, $J$ is the half saturation constant and $h$ is the Hill number. While the decision is still considered noisy, this function has been shown to give a good prediction of the decision on average. The functional form for the rate of lysogeny allows us to observe conditions where the rate of lysogeny is highest. This information may allow us to infer the conditions that temperate phage is best adapted to. To understand the role of this form of decision making we attempt to relate the population densities to the quantity of interest, the MOI. For this analysis we will assume that in the well-mixed chemostat environment, the MOI is proportional to the ratio of free phage ($T$) to the density of uninfected bacteria ($B$). This gives us

$$M = \kappa \frac{T}{B}$$

for some constant $\kappa$, where $0 < \kappa < 1$. In a spatially structured environment, we would expect the MOI to be proportional to the concentration of free phage only.

Initially, we perform a bifurcation analysis of the system. This enables us to understand the steady states of the system and the parameter regimes in which all players are present at the steady state and regimes where players are not able to survive. Steady state solutions were found using Matlab. The bifurcation analysis was performed using AUTO, and by manually calculating eigenvalues to give stability information on the solutions. We firstly conduct this analysis for temperate phage with a constant rate of lysogeny, and then for temperate phage with a responsive rate.
4.4 Results

4.4.1 Lytic Phage and Bacteria Dynamics

Phage replicate by infecting bacteria, therefore their population size is influenced directly by the host population size. This can be seen clearly by observing the sign of the equations 4.1 at different host population sizes. In particular, the phage population increases only at a sufficiently high host density, and the host density decreases at sufficiently high phage density. From observing the dynamics of a simulation involving bacteria and lytic phage it can be seen that the phage population both increases and decreases more rapidly than the host population (Figure 4.1).

![Figure 4.1: Illustration of dynamics of Bacteria (top) and Phage dynamics (bottom) in a well mixed environment based on parameters in Table 4.1. This observation represents lytic phage, however a similar figure can be obtained for temperate phage with appropriate parameters.](image)

To find out when temperate phage are advantageous over lytic phage we search for conditions that lytic phage cannot survive in using simulations. At low resource levels, the host density is not large enough to support phages. In addition, if the resource concentration supports both bacteria and phage but then drops below this level, the phage die out rapidly. This relates to the situation described in the Stewart and Levin (1984) paper as the phage undergoing hard times. This can also be interpreted as the bacteria being malnourished and having a low growth rate. Another situation that can lead to the elimi-
nation of lytic phage is at high resource concentrations. In the absence of a resistant host the phage can kill off the hosts, and hence eliminate themselves. This is due to the high density that the phage can reach in a sufficiently high productivity environment. While in nature we expect a resistant host to evolve (Clement et al., 1983; Meyer et al., 2012), this is at least theoretically possible in fluctuating environments where many phage could enter the system over a short period of time. This phenomenon is utilised in phage therapy (Thiel, 2004).

4.4.2 Bifurcation Analysis - Temperate Phage with Constant Rate of Lysogeny

Bacteria Growth Parameters

To analyse the potential for this elimination more thoroughly, we extend the work of Stewart and Levin (1984) and perform a bifurcation analysis of the system. In keeping with their work we first focus on the system where the rate of lysogeny is constant. In an attempt to learn more about the functional form of the rate of lysogeny described in Zeng et al. (2010) we then perform a similar analysis for the case where temperate phage have a responsive rate of lysogeny (Section 4.4.3).

An important parameter in the system is the growth rate of the susceptible bacteria $B$ ($\mu_{Max}$). It can be easily seen by inspection that the sign of the right hand sides of equations 4.1d and 4.1e, that given a particular burst rate ($\beta_T$ or $\beta_P$) and absorption rate ($\Phi_T$ or $\Phi_P$), phage ($T$ or $P$) need a set amount of bacteria ($B$) to grow. Therefore, the growth rate of bacteria affects the growth of phage in the system. We initially set up the system by choosing appropriate parameters so that all players coexist in the system. The parameters used to find initial steady states are shown in Table 4.1. Here it should be noted that we start with a higher burst rate for temperate phage $\beta_T$, than for virulent phage $\beta_P$. The reason for this is that it allows both lytic and temperate phage to coexist. This is explained in the section that follows concerning the infection parameters. To investigate the role of $\mu_{Max}$ we initially varied this parameter only. From Figure 4.2 it can be seen that this parameter has a marked effect on the system. There are 5 branches, which are labeled in relation to the players that coexist along them. The points at which the branches change are labeled by number 1-4. It can be seen for large $\mu_{Max}$ that both $T$ and $P$ can coexist. While there are other solutions possible at large values of $\mu_{Max}$, the solution with all players coexisting is the stable one. The other solutions can be obtained with different initial conditions. Decreasing $\mu_{Max}$, at branch point 1, $P$ can no longer
exist. This branch then becomes the stable solution. However, it must also be noted that this does not imply superior survival qualities for temperate phage $T$. This is due to the value of the burst parameter $\beta_T$, which is greater than $\beta_P$ in this case. The effect of changing this parameter can be seen in the section on infection parameters. If we change $\beta_P$ equal to $\beta_T$, then we find that the lowest value of $\mu_{Max}$ for which $P$ exists is equal to that for $T$ (found by inspection of AUTO generated data sets). Decreasing further to branch point 2, $T$ and $L$ can no longer exist. This leaves $B$ only existing as the single stable solution. Increasing $\mu_{Max}$ along this branch, point 3 is reached at which $P$ can now exist in the absence of $T$. This is an unstable branch. Decreasing $\mu_{Max}$ from point 2 further, point 4 is reached. At this point the none of the players are able to exist as the growth rate is too low to support $B$. It can be seen that for all points where $P > 0$ that also $P > T$. We also note that equivalently the initial resource concentration $S_0$ can be changed instead of $\mu_{Max}$ to analyse the effect of growth of bacteria on the respective phage. Figure 4.2 shows the density of $B$ for different solutions, the densities of $L$, $T$ and $P$ can be found in Appendix Figure A.1.
Chapter 4. Population Scale Model of Lambda Phage

Figure 4.2: Bifurcation Plot for system of equations 4.1 changing parameter $\mu_{Max}$ where $T$ has a constant rate of lysogeny $f = f_c$. The solution is shown for population of bacteria $B$. Solid lines denote stable solutions, dashed lines represent unstable solutions. Numbers 1-4 denote branch points. At 1, the solution with all 4 players coexisting changes to the one where $P$ is eliminated. At 2, $T$ and $L$ are eliminated leaving $B$ only. Increasing $\mu_{Max}$ we reach point 3, at which $P$ can again survive. For values of $\mu_{Max}$ below that at branch point 4, nothing is able to survive.

From Figure 4.2 we believe that the most interesting branch point to focus on is point 1. This is the point at which lytic phage $P$ is not able to survive, but temperate phage $T$ and lysogens $L$ remain. If this point came at a higher value of $\mu_{Max}$ than the point at which temperate phage $T$ and lysogens $L$ are no longer able to survive across a wide range of parameters then we could possibly conclude that temperate phage are able to survive where the growth rate of the susceptible host is low. To investigate this, we change other parameters of the system at these branch points. One parameter of interest is the cost parameter $q$, and the effect of changing this at branch points 1 and 2 is shown in Figure 4.3. Changing $\tau$ and $ind$ at the branch points obtained by varying $\mu_{Max}$ was also investigated. However, it is not believed that these have any additional qualitative effect on the system across a realistic parameter range.
4.4 Results

![Graph showing the location of branch points in system 4.1 when changing bacteria growth parameter $\mu_{Max}$ and cost to growth of being a lysogen $q$ where $T$ has a constant rate of lysogeny $f = f_c$]

Figure 4.3: Location of branch points in system 4.1 when changing bacteria growth parameter $\mu_{Max}$ and cost to growth of being a lysogen $q$ where $T$ has a constant rate of lysogeny $f = f_c$

From Figure 4.3 we infer that a cost to growth of being a lysogen $L$ actually makes it more difficult for virulent phage $P$ to survive. However, we note that this is only in the presence of an imposed cost on growth of $P (\beta_P < \beta_T)$. If there were no cost, the virulent phage would eliminate the temperate phage $T$, regardless of any cost to being a lysogen. This possibly implies an advantage in being temperate when the resource is limiting and there is coexistence of temperate phage $T$ and virulent phage $P$. In fact, increasing the cost gives less competitive advantage to virulent phage $P$ by reducing the number of lysogens $L$ and hence makes $T$ more similar in behaviour to $P$. While we observe that branch points 1 and 2 cross when $q < 0.6$, this would give lysogens $L$ a growth advantage over susceptible bacteria $B$, and thus make virulent phage $P$ more competitive despite having a lower burst rate $\beta_P$ in comparison to $\beta_T$. However, we have no reason to suggest that infection from temperate phage $T$ has a positive effect on growth rate of $L$ and are therefore not concerned with this behaviour. We also looked at changing $q$ only and keeping all other parameters the same. This revealed no qualitative changes in the system. Another important parameter to change at this point is the burst rate of virulent phage $\beta_P$. The initial value of this parameter was smaller than that of the temperate phage $\beta_T$ to ensure coexistence. It is therefore interesting to look at the location of the branch point, and the behaviour of the system when $\beta_P$ approaches $\beta_T$.
and also when $\beta_P > \beta_T$. This behaviour is shown in Figure 4.4. It can be seen that branch points 1 and 2 cross at the point where $\beta_P \simeq \beta_T$. This therefore confirms that any advantage in being temperate at this low growth rate was simply due to the lower value of $\beta_P$. At this point virulent phage $P$ eliminate temperate phage $T$. For lower values of $\mu_{Max}$, both species of phage die out.

![Figure 4.4: Location of branch points in system 4.1 when changing bacteria growth parameter $\mu_{Max}$ and burst rate of virulent phage $\beta_P$, where $T$ has a constant rate of lysogeny $f = f_c$.](image)

**Phage Infection Parameters:**

The burst rates of the phage $\beta_T$ and $\beta_P$ have an important effect on the system. We investigate the effects of changing both of these parameters. We only show the effect of changing one of them here, arbitrarily $\beta_P$, as it is the ratio of these parameters that is important. In Figure 4.5 we see the effect of changing $\beta_P$ on the system. Here it appears that if $\beta_P > \beta_T$ then the virulent phage eliminate the temperate phage.
4.4 Results

Figure 4.5: Bifurcation plot for system of equations 4.1 changing parameter $\beta_P$, where $T$ has a constant rate of lysogeny $f = f_c$. The solution is shown for population of bacteria $B$. Solid lines denote stable solutions, dashed lines represent unstable solutions. Numbers 1-3 denote branch points. At 1, the solution with all 4 players coexisting changes to the one where $P$ is eliminated. At 2, $\beta_P$ increases sufficiently to eliminate $T$ and $L$. Decreasing $\beta_P$ we reach point 3, at which $B$ only can survive.

By exploration of the effect of changing other parameters, it appears that the only parameters that affect this conclusion are $\phi_T$, $\phi_P$ and $\text{lys}$. At branch point 1, $P$ is eliminated due to its low burst rate $\beta_P$. For lower burst rates, the solution where $B$, $L$ and $T$ coexist is the stable branch. Increasing $\beta_P$ from branch point 1 allows all species to coexist. For further increases we see one branch becoming negative when $\beta_T \simeq \beta_P$. In reality, $T$ and $L$ approach 0 and $P$ approaches the value it has in the absence of temperate phage. This change occurs at branch point 2, and for larger values of $\beta_P$ this is the stable solution. Decreasing $\beta_P$ along this branch we arrive at branch point 3, where $P$ is eliminated and only $B$ remains due to low $\beta_P$. This indicates that virulent phage $P$ are better competitors than temperate phage $T$ under fixed resource conditions. In Figure 4.5 it appears that $T$ is eliminated for $\beta_P > \beta_T$ but this is due to the small value of $\text{lys}$, the rate of lysogeny in our initial parameter set. In Figure 4.6 we change $\text{lys}$ at branch point 2, where $P$ eliminates $T$ to demonstrate the effect that this parameter has on the system. It can be seen for larger $\text{lys}$ that $P$ can eliminate $T$ and $L$ at lower values of $\beta_P$. 

Figure 4.6: Location of branch points in system 4.1 when changing the burst rate of virulent phage $\beta_P$ and rate of lysogeny $\text{lys}$, where $T$ has a constant rate of lysogeny $f = f_c$.

4.4.3 Bifurcation Analysis - Temperate Phage with Responsive Rate of Lysogeny

Bacteria Growth Parameters

Here we perform a similar analysis to the above, but now we set $f = f_r$. We note that while in Zeng et al. (2010) the responsive function $f_r$ has a measured Hill coefficient of 1.08, we approximate it to 1 for simplicity of analysis.

Similarly to Section 4.4.2 we start to observe the system by analysing the effect on changing $\mu_{Max}$. This is shown in Figure 4.7. The behaviour is similar to that for the system with the constant rate of lysogeny (Figure 4.2), with a slight difference for the location of branch points. However we observe that they can be made equal by changing $\kappa$ for responsive $T$ or $\text{lys}$ for $T$ with constant rate of lysogeny. The effect of changing $\kappa$ can be seen in Figure 4.8. The important point here is that while we may be unsure as to the true value of $\kappa$, its value here does not affect the system qualitatively. We also note that in Figure 4.7 (branches 1 and 2), that both $T$ and $L$ have a higher equilibrium density in the case when $P$ has been eliminated, than when everything coexists (see Appendix
4.4 Results

This is in contrast to Figure 4.2 and Appendix Figure A.1, where only $L$ is greater when $P$ is eliminated than under coexistence conditions. This is possibly due to how $f_r$ now adjusts to the particular situation. Further changes of parameters at branch points of interest gave similar qualitative results to the case with constant rate of lysogeny. However, it is interesting to observe the behaviour when we vary $q$ at branch points in Figure 4.7. This is shown in Figure 4.9. In contrast to Figure 4.3, we find no additional benefit to temperate phage $T$ by varying $q$. We find that this is also the case for changing other parameters such as $ind$. The system therefore appears to be more robust to parameter changes than for the case with a constant rate of lysogeny. This implies that in lambda phage at least, there is no benefit to being temperate at low resource levels under equilibrium (and coexistence) conditions.

![Figure 4.7: Bifurcation Plot for system of equations 4.1 changing parameter $\mu_{Max}$, where temperate phage has responsive rate of lysogeny $f = f_r$. The solution is shown for population of bacteria $B$. Solid lines denote stable solutions, dashed lines represent unstable solutions. Numbers 1-4 denote branch points. At 1, the solution with all 4 players coexisting changes to the one where $P$ is eliminated. At 2, $T$ and $L$ are eliminated leaving $B$ only. Increasing $\mu_{Max}$ we reach point 3, at which $P$ can again survive. For $\mu_{Max}$ smaller than at branch point 4, nothing is able to survive.](image-url)
Figure 4.8: Location of branch points in the system with responsive temperate phage $T$ when changing bacteria growth parameter $\mu_{\text{Max}}$ and population MOI parameter $\kappa$.

Figure 4.9: Location of branch points in the system with responsive temperate phage $T$ when changing bacteria growth parameter $\mu_{\text{Max}}$ and cost to growth parameter $q$. 

4.4 Results

Phage Infection Parameters

We next investigated the effect of changing $\beta_P$ in the system with a responsive rate of lysogeny. The effect of this is shown in Figure 4.10. Here we see similarity with Figure 4.5 in that virulent phage $P$ eliminate temperate phage $T$ for cases when $\beta_P > \beta_T$.

![Figure 4.10](image)

Figure 4.10: Location of branch points in the system with responsive temperate phage $T$ when changing the burst rate of virulent phage $\beta_P$.

For the case of temperate phage with constant rate of lysogeny, we observed that the survival of temperate phage is dependent on the interaction of $\beta_T$ with $\text{lys}$. For temperate phage with a responsive rate of lysogeny, this parameter is redundant. The rate function $f_r$ does have an additional parameter $\kappa$ which relates the density of $B$ and $T$ to MOI. We therefore look at the effect of changing this parameter on the location of the branch points found in Figure 4.10. This analysis is shown in Figure 4.11. It can be seen that this parameter has no effect on the capacity of the temperate phage to compete with lytic phage under constant resource conditions. From this, it can be inferred that the responsive temperate phage is able to raise its rate of lysogeny to 1 without putting it at a disadvantage in competing with lytic phage. In contrast, in the previous section it was found that for temperate phage with a constant rate of lysogeny, a higher rate of lysogeny gave the phage a disadvantage when competing with lytic phage.
4.4.4 Fluctuating Conditions

In response to the previous hypothesis of the reason for being temperate (Stewart and Levin, 1984), we compare what happens when bacteria and phage are in the same environment and the bacteria suddenly drop to very low numbers. From comparison with the dynamics of lytic phage, we observed that phage density changes more rapidly than the hosts (section 4.4.1). Therefore at a given time point, from the point of view of the phage the host population is constant. In a high productivity environment the phage population is increasing and we would expect the rate of lysogeny to increase for temperate phage with a responsive rate of lysogeny. In a low productivity environment the phage population is decreasing, therefore we would expect the rate of lysogeny to be decreasing. If there is no other advantage in being temperate then it is interesting that when faced with this situation that the phage decrease their rate of lysogeny. From this, we could infer that temperate phages are best adapted to protecting themselves from dying out in high productivity environments by utilising inhibition of superinfection. Moreover, while they are able to help in hard times, this could be seen as an additional advantage. To test these statements we observe the dynamics using the same model, but allowing the resource level $S_0$ to fluctuate. In this analysis we compete not only lytic and tem-
perate types, but also temperate types with constant rate of lysogeny, denoted by $T_C$ and producing lysogens $L_C$, against temperate type with responsive rate if lysogeny, denoted by $T_R$ and producing lysogens $L_R$. Specifically, we allow a population of bacteria and temperate and lytic phages to grow to a steady state and then reduce the resource level $S_0$ to $1 \mu g ml^{-1}$. This is below the level that can sustain bacteria or any phage for our parameter range. This is reduced for a period of time $T_{fluc}$. The resource level $S_0$ is then increased again to a level at which all species can coexist for period $T_{fluc}$. This process is repeated many times. We then observe the dynamics that follow and measure the time taken until the each phage population ($T_R$, $T_C$, and $P$) goes below 1. We call this the time until death (TUD). In the case of temperate phages we require that both the phage and the corresponding lysogen population drop below this level. An illustration of the resulting dynamics is shown in Figure 4.12. We test this for different rates of lysogeny and different dependencies of the global MOI on $\kappa$. All other parameters are determined from those which allowed all phage to coexist under fixed resource conditions and are shown in Table 4.1. These are similar to those used in Stewart and Levin (1984).

![Figure 4.12: Dynamics of System Under Fluctuating Resource Conditions. A) Resource concentration over time. B) Bacteria $B$ and lysogen densities $L_C$ and $L_R$ over time. C) Lytic phage $P$ and temperate phages $T_C$ and $T_R$ densities over time.](image-url)
First we show that for smaller fluctuations, both in terms of resource concentration and $T_{Fluc}$, we observe that everything survives across the different parameters we choose in relation to rate of lysogeny. This is displayed in Figure 4.13. When changing the resource concentration or $T_{Fluc}$ or both, we observe different behaviours. Each player can eliminate the other, dependent on the parameters and we outline each case below.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{max}$</td>
<td>Maximal growth rate of susceptible bacteria $i$</td>
<td>$\mu_{max} = 0.7h^{-1}$</td>
</tr>
<tr>
<td>$K$</td>
<td>Bacterial half-saturation constant</td>
<td>4$\mu g/ml$</td>
</tr>
<tr>
<td>$c$</td>
<td>Resource conversion rate</td>
<td>$2.3 \times 10^{-5}\mu g/cell$</td>
</tr>
<tr>
<td>$\beta_T$</td>
<td>Burst size of Temperate Phage</td>
<td>$\beta_T = 200\text{virions/cell}$</td>
</tr>
<tr>
<td>$\beta_P$</td>
<td>Burst size of Virulent Phage</td>
<td>$\beta_P = 100\text{virions/cell}$</td>
</tr>
<tr>
<td>$D$</td>
<td>Chemostat dilution rate</td>
<td>$0.2h^{-1}$</td>
</tr>
<tr>
<td>$\Phi_T$</td>
<td>Adsorption rate of Temperate Phage</td>
<td>$\Phi_T = 10^{-9}\text{ml/(virions \cdot h)}$</td>
</tr>
<tr>
<td>$\Phi_P$</td>
<td>Adsorption rate of Virulent Phage</td>
<td>$\Phi_T = 10^{-9}\text{ml/(virions \cdot h)}$</td>
</tr>
<tr>
<td>$ind$</td>
<td>Rate of induction</td>
<td>$ind = 10^{-3}\text{cells/hour}$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Rate of vegetative segregation</td>
<td>$\tau = 10^{-3}\text{cells/hour}$</td>
</tr>
<tr>
<td>$h$</td>
<td>Hill number</td>
<td>1.08</td>
</tr>
<tr>
<td>$J$</td>
<td>Lysogeny rate half-saturation constant</td>
<td>2.64</td>
</tr>
<tr>
<td>$l$</td>
<td>Normalised Cell Length</td>
<td>1</td>
</tr>
<tr>
<td>$q$</td>
<td>Cost to growth rate for Lyso-gens</td>
<td>1.1</td>
</tr>
</tbody>
</table>
Figure 4.13: Parameters at which all species survive under fluctuating conditions. It can be seen that for small fluctuations of $T_{fluc} = 10$ hours or $T_{fluc} = 50$ hours and small resource fluctuations $S_0 = 5$ or $S_0 = 10$, that everything can coexist under fluctuating resources at all rates of lysogeny and MOI dependencies investigated.

When $T_{Fluc}$ is large we observe that $P$ is eliminated, while $T_R$ and $T_C$ are able to survive for longer and in some cases remain indefinitely. Evidence for this is shown in Figure 4.14. Here we see that as described in Stewart and Levin (1984), the temperate phages are able to sequester resources and outlast the lytic phage. We also see that for sufficiently high $\kappa$, the $T_R$ can outlast $T_C$. This is demonstrated in Figure 4.15. It can be seen that $T_C$ is also able to eliminate $T_R$ for low $\kappa$. These observations are consistent across all values of $S_0$ and all sufficiently large values of $T_{fluc}$. However, this low value of $\kappa$ is unlikely to be observed. While it was noted at the beginning of this section that phage with a responsive rate of lysogeny on the surface do not appear to be well adapted to conditions where resource oscillates to a low level, in practice it turns out to be a good competitor. The reason for this is that temperate phage under decreasing resource conditions are unable to make sufficient lysogens to preserve the phage population. The few lysogens produced at this stage will soon die out due to the limited resource level and dilution. Therefore, increasing the rate of lysogeny under such conditions will not extend...
the TUD. The rate of lysogeny is increased as resource increases, and here the number of lysogens can increase greatly. This leaves the temperate phage population in a good position should the resource level be suddenly decreased. This is a more efficient method of acquiring lysogens. This can be seen by analysis of the equation 4.1c when considering population densities observed in simulations. When the resource concentration is decreased the growth terms are small in comparison to the dilution term. Therefore, the best strategy is to maximise the number of lysogens before the host density starts to decrease.

Figure 4.14: Parameters at which lytic phage is eliminated under fluctuating conditions. Observations show difference in TUD for the weakest temperate phage and lytic phage which shows how much longer temperate phage is able to survive than lytic phage. A white square indicates that the phage were not eliminated.
4.4 Results

Figure 4.15: Parameters at which temperate phage is eliminated under fluctuating conditions. Here lytic phage is eliminated in all cases and the observations give an indication of advantage to $T_C$ or $T_R$. Positive values of TUD indicate $T_R$ was the best survivor.

We also note that it is possible that a high productivity environment can pose a problem for lytic phage. When the resource concentration increases greatly ($S_0 = 100$) and $T_{fluc}$ is short, that lytic phage can eliminate all players. Figure 4.16 shows parameters at which high densities of virulent phage can kill off all of the bacteria, in this case due to fast, high rises in resource concentrations. The reason for this is that high resource environments allow bacteria to increase density to a high level. This causes the phage population to increase to such a point that the rate of infection is greater than the growth rate of the bacteria. If the phage population is great enough so that this rate of infection is high for a long enough period of time, then bacteria never recovers and the bacteria are killed off. Performing a similar simulation with no lytic phage, we do not see the same behaviour from temperate phage alone. This does not happen in temperate types due to inhibition of superinfection and the number of lysogens that are produced under high...
Figure 4.16: Parameters at which lytic phage eliminates bacteria under fluctuating conditions. Observations of TUD indicate for how long the bacteria was able to survive. At the point of elimination, the simulations stop.

growth conditions. Therefore, temperate types are protected from circumstances when conditions are too good.

4.5 Conclusion

In this work we set out to use recent observations on factors influencing the lysis-lysogeny decision in lambda phage to gain further insight into the ecology of temperate phage. While there is no clear reason as to why it is necessary to be temperate, we do find different situations in which it has an advantage over lytic phage and also why a temperate phage with responsive rate of lysogeny may have an advantage over one with a constant rate of lysogeny. The inferences were made using a combination of bifurcation analysis and simulations with fluctuating conditions. The bifurcation analysis followed the work of Stewart and Levin (1984) and found no significant advantage in being tem-
perate at stable resource conditions. However, this did provide insight into parameter values necessary for coexistence of temperate phage alongside lytic phage and the stability of different steady states. We also see that being responsive allows the phage to potentially raise their rate of lysogeny without being at a disadvantage in competing with lytic phage. Analysis of the system under fluctuating conditions revealed an advantage of being temperate when resource levels drop to a low level for a sufficient amount of time. This reinforces the thoughts of previous studies (Stewart and Levin, 1984; Mittler, 1996). It can also be seen that there is an advantage when the resource level rises very quickly, leading to a high density of lytic phage that destroys all the bacteria. This does not happen in the temperate case and implies that this is a means of coexistence in the absence of a refuge (Heilmann et al., 2010). However, a study regarding the evolution of super-infecting temperate phage (Berngruber et al., 2010, 2013) suggest that this advantage could lost over evolutionary timescales. The advantage of being a responsive temperate phage is less obvious. It appears that this is of advantage when the oscillations in resource level have long fluctuation times.

While we have not checked our fluctuating conditions observations for all parameter sets, the bifurcation analysis gives us some indication of which sets are interesting. The results therefore have some generality. We cannot guarantee that the system was at a true steady state under the oscillating conditions. The equations were solved until the oscillations in all of the players appeared to have converged to a limit cycle. We therefore have some confidence that these are steady states. The nature of the simulations means that it is difficult to perform a bifurcation analysis under fluctuating conditions. Neither would this take into account times when players oscillated to very low levels and would die out in silico. It could also be argued that eliminating phage when they have oscillated to a low level is crude, this methodology has been used before (Fussmann et al., 2000). It is suggested in Kussell and Leibler (2005) that bet-hedging approaches in persister cells are beneficial over responsive types when the environment fluctuates rapidly; this gives the species of interest a higher long-term growth rate than a responsive strategy. However, it may not be of interest to analyse the long term growth rate in our case as at some point we would expect the bacteria to evolve a resistant mutant and thus change the system. As stated in Section 4.4.1, we are only interested in this short period of time. We therefore concentrate on the short term dynamics of the system. It is also noted that the functional form of rate of lysogeny in lambda phage was found from experiments under one set of conditions. It is possible that under limiting resource conditions for example, that the function somehow changes or is biased towards either lysis or lysogeny. As such, a
conclusive study of the behaviour cannot be carried out theoretically until more is known about how this can change under different conditions. As we discuss in Chapter 6, growth rate can affect cell volume. Since volume affects the rate of lysogeny, then accounting for this may also reveal benefits for being temperate with a responsive rate of lysogeny. We are also careful to note that while we demonstrate conditions under which temperate phage and in particular responsive temperate phage are advantageous over lytic phage that our observations are not sufficient to suggest that they are the only reasons why they have evolved in such a way. Other possible reasons for being temperate are discussed in Stewart and Levin (1984). In contrast to the game theoretic approach (Avlund et al., 2009), our work considers that a single phage infecting can give a stochastic decision. This complicates the argument that a single infecting phage give a deterministic outcome and more than one phage gives a stochastic outcome. However, it can be seen for the responsive rate $f_r$ (Zeng et al., 2010) that the range of viral concentration over which the decision is stochastic, is larger for higher MOI.

To continue this work it would be useful to carry out a more thorough analysis of the case with fluctuating conditions. For example, by looking at the results over a wider range of parameters. In particular, there may be a cost to being responsive (Kussell and Leibler, 2005). The effects of this cost, when competing responsive phage in competition with temperate phage with a constant rate of lysogeny could be useful to investigate. However, as we show in Chapter 5, responsiveness in phage decision making may be a consequence of stochastic gene expression and it may not introduce additional cost. Moreover, there may be effects of environmental stress, such as under low resource conditions, on rate of induction due to effects on host fitness (Riefardt and Rainey, 2010). Experiments using lambda phage could also be used to replicate this mathematical analysis in both the steady conditions using a chemostat (Berngruber et al., 2013) and also the fluctuating conditions. This could possibly give data that would lead to improvements that need to be made to the current model. For example under rapidly fluctuating conditions, accounting for a delay from infection to lysis-lysogeny decision may be necessary in the model to more accurately reflect the dynamics. Recent work has shown that there exist different phenotypes of temperate phages that can displace one another when in the latent form (Riefardt, 2011). While it is not known whether these types are found in the same places in nature, exploration of this may also lead to interesting questions. For example, does a responsive temperate phage allow one to reduce its rate of lysogeny when there is a competitor that can readily displace it? It may also be the case that temperate phages are important in environments with spatial structures. It would therefore be interesting to
perform analysis that doesn’t assume spatial homogeneity (Heilmann et al., 2010, 2012). This, coupled with more information on the ecology of lambda phage would give more depth to this analysis.
Chapter 5

Stochastic Cellular Fate Decision Making by Individual Lambda Bacteriophage

5.1 Overview

Lambda phage is a virus that infects *Escherichia coli*. After infection, the phage chooses either to kill the cell by replicating actively (lysis) or laying dormant and integrating itself in the bacterial genome (lysogeny). The lysis-lysogeny decision circuit is a paradigm for cellular decision making. Previous work has shown that the number of infecting phage and the cell volume have a large influence on the decision outcome. Combining these factors, the viral concentration was largely believed to govern the observed proportion of cells choosing lysogeny. However, a recent study showed that the rate of lysogeny in a small cell with one infecting phage can be nearly double that observed in a cell of twice the volume with two phage. Here, we investigate the mechanistic basis of this observation using a simple stochastic model of the underlying genetic switch. Our results suggest that higher gene expression noise apparent with less infecting phage is an important factor in explaining the observations. There may also be small contributions from cell growth effects due to number of infecting phage and spatial segregation of phages in the cell. Ido Golding and Lanying Zeng kindly shared the original experimental movies from their study, which enabled the research into cell growth effects. The results are general and could be relevant in understanding of gene copy number effects in other biological systems.
5.2 Introduction

As described in Chapters 1 and 4, Bacteriophage Lambda is a temperate virus that infects *Escherichia coli* bacteria. Infection of the cell leads to one of two fates. The phage can replicate quickly and kill the cell, this is called the lytic fate. Alternatively the phage can lay dormant and replicate more slowly with cell division, this is the lysogenic fate. The lysogenic state is stable but can enter the lytic cycle at a later time by chance or due to ultra violet stress (Aurell et al., 2002; Morelli et al., 2009; Zong et al., 2010). Its genetic circuit is well studied (Arkin et al., 1998; Shea and Ackers, 1985; Ptashne, 2004) and serves as a paradigm for biological switches under gene regulation (Ptashne, 2004; Little et al., 1999; Dodd et al., 2005; Oppenheim et al., 2005). It is thought that the decision is a response to prevent extinction by lying dormant in malnourished cells or when there is an overabundance of phage (Stewart and Levin, 1984). While the understanding of the decision and stability is not complete, modelling of the system has provided much mechanistic insight (Ackers et al., 1982; Shea and Ackers, 1985; Arkin et al., 1998; Aurell et al., 2002; Dodd et al., 2005; Oppenheim et al., 2005; Weitz et al., 2008; Avlund et al., 2010; Joh and Weitz, 2011; Golding, 2011).

It is evident from a wide range of systems that biochemical networks are noisy and that this noise can influence decision making (Eldar and Elowitz, 2010; Balázsi et al., 2011). This can be due to extracellular fluctuations or low copy numbers (Shahrezaei et al., 2008). This has been utilised previously in studies of lambda phage, and one of the early studies on stochastic processes in genetic networks showed how this can influence the fate of a clonal population (Arkin et al., 1998). It is thought that this probabilistic cell fate decision can be an advantage in preventing phage from dying out (Avlund et al., 2009).

The decision is known to be noisy (Avlund et al., 2009) and the understanding of the decision is not complete. However, much of the decision can now be accounted for. Classic experiments have revealed (Kourilsky, 1973) that there is some effect of the physiological state of the cell and multiplicity of infection (MOI) on the decision. It is believed that increasing MOI increases the probability that the cell undergoes lysogeny. While it has also been observed that cell volume (V) has a large effect (St-Pierre and Endy, 2008). It is thought that decreasing V leads to an increase in the probability of choosing lysogeny. The effect of the combination of these factors, the viral concentration (VC, ratio of MOI to V) has been explored further in recent years (Weitz et al., 2008) using deterministic models. Most recently experiments by Zeng et al. (2010) showed that the decision depends not only on the VC but also directly on MOI. This suggests
that there may be a gene compensated dose mechanism (Joh and Weitz, 2011) in effect. Furthermore, this implies that the lysogeny decision requires all of the infecting phage to choose lysogeny. This individual decision making has yet to be fully characterised mechanistically.

In this study, we use the framework of an existing model of the genetic switch (Weitz et al., 2008) to investigate different mechanisms which could contribute to the observed gene compensated dosage. In particular, we observe the role of intrinsic noise, extrinsic noise, spatial and cell cycle effects on lytic-lysogenic decision making in lambda phage. To do this we implement the simple model using the Gillespie algorithm (Gillespie, 1977). We find that intrinsic noise is and its interplay with a threshold protein concentration is largely responsible for this discrepancy between observed rates at the same VC and different MOI, while there may also be contributions from spatial and cell cycle effects. We also find that the choice of a threshold level is critical in determining the outcome.

5.3 Model

To determine the role of MOI on decision making at the same VC we use a simple model of the lambda phage genetic circuit (Weitz et al., 2008). The genetic regulatory network contains interlocking positive and negative feedback loops. It consists of three viral genes that are important in the early part of the decision making process. These are CI, Cro and CII. It is described in Arkin et al. (1998) that the decision consists of multiple steps, but it is believed that the CII protein has a pivotal role (Oppenheim et al., 2005; Weitz et al., 2008). High levels of CII promote CI production and lysogeny, while low levels promote production of Cro and lysis. The model is outlined in Figure 5.1 and the full model is shown below

\[
CI + CI \xrightleftharpoons[\kappa_{CI}]{\kappa_{CI}^{+}} CI_{2} \\
Cro + Cro \xrightleftharpoons[\kappa_{Cro}]{\kappa_{Cro}^{+}} Cro_{2} \\
CII + CII \xrightleftharpoons[\kappa_{CII}]{\kappa_{CII}^{+}} CII_{2} \\
CI \xrightarrow[\gamma_{CI}]{\gamma_{CI}} \emptyset \\
Cro \xrightarrow[\gamma_{Cro}]{\gamma_{Cro}} \emptyset
\]
\[ CII \xrightarrow{\gamma_{CI}} \emptyset \]

\[ P_{RM/R,0} + CI_2 \xleftrightarrow[k_CI^+]{k_CI^-} P_{RM/R,CI} \]

\[ P_{RM/R,0} + Cro_2 \xleftrightarrow[k_{Cro}^+]{k_{Cro}^-} P_{RM/R,Cro} \]

\[ P_{RE,0} + CII_2 \xleftrightarrow[k_{CII}^+]{k_{CII}^-} P_{RE,CII} \]

\[ P_{RM/R,CI} \xrightarrow{\beta_{CI}} P_{RM/R,CI} + m_{CI} \]

\[ P_{RE,CII} \xrightarrow{\delta_{CI}} P_{RE,CII} + m_{CI} \]

\[ P_{RM/R,0} \xrightarrow{\alpha_{Cro}} P_{RM/R,0} + m_{Cro} \]

\[ P_{RM/R,0} \xrightarrow{\alpha_{CII}} P_{RM/R,0} + m_{CII} \]

\[ m_{CI} \xrightarrow{\gamma_{m}} \emptyset \]

\[ m_{Cro} \xrightarrow{\gamma_{m}} \emptyset \]

\[ m_{CII} \xrightarrow{\gamma_{m}} \emptyset \]

\[ m_{CI} \xrightarrow{\sigma} m_{CI} + CI \]

\[ m_{Cro} \xrightarrow{\sigma} m_{Cro} + Cro \]

\[ m_{CII} \xrightarrow{\sigma} m_{CII} + CII \]

where \( CI_2, Cro_2 \) and \( CII_2 \) represent the dimers of \( CI, Cro \) and \( CII \) respectively, while \( m_{CI}, m_{Cro} \) and \( m_{CII} \) represent their corresponding mRNA. The unbound promoters are represented by \( P_{RM/R,0} \) and \( P_{RE,0} \), and \( P_{RM/R,CI}, P_{RM/R,Cro} \) and \( P_{RE,CII} \) denote the bound configurations of the promoters. The parameters used are shown in Table 5.1 and are the same as used in Weitz et al. (2008). Dimerisation rates are given by \( \kappa_{CI}^+, \kappa_{Cro}^+ \) and \( \kappa_{CII}^+ \), while the corresponding separation rates are given by \( \kappa_{CI}^-, \kappa_{Cro}^- \) and \( \kappa_{CII}^- \). Transcription rate of \( CI \) from \( P_{RM/R} \) is given by \( \beta_{CI} \) and from \( P_{RE} \) it is given by \( \delta_{CI} \). Transcription rates of \( Cro \) and \( CII \) are given by \( \alpha_{Cro} \) and \( \alpha_{CII} \) respectively. The decay rates of the proteins are given by \( \gamma_{CI}, \gamma_{Cro} \) and \( \gamma_{CII} \) respectively. Translation rates of all proteins are given by \( \sigma \). The binding rates of dimers to promoters are given by \( k_{CI}^+, k_{Cro}^+ \).
Figure 5.1: Lysis-lysogeny Genetic Switch. A) Outline of the role of the $P_{RM}/P_R$ and $P_{RE}$ promoters involved in decision making in lambda phage. Dashed lines denote transcriptional events that require no activation while solid lines denote transcriptional events that require activation. B) Schematic of the simplified version genetic network involved in the lysis-lysogeny decision. CI gene promotes itself and represses the other genes. Cro represses everything, while CII promotes CI.

The model was originally implemented deterministically, but here we simulate these equations stochastically using Facile (Siso-Nadal et al., 2007) and Easystoch (Shahrezaei et al., 2008). This uses the Gibson-Bruck (Gibson and Bruck, 2000) version of the Gillespie algorithm. We also carry out spatial modelling using Smoldyn (Andrews and Bray, 2004). Diffusion constants were found using Bionumbers (Milo et al., 2010) and are shown in Table 5.2. For both spatial and non-spatial simulations, data points were output at every 0.1 minutes. The analysis of the generated data was carried out in Matlab (version R2012a).
### Table 5.1: Model Parameters for Stochastic Simulation of Lambda Phage Genetic Switch

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_{CI}$</td>
<td>Backward dimerisation rate of CI</td>
<td>$\kappa_{CI} = 0.1 MOL^2 min^{-1}$</td>
</tr>
<tr>
<td>$\kappa_{CI}^+$</td>
<td>Dimerisation rate of CI</td>
<td>$\kappa_{CI}^+ = 1 \times 10^6 MOL min^{-1}$</td>
</tr>
<tr>
<td>$\kappa_{Cro}$</td>
<td>Backward dimerisation rate of Cro</td>
<td>$\kappa_{Cro} = 0.1 MOL^2 min^{-1}$</td>
</tr>
<tr>
<td>$\kappa_{Cro}^+$</td>
<td>Dimerisation rate of Cro</td>
<td>$\kappa_{Cro}^+ = 1 \times 10^6 MOL min^{-1}$</td>
</tr>
<tr>
<td>$\kappa_{CII}$</td>
<td>Backward dimerisation rate of CII</td>
<td>$\kappa_{CII} = 0.1 MOL^2 min^{-1}$</td>
</tr>
<tr>
<td>$\kappa_{CII}^+$</td>
<td>Dimerisation rate of CII</td>
<td>$\kappa_{CII}^+ = 1 \times 10^6 MOL min^{-1}$</td>
</tr>
<tr>
<td>$\gamma_{CI}$</td>
<td>Decay Rate of CI</td>
<td>$\gamma_{CI} = 0.04 min^{-1}$</td>
</tr>
<tr>
<td>$\gamma_{Cro}$</td>
<td>Decay Rate of Cro</td>
<td>$\gamma_{Cro} = 0.05 min^{-1}$</td>
</tr>
<tr>
<td>$\gamma_{CII}$</td>
<td>Decay Rate of CII</td>
<td>$\gamma_{CII} = 0.12 min^{-1}$</td>
</tr>
<tr>
<td>$k_{CI}^-$</td>
<td>Protein Unbinding Rate of CI</td>
<td>$k_{CI}^- = 0.1 MOL^2 min^{-1}$</td>
</tr>
<tr>
<td>$k_{CI}^+$</td>
<td>Protein Binding Rate of CI</td>
<td>$k_{CI}^+ = 1 \times 10^6 MOL min^{-1}$</td>
</tr>
<tr>
<td>$k_{Cro}^-$</td>
<td>Protein Unbinding Rate of Cro</td>
<td>$k_{Cro}^- = 0.1 MOL^2 min^{-1}$</td>
</tr>
<tr>
<td>$k_{Cro}^+$</td>
<td>Protein Binding Rate of Cro</td>
<td>$k_{Cro}^+ = 1 \times 10^6 MOL min^{-1}$</td>
</tr>
<tr>
<td>$k_{CII}^-$</td>
<td>Protein Unbinding Rate of CII</td>
<td>$k_{CII}^- = 0.1 MOL^2 min^{-1}$</td>
</tr>
<tr>
<td>$k_{CII}^+$</td>
<td>Protein Binding Rate of CII</td>
<td>$k_{CII}^+ = 1 \times 10^6 MOL min^{-1}$</td>
</tr>
<tr>
<td>$\beta_{CI}$</td>
<td>Transcription rate of mRNA for CI from promoter 1</td>
<td>$\beta_{CI} = 1.6 min^{-1}$</td>
</tr>
<tr>
<td>$\delta_{CI}$</td>
<td>Transcription rate of mRNA for CI from promoter 2</td>
<td>$\delta_{CI} = 1.2 min^{-1}$</td>
</tr>
<tr>
<td>$\alpha_{Cro}$</td>
<td>Transcription rate of mRNA for Cro</td>
<td>$\alpha_{Cro} = 0.8 min^{-1}$</td>
</tr>
<tr>
<td>$\alpha_{CII}$</td>
<td>Transcription rate of mRNA for CII</td>
<td>$\alpha_{CII} = 0.8 min^{-1}$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Translation rate</td>
<td>$\sigma = 0.5 min^{-1}$</td>
</tr>
</tbody>
</table>

### Table 5.2: Diffusion Model Parameters for Stochastic Simulation of Lambda Phage Genetic Switch

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time step</td>
<td>$1 \times 10^{-5}$</td>
</tr>
<tr>
<td>Protein Diffusion Coefficient</td>
<td>$6 \times 10^6 nm^2 min^{-1}$</td>
</tr>
<tr>
<td>mRNA Diffusion Coefficient</td>
<td>$3 \times 10^6 nm^2 min^{-1}$</td>
</tr>
<tr>
<td>Binding Radius</td>
<td>$\sim 1 nm$</td>
</tr>
<tr>
<td>Cell Width</td>
<td>250 nm</td>
</tr>
</tbody>
</table>
5.4 Results

5.4.1 Motivation

The observations in Zeng et al. (2010) suggest that the probability of lysogeny is well described by

\[ f(M, h, J, V)_{\text{Cell}} = \left( \frac{(M/V)^h}{J^h + (M/V)^h} \right)^M \]

where \( M \) is the MOI, \( V \) is the cell volume, \( J = 1.17 \) is the half saturation constant and \( h = 2.07 \) is the Hill number. We plot this function as a function of VC for different MOI in Figure 5.2. It demonstrates that for a given VC, the probability of lysogeny can be quite different depending on the MOI. In the most striking case, a small cell with MOI = 1 can have a much larger probability of lysogeny than a cell of twice the volume and MOI = 2 despite both cases having the same VC. While it can be seen in Figure 5.2B that in some cases there can be little effect for different MOI at constant cell volumes. In this section we focus on this case to investigate mechanisms that can explain this dependency of probability of lysogeny on MOI.
To understand the effect that different factors have on the probability of lysogeny, we use stochastic simulations (Gillespie, 1977). They are used to simulate realisations of the viral gene expression soon after infection by the phage. To quantify the cell fate decision we need to choose a criterion to classify whether an individual simulation undergoes lysis or lysogeny. Work by Arkin et al. (1998) outlines the steps involved in this complex decision that could depend on transient or steady state dynamics of multiple proteins. Here we assume for simplicity that the decision is mediated by the transient dynamics of the CII gene (with concentration $[CII]$). The decision process is explained in Figure 5.3. The time averaged levels of CII concentration up to a point $\tau$ ($\langle [CII] \rangle_\tau$) alone determine the outcome. If this level is above a threshold $\Phi$ then the fate is lysogeny, or if it is under
then the fate is lytic. The simulations gave output every 0.1 minutes and mean protein numbers were calculated over all output times between $t = 0$ and $t = 60$ minutes. The concentrations were calculated by dividing the protein numbers by the standardised cell volumes where $V = 1$ corresponds with a volume of $1 \times 10^{-15} L$.

Figure 5.3: Decision criteria for lysis-lysogeny decision. A) Illustration of a stochastic trajectory of CII proteins from time of infection to decision point $\tau$ (yellow dashed line). We consider the decision criteria as to whether the area under the curve of CII (blue area) is above $\Phi \tau$, or equivalently average $\Delta_\tau$ is above $\Phi$ (red dashed line) over a particular length of time $\tau$. B) Flow diagram outlining the decision criteria.
5.4 Results

5.4.2 Threshold and Noise

In order to find reasons for the difference in the rate of lysogeny for different MOI at the same VC, we focus initially on the difference between cells with MOI = 1 and MOI = 2 at VC = 1. We find similar mean [CII] for both MOI = 1, V = 1 (⟨[CII]⟩_{60} = 18.61, standard error se = 0.0614) and MOI = 2, V = 2 (⟨[CII]⟩_{60} = 18.85, se = 0.0441). Traces of the average [CII] and a single stochastic simulation are shown in Figure 5.4. While it is clear that the mean concentration is similar in both cases, it is also evident that the variation about the mean is larger in the case where MOI = 1 compared with the case where MOI = 2. The reason for this is that there is more intrinsic noise in the case with MOI = 1 due to lower protein copy number.

Figure 5.4: Average stochastic trajectories of [CII] over time interval τ for stochastic non-spatial model. Mean trajectories shown (blue line) ± standard deviations (cream shaded region), alongside the deterministic trajectory (turquoise line). One stochastic trajectory from the data is also shown in green. A) MOI = 1, V = 1 (⟨[CII]⟩_{60} = 18.61, se = 0.0614). B) MOI = 2, V = 2 (⟨[CII]⟩_{60} = 18.85, se = 0.0441). Results calculated based on n = 5000 simulations.
To determine the rate of lysogeny, we need to choose an appropriate threshold level $\Phi$ and time point $\tau$. For all of the analysis here we keep $\tau = 60$ minutes, which is reasonable from our experience with the experimental data. On analysing initial results, we find that the choice of $\Phi$ is crucial in determining the outcome. In particular, we find that for $\Phi > \langle [CII] \rangle_{60}$ the rate of lysogeny is higher for MOI = 1 than for MOI = 2. In contrast, for $\Phi < \langle [CII] \rangle_{60}$ the rate of lysogeny is higher for MOI = 2 than for MOI = 1, while the rates are similar for $\Phi \simeq \langle [CII] \rangle_{60}$. This is shown in Figure 5.5. We believe that these observations reflect noise in protein concentrations in the different cases. This is reflected in the respective standard errors of $\langle [CII] \rangle_{60}$.

![Figure 5.5](image.png)

Figure 5.5: How the relationship between noise and threshold level affect the decision outcome. Rate of lysogeny across different threshold values for basic model. Blue line: MOI = 1, V = 1; green line: MOI = 2, V = 2; orange dashed line: $\langle [CII] \rangle_{60}$; pink and yellow dashed lines: Theoretical approximations for the rate for MOI = 1, V = 1 and MOI = 2, V = 2 respectively.

While we have chosen the method outlined in Figure 5.3B to determine the decision
outcome of the simulations, we could have chosen a simpler decision for similar qualitative results. For example, we could have simply evaluated the concentration of CII at \( t = \tau \). For mathematical simplicity we use this criteria to provide insight into the observations. If we assume a Gaussian distribution for \([CII]_\tau\), with mean \( \mu \) and standard deviation \( \sigma \) then

\[
P([CII]_\tau > \Phi) = \frac{1}{2} \left(1 - \text{erf} \left( \frac{\Phi - \mu}{\sqrt{2\sigma^2}} \right) \right).
\]

This inverse cumulative function is equivalent to the rate of lysogeny in this case. The error function changes sign with that of \( \Phi - [CII] \). This accounts for the change in observed dependence on \( \Phi \) in relation to the the mean. The size of this difference is then dictated by \( \sigma \) which is dependent on the copy number and hence MOI. Therefore, it can be seen for an increase in MOI that \( |\Phi - [CII]| \) remains constant, but the denominator decreases. This increases the magnitude of the difference in rate from \( \frac{1}{2} \).

In order to check that the experimental observations are directly dependent on noise, independent of MOI and volume, we adjust the parameters to make the system more or less noisy. We therefore look only at the case where MOI = 1, V = 1, but specifically increase (decrease) noise by increasing (decreasing) translation and mRNA degradation rates. This increases the burstiness of the gene expression (Shahrezaei and Swain, 2008a). The results are displayed in Figure 5.6A. It can again be seen that for higher levels of noise we see higher rates of lysogeny for \( \Phi > \langle [CII] \rangle_{60} \) and lower rates of lysogeny for \( \Phi < \langle [CII] \rangle_{60} \) in accordance with the above reasoning. Comparing results from our simulations to those observed in Zeng et al. (2010), it appears that intrinsic noise contributes about half of the observed dependence of rate of lysogeny on MOI.
Figure 5.6: How changing the level of noise further affects the decision outcome. A) How changing the level of noise for same MOI affects the outcome. B) How the MOI affects the rate of lysogeny. Results shown for a high threshold (green line), threshold at $\langle [CII] \rangle_{60}$ (yellow line) and low threshold (blue line). The phenomenological rate $f(M, h, J, V)$ is shown for a typical cell volume using cell length $l = 1.11$ (orange dashed line).

If this observed relationship is due to the interaction between $\Phi$ and the amount of noise, then it follows that this difference should be even stronger for higher MOI at the same viral concentration. We therefore perform additional simulations for MOI= 3, 4, 5 at VC= 1. We find that this indeed holds for higher MOI (Figure 5.6B). This reduction in noise at larger MOI is due to the increase in copy number for mRNA and proteins.
5.4 Results

5.4.3 Extrinsic Noise

From the above results we believe that noise is an important driving factor in the observed results. We therefore investigate whether extrinsic noise can enhance these effects further. We use the methods outlined in Shahrezaei et al. (2008) to model extrinsic noise by varying the transcription rates, as outlined in Chapter 2. Again, we compare the effects for a small cell with MOI = 1 and a larger cell with MOI = 2 and in this case we also compare the effect of anti-correlated extrinsic noise. In the first case we assume that extrinsic noise works equally on the two promoters, whereas in the anti-correlated case we assume that the effects on each promoter work against each other. The results are shown in Figure 5.7. It can be seen that while modelling extrinsic noise increases the total noise and thus the rate of lysogeny, the difference between the rates for MOI = 1 and MOI = 2 does not change. The anti-correlated case does not increase noise as much in comparison to the correlated case and does not contribute towards the observed dependence.

![Figure 5.7: The effect of extrinsic noise on decision outcome. Rate of lysogeny across different threshold values for basic model. Blue dashed line: MOI= 1, V= 1 with effects of intrinsic noise only ($\langle [CII] \rangle_{60} = 18.61$, se= 0.0614); yellow dashed line: MOI= 2, V= 2 with effects of intrinsic noise only ($\langle [CII] \rangle_{60} = 18.85$, se= 0.0441); green line: MOI= 1, V= 1 with extrinsic noise ($\langle [CII] \rangle_{60} = 18.65$, se= 0.0636); brown line: MOI= 2, V= 2 with extrinsic noise ($\langle [CII] \rangle_{60} = 18.90$, se= 0.0472); purple line: MOI= 2, V= 2 with anti-correlated extrinsic noise ($\langle [CII] \rangle_{60} = 18.89$, se= 0.0445).](image-url)
5.4.4 Spatial Effects

Phage can infect the bacteria at any position along the cell surface but recent observations suggest that they are more likely to infect at the cell pole (Edgar et al., 2008; Zeng et al., 2010). It is therefore possible that multiple phages infecting a cell are spatially separated. If this infection location is coupled with slow diffusion then this could lead to an effect on the biomolecule dynamics inside the cell. This in turn could affect the rate of lysogeny, particularly when there are multiple infecting phages. The effect of infection site on infection success was previously analysed in Zeng et al. (2010). However, we are not aware of any data relating infection position to the decision outcome. To investigate the effect of infection site on rate of lysogeny we use spatial simulations that track individual molecules as they diffuse and react (Andrews and Bray, 2004). We compare the case where there is 1 infecting phage positioned at the centre of the cell with a cell with 2 phage arranged in 3 different ways. In these cases the phage will be positioned either with both in the centre, equally spaced or at the cell poles (Figure 5.8). The results were also compared with the non-spatial case outlined in the previous section. Due to the computational power required to process these simulations we perform a significantly lower number in comparison to the non-spatial case ($n = 200$). However, the standard deviation of the observed rates are still relatively small.
Figure 5.8: The role of spatial effects on CII average. Average CII concentrations for different phage infection positions. A) MOI = 1, V = 1, phage at centre (\(\langle[CII]\rangle_{60} = 19.02, \text{se} = 0.316\)); B) MOI = 2, V = 2, phage at centre (\(\langle[CII]\rangle_{60} = 19.35, \text{se} = 0.222\)); C) MOI = 2, V = 2, phage equally spaced (\(\langle[CII]\rangle_{60} = 18.91, \text{se} = 0.220\)); D) MOI = 2, V = 2, phage at cell poles (\(\langle[CII]\rangle_{60} = 18.72, \text{se} = 0.202\)). Average [CII] with for non-spatial model (blue line) shown alongside spatial model (yellow lines). Results based on \(n = 200\) simulations.

It can be seen in Figure 5.8 that the effect of phage positioning and diffusion rate on the mean [CII] is negligible. These results are also comparable with the non-spatial and deterministic models. However, it can be seen in Figure 5.9 that there are some effects on rate of lysogeny due to phage positioning. Specifically the rate is lower (higher) for the case when phages are at the cell poles for \(\Phi > \langle[CII]\rangle_{60}(\Phi < \langle[CII]\rangle_{60})\). While the rate is highest (lowest) for cells with phages at the centre when \(\Phi > \langle[CII]\rangle_{60}(\Phi < \langle[CII]\rangle_{60})\). The case where phages are equally spaced is in general an intermediate of the other results and follows the non-spatial results the closest. This difference could be due to the small differences in mean [CII]. To emphasise any spatial effects we also performed simulations with slower diffusion rates for mRNA and proteins. The effects
of this on rate of lysogeny are shown in Figure 5.10. It can be seen that again there is some difference between the rates for different phage positioning, but there is not a clear significant difference across all thresholds and therefore no clear evidence that spatial effects are the main driving force behind the observed results.

Figure 5.9: The role of spatial effects on rate of lysogeny. Rate of lysogeny across different threshold values for spatial model with MOI = 1, \( V = 1 \) (blue line) and MOI = 2, \( V = 2 \) with phage at centre (yellow line), phage equally spaced (green line) and phage at cell poles (red line). Results are shown alongside the non-spatial results (dashed lines).
5.4 Results

Figure 5.10: The role of spatial effects on rate of lysogeny for slower diffusion rate. Rate of lysogeny across different threshold values for spatial model with slower diffusion rates for MOI= 1, V= 1 (blue line) and MOI= 2, V= 2 with phage at centre (yellow line), phage equally spaced (green line) and phage at cell poles (red line). Results are shown alongside the non-spatial results (dashed lines).

In comparison with the non-spatial results, the effects of phage positioning are less crucial than intrinsic noise. In many cases the spatial results are not significantly different from their non-spatial counterparts. It is also noted that taking an average of the three cases when MOI= 2 gives rates of lysogeny that are very close to the non-spatial case. Therefore, this is only likely to have a substantial effect if there were any bias in the positioning of the phage. While previous results (Edgar et al., 2008; Zeng et al., 2010) display some infection site bias to the cell pole, it is unlikely to be enough to explain the observed difference in rate of lysogeny.

5.4.5 Cell Growth Effects

It is possible that the observed effect of MOI on the rate of lysogeny is due to the way the number of phage affect the physiology of the cell. We have performed more analysis of the movie data from Zeng et al. (2010) courtesy of Ido Golding and Lanying Zeng. Time lapse cell imaging analysis, and subsequent statistical analysis was performed in Schintzcell (courtesy of Michael Elowitz, Rosenfeld et al. (2005)) and Matlab respec-
Chapter 5. Stochastic Cellular Fate Decision Making by Individual Lambda Bacteriophage

For details of the time lapse microscopy and other experimental details, see Zeng et al. (2010). From these movies we were able to investigate any potential influence of MOI on cell cycle effects. Specifically, we looked at the effect of MOI on elongation rate of the cell. The methods used in this calculation are explained in Figure 5.11A. Cells were observed from time of infection until a decision event or first cell division. It was observed that cells with higher MOI have lower growth rate (Figure 5.11B). This is logical since a cell supporting increased levels of phage replication is likely to have less resource available for cell growth. We did not take into account non-growing cells in the analysis while the higher proportion of such cells observed with higher MOI is consistent with Zeng et al. (2010).

Figure 5.11: The role of cell cycle effects - experiments. A) Method for determining cell growth rate. Measurements are taken at 2 time points, the first frame and final frame. The final frame is determined by the event. For an uninfected cell (grey) this is the point at which the cell divides. For an infected cell it is the point at which a decision of lysis (green) or lysogeny (red) has been made. The rate was calculated using Growth Rate = \( \frac{\log(L_2) - \log(L_1)}{T} \) where \( T \) is the time between the first frame and the final frame. B) Observed growth rates at different MOI.

This information on growth dependency and MOI was then incorporated into our model. To take the effect of growth rate into account we adjust the dilution rates (Eden et al., 2011) to mimic cells growing at different rates. The results are compared with those of the non-growing cells and also growing cells with no effect of MOI on growth rate. It can be seen in Figure 5.12A that when accounting for the growth effects observed
experimentally, the results actually go against the dependence on growth rate. This is likely due to the increased relative volume in accordance with the dependence of rate of lysogeny on volume (St-Pierre and Endy, 2008).

Figure 5.12: The role of cell cycle effects - modelling results. A) Effect of cell growth on rate of lysogeny by considering dilution only. Fast rate $= 0.007 \text{min}^{-1}$ and slow rate $= 0.0055 \text{min}^{-1}$. B) Effect of growth rate on rate of lysogeny when also considering possible changes in transcription rates.

Studies have shown that growth rate has an effect on protein synthesis (Klumpp et al., 2009) and this is largely due to an effect on the rate of mRNA transcription (Liang et al., 1999; Marguerat and Bähler, 2012). We therefore looked at the effect of changing the rates of mRNA transcription, dependent on the growth rates, on the decision outcome. It can be seen in Figure 5.12B that this increases the average protein concentration and therefore affects the rate of lysogeny. If the growth rate for cells with MOI= 1 is higher than those of MOI= 2, then $\langle [CII] \rangle$ would be larger in the case where MOI= 1 due to the increased mRNA transcription. This increases the rate of lysogeny. This difference in rates concurs with the experimental results. Therefore, if the transcription rate increases sufficiently at higher growth rates then this could also be contributing factor. However, we don’t believe that the effect would be large enough to have a major effect on rates at different viral concentrations.
5.4.6 Model Robustness

While we believe that our model captures the behaviour sufficiently to demonstrate that noise and other factors explored affect the rate of lysogeny, due to the simplicity of the model we perform some additional analyses to support this. In particular, in the genetic circuit CII is tetrameric (Parua et al., 2010). We therefore look at the effect of allowing dimers to bind and form tetramers, which can then bind to the promoter. The effect on rate of lysogeny when adding this mechanism into the model is shown in Figure 5.13A. It can be seen that this has little effect on our results. Secondly, upon infection phage promoters double in the cell every 2-3 minutes for the first 15 minutes (Better and Freifelder, 1983). This has not been included in our model. We therefore investigate the effect of phage replication on our results. This has the effect of increasing the mean \([CII]\). While increasing the concentration reduces the protein noise, we still observe similar qualitative results. This is demonstrated in Figure 5.13B. Finally, we performed a sensitivity analysis on the system. All parameters were varied randomly within magnitude of half, to double the original values. This was performed for 100 parameter sets, and 500 replications were carried out for each set. The threshold for each parameter set was taken to be 20% above the mean \([CII]\) value for those parameters. The results are shown in Figure 5.13C. It can be seen that our results are consistent across all parameter sets tested.
5.5 Discussion

It was observed in a study by (Zeng et al., 2010) that the rate of lysogeny in lambda phage depends not only on $VC$, but also directly on MOI. To try and explain this dependence we have used a stochastic model of the genetic regulatory network to investigate the effect of possible contributing factors. We have observed that intrinsic noise and its interaction with a threshold is a determining factor in the observed dependence of rate of lysogeny in lambda phage. This is due to the difference in copy number at different MOI where for fewer phage, the protein levels are more likely to fluctuate above a threshold level due
to noise. This was found by analysing the dynamics in protein concentrations over time under different MOI and varying appropriate threshold parameters. While we observed some spatial effects, we do not believe that these are sufficient alone to drive the observed differences between different MOI at the same $VC$. Cell image analysis has allowed us to infer some dependance on growth rate of the cell prior to the decision taking place. Specifically, we observed that increases in MOI reduce the growth rate of the cell between the time of infection and a decision event. It is possible that if this growth rate difference affects the transcription rates adequately that this would also contribute to the observed dependence. This work suggests that while there are determining factors in the decision, how these factors affect the cell dynamics can also affect the decision. The combination of all of the factors could contribute to the observations of individual decision making in lambda phage.

The model we use is simple, but not too simple (Avlund et al., 2010) and we believe that our claims would hold using more detailed models. We have assumed that the decision is dependent or at least indicated by the transient dynamics in the cell and an appropriate threshold. It could be argued that our choice of threshold is crude in terms of a decision criteria. However, our results agree qualitatively with work that use other criteria (Joh and Weitz, 2011). Furthermore, investigations have shown that using other criteria and different time periods for the decision or different proteins as indicators that the qualitative results hold. Moreover, we have only considered the parameters given in Weitz et al. (2008) for much of the work. However, we do not believe that the inferences that we have made are parameter dependent. This is because the differences in decision rate occur despite having the same deterministic steady state and the main differences we observe are due to copy number. This inference is therefore general and likely to hold in other models. We have addressed this briefly using global sensitivity analysis. While the mean protein concentration is largely controlled by $VC$, the amount of noise is largely controlled by the MOI. Such effects on cell fate decision making have been previously explored in other systems (Süel et al., 2007; Chang et al., 2008; Mehta et al., 2008; Kalmar et al., 2009; Singh and Weinberger, 2009; Johnston et al., 2012).

We investigated the effect of spatial separation of the infecting phage on the decision making process using a particle based Monte Carlo method. We observed some effect, but this was small in comparison to the effect of intrinsic noise. It is also possible that the spatial effects can also have an effect on the total level of noise which could further influence the decision (Cottrell et al., 2012). While the diffusion coefficients used (Table 5.2) allowed the protein to travel the length of the cell in seconds and mRNA in minutes,
slowing both down did not increase these spatial effects to the level of intrinsic noise. If there were spatial barriers in the cell, then this could further increase any small spatial effects on the probability of lysogeny.

The original movie data used in Zeng et al. (2010) was reanalysed using cellular image processing and this analysis revealed that there is an effect of MOI on cell elongation rate. In particular, we observed that cells with more infecting phage have a lower elongation rate. This is likely due to the extra resource needed to support multiple phage. If this affects transcription rates sufficiently then we predict from the simulations that this could also be a contributing factor to the observed dependence of decision outcome on MOI.

Of the other factors we investigated, while extrinsic noise affected the total level of protein noise it did not appear to have any effect on the difference between cases at MOI= 1 and MOI= 2. In the experiments reported in Zeng et al. (2010) the phages were synchronised to infect the cells at the same time by a change in temperature (see Zeng et al. (2010) for details), we investigated the effect of a slight delay in phages infecting. Using stochastic simulations with one phage infecting slightly after the other in the case with 2 infecting phages it was observed that this had little effect on the rate of lysogeny. Finally, it could be argued that the difference in observed rate in the experiments was due to the length not being a true measure of the volume when comparing cells. This is due to the curvature of E. coli, and this will make a cell of double length not quite double the volume of the smaller cell. However, simulations revealed that while taking this into account has some effect on the mean, noise is still the dominant factor in the decision outcome.

Partial gene dosage compensation mechanisms have been observed in many different systems and networks (Gilfillan et al., 2006; Gallegos Ruiz et al., 2008; Acar et al., 2010). However, a complete general mechanism for this phenomenon has not been established. Previous studies have suggested that volume changes (Galitski et al., 1999), nonlinear interactions (Veitia et al., 2013) and network structure (Acar et al., 2010) may be important in contributing to this compensating effect. It was suggested by Joh and Weitz (2011) that gene dosage compensation can explain the observed dependence of probability of lysogeny on MOI. The factors that influenced the decision in our study may also contribute to such a general mechanism. Intrinsic noise depends on gene dosage and is present in all cellular systems and thus has the potential to play a part in all decision making. The structure of the genetic network has a large effect on the how much noise affects the decision outcome (Becskei and Serrano, 2000). We have observed how the
Chapter 5. Stochastic Cellular Fate Decision Making by Individual Lambda Bacteriophage

Cell elongation rate can change with gene dosage, in this case the MOI. The effect of gene dosage on cell size has also been observed in other systems (Galitski et al., 1999). While we did not find any role of spatial effects in our system, there is potential that it could have some effect in other systems where either diffusion is slow or there is some form of spatial partitioning in the cell. It is possible that all of these factors can contribute in combination towards a partial gene dosage compensation mechanism.
Chapter 6

The Effects of Cell Partitioning on Noise in Genetic Regulatory Networks

6.1 Overview

Recent work by Huh and Paulsson (2011) has shown that the effects of noise from cell partitioning while often overlooked, are important and can be difficult to separate from intrinsic noise. This work implies that for a constitutively expressed protein, noise increases when the doubling time is decreased due to the higher frequency of imperfect partitioning at cell division. It has also been observed that growth rates control RNA polymerase and ribosome numbers, which can lead to parameter changes that affect protein copy number (Klumpp et al., 2009). Furthermore, growth rate can also affect cell volume (Klumpp et al., 2009). Do the parameter and volume changes compensate for the increased noise at faster cell division? Here we use simulations and the linear noise approximation to show how the effects of cell partitioning on protein noise, coupled with effects of growth rate on global cell parameters affect the total noise in simple gene regulatory networks. The work was carried out with Vahid Shahrezaei and is in preparation for submission.

6.2 Introduction

As described in Chapters 1 and 5, noise is important in cellular systems and can be due to bursty processes, low copy numbers of biomolecules and extracellular fluctuations. It is known that noise can have a significant effect in decision making (Kalmar et al., 2009; Balázsi et al., 2011; Golding, 2011), phenotypic switching (Blake et al., 2006) and is important for signalling accuracy (Andrews et al., 2006). Due to the ways in which cells
Chapter 6. The Effects of Cell Partitioning on Noise in Genetic Regulatory Networks

replicate, additional noise is incurred on biomolecules upon division. During the cell cycle, biomolecules are created and destroyed and upon mitosis they undergo the nonlinear process of cell division, where their numbers are split between the daughter cells. The partitioning is random between the cells, and therefore the difference between proteins for example, split between two daughter cells can be important when copy numbers are low. Moreover, in *E. coli* doubling times can vary from around 20 minutes to a few hours, and so the frequency of partitioning error across a population of cells could vary significantly, depending on the growth media. When modelling genetic regulatory networks, it is often assumed that the effects of cell division are negligible such that they are ignored. More recently, it is shown that these effects are not negligible (Huh and Paulsson, 2011), but it is difficult to differentiate partitioning noise from intrinsic birth and death events.

It has also been demonstrated that growth rate can affect global cell parameters such as cell volume and transcription rates and thus have an effect on protein concentration (Liang et al., 1999; Zhurinsky et al., 2010). This also needs to be taken into account when modelling growing cells. Specifically, increasing growth rate in *E. coli* approximately maintains protein copy numbers but increases cell volume and decreases protein concentration for a constitutively expressed protein (Klumpp et al., 2009). Does this increase in copy number lead to a reduction in noise sufficient to mask an increase in noise from faster cell division? Are the cells able to maintain the level of noise when exposed to fast or slow growth conditions? Multistability of phenotypes can also be a result of network architecture (Gardner et al., 2000; Atkinson et al., 2003). It is therefore also interesting to investigate how this relationship between growth rate and noise affects the dynamics of genetic networks.

While recent advances in experimental and cell imaging techniques allow us to measure protein copy number in cells (Golding et al., 2005; Cookson et al., 2010), there is not a wide range of data to our knowledge that shows how the distribution of protein copy number and cell volume changes with growth rate across a range of single cell organisms. Some studies have looked briefly at noise effects in different growth media (Bar-Even et al., 2006; So et al., 2011) but not extensively enough to make any clear inference. Several studies suggest that cell volume increases with decreasing doubling time in prokaryotes and eukaryotes (Chien et al., 2012; Turner et al., 2012; Marguerat and Bähler, 2012). There are different explanations of why this occurs. In bacteria it is thought that this is due to multifork replication and higher DNA content at faster growth rates (Chien et al., 2012), while less is known about the reasons in yeast (Turner et al., 2012). For protein copy number numbers, experimental data show us that copy number
increases slightly with decreasing doubling time (Klumpp et al., 2009), but we have no data available to demonstrate how the variation in the copy number also changes. In certain cases it is difficult to establish the source of the noise. While the effects of cell division on cell dynamics have been modelled for gene regulatory networks (Chen et al., 2004), the effect on noise is less studied.

Antibiotic resistance is a growing problem in modern medicine and one way that this resistance is gained is attributed to persister cells (Balaban et al., 2004). One characteristic of these persister cells is their biphasic killing curves, where in a population of genetically identical cells many are killed quickly and the remainder are more difficult to kill. These different phenotypes are characterised by their growth rates (Klumpp et al., 2009; Rocco et al., 2013). This can be explained by the hipBA toxin system (Rotem et al., 2010). While toxins can kill cells, in low concentrations they can simply reduce the growth rate of the cell. This can be modelled as two different switching populations (Balaban et al., 2004), while mechanistic explanations for the bistability of the cell growth rates in identical cells are slow protein fluctuations and cell growth effects on transcription rates (Klumpp et al., 2009; Rocco et al., 2013).

In this study we use stochastic simulations to establish the effect of cell division on noise in simple genetic regulatory circuits. We use available data concerning how transcription rates, gene dosage and cell volume change at different growth rates to observe whether this offsets the additional noise at faster growth rates. We find for constitutively expressed proteins that the parameter changes ensure that the level of noise does not increase for faster cell division. For negative feedback we find that the global changes allow the cell to maintain the level of noise at different growth rates, while positive feedback can lead to increased noise at higher growth rates. For a protein complex formed from the binding of two constitutively expressed proteins, increasing cell division has the reverse effect and leads to increased noise. While we also find that feedback expressed through growth rate for a toxin, that noise from cell partition can contribute to complex effects on switching rates between two stable growth states.

6.3 Methods

As a starting point for the work we use the linear noise approximation (LNA) following Huh and Paulsson (2011) to analyse the effect of growth rate on protein noise under constitutive gene expression. The derivation of these equations is outlined in Chapter 2. To explore more complicated genetic networks we use computer simulations using the Gillespie algorithm (Gillespie, 1977) to investigate the growth effects under different network
motifs. To do this we use Facile (Siso-Nadal et al., 2007) and Easystoch (Shahrezaei et al., 2008) software tools and the subsequent analysis of the generated data is carried out in Matlab (version 2012a). The simulations are studied for 15,000 cell generations. The circuits for the positive and negative feedback models are shown in Figure 6.1.

Figure 6.1: Outline of negative (A) and positive (B) feedback regulatory networks. Transcription rates are $\lambda_1$, $\lambda_{1a}$, $\lambda_{1b}$ and $\lambda_2$. Degradation rates are given by $\beta_1$ and $\beta_2$. Binding and unbinding rates are given by $\gamma_1$ and $\gamma_2$, while $P$ represents the protein concentration. For the positive feedback case $\lambda_{1a} > \lambda_{1b}$. 
6.4 Results

6.4.1 Motivation

As motivation for the work that follows, we look at existing data to analyse how cell volume changes across doubling time for a range of different organisms. Through a literature search we find data for *E. coli* (Bremer and Dennis, 1996), *Salmonella* (Schaechter et al., 1958), budding yeast (Johnston et al., 1979) and fission yeast (Fantes and Nurse, 1977). This is shown in Figure 6.2. The volume data is given by mass per cell which is thought to be a good indicator of cell volume. It can be seen that cell volume increases as doubling time decreases consistently in all the organisms. The work here therefore has generality and may have relevance to many systems. We look to relate this finding to the effects of additional noise in biomolecules from shorter doubling times. We do not have much data relating protein noise to growth rate. We do however have information on how cell parameters can change with growth rate and therefore, we attempt to relate these factors using simulations. We focus on parameters from *E. coli* originally found in Bremer and Dennis (1996).

Figure 6.2: Reported data relating relative volume to relative doubling time. Data is found for *E.coli* (Bremer and Dennis, 1996), *Salmonella* (Schaechter et al., 1958), budding yeast (Johnston et al., 1979) and fission yeast (Fantes and Nurse, 1977). Relative volume is calculated by dividing by the largest volume in the available data, and similarly for relative doubling time.
Firstly, we check our hypothesis that increased partitioning frequency alone can increase the level of noise in a cell. To do this we use the LNA (equation 2.8 and 2.9) derived in Huh and Paulsson (2011) for mean and coefficient of variation (CV) as outlined in Chapter 2. These are as follows

\[
\langle y \rangle_t = \lambda_1 T c_{1,t}
\]

\[
\langle x \rangle_t = \lambda_1 \lambda_2 T c_{1,t} - \frac{c_{2,t}}{\beta_2 - \beta_1}
\]

\[
CV_{y,t}^2 = \left[ 1 + k_{11,t} \frac{c_{1,t}}{c_{1,t}} (A_y - 1) \right] \frac{1}{\langle y \rangle_t}
\]

\[
CV_{x,t}^2 = (S_{x,t} + U_{x,t} A_x) \frac{1}{\langle x \rangle_t} + (S_{y,t} + U_{y,t} A_y) \frac{1}{\langle y \rangle_t}
\]

where \( x \) is a constitutively expressed protein with translation rate \( \lambda_2 \) and degradation rate \( \beta_2 \), \( y \) is mRNA with transcription rate \( \lambda_1 \) and degradation rate \( \beta_1 \). The division time is \( T \) and \( t \) is the time point in the cell cycle. The protein variation is split into that from the birth death process \( (S_{y,t} \text{ and } S_{x,t}) \) and the part due to cell partitioning \( (U_{y,t} \text{ and } U_{x,t}) \). These terms along with the functions \( c_{1,t}, c_{2,t} \) and \( k_{11,t} \) are further explained in Chapter 2. The \( A_x \) and \( A_y \) terms describe the type of partitioning. For binomially dividing mRNA and protein \( A_x = A_y = 1 \), while for perfectly partitioned mRNA and protein \( A_x = A_y = 0 \). The effects of growth rate on mRNA and protein copy numbers along with their respective coefficient of variation (CV) and covariance are shown in Figure 6.3. It can be seen that the copy number decreases with increasing growth rate due to the increased dilution. It can also be seen that increased partitioning due to faster growth rates leads to a higher CV for protein and mRNA. While this is in part due to the decreasing copy number, it can also be seen that the contribution of the partitioning noise to the total noise also increases.
Figure 6.3: Approximation of effects of growth rate on noise. The effects on mRNA and protein copy number and coefficient of variation are shown (blue line), along with the contribution of cell partitioning for protein CV (green line). A) Approximation at $t = 0$, just after cell division. B) Approximation at $t = T/2$ at the mid point of the cell cycle. C) Approximation at $t = T$, just before cell division.

We then compare simulations for a constitutively expressed protein under binomial partitioning for both mRNA and protein with the approximation derived in Huh and Paulsson (2011). This is shown in Figure 6.4. It can be seen that as expected, protein noise increases as cell doubling time decreases. We observe that while the approximation and the simulations agree qualitatively, there is a quantitative difference observed in protein noise. The protein copy number is found to be the same for the approximation and the simulations (simulations copy number = 360.78 for $A = 1$ at $t = T/2$, $T = 50$ and approximation copy number = 364.99). It can be seen that the linear noise approximation underestimates the protein noise. We also look at the effect under perfect partitioning, and find that as expected, overall noise in mRNA and protein is less than for binomial partitioning, while the means are equal (for simulations copy number = 365.14 for $A = 0$ at $t = T/2$, $T = 50$, for approximation copy number = 365.00). Again, we find that protein CV is underestimated by the linear noise approximation. We have checked the assumption that there is no correlation between protein and mRNA copy numbers and found no evidence to invalidate the assumption. We believe the discrepancy could be due to the non-linear jump in copy numbers at cell division and that this makes the ap-
proximation for variance less accurate. We find that by increasing transcription rates to increase the protein copy number, the approximation for the protein CV improves. This is demonstrated in Figure 6.5. Through the remainder of the work we rely on stochastic simulations to make new inferences and focus on results at the mid-point of the cell cycle ($t = T/2$).

Figure 6.4: Comparison of LNA for growing cells to our simulations for binomial partitioning $A = 1$ and perfect partitioning $A = 0$. A) Mean and CV for mRNA and protein at $t = 0$, just after cell division. B) Mean and CV for mRNA and protein at $t = T/2$, the mid point of the cell cycle. C) Mean and CV for mRNA and protein at $t = T$, just before cell division.
6.4 Results

Figure 6.5: Convergence of LNA for growing cells to our simulations for binomial partitioning at different mean protein copy numbers. A) Mean and CV for mRNA and protein at \( t = 0 \), just after cell division. B) Mean and CV for mRNA and protein at \( t = T/2 \), the mid point of the cell cycle. C) Mean and CV for mRNA and protein at \( t = T \), just before cell division.

6.4.2 Constitutive Gene Expression

In Klumpp et al. (2009) a deterministic model with parameters from experimental data (Bremer and Dennis, 1996) is used to observe how growth rate affects protein copy numbers and concentration for simple regulatory circuits. Using stochastic simulations we can now analyse the effect that these parameter changes have on the level of noise in the cell. While it is clear from Figure 6.3 that increasing growth rate increases the level of protein noise, it is possible that the changes in transcription rate, gene copy number and cell volume offset this additional noise. To investigate this we perform simulations to gain insight into how changing transcription rates and gene dosage affects noise at different doubling times. We use parameters found in Klumpp et al. (2009) and Bremer and Dennis (1996) and these are shown in Table 6.1. We assume that translation rate and mRNA degradation rate are constant across different doubling times. They are found to change a small amount in Bremer and Dennis (1996), and we note that including these minor differences does not change our results significantly. Gene dosages reported are averaged and not whole numbers, therefore for simplicity we simulate a difference in gene dosage by multiplying the transcription rate by the gene dosage. We compare these simulations
Chapter 6. The Effects of Cell Partitioning on Noise in Genetic Regulatory Networks

against those where parameters remained constant across doubling times. Specifically we change transcription rate, gene copy numbers and cell volume in turn. We first compare how copy numbers of mRNA and protein and their respective concentrations change with doubling time. This is shown in Figure 6.6. It can be seen that protein copy number remains relatively constant across doubling time when the transcription rate is increased at shorter doubling time. While the protein concentration decreases due to the increasing volume. This is in agreement with Klumpp et al. (2009). It can also be seen that changing the volume in addition to transcription rate and gene copy number maintains the concentration observed in the case when strictly no parameters change. While the changes in parameters allow the cells to maintain a constant mRNA concentration across doubling times. We compare the coefficient of variation for mRNA, protein and their covariance for these cases in Figure 6.7. It can be seen that if there were no increase in transcription rate or gene copy number that the noise would increase with decreasing doubling time. In fact, the protein CV is doubled from doubling time of 100 minutes to doubling time of 20 minutes. However, the increase in transcription rate leads to noise decreasing with decreasing doubling time in both mRNA and protein. Here, the protein CV is reduced by 24.8% from doubling time of 100 minutes to doubling time of 20 minutes. In summary, the increase in transcription rates, gene copy number and cell size help to maintain the level of protein noise and protein concentration when doubling time changes.
Figure 6.6: Effect of growth rate on mean and copy number of constitutively expressed proteins. A) mRNA copy number. B) Protein copy number. C) mRNA concentration. D) Protein concentration. Blue line - No change in transcription rate, gene copy number or cell volume (protein copy number = 316.29 for T = 60 minutes). Green line - Increasing transcription rate with decreasing doubling time (protein copy number = 178.09 for T = 60 minutes). Yellow line - Increasing transcription rate and gene copy number with decreasing doubling time (protein copy number = 316.81 for T = 60 minutes). Purple line - Increasing transcription rate, gene copy number and cell volume with decreasing doubling time (protein copy number = 315.92 for T = 60 minutes).
Figure 6.7: Effect of growth rate on coefficient of variation for constitutively expressed proteins. A) CV mRNA. B) CV Protein. C) CV mRNA and protein. Blue line - No change in transcription rate, gene copy number or cell volume (CV protein = 0.126 for T = 60 minutes). Green line - Increasing transcription rate with decreasing doubling time (CV protein = 0.188 for T = 60 minutes). Yellow line - Increasing transcription rate and gene copy number with decreasing doubling time (CV protein = 0.130 for T = 60 minutes). Purple line - Increasing transcription rate, gene copy number and cell volume with decreasing doubling time (CV protein = 0.132 for T = 60 minutes).

Table 6.1: Model Parameters - Constitutive Gene Expression in Dividing Cells

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Doubling Time (Minutes)</td>
<td>24</td>
</tr>
<tr>
<td>$\lambda_{1r}$</td>
<td>Transcription Rate Per Gene ($min^{-1}$)</td>
<td>1.51</td>
</tr>
<tr>
<td>$\lambda_{1c}$</td>
<td>Gene Copy Number</td>
<td>3.8</td>
</tr>
<tr>
<td>$\lambda_{2}$</td>
<td>Translation Rate ($min^{-1}$)</td>
<td>1</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>mRNA Degradation Rate ($min^{-1}$)</td>
<td>0.5</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>Protein Degradation Rate ($min^{-1}$)</td>
<td>$1 \times 10^{-9}$</td>
</tr>
<tr>
<td>$V$</td>
<td>Relative Cell Volume</td>
<td>0.85</td>
</tr>
</tbody>
</table>
6.4.3 Simple Genetic Networks

It has been observed that negative feedback can decrease noise in gene expression (Thattai and van Oudenaarden, 2001; Swain, 2004). Here we investigate whether this holds with changing growth rates. It was observed in Huh and Paulsson (2011) that negative feedback can increase noise in comparison to constitutively expressed cells when protein half-life is long in comparison to the cell cycle. This was based on a system where expression of mRNA and protein were deterministic and noise was added by partitioning only. In this work we assume that expression of mRNA and proteins are stochastic, in addition to cell partitioning. The effects of negative feedback across growth rates on copy number and noise are shown in Figure 6.8. It can be seen that our system with negative feedback has a higher level of noise than a constitutively expressed protein. This is observed for sufficiently high transcription rate (Swain, 2004). We chose binding parameters to match protein copy numbers at doubling time of 60 minutes to minimise effect of copy number on observed noise. While increasing growth rates can lead to lower noise in constitutively expressed proteins, the noise is maintained across doubling times for the negative feedback model. We performed a similar analysis for a network with positive feedback and this is shown in Figure 6.9. We expect positive feedback to increase protein noise in comparison to a constitutively expressed protein. We observe that not only is the noise greater for the positive feedback model, but that noise increases greatly as doubling time decreases.
Figure 6.8: Effect of growth rate on system with negative feedback. A) mRNA concentration. B) Protein concentration. C) Average gene on rate. D) mRNA CV. E) Protein CV. Blue solid line - Negative feedback model (protein copy number = 320.29, CV protein = 0.223 for T = 60 minutes). Green dashed line - constitutive protein expression (protein copy number = 286.40, CV protein = 0.186 for T = 60 minutes).

Figure 6.9: Effect of growth rate on system with positive feedback. A) mRNA concentration. B) Protein concentration. C) Average gene on rate. D) mRNA CV. E) Protein CV. Blue solid line - Positive feedback mode (protein copy number = 359.22, CV protein = 0.541 for T = 60 minutes). Green dashed line - constitutive protein expression (protein copy number = 475.06, CV protein = 0.154 for T = 60 minutes).
Next we consider a protein, C, which is only produced by the binding of two constitutively expressed proteins A and B. We assume that C is diluted by cell division only and can unbind relatively quickly. The A and B proteins are produced and diluted at the same rate. We observe how the concentration and noise are affected by growth rate in contrast to the constitutively expressed proteins. The results are displayed in Figure 6.10. In contrast to A and B, the copy number of C decreases with increasing growth rate. This is due to the dependence of the expression of C on the concentration of A and B, coupled with the relationship between concentration of A and B on growth rate. It can also be seen that the relationship between the noise and doubling time for the simple proteins A and B is the opposite of that for the complex protein C. While noise decreases with increasing growth rate for A and B, noise increases for with increasing growth rate for C. That is, noise increases as doubling time decreases. This is likely due to the decreasing copy number of C with decreasing doubling time.

![Figure 6.10: Effect of growth rate on a bound protein. A) Protein copy number. B) Protein concentration. C) Protein CV. Blue and Green lines - constitutively expressed proteins A and B (protein copy number = 381.57, CV protein = 0.0927 for T= 60 minutes). Yellow line - bound protein (protein copy number = 252.97, CV protein = 0.0839 for T= 60 minutes).](image)

Other networks of significant interest are those that cause bistability (Gardner et al., 2000). For example, it is useful to study the reasons for growth bistability in persister cells. One source of bistability can be a constitutively expressed toxin. We have shown the concentration of such proteins decrease with increasing growth rate. This can in turn affect the growth rate of the cell negatively. Therefore, there is an overall positive feedback acting through the growth rate (Klumpp et al., 2009). The effect of toxin concentration on growth rate is modelled by the following function
Chapter 6. The Effects of Cell Partitioning on Noise in Genetic Regulatory Networks

\[
\text{Growth Rate} = \frac{\mu_{Max}}{\left( 1 + \frac{<P>}{P_k} \right)}
\]

where \( \mu_{Max} \) is the maximum growth rate, \( <P> \) is the average toxin concentration of the previous generation and \( P_k \) is a half saturation constant. While transcription rates and cell volume are adjusted accordingly by interpolating from the parameters in Table 6.1. It is possible that the level of noise due to the growth rate can have an effect on the rate of switching and the state of the cell. We therefore investigate the effect of growth rate on the rate of switching. The simulation runs for a cell lifetime, and the growth rate of the subsequent generation is determined by the average toxin concentration of the previous generation. This is demonstrated in Figure 6.11. We then change the maximum growth rate across the bistable region to see the effect that this has on switching rate. It can be seen in Figure 6.12 that the growth rate does affect the rate of switching between two growth rates. However, the interaction between growth rate and the effects of noise on switching rate are complex. It appears that the switching rate is greatest at the middle of the bistable region and is therefore more to do with the maximum growth rate than the noise. However, if we also look at the switching rates when the partitioning of mRNA and protein are perfect (\( A = 0 \)), then the effects of partitioning can clearly be seen on switching rates (Figure 6.12).
Figure 6.11: Bistable growth rate due to a constitutively expressed toxin. A) Histogram showing number of cells with a particular growth rate. B) Changes in growth rate over successive generations. C) Change in mean toxin concentration over successive generations. This is what triggers the change in growth rate in the generation that follows. D) Toxin copy number over successive generations.

Figure 6.12: Effect of maximum growth rate on switching rate between two stable growth rates. Blue line - Binomial partitioning of mRNA and protein; green line - perfect partitioning of mRNA and protein
6.4.4 Decision Making in Lambda Phage

Finally, we look at the effect of growth media on the lysis-lysogeny decision in lambda phage. In Chapter 5 we considered the effect of difference in elongation rate where cells are infected by different numbers of phage, but here we model what might happen if the experiments were conducted in different growth media. We use the same model as used in Chapter 5 and adjust the relative parameters at different growth rates as per the parameters in Table 6.1. While the faster doubling times we analyse in this chapter are shorter than the 60 minutes in which we model the decision in Chapter 5, further analysis of the experimental data finds that when cells are infected by phage that their division is delayed. This is demonstrated in Figure 6.13. It can be seen that the time until division is longer for infected cells but that the time is not further affected by MOI. We therefore still analyse the average concentration of CII for 60 minutes after infection, even if the division time of the cell in the absence of infection is less than 60 minutes. The rate of lysogeny as a function of growth rate when MOI = 1 and MOI = 2 is shown in Figure 6.14. It can be seen that as growth rate is decreased that the rate of lysogeny increases. This is due to the increase in protein concentrations at slower growth rates. This is the case for both MOI = 1 and MOI = 2.

![Figure 6.13: Effect of MOI on division time in lambda phage. The time from infection of the cell until division for cells undergoing the lysogenic fate, or time from 1st frame to division for uninfected cells.](image)

Figure 6.13: Effect of MOI on division time in lambda phage. The time from infection of the cell until division for cells undergoing the lysogenic fate, or time from 1st frame to division for uninfected cells.
6.5 Discussion

In this work we investigated the effects of cell growth on the protein noise for simple genetic regulatory circuits. It was found that for constitutively expressed proteins, while increasing growth rate alone increases noise, additional increases in transcription rates, gene copy number and cell volume counteract this additional noise. It is therefore observed that increasing growth rates can lead to less noise for a constitutively expressed protein. This is due to the higher copy number of proteins at faster growth rates. For a negatively regulated protein it was found that the level of noise is maintained across growth rate, while for positively expressed protein noise increases for faster growth rate. We also observed that if two constitutively expressed proteins bind to form a more complex protein, then this protein experiences greater levels of noise at faster growth rates. This is largely due to the relationship between protein concentration and growth rate. For a circuit with a constitutively expressed toxin which reduces growth rate, we find that growth is bistable and the rate of switching between steady states in bistable regions depends on growth rate in a complex manner. Finally, we modelled the effect of growth media on decision making in lambda phage. It was found that at slower growth rates the rate of lysogeny is increased. Together these results show that the effects of noise due to cell partitioning and global cell parameters across different growth rates can be important in cellular decision making and signalling.

The basis of our parameters used are from E. coli, and for deeper analysis, different

Figure 6.14: Effect of growth rate on rate of lysogeny in lambda phage. Blue line - MOI= 1, green line MOI= 2.
parameter dependencies on growth rate should also be investigated. It is also possible that different genes are affected by growth rates in different ways, some may not be affected. The data we have on the relationship between growth rate and cell volume in different organisms suggest that our main conclusions may have some generality. For the toxin affecting growth model, we have assumed that the concentration of toxin in one generation affects the growth rate for the following generation. There are other models (Rocco et al., 2013) that assume that the growth rate is affected instantaneously. We have no evidence for support of one method or the other. However, we suggest that some time delay is reasonable.

It was observed in negative feedback loops that the level of noise was maintained across growth rates. While we show this for parameters such that negative feedback increased noise in comparison to constitutive gene expression, by decreasing transcription rate we expect to observe that negative feedback decreases the level of noise (Swain, 2004). The additional factor of growth rate dependent noise in this case will be explored in further work. For a positive feedback loop, we observed that noise increased with faster doubling times.

It is possible that the global dependencies of gene expression and cell size exist to maintain the level of both protein copy number and noise under different growth conditions. In particular, based on our results we hypothesise that the observed increase of cell size observed in bacteria and yeast is to increase biomolecular copy numbers at cell division, which reduces the partitioning noise. Likewise, some systems may use the changes in levels of noise at different growth rates to drive phenotypic changes to enhance survival. If phenotype is largely determined by growth conditions and it is clear that concentration of a biomolecule controlling the phenotypic switch is constant, then differences in behaviour could be explained by the dependency of noise on growth rate.

In further work it would be interesting to see the effects of different transcription rate dependencies on simple genetic regulatory networks. More experimental data for protein noise across different growth rates could confirm our predictions and lead to further developments in our simple model. It has been demonstrated previously that it is possible to acquire this data in both bacteria and yeast (Bar-Even et al., 2006; So et al., 2011). Having a wider range of data across different systems on growth rate dependent parameters would be useful in understanding the wider implications of our results in other systems. Furthermore, we have only analysed growth bistability for a relatively simple model with a constitutively expressed toxin. Models exist that demonstrate the complexity of toxin - anti-toxin systems (Rotem et al., 2010). It would therefore be
useful to look at the effects of growth rate on switching rates in a more complex form of the toxin - anti-toxin network.
Chapter 7

Conclusion and Outlook

In this thesis, mathematical models were used to develop further understanding of biological systems at both the population and single cell level. The main features of the work focused on using genetic detail for population scale interactions and taking into account the effects of noise at the single cell level. The majority of the work was motivated by biological data and this data was integrated into the models, which were used to make further inferences. This work can now be used to inform more experiments and models.

Specifically, we studied the coevolution of bacteria and phage, the ecology of lambda phage, decision making by individual lambda phage and the effects of cell partitioning on noise. This was carried out using a mixture of deterministic ODEs and stochastic models. Throughout the work we have attempted to provide a mechanistic understanding for the observations that inspired the research. Some of the measured outcomes were determined by time-averaged quantities. While it is difficult to know whether this form of measurement is the important factor in a biological outcome or it is a single time point measure, some of our results are likely to hold under such measures. The complexity of the systems meant that most inferences were made using computational work as opposed to analytical results. However, where possible theory has been used to connect the work to mathematical relations.

The models of coevolving bacteria and phage were motivated by previous experiments and models in Forde et al. (2008). Experiments were used to analyse the effect of gene flow across productivity gradients on diversity. The data showed that bidirectional gene flow could have a negative effect on diversity. The data was then used to inform models and make predictions for different rates of gene flow and different genetic interactions. It was found that the effects of gene flow on diversity are not consistent across rates and genetic interactions. In further work it would be useful to perform experiments across a wider range of rates of gene flow to test the predictions and refine the model.
One shortcoming of the modelling used in Chapter 3 is that it is unable to capture data giving the number of coexisting phenotypes and is restricted to the number set out in the model. One possible way of dealing with this would be to develop a model with more phenotypes. While another method that can capture numbers of coexisting phenotypes is adaptive dynamics, this requires the assumption of arriving at steady state while our current results do not require this assumption. The complexity in this particular system could prove difficult to model under this framework.

Work regarding the ecology of lambda phage was motivated by recent experimental results (Zeng et al., 2010) that give a functional form of the dependency of rate of lysogeny on MOI and cell volume in lambda phage. We used this information along with bifurcation analysis of a set of ODEs that compete temperate phage with lytic phage to make new inference on the reasons for being temperate. Analysis of the system with a responsive rate of lysogeny did not provide clear new insight onto the problem in contrast to temperate phage that have a constant rate of lysogeny. While it was observed that responsive phage are able to raise their rate of lysogeny to a high level while still being able to coexist with lytic phage, in contrast to temperate phage with a constant rate of lysogeny. Exploratory simulations demonstrated that temperate phage that responds to fluctuating conditions is able to survive longer than phage with a constant rate of lysogeny, although this observation is parameter dependent. It would also be useful to understand any changes to the rate of lysogeny that may occur under different conditions. If there is any change in the function at different resource conditions, then the bifurcation work may need more detail to give a true understanding of the ecology. There are now examples of experiments that model the ecology of lambda phage (Berngruber et al., 2013). Incorporating fluctuating resource conditions to the experimental setup would not add unreasonable complexity and would provide vital information for further model development and inferences.

The study by Zeng et al. (2010) suggested that the reasons for this functional form of rate dependency on MOI and viral concentration was due to individual decision making. We attempted to provide mechanistic insight to this individual decision making. To do this we used a simple model of the genetic switch and stochastic simulations. We found that the difference in observed rate at different MOI and same viral concentration is in some part due to intrinsic noise and the difference in copy number due to the difference in MOI between the studied cases. There may also be contributions of spatial effects and cell cycle effects. This was inferred after re-analysing the time lapse movies from Zeng et al. (2010) using cellular image analysis to observe the effect of MOI on cell elongation.
rate. In further work it would be useful to observe whether the functional response of temperate phage differs in different growth media. This would not require much alteration to the experimental methods used in Zeng et al. (2010), and would challenge the current dependency and possibly necessitate improvements of our model. We put forward that the important factors we identified could be sources of a general gene dosage compensation mechanism and it would also be useful to carry out similar analysis using both models and experiments to see if these observations are consistent in other systems. While our observations are general, it would be useful to investigate these properties in a more complicated representation of the genetic switch.

Finally, we analysed the effects of cell partitioning on noise in cells using stochastic simulations. Data we have from previous studies suggests that growth rates can affect cell volume and global cell parameters such as transcription rate. While theoretical studies have shown that noise due to cell partitioning is significant, we use this to show that this can change at different growth rates. We find that the combined effects of growth rate on noise and global cell parameters work to cancel each other out under constitutive gene expression, while the level of noise can be maintained under negative feedback models. If there is a constitutively expressed toxin that affects growth rate then it is possible that the growth rate could affect the rate of switching between two bistable growth rates. However, this effect may be coupled with the size of the bistable region and is therefore difficult to untangle. This observation has relevance in study of persister cells. While we studied the effects of cell growth rate on a constitutively expressed protein, networks involving toxins can be more complex. It would therefore be interesting to study the effects for a more detailed network. The lack of experimental data relating protein noise to growth rate means that our model and observations are not challenged by data. Therefore, experimental study of growth rate dependence of stochastic gene expression in bacteria or yeast would be a logical next step in validating our predictions.

The thesis attempted to show the importance of incorporation of biological detail in mathematical modelling. Deeper understanding of the mechanisms underlying a biological system can make for a more useful model, and provide clearer insight. This insight can be used for the next set of experiments on the systems described and may have relevance in other systems. The insight here can be useful for work in ecology, decision making and antibiotic resistance.
References


REFERENCES


Mariëlle I. Gallegos Ruiz, Karijn Floor, Paul Roepman, José A. Rodriguez, Gerrit A. Meijer, Wolter J. Mooi, Ewa Jassem, Jacek Niklinski, Thomas Muley, Nico van Zandwijk, Egbert F. Smit, Kristin Beebe, Len Neckers, Bauke Ylstra, and Giuseppe Giac-


REFERENCES


M. E. Hochberg and M. van Baalen. Antagonistic coevolution over productivity gradi-)

Jason D. Hoeksema and Samantha E. Forde. A Meta-Analysis of Factors Affecting Local
2008.

R. Stephen Howard and Curtis M. Lively. Parasitism, mutation accumulation and the

Bo Huang, Mark Bates, and Xiaowei Zhuang. Super-resolution fluorescence microscopy.

Dann Huh and Johan Paulsson. Non-genetic heterogeneity from stochastic partitioning

Khuloud Jaqaman, Dinah Loerke, Marcel Mettlen, Hirotaka Kuwata, Sergio Grinstein,
Sandra L. Schmid, and Gaudenz Danuser. Robust single-particle tracking in live-cell

Christine M. Jessup and Brendan J. M. Bohannan. The Shape of an Ecological Trade-off

Christine M. Jessup, Rees Kassen, Samantha E. Forde, Ben Kerr, Angus Buckling,
Paul B. Rainey, and Brendan J.M. Bohannan. Big questions, small worlds: micro-
bial model systems in ecology. *TRENDS in Ecology and Evolution*, 9(4):189–197,
2004.

Richard I. Joh and Joshua S. Weitz. To lyse or not to lyse: Transient-mediated stochastic
fate determination in cells infected by bacteriophages. *PLoS Computational Biology*,
7(3), 2011.

G.C. Johnston, C.W. Ehrhardt, A. Lorincz, and B.L. Carter. Regulation of cell size in the

Iain G. Johnston, Bernadett Gaal, Ricardo Pires das Neves, Tariq Enver, Francisco J.
Iborra, and Nick S. Jones. Mitochondrial variability as a source of extrinsic cellular

Graham Joyce, Kerstin J. Williams, Matthew Robb, Elke Noens, Barbara Tizzano, Vahid
Shahrezaei, and Brian D. Robertson. Cell division site placement and asymmetric


REFERENCES


Appendix A

Figure A.1: Population densities of equations 4.1 along different branches for temperate phage $T$ with constant rate of lysogeny when changing bacteria growth parameter $\mu_{Max}$. A) Bacteria; B) lysogens; C) temperate phage; D) lytic phage.
Figure A.2: Population densities of equations 4.1 along different branches for responsive temperate phage $T$ when changing bacteria growth parameter $\mu_{Max}$. A) Bacteria; B) lysogens; C) temperate phage; D) lytic phage.