Electronic Thesis and Dissertation Repository

8-24-2020 11:00 AM

Mathematical modelling of prophage dynamics

Tyler Pattenden, *The University of Western Ontario*

Supervisor: Wahl, Lindi M., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Applied Mathematics

© Tyler Pattenden 2020

Abstract

We use mathematical models to study prophages, viral genetic sequences carried by bacterial genomes. In this work, we first examine the role that plasmid prophage play in the survival of *de novo* beneficial mutations for the associated temperate bacteriophage. Through the use of a life-history model, we determine that mutations first occurring in a plasmid prophage are far more likely to survive drift than those first occurring in a free phage. We then analyse the equilibria and stability of a system of ordinary differential equations that describe temperate phage-host dynamics. We elucidate conditions on dimensionless parameters to determine a parameter regime that guarantees coexistence of all populations. We then develop a resource-explicit model to investigate further the lysis-lysogeny decision in variable environments. A novel feature of our model is the inclusion of a distinct stationary phase for the hosts and lysogens. Through the application of evolutionary invasion techniques, we determine that as variability increases, bacteriophage populations tend to evolve to a fully lysogenic state, so long as the hosts and lysogens are able to enter stationary phase. This lead us to question the evolutionary fates of prophage in fast- and slow-growing bacterial species. Using a partial differential equation model developed in [94], we fit distributions of prophage lengths for both growth classes and observe several significant differences in strategies of the phage that infect both growth classes. Specifically, we demonstrate that phages infecting fast-growing hosts have a much higher rate of lysogeny. Our work sheds light on the long-standing question, "why be temperate?" [176], offering novel explanations regarding the evolution of temperate bacteriophage.

Keywords: Bacteriophage, bacteria, stationary phase, prophage, plasmid prophage, mathematical model, equilibria, stability, chemostat, bioinformatics

Summary for Lay Audience

Bacteriophages (or "phages" for short) are viral particles that predate on bacteria. Phages are pivotal players in the evolution of bacteria, and as such provide interesting avenues for experimental and theoretical work. These microscopic predators are able to infect through two main methods. The first has been studied extensively, wherein the phage infects and immediately kills the host. Phages that utilise this first strategy are often used in "phage therapy" treatments, where they eliminate antibiotic resistant bacteria from an infected human. The second involves the phage becoming integrated into the bacterial genome, and laying dormant for some time. The second strategy has puzzled researchers, as it does not seem to directly benefit the phage. The question as to why the phage evolved such a strategy is the focus of our work. Although some biologists have offered insight into this evolutionary question, we further their work by developing mathematical models of phage-bacteria dynamics. We identify that mutations beneficial to the phage are more likely to propagate forward if they occur when the phage is integrated. We also find that this integration may be an important strategy for the survival of the phage, if the host is living in a highly variable environment. We, therefore, hope our work has generated new answers to this biological puzzle.

Co-Authorship Statement

I, Tyler Christopher Pattenden, declare that this thesis titled, "Mathematical modelling of prophage dynamics" has been written by me under the supervision of Dr. Lindi M. Wahl.

Chapter 2: The fate of beneficial mutations in plasmid prophage, has been published in part in an article entitled "*Prophage provide a safe haven for adaptive exploration of temperate virus*" co-authored with L.M. Wahl in Genetics. LMW developed the original life-history model of chromosomal prophage, while TCP developed the model of plasmid prophage with supervisory input from LMW. TCP performed the analysis of the model of plasmid prophage. LMW and TCP drafted and finalized the manuscript.

Chapter 3: The model used in Chapter 3 was developed for the article "*Evolutionary stability of the lysis-lysogeny decision: Why be virulent?*" co-authored with L.M. Wahl, M.I. Betti, D.W. Dick, and A.J. Puccini in Evolution. The model was developed by all co-authors. AJP and TCP developed the dimensionless version of the model. TCP developed all analysis in Chapter 3 with supervisory input from LMW.

Chapter 4: The model, analysis and manuscript were all developed by TCP with supervisory input of LMW.

Chapter 5: The effect of life-history traits on the distribution of prophage length has been part of the NSERC Undergraduate Student Research Award work by Christine Eagles, supervised by Lindi M. Wahl and TCP. The original data set, and bioinformatics pipeline to compile data was developed by TCP with supervisory input from LMW. CE implemented and ran the code to find the best fit parameters. The statistical analysis of the best fit parameters was developed by TCP with supervisory input from LMW.

Acknowledgements

First and foremost, I would like to thank my supervisor, Lindi Wahl. Lindi has continually pushed me to succeed, challenged me to grow, and supported me in all my crazy ideas over the past years. She has been a role model for me and so many others. Lindi is truly one of a kind, the superstar of superstars, and I cannot imagine completing my graduate studies with anyone else. To the members of our so called "Wahl Lab", past and present, thank you. Our group is, and always will be, a family, and it's been amazing getting to know everyone personally throughout our time together. We help, guide and support each other, and I hope this never changes.

I would like to thank all those in the Department of Applied Mathematics at Western, as well. From the administration team, dealing with my overly excited running around, to the members of the faculty that have taught me so much. I would like to specifically thank Rob Corless, Allan MacIsaac, and Greg Reid; all of whom I've had as instructors for various courses throughout my graduate studies. Each of these courses taught me something that is applied within this work. In addition, I would like to thank Geoff Wild, who has been a guiding figure throughout my time at Western. From his belief in me as an instructor, to his insights into my main project, Geoff has been such a wonderful person to have around.

Finally, to my friends and family, I could not have done this without you. My parents, Cindy and Larry, supporting me throughout all of my schooling, being there when I needed a helping hand, or a pat on the back, you've done more than you can ever know. My brother, Scott, and his family. The laughs (and beverages) we share while I was going crazy meant the world. To my closest friend, Josh LeClair. Thank you for always being there; for being a shoulder to lean on, a friendly face to help push past adversity, and an amazing friend that anyone would be lucky to have. Finally, thank you to so many others, the list too long to try to create, that have helped along the way, been there when I was down, and helped me get back on my feet.

Without all of you, I would not have made it this far. Thank you.

Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction

1.1 Bacteriophage

Bacteriophage (or "phage" for short) are viral particles that infect bacterial hosts, and are estimated to be the most abundant group of organisms on earth [37]. Although phage were first identified over 90 years ago [80, 186], they have become pivotal organisms in scientific research in recent years due to their role in medical treatments [100, 124], ecosystem health and cycling of nutrients [89, 135], and in bacterial evolution [12, 21, 34, 38, 60].

Phage are used in the medical treatment of bacterial infections in procedures called phage therapy [19]. First utilised shortly after the discovery of bacteriophage, phage therapy is often used to eliminate antibiotic-resistant bacteria from an infected patient [27, 100]. Although phage therapy has been practiced over the last eight decades in France, Russia and Eastern Europe [40, 69], the treatment has become more prevalent in recent years, due to the emergence of multi-drug-resistant bacteria [110, 209].

As was previously mentioned, bacteriophage are very abundant and often outnumber all

other organisms within an ecosystem, thus providing substantial evolutionary pressure on prokaryotes [205]. Due to the phage's ability to influence the community present in an ecosystem, for example due to the "kill-the-winner" strategy they employ [182], phage play an integral role in determining ecosystem health [157]. In particular, phage are important components of the cycling of nutrients, such as nitrogen, phosphorous and carbon, throughout the ecosystem [202, 204].

Many bacterial hosts have evolved an immune system of sorts in the CRISPR-Cas system, which targets infecting phage genomes and reduces the possibility of horizontal gene transfer [180]. It should be noted though that co-evolution has also occurred, that is, the phage themselves have evolved in an attempt to evade the immune response of the CRISPR system [161]. Although we often view the interactions between phage and host as antagonistic (or as a more traditional predator-prey dynamic), it is often observed that phage and host may have a mutualistic relationship [140]. Infecting phage may provide various benefits to their host, including virulence factors [26], stress tolerance [197] and most commonly super-infection exclusion [115, 121].

Phage adopt a variety of different infection strategies. Often, phage are classified by their use of two key infection strategies, lytic or lysogenic [114] (see Figure 1.1), that will be explained in detail below. However, these two strategies are not the only classifications, and recently it has been discussed that there is a plethora of classifications that are useful in the identification of bacteriophage [64, 81].

Phage that infect a host through the lytic strategy, called virulent phage, are characterized by bacterial adsoprtion, the hijacking of the host's machinery and the lysis (death) of the host releasing a multitude of new progeny into the environment [4]. In contrast, a phage employ-
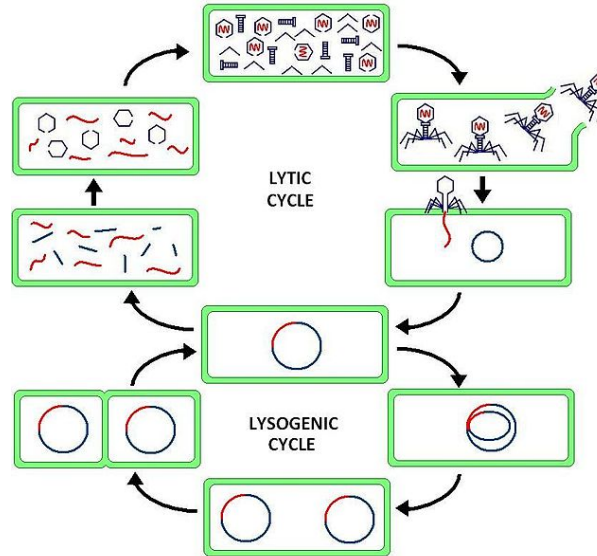
Figure 1.1: Cartoon diagram displaying the two modalities of infection strategy for temperate bacteriophage. Within the lytic life cycle, a phage attaches, and hijacks the host's machinery to produce a large number of progeny. The new phage then burst the cell open, killing it in the process. The lysogenic life cycle is characterised by the integration of the phage genome into the host's genome after adsorption. Generally, the integrated genome will stay within the host's genome for some time before inducing (i.e., excising from the host's genome) and then entering the lytic cycle.

ing a lysogenic strategy has the ability to integrate its genome into the host's genome, without killing the host [41]. These phage, called temperate phage, undergo a decision to determine if they will integrate within the host or undergo the lytic strategy upon attachment [114]. The integrated phage genetic material, called a prophage, may exist for numerous bacterial generations, until induction of the prophage occurs. This process typically excises the prophage from the host's genome, allowing the phage to then enter into the lytic cycle [111]. During the bacterial generations when the prophage is integrated in the bacterial genome, there is mutational

risk involved in the replication process (i.e., the prophage relies on perfect replication by the host) [35]. Due to this gradual mutational decay, the prophage may become unable to perform key functions such as cell lysis, or progeny production; such a prophage is deemed "cryptic" [33, 196]. Temperate bacteriophage will be discussed in more detail in Section 1.1.1.

Although we commonly refer to phage as either lytic or lysogenic, multiple other exploitation strategies exist. One such strategy is that of filamentous phage. Unlike its lytic or lysogenic counterparts, a filamentous phage causes a chronic infection of the host, continually secreting progeny from the host without causing cell death [120]. A filamentous phage, such as the M13 phage, will reduce the host's growth rate due to metabolic strain from replication of the phage DNA and the secretion process, as well as the embedded proteins the phage encodes on the membrane of the host [117, 118]. In addition to secreting progeny into the environment, filamentous phage may also be transferred vertically through generations [107]. These phage have also been of interest to experimental evolution studies that investigate the switching behaviour between a mutualistic and parasitic relationship with the host bacterium [30, 168, 169].

Another prevalent exploitation strategy by bacteriophage is a version of lysogeny wherein the phage forms a plasmid-like structure (a separate circle or rod of DNA), as opposed to integrating within the host's genome [154]. Plasmid prophage, as they are known, often exist as low copy number plasmids within the host's cytoplasm, allowing them to be horizontally and vertically transferred [152, 153]. Plasmid prophage will be discussed in greater depth in Section 1.1.2.

1.1.1 Temperate Bacteriophage and Prophage

As was briefly discussed in Section 1.1, phage have two main life cycles (or infection strategies). Here, we focus solely on the infection strategy of temperate phage. These phage undergo bacterial adsorption, after which the phage may either immediately lyse the host, or may integrate its genome into the host's. The phage then replicates along with the host, and may stay integrated for many bacterial generations [36, 37, 114].

As described in the previous section, once integrated within the host's genome, the phage genetic material is called a prophage [114]. Through genome sequencing studies, it has been discovered that bacterial genomes include anywhere between zero and 15 prophage (in the latter case, prophage account for about 16% of the host's genome) [183]. The large number of prophage found in many different bacterial species suggests that lysogeny is quite prevalent in bacteria [35, 36, 61, 66]. Although it may seem detrimental to harbour viral genetic material, the prophage often confers some benefit to the host bacterium [16]. Prophage are also responsible for repressing phage genes, thus eliminating the possibility of superinfection by the same phage [51].

Several temperate phage have been closely studied. One such phage is the lambda phage. Lambda has been extensively studied, and is the most common example of a tailed, temperate phage [78]. Lambda is often used in research studies to determine factors that influence the decision between immediately lysing the host, or integrating into the genome to form a prophage [98, 151]. Another closely studied temperate phage is phage Mu. Unlike lambda, which only infects *Escherichia coli*, Mu has a wide range of possible hosts [134]. Mu-like phage have a key difference from lambda-type phage in that they insert at apparently random positions in the

host's genome, often leading to noticeable mutations of the host [195].

Temperate phage face a lysis-lysogeny decision upon infecting a susceptible host. This decision is regulated by bistable genetic switches [50, 141, 151, 171]. Dependent on the specific phage, the switch may be controlled by various factors; for instance, [50] suggests the controlling feature is the existence of overlapping face-to-face promoter regions on the phage genome, while [141] suggests that environmental signals and the number of infecting phage control the genetic switch. It is important to note, however, that it has been shown that the propensity, or probability as we will call it, for lysogeny can evolve in response to selection [155].

The link between bacterial life-history traits and the distribution of prophage within the bacterial genome has been recently discussed [183]. One of the life-history traits that was most highly correlated with prophage content was the bacterial minimal doubling time. This doubling time refers to the minimal time (on average) it takes a certain bacterial population to double in size (not to be confused with generation time, which is much more difficult to determine) [67]. Fast-growing bacteria are often found in environments where the resources are highly variable [23, 95], which presents an interesting question: is lysogeny advantageous in highly variable environments? It has also been suggested that the number of prophage carried by a bacterium may be linked with bacterial pathogenicity. Therefore, it has been hypothesized that more prophage may be present in pathogenic bacteria because the prophage confer an evolutionary benefit to their hosts [26, 189].

1.1.2 Plasmid Prophage

As previously discussed, in certain cases the temperate phage may be incorporated into plasmid sequences, freely existing within the host cell, but outside the host's genome [34, 159]. Typical plasmid prophage exist as relatively large genetic elements that are kept at low copy numbers to reduce the metabolic burden on the host [112].

Due to the low copy number of the plasmid prophage, the transmission of the viral genetic material vertically may be compromised [137]. To ensure the virus's long term survival in the population, plasmid prophage encode maintenance strategies to ensure transmission [166]. One such strategy is plasmid addiction, or post-segregational killing [210]. When utilizing this strategy the virus ensures its survival by encoding two proteins, a "toxin/anti-toxin" system, that are released into the host. This system will kill a post-segregation plasmid-free host, while if the plasmid is present the anti-toxin will clear the toxin from the host ensuring that only those that carry the viral DNA are alive in the next generation [76]. Similarly, a plasmid prophage may inhibit the growth of a daughter cell that does not contain a copy [104].

A secondary maintenance strategy that plasmid prophage encode is called a partition system. Here, the plasmid encodes proteins that will ensure that, even at low copy numbers, each daughter cell will receive at least one copy of the prophage [68]. Partition systems direct the segregation of the plasmids by moving them to either side of the cell, minimizing the probability of loss during division [166]. Most low copy number plasmids seem to encode for a partition system, including plasmid prophage [68, 152, 212].

The role of plasmid prophage on the evolution of temperate phage is the focus of discussion in Chapter 2. Within this chapter, we often mention "chromosomal" prophage and plasmid

prophage. For clarity, we call any prophage that is integrated within the host's genome a chromosomal prophage, while those that exist as a plasmid-like structure as plasmid prophage.

1.2 Their Hosts

In most studies (both *in vivo* and *in silico*), the bacterial host that is of interest in researching host-phage dynamics is *Escherichia coli* [92, 109, 176, 211]. However, most, if not all, bacteria have been identified as being able to be infected by some phage [71, 87, 200]. Phage do indeed have specific hosts that they can infect, as evolutionary pressures have forced them to become niche predators of sorts (see [87] for discussion on host range of bacteriophage).

1.2.1 Stationary Phase

Although the main focus of Chapter 4 is the lysis-lysogeny decision of temperate phage, the bacterial stationary phase may play a role in this decision. The stationary phase for a bacterial population is defined as a period of limited or no growth when resources are depleted [96] (see Figure 4.1 in Chapter 4 for a clear diagram). Many bacteria-phage experimental protocols predominantly study the exponential growth phase, when phage and hosts can freely propagate [4, 208]. This experimental design has also been challenged by several researchers, who discovered that phage may persist even through "famine" periods of diminished resources [91, 130, 164].

The common example of a bacterial host with a distinct stationary phase is *E. coli* [62, 95, 146]. As previously mentioned, these hosts are often used in experimental research regarding the lysis-lysogeny decision, however, little is understood about how the stationary phase may

play a role in the switch between the two strategies.

Bacterial stationary phases are often initiated when a population faces less favourable environmental conditions (such as the "famine" periods suggested in [91] and [164]). An interesting feature of the stationary phase, other than its characteristic of inhibition of growth, is that the host cells often become immune to infection by bacteriophage [28, 163]. A distinct stationary phase for the host bacteria population has also been suggested to play a role in the coevolution of phage and host [63]. We conclude, then, that due to its possible role in the coevolution of phage and host, the stationary phase may in fact play a pivotal role in the lysis-lysogeny decision, and has oft been dismissed from both models and experiments.

1.3 Previous Models of Host-Phage Dynamics

Within this section, we will be focusing solely on models of temperate bacteriophage and their bacterial hosts. Extensive literature is available on the dynamics of strictly lytic phage and hosts, but this will not be the focus of our discussion. For an in-depth review of such literature, see [198].

The models used to describe the dynamics of host and phage can often be broken into two main categories: (i) resource-explicit models and (ii) resource-implicit models. We will briefly review important literature of both types of models, which give some insight to our ventures in modelling these dynamics in Chapters 3 and 4.

1.3.1 Resource-Explicit Models

In models where resources are explicit, meaning they are modelled alongside the population densities of the species in question, models of chemostats are often used. The chemostat is often considered the idealization of a laboratory experiment in population studies [179], first described, by all accounts, by Novick and Szilard [139]. A simple chemostat is pictured as a container of a well-mixed solution that has an inflow of a limiting resource for some growing species, in our case a population of bacteria, and an outflow of mixed solution. Smith and Waltman [173] describe the chemostat model in detail, and provide an extensive discussion on methods to analyse mathematical models of chemostats.

A chemostat is often described by a set of ordinary differential equations, that contain two major components: a resource and at least one species. The resource is often modeled with a constant input to the chemostat, but inflow may be altered by some input function, denoted hereafter as $C(t)$. The resource is lost to the system from the obvious outflow, but also through the species in question consuming it. The species then grows dependent on the resource concentration, and is lost due to the outflow of the system or by a death/clearance rate. If the species of interest is a bacterium, the growth rate is often given by the Michaelis–Menten equation [6, 127, 172, 173]; this function was shown to fit empirical bacterial growth results by Jacques Monod [133].

Interestingly, the article that first brought forth the mathematical description of temperate phage and host interactions was a chemostat model developed to address the question, "why be temperate?". Developed by Stewart and Levin [176], the authors sought to elucidate important conditions that may play a role in understanding the presence of lysogeny. Here, the model

contains both temperate and strictly lytic phage, infecting a single host. By breaking down the model into various cases (i.e., excluding certain populations), Stewart and Levin champion two hypotheses on why a phage may be temperate, over its virulent (i.e., lytic) counterpart. Their hypothesis that lysogeny is a method for phage to survive through difficult times is still well-regarded, and is often promoted as the predominant reason for the evolution of temperate phage [90, 122, 144, 145].

Not only does Stewart and Levin's model provide a basis for our work, but it also provides a base model structure for many other research studies. For instance, Mittler bases his initial model structure of two phage strains on Stewart and Levin's, but switches focus to analyse the evolution of the genetic "switch" (described previously) of temperate phage [129]. Evans *et al.*, in fact, extend Mittler's work to include the continuous evolution of temperate phage via small mutations; here, they utilise adaptive dynamics to determine an evolutionarily stable strategy for temperate phage [57]. Schoustra *et al.* make simplifications to the Stewart and Levin model in order to demonstrate that temperate phage may act as allelopathic agents, and as such be favoured by selection [162].

In all the previous work cited thus far, the resource input, $C(t)$, was assumed to be a constant function. An interesting avenue for chemostat models of phage-host interactions is utilising a chemostat to simulate environmental conditions, as is done in [8] for instance. Although some have demonstrated interesting conditions regarding the growth and competition of bacterial species [84, 172], comparitively little work has addressed temperate phage-host dynamics in a variable (often periodic) environment.

1.3.2 Resource-Implicit Models

In contrast to resource-explicit models that utilise chemostats, researchers may also model phage-host interactions without an explicit resource. Here, the bacteria population is allowed to grow at a rate that is a function of population density, instead of growing dependent on resource availability. Typically, so that the population does not grow exponentially large, a logistic growth rate is employed. However, other density-dependent growth models such as the Beverton-Holt or Ricker model, could also be used [14, 156].

A model of resource-implicit phage-host interactions was developed by Berngruber *et al.* in 2013 [12]. Here, the authors build a model of hosts, phage and "infected" hosts (what we later call lysogens) and compete a virulent with a temperate phage to elucidate evolutionary trends of infection. Although they initially dominate, the virulent phage are shown to lose the competition in the long-term, suggesting that as the hosts become less abundant it is more beneficial to the phage to act temperately. Berngruber *et al.* improve upon their model to then include spatial structure in 2015 [13]. Here, they further discuss similar results to those found previously, but also determine that as the population increases its structure, virulence tends to decrease.

A recent model was developed by Doekes *et al.* to determine the role of a "communication" protein on the propensity of lysogeny in a bacteriophage population [52]. In experimental work, Erez *et al.* observed the hijacking of the host bacterium's quorum sensing machinery by a temperate phage [56]. Upon integration into the host's genome, the lysogen secretes a protein into the environment, which experimentally altered the propensity of lysogeny. Doekes *et al.* model such behaviour and determine that a switch between lysogenic and lytic behaviour at

certain levels of the protein confers the most benefit to the phage.

1.4 Evolutionary Invasion Analysis

A common notion from evolutionary theory is that the "fittest will survive", an idea coined by Herbert Spencer in 1864 [175] after reading the prolific work of Darwin. An obvious question then presents itself, namely how does one determine who is the fittest in various scenarios of competition? This is the underlying idea of evolutionary invasion analysis, and the bulk of the proceeding discussion.

According to Diekmann, evolutionary invasion analysis comes in two brands: a game theoretical approach popularized by Maynard-Smith [174] and the more recent approach using adaptive dynamics [47]. The latter approach focuses on the long-term growth of a mutant invading a population in the environmental conditions that are set by the current wild-type (or resident) species [45]. We define this resident species as the population with the specific allele that is currently fixed, while the mutant or invading population is that with a different allele that has arisen by mutation [142]. The goal of undergoing an evolutionary invasion analysis is two-fold: (1) determine what types of mutant alleles can invade and (2) determine if there is a resident that can resist invasion [142].

To determine the answers to these questions, we undergo an invasion analysis. We outline the steps for a one-dimensional trait given by Diekmann in [47]. After determining the allele or trait in question, we allow the population to become monomorphic in that trait, which as defined by Otto and Day would be the resident population [142]. An invasion analysis involves seeding a rare mutant that differs from the resident only slightly in the specific trait and determining

if it begins to grow. To determine the long-term outcome, a non-linear analysis is generally

necessary. The tools necessary to complete such an analysis are not typically generalizable

across models, although several researchers have determined patterns in their results [47, 65].

Using the information discerned from this non-linear analysis and invasion of a mutant, we

finally can determine the trait value that will be invulnerable to invasion, *i.e.,* cannot be driven

to extinction by a subsequent mutant. This strategy is often referred to as an evolutionarily

stable strategy (ESS) [174].

Conveniently, the results of an adaptive dynamic analysis may be plotted in a pairwise

invasibility plot, or PIP. An example PIP that has been described many times preivously, for

instance in [46, 65, 142], is shown in Figure 1.2. A line is drawn at a 45° angle, indicating

neutrality between the resident and mutant trait; intuitively, a mutant and resident that have

the same trait value are neutral to invasion. This line separates regions of + and − indicating

a successful and unsuccessful invader respectively. We can think of these regions as meaning

that a successful (unsuccessful) invader will have offspring greater than (less than) one in the

+ (−) region. [142].

A potential ESS corresponds to the intersection of the two lines on our simple PIP; indicated

with a dashed circle in Figure 1.2. In our cartoon PIP, the indicated point of intersection is in

fact an ESS, as no other mutant strategy will be able to invade such a point; equivalently the

growth rate of any other mutant will not be higher once the population has reached this trait

value [142, 47].

Each PIP has a corresponding coexistence plot, that displays regions of parameter space

where both the mutant and resident traits may persist resulting in a polymorphic population

[125]. The coexistence plot is constructed by mirroring a PIP over its diagonal and overlaying

the resulting mirrored PIP over the original PIP [22]. For a clear illustration of this construction, we refer the reader to Figure 6 of Brännström *et al.* [22].

Although we only show one PIP with a clear ESS, these graphical representations may take many other forms dependent on the model in question and, as such, may indicate a variety of evolutionary trajectories. Such trajectories and a discussion on their biological meaning are reviewed in [46], [125], and [142].
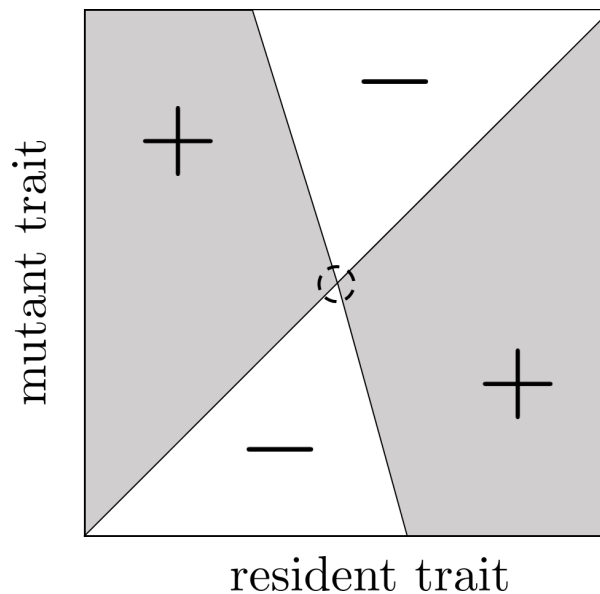
Figure 1.2: Cartoon diagram of a common pairwise invasibility plot shown in most adaptive dynamics literature. Resident traits are plotted on the horizontal axis, with invading mutant traits on the vertical. Regions where a mutant trait can invade a resident trait are coloured in grey and marked with a "+" symbol. Regions where a mutant trait cannot invade are coloured white and marked with a "−" symbol.

1.5   Extinction Probabilities

De novo mutations are the raw material for adaptation. But even if they confer a benefit, *de novo* mutations that occur as a single copy in a large population do not usually succeed in producing a lineage that persists. This loss of rare mutations is one aspect of what is known as genetic drift [102]. Therefore, the study of the long-term fate of rare mutations has become a focus of interest. A typical approach in determining the long-term fate is through the utilisation of a branching process [88] which tracks birth and deaths of a focal lineage, assuming all else in the population is constant [9, 72].

Mathematically, branching processes are often modelled using probability generating functions (PGF). Consider a discrete random variable R, with a probability mass function p_i, where the probability mass function is defined as a function that gives the probability that a discrete random variable is exactly equal to some value (i.e., $P(R = i)$) [177]. If R can only take non-negative integer values, the PGF for R is defined as

$$f(x) = \sum_{i=0}^{\infty} p_i x^i.$$

Note that $f(0) = p_0$. Suppose that R defines the offspring distribution of a single generation in a lineage, then $f(0)$ gives the probability that there are zero offspring in that generation.

If we continue with our assumption that R defines the number of offspring in the current generation, we can then use the PGF, $f(x)$ to determine the offspring distribution after two generations or further. To do so, we simply need to take the composition of $f(x)$ with itself. For example, if we wish to determine the offspring distribution after two generations of reproduction, we would need $f(f(x))$ or $f^2(x)$ [58]. Then $f(f(0))$ gives the probability of zero individuals in the lineage after generation two.

To find the ultimate probability that the lineage goes extinct, we need to evaluate the limit as the number of generations, n, goes to infinity. We define this probability as X. Mathematically, we are looking to determine

$$X = \lim_{n \to \infty} f^n(0)$$

under the condition $0 \le X \le 1$.

If we apply the function f to both sides of the equation above, we find

$$f(X) = f(\lim_{n \to \infty} f^n(0)) = \lim_{n \to \infty} f(f^n(0)) = \lim_{n \to \infty} f^{n+1}(0) = X,$$

which is a powerful result that allows us to determine the extinction probability of a lineage given the probability generating function [58, 72]. We utilise this idea of calculating the extinction probability of a *de novo* mutation in Chapter 2.

1.6 Basic Reproduction Number, R_0

The basic reproduction number is an extremely important quantity in mathematical biology in the field of disease spread and the identification of the invasiveness of a mutant species. This scalar quantity is defined in a variety of ways, however, we will focus on its definition as the average number of secondary infections an infected individual will produce over its lifetime in a completely susceptible population [44].

In 1927, Kermack and McKendrick developed the first epidemic model, utilising a threshold quantity made from model parameters to predict an epidemic outbreak [93]. A general formula, lacking a symbol and name at the time of discovery, was first given by Sharpe and Lotka in 1911 [170]. The first noted naming and symbolizing of the basic reproduction num-

ber[1] as R_0 is credited to Lotka in 1939 [113].

A common approach to identify the basic reproduction number is the usage of a next gen-
eration matrix (NGM). This technique offers both an intuitive biological basis and algorithmic
framework to derive R_0 for compartmental disease models [49, 187]. The method of the NGM
applies to a system of n ordinary differential equations of which $m < n$ define the infected
classes. This concept of "infected" classes can also be applied to the idea of "invading" classes
(or "mutant" class) as described in Section 1.4. The equations for the infected classes are
linearized about the disease-free equilibrium and divided into two matrix components. The
first matrix contains the terms which define the influx into the infected classes from uninfected
classes. This matrix is often denoted as F. The remaining terms, such as the remaining influx
terms and the outflux from each class, are used in the second matrix, denoted by $-V$. The next
generation matrix is then given by $NGM = FV^{-1}$.

The main function of this matrix is to provide the basic reproduction number for a popula-
tion model [49]. The basic reproduction number is given by the spectral radius of the NGM,
i.e.,

$$R_0 = \rho(FV^{-1}),$$

also known as the modulus of the largest eigenvalue of the NGM [187]. To illustrate the
importance of the basic reproduction number, see [77] for a number of studies utilizing this
method for stability analysis of epidemic models.

In 2009, Diekmann *et al.* developed an intuitive and biologically relevant methodology for
the construction of the NGM [48]. A discrepancy often arises in the construction of the NGM
as it may be constructed in a multitude of ways dependent on the scientist's views on what

---

[1]Although, it was called the "net reproduction number" at the time.

constitutes a "new infection". In this article, Diekmann *et al.* validate various definitions of the NGM by proving conditions under which the R_0 approximations are the same [48].

Although this methodology is elegant and often somewhat simple, it can falter when we extend our models to anything other than autonomous ordinary differential equations. As biological processes are often multi-layered and quite complex, the systems we use to model them are often non-autonomous. Specifically, we will discuss some methodologies to deal non-autonomous models with with time-periodic parameters.

1.6.1 Basic Reproduction Number for Time-Periodic Models

The NGM method of calculating the basic reproduction number for biological models is a long-standing approach that unfortunately falters in non-autonomous environments [48, 201]. Here, we will review three methods that may be employed to calculate R_0 when dealing with time-periodic, non-autonomous models: time-averaging, root finding and a numerical discretization of the infection operator. We note that time periodic models are often used for species that are affected seasonal, such as epidemics through the course of a year, or a bacterial population in a fluctuating environment (as is modeled in Chapter 4).

Method of Time-Averaging

In some periodic non-autonomous epidemic models, the basic reproduction number may be obtained from the corresponding autonomous model determined by time-averaging the coefficients. Ma and Ma develop this method for SEIR (susceptible-exposed-infected-recovered) models (and their variants) with and without recruitment for the susceptible class; recruitment is defined as births or loss of immunity by the authors [116]. The authors use a simple

SEIR model to demonstrate sufficient conditions for extinction of the infected class, wherein

only the contact rate, β, varies periodically. It is found that if no recruitment is allowed, then

$\max_t\{R_0(t)\} < 1$, meaning the maximum that the time-varying R_0 reaches over time is less than

one, is a sufficient condition for extinction [116]. Similarly, the authors determine that when

recruitment is allowed, a sufficient condition for extinction of the disease is the time-average

R_0, denoted in the article as $\bar{R}_0$, being less than one. Concurrently, Greenhalgh and Moneim

determine the same result, regarding a time-averaged basic reproduction number, for models

without an exposed class but with loss of immunity (i.e., for SIRS models) [70, 132].

A caveat to the results of the aforementioned articles was the need for a constant population,

and that there was no delay in infection. Wesley and Allen extend the results of Ma and Ma, and

Greenhalgh and Moneim by demonstrating necessary conditions to calculate a time-averaged

R_0 when the population is not constant [201]. By first proving the existence of a periodic

solution to a logistic growth differential equation with periodic birth and death rates, the authors

elucidate sufficient conditions on multiple variants of the SIRS model where the time-averaged

basic reproduction number is equivalent to our understanding of R_0 from the corresponding

autonomous system [201].

We note that all of the work discussed above applies only to specific classes of either SEIR

or SIRS models, and cannot be widely applied to all time-periodic models.

Method of Root-Finding

Although the time-averaged basic reproduction number is simplistic and intuitive, it fails when

models break the basic assumptions outlined in all of the above examples. Extrapolating these

ideas further, Wang and Zhao demonstrate more broad conditions under which time-averaging

approaches may work, but also note that this is not always applicable to every model [194]. To alleviate some of these constraints, Safi *et al.* develop a numerical method akin to a root-finding method, such as Newton's Method [160].

Their method utilizes theory derived in [194] (see Theorem 2.1 in that article) by considering the ω-periodic system,

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \left[-V(t) + \frac{F(t)}{\lambda} \right] w, \tag{1.1}$$

where $\lambda > 0$ and $V(t)$ and $F(t)$ are $n \times n$ matrices constructed in the traditional NGM approach described by Diekmann *et al.* [48]; however, some of the parameters in such construction may now vary with time. We then let $W(t, s, \lambda)$ where $t \geq s$, be the evolution operator of this system on $\mathbb{R}^n$. Safi *et al.*'s method can then be summarized as an interative approach, outlined as follows in [147],

1. Determine an appropriate range for λ. Pick a value $\lambda_1 > 0$ that is within this range.

2. Compute the evolution operator, $W(\omega, 0, \lambda_1)$ by solving (1.1) through the utilisation of a differential equation solver.

3. Determine the spectral radius of the evolution operator, *i.e.*, $\rho(W(\omega, 0, \lambda_1))$.

4. Calculate $f(\lambda_1) = \rho(W(\omega, 0, \lambda_1)) - 1$. If $f(\lambda_1) = 0$ (or within a specified tolerance ε), then $R_0 = \lambda_1$; otherwise, pick another value for λ_1 within the appropriate range and repeat.

This root-finding iterative method was inspired by the results of Bacaër in [10], based on Floquet theory.

We note, however, although root-finding is quite simplistic in its underlying mechanics, it can be computationally expensive. There is also no guarantee of convergence given by Safi *et*

al. or any literature on the subject since [160]. In addition to this problem, the computational

cost of repeatedly computing a solution to an ODE can be quite high, and suggests a better

method may be needed. This improvement on Safi *et al.*'s method is discussed below.

Numerical Discretization of the Infection Operator

This final approach to calculating the basic reproduction number of an infected class is de-

veloped by Posny and Wang in [147]. The authors comment on the cost and implementation

struggles one might face when implementing the aforementioned root-finding method, and

seek to construct a simpler algorithm. As with Safi *et al.*'s method, Posny and Wang consider a

system of differential equations that is ω-periodic (i.e., $f(t) = f(t + \omega)$), for some $\omega > 0$. They

define an infection operator, as is done in [194], with an evolution operator $Y(t, s)$ defined from

the linear system

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -V(t)y,$$

with $V(t)$ being the same construction as the V matrix in the NGM approach for a non-periodic

system. The authors go onto redefine their next generation operator as,

$$(L\phi)(t) = \int_0^\omega G(t, s)\phi(t - s)\,\mathrm{d}s,\tag{1.2}$$

where

$$G(t, s) \approx \left(\sum_{k=0}^M Y(t, t - s - k\omega)\right) F(t - s).$$

In the approximation of G, the matrix $F(t)$ is from the construction of the next-generation

matrix (as was done with $V(t)$) and M is some integer larger than zero. An interesting feature

of this approximation, that the authors note, is that M need not be excessively large for accurate

approximations, due to the exponential decay of the terms in the summation. By writing the infection operator in this way, it is possible to numerically integrate Equation (1.2).

Consider the interval $[0, \omega]$ that is uniformly partitioned over n nodes, hereby denoted as $t_i = i \cdot \frac{\omega}{n}$ for $i = 0, \ldots, n-1$. The authors then approximate (1.2) using the trapezoidal rule (due to its second-order accuracy):

$$(L\phi)(t) \approx \frac{\omega}{n} \left(\sum_{i=0}^{n-1} G(t, t_i)\phi(t - t_i) + \frac{1}{2} \left[G(t, t_0) + G(t, t_n) \right] \phi(t - t_0) \right).$$

By denoting,

$$\widetilde{G}(t, t_0) = \frac{1}{2} \left[G(t, t_0) + G(t, t_n) \right]$$

we can rewrite the approximation as,

$$(L\phi)(t) \approx \frac{\omega}{n} \left[\widetilde{G}(t, t_0)\phi(t - t_0) + \sum_{i=0}^{n-1} G(t, t_i)\phi(t - t_i) \right].$$

This operator equation may be simplified as the matrix equation $(L\phi)(t) = \lambda\phi(t)$ through two vector multiplications (see Equation (5) of [147] for the details). This rewriting of the infection operator generates a matrix system that has coefficient matrix, A. This matrix A has dimension $(nm) \times (nm)$, where n is the number of partitions and m is the size of the the matrix G. Simplifying the structure of A to an $n \times n$ matrix, where each entry is given by an $m \times m$ matrix, we can construct A as follows:

1. Diagonal entries of A are given by $\widetilde{G}(t_{j-1}, t_0)$, where j is the row of the $n \times n$ simplification of A, and

2. Non-diagonal entries are given by $G(t_j, t_u)$, where $u = (j - k) \mod n$ with j being the row and k being the column of the $n \times n$ simplification of A.

Therefore, the matrix equation of the infection operator may be rewritten as

$$(L\phi)(t) = \lambda\widetilde{\phi} = \frac{\omega}{n}A\widetilde{\phi},$$

where $\widetilde{\phi}$ is a column vector of dimension $(nm) \times 1$.

This rewriting of the infection operator as the product of A with $\widetilde{\phi}$ allows one to find the basic reproduction number as,

$$R_0 \approx \frac{\omega}{n}\rho(A),$$

where ρ is the spectral radius operator as defined in Section 1.6. This idea for the basic reproduction number may be numerically implemented, and is more computationally efficient when compared with Safi *et al.*'s root finding approach [147].

The authors demonstrate their method by comparing it with both previous methods, and that outlined in the basis for Safi *et al.*'s approach. They find that their algorithmic approach gives essentially the same basic reproduction number as the root-finding method; as mentioned, however, it is found to be much more computationally efficient. Both the root-finding and numerical discretization methods determined that the time-averaging method underestimates the value of R_0, while Floquet theory tends to overestimate. The authors of both articles point out that their approaches are more broadly applicable, and are a better approximation of R_0 for time-periodic systems of ordinary differential equations [147, 160, 194].

Chapter 2

The fate of beneficial mutations in plasmid prophage

It has been well-studied that prophage sequences (viral sequences within a bacterial host) can undergo significant genetic degradation, since prophage are temporarily shielded from selection on their lytic function. In addition, the host stands to benefit from the inactivation of potentially lethal prophage. For instance, [15] observed a clear pattern of purifying selection on prophage genes, suggesting that there may be a rapid inactivation of these sequences by their host. Thus, genetic deletions and other deleterious mutations should ultimately accrue and, therefore, impair their function. This has been validated by identifying a large number of defective or degraded prophage elements in various bacterial genomes [35, 36].

In contrast, however, several recent studies have elucidated the adaptive potential of prophage sequences. Likewise, [196] demonstrated that such defective prophage sequences confer a range of beneficial functions to their host (in this case, *Escherichia coli*) when encountering adverse environmental conditions. We can conclude then that prophage sequences clearly im-

prove the fitness of their host, in some situations. In addition, it has been shown that adaptive recombination between prophage and infecting phage occurs quite frequently, demonstrating the role prophage play as an accessible reservoir of phage genetic material [20, 101, 131, 136].

It has been found that prophage sequences within the bacterial genome play a large role in the adaption of temperate phage [192]. In particular, new mutations are more likely to survive if they first occur in a prophage sequence, rather than a free phage. But, do these results still hold when examining a temperate phage that creates a plasmid prophage, in lieu of integration into the genome? This is the question we will explore in this chapter. These results have also been published, in part, in [192].

2.1   Background Biology

Although the vast majority of temperate bacteriophages integrate themselves into the host's genome directly, some, like the P1-phage or n15-phage, are known to form a plasmid-like structure in the cytoplasm of the host [38, 159]. As discussed in Chapter 1, these plasmid-like structures are generally large genetic elements that are kept at low copy numbers in order to reduce the possible metabolic strain on the host [38, 112]. Due to plasmid-prophage keeping low copy numbers, during partition, one or both daughter cells may inherit a single copy of the prophage. In an attempt to ensure the survival of its lineage, a plasmid prophage often codes for some maintenance strategy [166]. These maintenance strategies may include coding for a toxin/anti-toxin system or some partition system that ensures daughter cells receive a copy of the plasmid [54, 213].

2.2 Model

In previous work, it was found that the survival probability of adaptive mutations can depend sensitively on the details of an organism's life history [85, 143, 191]. We, therefore, begin by developing a life-history model for temperate viruses that includes the lytic and plasmid prophage replication cycles (for details on the life history of chromosomal prophages, see [192]). Our model considers free bacteriophages and lysogens, tracking which events occur in their life cycles, and in what order they occur. It should be noted, however, that time is not modelled explicitly, which greatly simplifies the analysis.

Figure 2.1 illustrates the approach. A free phage successfully attaches to a host bacterium with probability A, or is lost with probability $1 - A$. After attachment has occurred, either the temperate phage will lysogenize, forming a plasmid structure inside the host with probability P, or will immediately enter the lytic cycle with probability $1 - P$. The outcome of undergoing lysogeny is a single new plasmid prophage; the outcome of undergoing lysis is a burst of B free phages. In our model we assume this process is immediate. However, in reality, the lysis process would take some time to produce B progeny viral particles within the host, and then release into the environment.

We note that the attachment probability A reflects the overall probability of attachment and successful infection of a susceptible host cell. The underlying assumption here is that the population of susceptible hosts for a given temperate phage is large and approximately constant. We also assume that having one copy of a plasmid prophage is sufficient to not be infected a second time by the same phage. This would be true whenever the prophage confers superinfection immunity (often the case, presumably to reduce the metabolic strain that the

plasmid confers on the host) [38, 53, 178].

A plasmid prophage has a very distinct life history, that has a multitude of avenues available. Cells that contain a plasmid prophage may undergo either induction (the process in which a prophage initiates the lytic cycle), conjugation (the process in which an infected cell horizontally transfers a copy of the prophage to an uninfected cell), cell fission (the process in which a cell produces a copy of itself) or cell death. The probability that induction occurs first is N, the probability of conjugation occurring first is J, fission occurring first is F and cell death occurring first is M, with the condition that $N + J + F + M = 1$. We then use copy fidelity, C, to quantify the probability that each subsequent daughter cell, after fission has completed, carries a functional copy of the plasmid prophage. Thus, the plasmid prophage may be carried by zero, one or two daughters with probabilities $(1 - C)^2$, $2C(1 - C)$, and C^2, respectively. When $C = 1$, both daughter cells are guaranteed to carry the plasmid prophage (suggesting a perfect partition maintenance strategy is being employed [54]), whereas when $C = 0$, neither daughter is a carrier.

It is important to mention that daughter cells that contain zero copies of the plasmid may be short lived due to a toxin/anti-toxin maintenance strategy often employed by such phages [104, 166]. This is immaterial, however, since such cells do not contribute to the prophage lineage. We do not distinguish between daughter cells that carry one or more copies of the prophage due to this being quite unlikely in reality, as described previously.
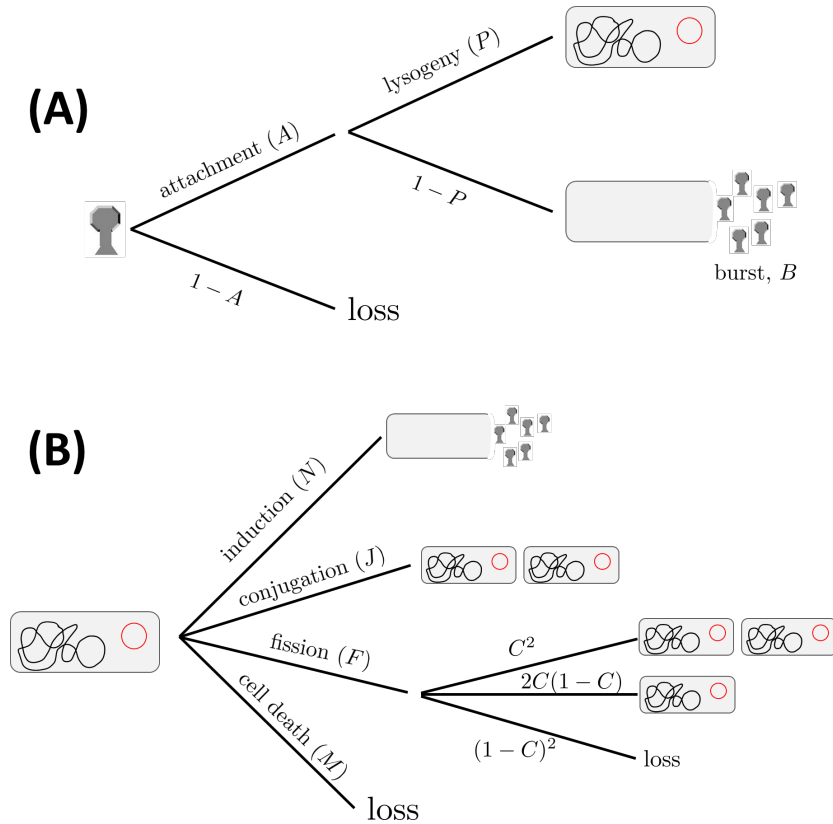
Figure 2.1: Diagram of the life-history model. The life-history events modeled for (A) free phage and (B) lysogens containing a plasmid prophage are shown (see Section 2.2 for details). This diagram is a modified version of Figure 1 from [192]. In both panels, the black curves represent the bacteria's genome and the red circle represents a plasmid prophage.

2.2.1   Wild-type Population

To study the adaptation of temperate phage, we would like to predict the survival probability of rare beneficial mutations. Thus, we consider a wild-type population that is initially at equilibrium. This translates to a wild-type phage population in which each phage produces one surviving offspring. This constraint, that the wild-type fitness is one, can be achieved by balancing parameter values in a multitude of ways.

We classify these ways of balancing parameters in three distinct regimes: neutral, lysogeny

advantage or lysis advantage. These three classifications were also used in the chromosomal

prophage life-history model examined in [192]. The basic idea is to determine the average

number of offspring produced during either the lytic or lysogenic life cycle. The lytic cycle

will have an average number of offspring given by the product of the attachment probability

and burst size, i.e., AB. If this product of parameters is equivalent to one, then we call this

the "neutral" regime. If the product is greater than one, then there is an advantage in this lytic

life cycle, placing us in the "lysis advantage" regime of parameter space. Finally, if we set the

product to be less than one, we have given the advantage to the lysogenic life cycle, meaning

we are in the "lysogenic advantage" regime.

We may think of our life-history model mathematically as a branching process in which a

single free phage produces A lysogens with probability P and AB free phages with probability

$1 - P$. A single plasmid prophage, however, would produce B free phage with probability N,

two lysogens with probability J and $2C(1-C)+2C^2$ lysogens with probability F. The lysogens

produced by a single plasmid prophage can be written as,

$$2J + F\left[2C(1 - C) + 2C^2\right].$$

We determine this equation by identifying the possible avenues a plasmid prophage may un-

dergo to produce new progeny prophage; namely, fission and conjugation. As mentioned

above, when undergoing conjugation, with probability J, a plasmid prophage will horizon-

tally transfer a copy of its plasmid to a non-infected host. Thus, after conjugation has been

completed, two plasmid prophage are present. Similarly, a plasmid prophage may be present

while the host undergoes a fission event, with probability F. As is clear from Figure 2.1, af-

ter a fission event the resulting daughter cells may both have the plasmid, one may have the

plasmid or neither may have the plasmid; where we denote the probability of both daughter cells acquiring a copy of the prophage as C. Thus, one daughter cell having a prophage after fission is represented by the probability $C(1 - C)$ (i.e., one daughter has a copy, and one does not), which may occur in two ways. In contrast, both daughter cells may have the prophage, which would occur with probability C^2. We multiply this term by two, as well, to represent that two prophage are present in the next time step. As mentioned, there is indeed a third event present after fission: no daughter cell has a prophage sequence present. This would occur with probability $(1 - C)^2$, however, much like the event where both daughter cells have a prophage, we multiply this term by the amount of resultant plasmid prophage, namely zero. Thus, this final event is not seen in our equation above.

Defining V_n and L_n as the number of free virus and lysogens, respectively, in generation n, we can express these concepts as a system of difference equations,

$$
\begin{aligned}
V_{n+1} &= A(1 - P)BV_n + NBL_n \\
L_{n+1} &= APV_n + \left(2J + F\left[2C(1 - C) + 2C^2\right]\right)L_n \,.
\end{aligned}
$$

This system of equations can then be rewritten as a matrix system, with the projection (or coefficient) matrix given by,

$$
\mathbf{G} = \begin{bmatrix} A(1 - P)B & NB \\ AP & 2J + F\left[2C(1 - C) + 2C^2\right] \end{bmatrix} . \tag{2.1}
$$

After determining this matrix $\mathbf{G}$, we can express the average number of individuals produced by a single wild-type individual as the dominant eigenvalue (see [39] for an in-depth explanation). For the matrix $\mathbf{G}$, the eigenvalues are given by

$$
\lambda_{\pm} = \frac{1}{2}A(1 - P)B + FC + J \pm \frac{1}{2}\sqrt{(A(P - 1)B + 2(FC + J))^2 + 4ABPN} \,, \tag{2.2}
$$

and it is clear that the dominant (i.e., largest) eigenvalue will be given by λ_+, hereafter referred to as simply λ.

For the wild-type phage to have dominant eigenvalue, λ, equal to one in each parameter regime as described above, we first set the product of AB to an appropriate value. It is worth noting that in the numerical examples to follow, we only give a slight advantage to either lysis or lysogeny by setting B as a constant estimated from the experimental literature and shifting the attachment probability just above or below the value of $1/A$. We then set the other parameters, with the exception of one parameter (in our case the probability of conjugation, J, but the same process could be applied to other parameters), to values consistent with the experimental literature, as chosen in the chromosomal prophage model [192]. We compute the missing parameter value using equation (2.2), to satisfy the condition that $\lambda = 1$. Table 2.1 gives the parameter values that were used to construct the figures below. Note that in [155], the probability of lysogenization, P, is reported to be much lower than our selected parameter values; for λ-phage P is in the range of 10^{-4} to 10^{-1}. Here, we use larger values of P for illustrative purposes, since beneficial mutations must reduce this value in the lysis advantage case. The probability of induction for temperate phages, N, is strongly dependent on environmental conditions and, as such, may vary widely [83, 129, 150].

2.2.2 Beneficial Mutations

Now that we have constructed and ensured our wild-type population has growth rate equal to one (or having, on average, one offspring per generation), we wish to consider the fate of rare adaptive mutations that may arise first in the free phage or the plasmid prophage genome.

Parameter	Lysogeny Advantage	LL Neutral	Lysis Advantage
A, attachment probability	0.008	0.01	0.011
B, burst size	100	100	100
M, cell death	0.2614	0.315	0.4092
P, lysogenisation probability	0.10	0.10	0.20
N, induction probability	0.15	0.15	0.10
F, fission probability	0.20	0.20	0.15
J, conjugation probability	0.3886	0.335	0.3408
C, fidelity probability	0.45	0.50	0.45

Table 2.1: Temperate phage wild-type parameter values for plasmid prophage life-history model.

It should be clear that increasing attachment probability or the burst size increases the phage growth rate. Suppose that the wild-type temperate phage has life-history trait x, we will then use $\widetilde{x}$ to denote the analogous parameter in the mutant phage. Thus, translating our above statement into mathematical conditions, we would say $\widetilde{A} > A$ and $\widetilde{B} > B$ always reflects a beneficial mutation. Correspondingly, mutations that would reduce the death of hosts containing plasmid prophage, and increase the probability of conjugation (i.e., $\widetilde{M} < M$ and $\widetilde{J} > J$, respectively) always benefit the plasmid prophage. The probabilities for induction, N, and lysogenisation, P, reflect transitions between the two life cycles for the phage. Depending on which parameter regime we are in, either lysis or lysogeny advantage, it may be beneficial to increase one or the other probability.

Similarly, the impact of changes to the probability of fission, F, is dependent on the value selected for the wild-type phage's copy fidelity. In Table 2.1, we have selected values of $C <$ 0.5, meaning that it is more likely, if only slightly, that when undergoing fission the wild-type prophage produces less than one offspring. Thus, if C has a value less than 50%, it would be beneficial to the mutant phage to reduce its fission probability, that is $\widetilde{F} < F$ if $C < 0.5$. In contrast, if $C > 0.5$ then it would be beneficial to the mutant to increase its probability of fission, as the number of offspring produced via fission will, on average, be greater than one. This second case, although not shown in the parameter values in Table 2.1, is demonstrated below. Although F has a condition on the direction of beneficial mutation, it is always beneficial to the mutant to increase its copy fidelity (i.e., $\widetilde{C} > C$).

To compare mutations affecting the various traits, we compute the selective advantage, s, of each mutation. If the generation times in the free phage and plasmid prophage life cycles are sufficiently similar, we can simply compute the dominant eigenvalue $\widetilde{\lambda}$ of $\widetilde{\mathbf{G}}$, which would be analogous to the matrix $\mathbf{G}$ from above, with each parameter replaced with the appropriate mutant parameter value. The selective advantage is then simply $s = \widetilde{\lambda} - 1$.

2.2.3   Extinction Probabilities

The model of a simple branching-process described above has an additional advantage in that extinction probabilities can be readily expressed in two implicit equations. Here, we will use X_V to denote the extinction probability of the mutant lineage that first has a mutation in the free phage, while X_L is the extinction probability of a lineage that begins in a single plasmid prophage.

The free phage lineage is lost if the original phage does not attach (with probability $1 - A$) or if the phage successfully attaches (with probability A), but the lineage that is then created is lost. If the phage creates a plasmid prophage, the the lineage is lost only if the resulting lineage, starting with a single plasmid prophage, is lost; by definition, this occurs with probability X_L. Similarly, if the successfully attaching phage immediately enters the lytic cycle, then the resulting lineage is lost if all of the burst of progeny phage are lost; this would be defined then as X_V^B. Putting this all together, we write:

$$X_V = 1 - A + A[PX_L + (1 - P)X_V^B].$$ (2.3)

The plasmid prophage lineage is a little more complex, however, we can logically determine the extinction probability through a similar approach. Initially the plasmid prophage has four avenues to consider; it may induce producing B free phage, it may conjugate with a susceptible host, it may be lost due to cell death or clearance, or it may undergo fission. If the prophage induces, then the lineage is lost only if all B progeny free phage are lost (i.e., X_V^B, as above). If the prophage congugates with another host bacterium, then two plasmid prophages are created, thus, to lose this lineage both must be lost, which we can write as X_L^2. As above, we treat cell loss or clearance just as we treated unsuccessful attaching, where the lineage goes extinct with probability M. Finally, if the prophage undergoes fission, with probability F, depending on the copy fidelity, the lineage is lost with probability

$$C^2 X_L^2 + 2C(1 - C)X_L + (1 - C)^2.$$

Once again, we can put this all together to write:

$$X_L = M + NX_V^B + JX_L^2 + F[C^2 X_L^2 + 2C(1 - C)X_L + (1 - C)^2].$$ (2.4)

For the wild-type population, elementary algebraic manipulation reveals that the only so-lution to these two equations is given by $X_V = X_L = 1$, as the expected number of offspring is equal to one [9]. For beneficial mutations, we once again replace all parameters with their analogous mutant parameter values, then find the solutions with both X_V and X_L constrained between zero and one. In order to solve these equations, we utilize an iterated root finding method, that uses an initial condition sufficiently close to $X_V = X_L = 0$. It is clear that due to the order of the equations, namely B (which should be noted is quite large), that there are many roots to these equations. We start at this end of our constrained probability space in order to ensure that we find the smallest possible real root to these equations. We look to determine this specific root, as opposed to any of the other possible real roots, as the extinction probability is always given by the smallest positive real root [9].

In all our results to follow, we plot survival probabilities, as opposed to the extinction probabilities we solve for. We define survival probability simply as the complement of the extinction probability, meaning $\Pi_V = 1 - X_V$ and $\Pi_L = 1 - X_L$. This illustrates the probability that *de novo* mutations survive drift when rare.

2.3   Results

Figure 2.2 plots the survival probability of *de novo* mutations affecting a single trait versus the trait value, for each of the eight life-history traits previously described. For each trait, one should observe the unmistakable difference between the fate of a *de novo* beneficial mutation that were to first occur in the prophage state (displayed with solid lines) and those that first occur as a free phage (dots). This figure shows the results for the lysogeny-advantage parameter

regime. The survival of the mutations occurring within the plasmid prophage state is always much more likely than those that begin in free phage. This coincides with our results for a traditional chromosomal temperate phage, as well.
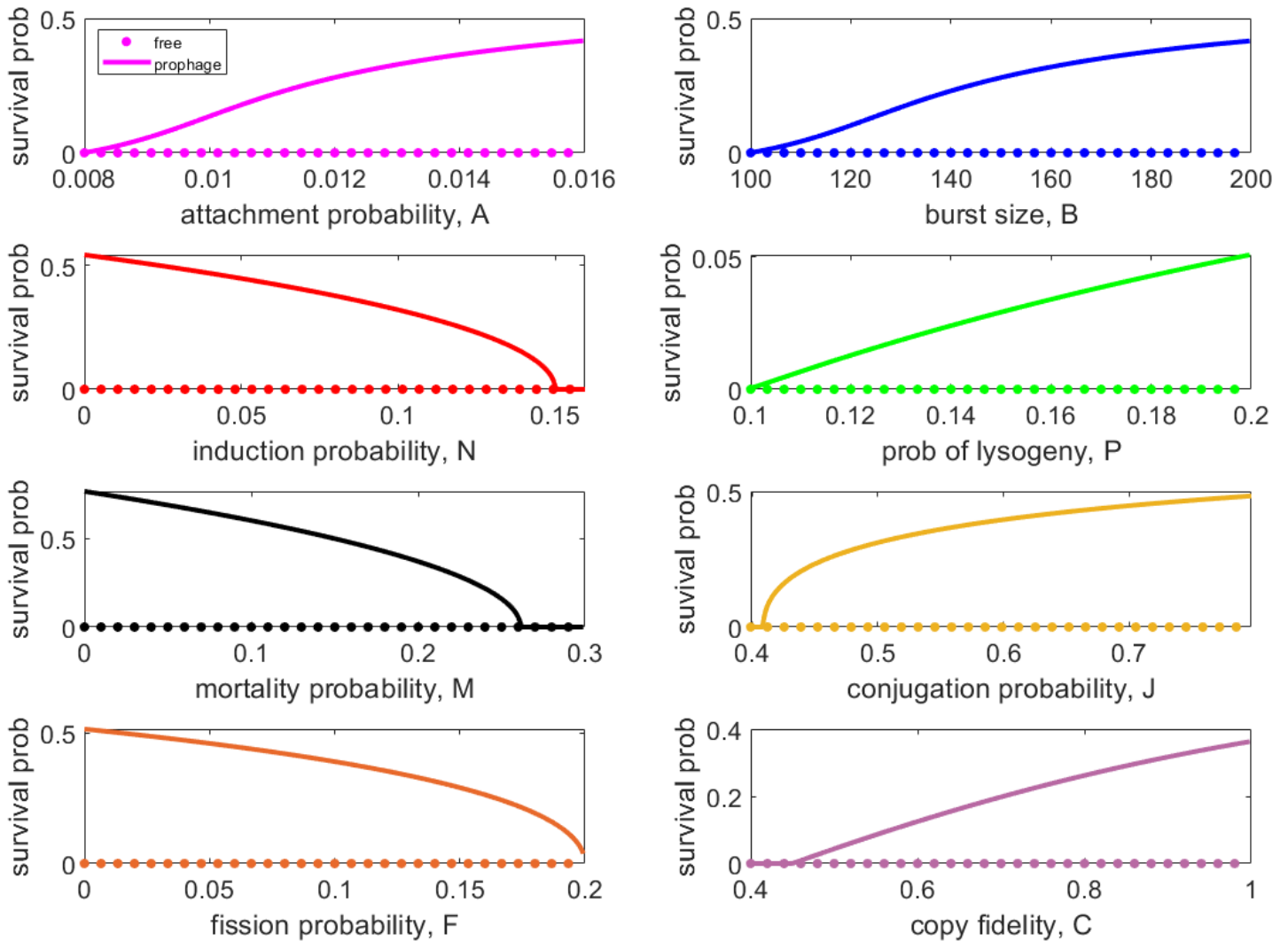
Figure 2.2: The survival probability of beneficial mutations affecting one of the eight life-history traits vs. the trait value. Results are shown for beneficial mutations that first occur in a single free phage (dots) or in a single plasmid prophage (lines). Wild-type paramater values are as provided for the lysogeny-advantage case in Table 2.1.

Another key feature to take note of in Figure 2.2 is the inflection points for mutations that increase attachment probability or burst size. These inflections occur where $\widetilde{A}B = 1$ and $A\widetilde{B} = 1$, respectively. This reflects an interesting phenomenon wherein either attachment or burst size has little effect on the mutant phage's survival, provided that the mutant lineage remains within this parameter regime where lysogeny is advantageous. For larger magnitude changes (i.e., those that make the product of $AB > 1$), we notice that substantial survival probabilities emerge.

Figure 2.3 shows the same results, plotted against selective advantage, s, for each mutation, and thus allows us to make comparisons across the life-history traits. We have computed s in this figure from the projection matrix eigenvalue λ, assuming that the generation times for both phases of the phage life cycle are similar. As we see in Figure 2.2, the y-axis in the top panel versus the bottom panel has a stark difference. This illustrates that beneficial mutations that were to first occur in the plasmid prophage are orders of magnitude more likely to survive than those which begin in a free phage. For mutations that occur first in the free phage, we see that attachment mutations are most likely to survive, followed by burst size mutations. For mutations that first occur in a plasmid prophage, however, a mutation that reduces the host mortality is most beneficial. Mutations that increase the probability of conjugation are also very likely to survive. This presumably follows from the fact that, in our model, conjugation guarantees a successful creation of a new copy of the plasmid prophage, where fission does not.

The fate of mutations that affect fission rates depends exclusively on copy fidelity, C, in the wild-type phage. If $C < 0.5$, as we've illustrated in Figure 2.3, plasmids are often lost during fission and the expected number of copies of the plasmid after fission is indeed less than one;
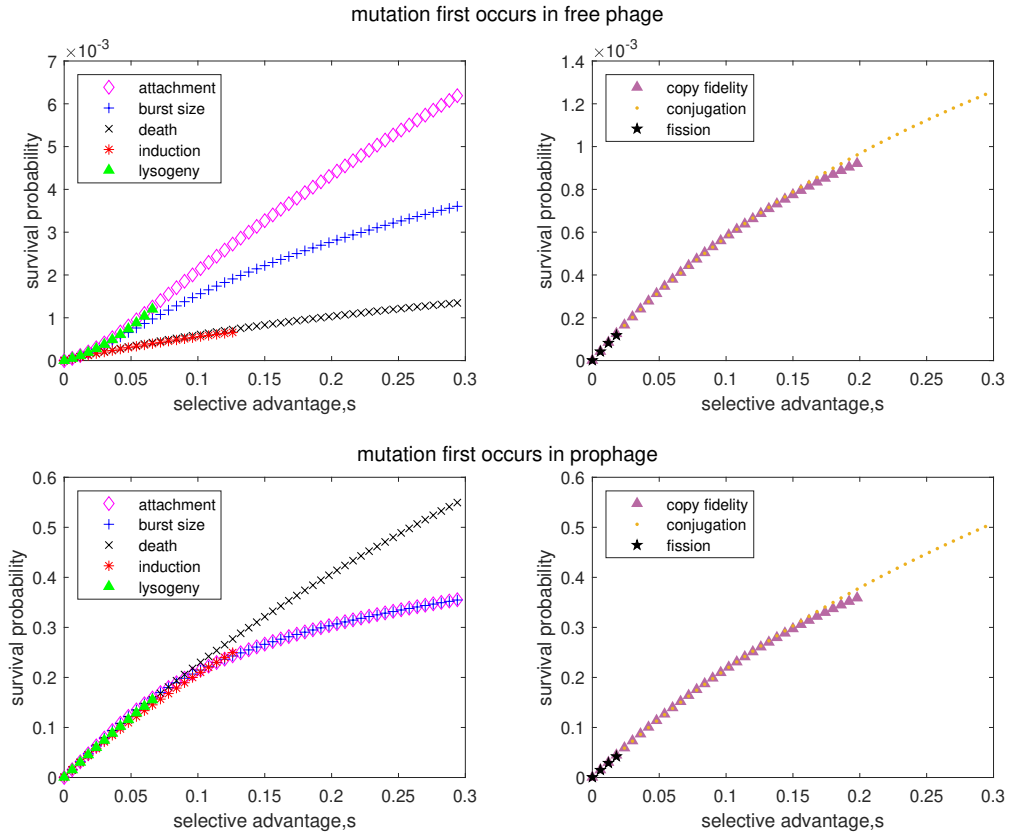
Figure 2.3: The survival of mutations on plasmid prophage vs. selective advantage. Panels on the left illustrate traits that are in common between plasmid and chromosomal prophage; panels on the right are traits unique to a plasmid's life history. We find that muations that first occur in plasmid prophages are far more likely to survive than those first occurring in free phage. Wild-type parameter values are $A = 0.008$, $B = 100$, $M = 0.2614$, $N = 0.15$, $P = 0.10$, $F = 0.20$, $J = 0.3886$, and $C = 0.45$.

thus, by reducing F, we are actually benefiting the phage. The reverse holds true if $C > 0.5$. In this case, mutations that increase the probability of fission, F, have a high survival probability, as shown in Figure 2.4.
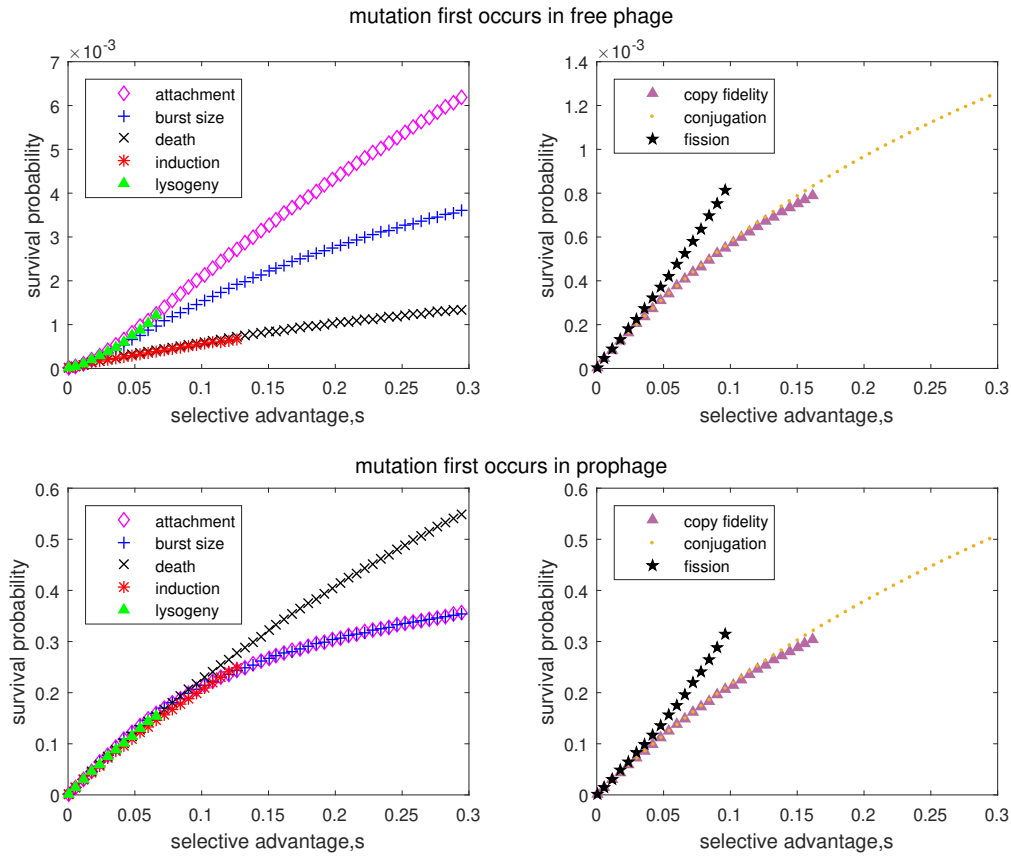
Figure 2.4: The survival probability of mutations on plasmid prophage, versus selective advantage. Panels on the left illustrate traits that are in common between plasmid and chromosomal prophage; panels on the right are for traits unique to the plasmid life history. This figure illustrates the lysogeny advantage regime, with $C > 0.5$. Parameter values are: $C = 0.55$, $M = 0.2814$, and $J = 0.3686$; all other parameter values are as described for Figure 2.3.

2.4   Discussion and Future Work

Prophage sequences constitute of the primary sources of genetic diversity in bacterial populations [26, 34]. Although integrated prophage are often the focus of these studies [34, 37, 79, 144], plasmid-forming temperate phage offer a new avenue in the explanation of prophage's

role in generating *de novo* genetic innovations.

Because phage sequences constitute the most diverse and pervasive gene pool on Earth, with extremely large population sizes and (in comparison) short generation times, phage arguably represent the "principle reservoir of genetic novelty" [97] for all life forms. For instance, it has been experimentally shown that phage can facilitate lateral gene transfer (either chromosomally or through a plasmid) between highly divergent species, like finding arthropod sequences within a bacteriophage [18]. While phage present a novel exploration of genotype space, most of these innovations are lost to genetic drift. Within this project, we sought to understand this process more deeply and look to forecast which factors affect the survival of mutations in this vast reservoir.

Principally, our results suggest that *de novo* beneficial mutations in temperate bacteriophage are much more likely to survive genetic drift if they first occur during plasmid prophage replication, as opposed to occurring during the lytic replication. It is reasonable to think this is due to the relatively low variance in offspring that a bacteriophage experiences when it has integrated within a host (either in the traditional sense, or as a plasmid-like structure as we've described here). In contrast, a bacteriophage is more likely to experience high variance in offspring success in the lytic cycle. We can see this by considering a population of bacteriophage at equilibrium that has a burst size B; on average only one of B daughter phages can successfully attach to and infect a new host cell. Thus the probability of immediate extinction of the bacteriophage of any new lineage produced during a burst of B new progeny is given by $1 - 1/B$ for the lytic phage, which is much higher than immediate extinction for a prophage. Roughly speaking, we can expect that prophage mutations are at least $B/2$ times more likely to survive genetic drift than their free phage counterparts (this can be seen clearly in Figure 2.2).

Our results, therefore, suggest that plasmid prophages may have played a disproportionate role in genetic innovation for their associated free phage lineages. Experimentally, rapid evolution and coevolution in temperate phages has been observed; namely, induction rates and the probability of lysogeny rapidly respond to selection [155], whereas temperate phage can adapt to host coevolution by exploiting their hosts [31, 64, 126]. The evolutionary potential of obligately lytic vs. temperate phage has not yet been explored; leading to the question: would a temperate phage evolve more rapidly if genes for lysis were knocked out, or with the genes for lysogeny knocked out, assuming all else is held equal? Our results suggest that genetic drift would strongly favour innovation if it were to occur first in a prophage (integrated or as a free formed plasmid).

Chapter 3

Analysis of the existence and stability of equilibria in a model of temperate bacteriophage

In this chapter, we will be developing and analysing a system of non-linear ordinary differential equations describing the dynamics of temperate phages, prophages and their hosts. The evolutionary dynamics have been studied, in part, in Wahl *et al.* (2019) and Berngruber *et al.* (2013, 2015) . However, no analysis of the equilibria or stability of such a model was completed in these papers. This analysis will establish a foundation for the more complex model treated in Chapter 4.

3.1　Introduction

As mentioned previously, temperate bacteriophage are viral particles that infect bacterial hosts, and may undergo two distinct propagation strategies: the lytic strategy and the lysogenic strategy [184]. In this chapter, we develop a simple model to examine the interaction between bacteria, a temperate phage species and a lysogen (an infected host with an integrated viral genome, called a prophage). We track the population densities of each of the three classes, hosts, H, free virus, V, and lysogens, L. Host bacteria reproduce at maximum rate r, and this rate is reduced to zero as the cell density approaches a fixed carrying capacity, K. Free virus adsorbs to host cells via mass action kinetics at rate α. Upon adsorption, the virus instigates lysogeny with probability p, or induces lysis with probability $1 - p$; the latter produces a burst of b viral particles. The free virus is cleared from the environment, denatured, or lost at an overall rate c. By assuming that an integrated viral genome gives no benefit or cost to the host bacteria, the lysogens will grow at the same rate as uninfected hosts. The lysogens are also induced at a rate v, producing a burst of b free virions. Finally, we assume that lysogens cannot be re-infected, since many prophages confer immunity to superinfection [17, 185]. These assumptions yield the following system of ordinary differential equations:

$$
\begin{aligned}
\frac{dH}{dT} &= r\left(1 - \frac{\kappa}{K}\right)H - \alpha HV \\
\frac{dL}{dT} &= r\left(1 - \frac{\kappa}{K}\right)L + p\alpha HV - vL \\
\frac{dV}{dT} &= \eta bL + b(1 - p)\alpha HV - cV \ .
\end{aligned}
\tag{3.1}
$$

where κ is the total cell density, i.e., $\kappa = H + L$ and all other parameters (namely α, b, c, v, r, K and p) are positive. If any of these parameters were zero, our model would not be sensible. For instance, if c or v happen to be zero, the free phage and lysogen may only grow. In addition,

p, the probability of lysogeny, is in the interval (0, 1); we do not include the extreme values of

the probability in order to ensure existence of both the free virus and lysogens in our model.

Two additional assumptions are present within the construction of System (3.1). First, we

have assumed that the lysis time is negligible on time scales we will consider, meaning we treat

lysis as an instantaneous event. Second, we have assumed that a carrying capacity, K, limits

the cell density, as opposed to explicitly including a resource concentration [176]. In Chapter

4, we analyse a more complex model that includes a resource compartment.

We also note that System (3.1) is very similar to the model developed by Berngruber *et al.*

(2013, 2015) in a study of the frequencies of virulent and avirulent phages over the course of

a single infection. In addition, our model presents a similar structure to the model developed

by Beretta and Kuang [11]. However, their model only examines the interplay of strictly lytic

phage and bacteria, while ours is investigating the interactions between temperate phage and

bacteria.

For the sake of simplicity, we rescale System (3.1) to write it in a dimensionless form. By

expressing the host and lysogen densities as fractions of the carrying capacity, and normalizing

the free virus density in units of burst size, i.e.,

$$h = \frac{H}{K}, \qquad l = \frac{L}{K} \quad \text{and} \quad v = \frac{V}{bK}$$

and re-expressing time in units of the generation time for lytic replication, $t = \alpha bKT$, we can

reduce the system to a dimensionless form,

$$
\begin{aligned}
\frac{\mathrm{d}h}{\mathrm{d}t} &= B(1 - h - l)h - vh \\
\frac{\mathrm{d}l}{\mathrm{d}t} &= B(1 - h - l)l + pvh - Al \\
\frac{\mathrm{d}v}{\mathrm{d}t} &= Al + (1 - p)vh - Cv \ .
\end{aligned}
\qquad (3.2)
$$

Here, we have relabelled various parameter groupings by,

$$A = \frac{\eta}{\alpha b K} \, , \qquad B = \frac{r}{\alpha b K} \, , \quad \text{and} \quad C = \frac{c}{\alpha b K} \, .$$

This reduction reveals that the underlying dynamical system has only four free parameters: the induction (or activation) rate, A; the bacterial growth rate, B; the free virus clearance rate, C; and the probability of lysogeny, p, where $p \in (0, 1)$ (as stated above). In addition to the condition on p, A, B and C are all positive parameters due to the positivity of parameters in System (3.1).

3.2   Preliminaries

3.2.1   Initial Condition and Positive Invariance

The initial condition, $\mathbf{x}_0 = (h(t_0), l(t_0), v(t_0))$, for System (3.2) may be any point in the non-negative octant of $\mathbb{R}^3$, i.e.,

$$\mathbf{x}_0 = \left\{ (h, l, v) \in \mathbb{R}^3 \,\middle|\, h \geq 0 \, , l \geq 0 \, , v \geq 0 \right\} \, . \tag{3.3}$$

Our argument of positive invariance for System (3.2) follows suit with the arguments of Beretta and Kuang [11]. Although the model developed in [11] focuses on strictly lytic phage infecting bacteria, the analysis can be applied to our model.

To prove positive invariance we will begin by rewriting System (3.2) in vector form by setting,

$$\mathbf{x} = [h \, , l \, , v]^T \in \mathbb{R}^3$$

and

$$
\mathbf{F}(\mathbf{x}) =
\begin{bmatrix}
F_1(\mathbf{x}) \\
F_2(\mathbf{x}) \\
F_3(\mathbf{x})
\end{bmatrix}
=
\begin{bmatrix}
B(1 - h - l)h - vh \\
B(1 - h - l)l + pvh - Al \\
Al + (1 - p)vh - Cv
\end{bmatrix}
\tag{3.4}
$$

where $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ and the vector function $\mathbf{F}$ is continuously differentiable. We may then simply write System (3.2) as

$$
\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) \,,
\tag{3.5}
$$

with initial condition in the Set (3.3). Suppose we choose $\mathbf{x}$ in the non-negative octant of $\mathbb{R}^3$ such that $x_i = 0$, then $F_i(x)|_{x_i=0} \geq 0$ for $i = 1, 2, 3$. It therefore follows from the standard existence and uniqueness theory for ordinary differential equations that the face $h = 0$ is invariant and for positive initial conditions the other faces repel into the interior.

3.2.2 Equilibria

Let us determine the equilibria of System (3.2). It is straight-forward to see that, for all parameter values there exists a vanishing equilibrium, $E_0 = (0, 0, 0)$. Furthermore, there exists a virus free equilibrium (or host only equilibrium) at $E_H = (1, 0, 0)$, where the index H stands for "host only". This system exhibits a third boundary equilibrium, $E_{VL} = (0, \hat{l}, (A/C)\hat{l})$, where $\hat{l} = (B - A)/B$; as with our previous equilibrium, we have indexed this equilibrium value with VL to indicate "only viral and lysogen densities". Finally, all three populations are able to co-exist at the equilibrium,

$$
E_\star = (h^*, l^*, v^*) = \left(C - l^*, \frac{Cpv^*}{A - (1 - p)v^*}, B(1 - C) \right) .
\tag{3.6}
$$

We note here that while the sum $h^* + l^*$ is fixed at the clearance rate of virus, C, h^* is a decreasing function while l^* is an increasing function of p. Biologically, we interpret this as meaning there are more lysogens and fewer susceptible host cells if the probability of lysogeny is increased.

Although E_0 and E_H are always present as equilibria for System (3.2), meaning no conditions on parameters are needed for existence, both E_{VL} and $E_\star$ have inherent conditions for existence. The "only viral and lysogen densities" equilibrium, E_{VL}, exists only if the induction rate A is less than the growth rate B. By fixing B, as A approaches and passes through B from below, we see that E_{VL} approaches and crosses through E_0. This intersection of E_{VL} with E_0 is shown numerically in Figure 3.1.
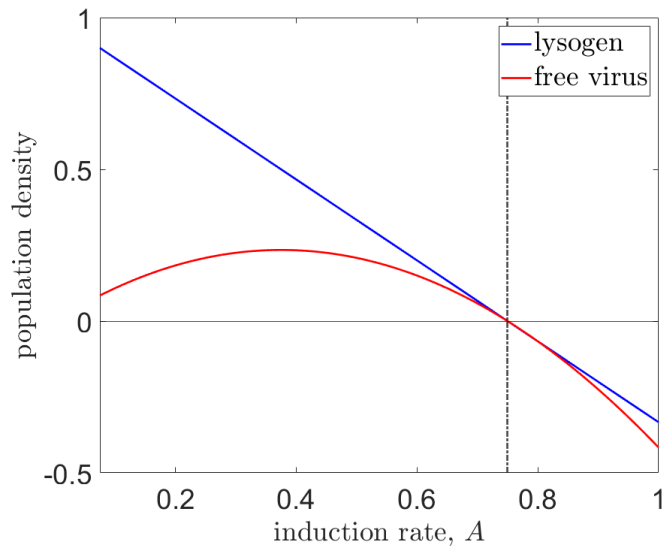
Figure 3.1: The biological feasibility of E_{VL} for System (3.2) as the induction rate, A, increases. As A increases and passes through B (plotted as a dash-dot line), we see both the lysogen and free virus populations become negative. Parameters are set at $B = 0.75$, $C = 0.8$ and $p = 0.5$.

For the coexistence equilibrium, $E_\star$, to be biologically feasible, several conditions must hold. The clearest condition is that on the clearance rate of the virus, C. If $C > 1$, then the

viral component of the equilibrium, v^*, would be less than zero. As C approaches and passes through one from below, we see $E_\star$ approaching and crossing through E_H. In addition to ensure positivity of all components of $E_\star$, we must also ensure that $A > (1 - p)B(1 - C)$; if this does not hold, then the denominator of l^* would become negative. Finally, to ensure h^* is positive, we must have the stronger condition, $A > B(1 - C)$. Thus, as A approaches and passes through $B(1 - C)$ from above, $E_\star$ approaches and passes through E_{VL}, as shown in Figure 3.2. This condition will be of even greater importance during our discussion of local stability of the equilibria.
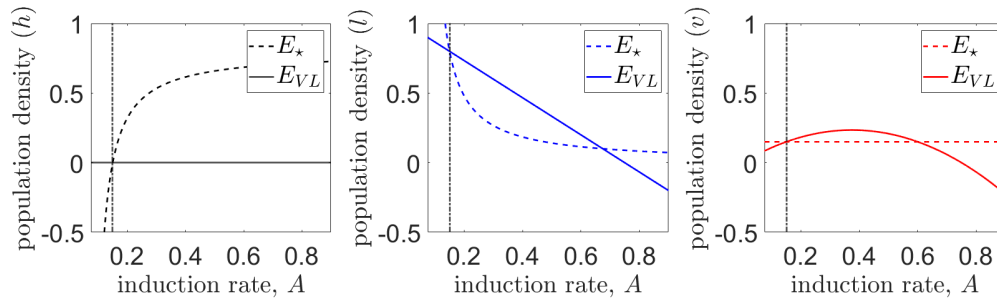
Figure 3.2: The biological feasibility of $E_\star$ and the crossing of $E_\star$ and E_{VL} at $A = B(1 - C)$ (dash-dot line), as the induction rate, A, decreases. As A decreases and passes through $B(1 - C)$, we see that $E_\star$ (dashed line), intersects with E_{VL}, (solid line). Consequently, the host equilibrium value of $E_\star$ becomes negative. The three panels show the host, lysogeny and free virus population densities from left to right. Parameters are set at $B = 0.75$, $C = 0.8$ and $p = 0.5$.

In summary, for the coexistence equilibrium, $E_\star$, two conditions are sufficient. First, we must have $C < 1$. Otherwise, the viral and lysogen populations would lose positivity. Second, we need $A > B(1 - C)$, otherwise the host population density will be negative. Based on these results, we can select A as a biologically relevant parameter on which the equilibrium

behaviour of System (3.2) depends.

To consolidate the ideas of existence of equilibria, let us denote

$$A^* = B(1 - C).$$

We may then summarize the results of this analysis in Proposition 3.2.1 and graphically in Figure 3.3. The phrasing that each equilibria exists in Figure 3.3 means that each equilibria is biologically feasible. This critical value of the induction rate will play a similarly important role when discussing the local stability of E_{VL} and $E_\star$ in Section 3.3.

Proposition 3.2.1. *For all positive values of A, B, C in System (3.2), the equilibria E_0 and E_H exist. If $C > 1$ and $A > B$, then these are the only two equilibria that exist. If $C > 1$ and $A < B$, then E_{VL} also exists. If $C < 1$ and $A < A^*$, then these are the only three equilibria that exist once again. If $C < 1$ and $A > B$, then E_{VL} no longer exists but $E_\star$ exists. Finally, if $C < 1$ and $A \in (A^*, B)$, all four equilibria exist.*

3.3   Stability of the equilibiria

Using local stability techniques as described in [142], we begin by finding the Jacobian matrix. The Jacobian matrix for the system of equations (3.2) is given by,

$$J(h, l, v) = \begin{bmatrix} B(1 - 2h - l) - v & -Bh & -h \\ pv - Bl & B(1 - h - 2l) - A & ph \\ (1 - p)v & A & (1 - p)h - C \end{bmatrix}. \tag{3.7}$$
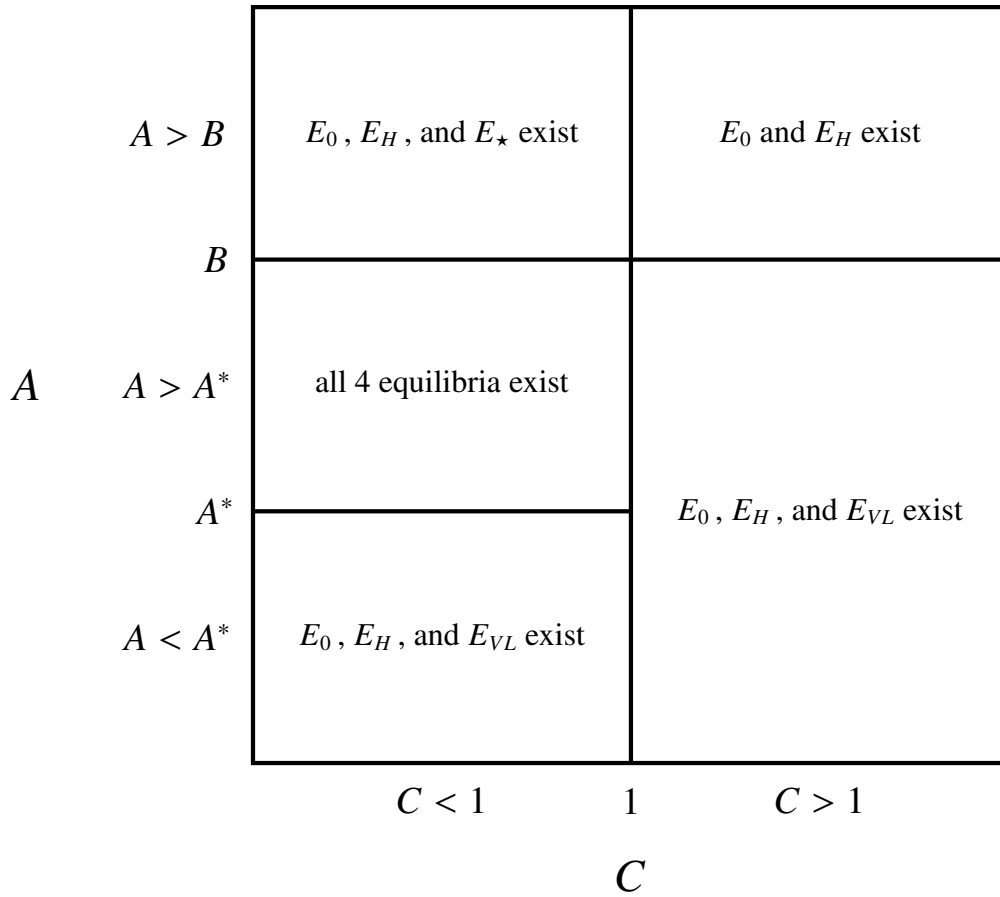
Figure 3.3: Summary of the existence of the equilibria of System (3.2). The parameter space can be divided into regions in which the clearance rate, C, is greater or less than one, while the induction rate, A, exhibits two critical points $A = A^*$ and $A = B$. Note that when $C < 1$, $A^* < B$ since $A^* = B(1 - C)$.

At the vanishing equilibrium, E_0, we have,

$$
J(0,0,0) =
\begin{bmatrix}
B & 0 & 0 \\
0 & B - A & 0 \\
0 & A & -C
\end{bmatrix}
$$

Due to the structure of $J(0, 0, 0)$ as a lower triangular matrix, we may quickly identify its three

eigenvalues as

$$\lambda_1 = B, \quad \lambda_2 = B - A, \quad \text{and} \quad \lambda_3 = -C.$$

If we wish to ensure the stability of E_0 (i.e., all eigenvalues of $J(0, 0, 0)$ must be negative), we must have $B < 0$. Biologically, this would mean the growth rate of the host and lysogen population is negative, which is not realistic. We can therefore conclude, for biologically relevant parameters (i.e., all positive and non-zero), that E_0 is always unstable.

Consider now the "host only" equilibrium, $E_H = (1, 0, 0)$. From Equation (3.7), we have

$$J(1, 0, 0) = \begin{bmatrix} -B & -B & -1 \\ 0 & -A & p \\ 0 & A & 1 - p - C \end{bmatrix},$$

whose eigenvalues, Λ_i, for $i = 1, 2, 3$, are the roots of

$$(\Lambda + B)(\Lambda^2 + (A + C + p - 1)\Lambda + A(C - 1)) = 0.$$

Clearly, one eigenvalue is always real and negative, denoted as $\Lambda_1 = -B < 0$. The other two eigenvalues are the roots of the quadratic equation,

$$\Lambda^2 + (A + C + p - 1)\Lambda + A(C - 1) = 0.$$

To ensure that both roots of this quadratic have negative real parts, that is all three eigenvalues have negative real parts, we must have the coefficients all be the same sign, namely positive (as the coefficient of Λ^2 is $1 > 0$) [24]. This means we must ensure that $A + C + p - 1 > 0$ and $A(C - 1) > 0$. Looking at the second condition, $A(C - 1) > 0$, and knowing that $A > 0$, we can deduce that $C - 1 > 0$ or $C > 1$. If this holds the former condition will also hold. Therefore, we conclude that a necessary condition for local stability of E_H is that $C > 1$. In fact, through the

utilization of Lyapunov function theory, we can show global asymptotic stability of E_H when

$C > 1$.

Theorem 3.3.1. *The equilibrium $E_H = (1, 0, 0)$ is globally asymptotically stable when $C > 1$.*

Proof. Set $\mathbf{x}_H = (1, 0, 0)$ and

$$G(h, l, v) = \ln(h) - h - l - v.$$

Clearly, $G(h, l, v)$ is not positive definite in a neighbourhood around $\mathbf{x}_H$. However, we build the

function $V(h, l, v)$ as

$$V(h, l, v) = G(\mathbf{x}_H) - G(h, l, v) = h + l + v - \ln(h) - 1, \tag{3.8}$$

then we satisfy the two necessary conditions for $V(h, l, v)$ to be a proposed Lyapunov function.

That is to say that $V(\mathbf{x}_H) = 0$ and that $V(h, l, v)$ is positive definite in a neighbourhood about

this equilibrium.

Taking the derivative of Equation (3.8) with respect to t, we find

$$\frac{dV}{dt} = \frac{dh}{dt} + \frac{dl}{dt} + \frac{dv}{dt} - \frac{1}{h}\frac{dh}{dt}. \tag{3.9}$$

By evaluating this derivative at the values given in System (3.2) and doing some simplification,

we find that

$$\frac{dV}{dt} = v(1 - C) - B(h + l - 1)^2.$$

We are aware that $B > 0$ from our original model construction, therefore the second term is

strictly less than zero, for all values of h and l. Thus, if $C > 1$ then $\dot{V} < 0$ for all $(h, l, v) \in$

$\mathbb{R}^3 \setminus \{\mathbf{x}_H\}$. Hence, Equation (3.8) is a Lyapunov function, and we can conclude that $\mathbf{x}_H =$

$(1, 0, 0)$ is globally asymptotically stable when $C > 1$.   □

Biologically speaking, the condition on global asymptotic stability of E_H corresponds to a large clearance rate of the virus. Consequentially, if we remove the virus at a high rate (namely greater than it is produced via induction or infection), we would see the virus and, in turn, the lysogens being removed completely from the population, even if the growth rate of the lysogens was large.

Next, we will consider the last boundary equilibrium, E_{VL}. Again, from Equation (3.7), we have

$$J(0, \hat{l}, (A/C)\hat{l}) = \begin{bmatrix} (A/BC)(A - B(1 - C)) & 0 & 0 \\ (1/C)(Ap - BC)\hat{l} & A - B & 0 \\ (A/C)(1 - p)\hat{l} & A & -C \end{bmatrix},$$

where $\hat{l} = (B - A)/B$. We may utilize the structure of this matrix, as a lower triangular matrix, to quickly identify its three eigenvalues as

$$\lambda_1 = -C, \quad \lambda_2 = A - B, \quad \text{and} \quad \lambda_3 = \frac{A}{BC}(A - B(1 - C)).$$

We note that due to the positivity of the parameters of System (3.2), and the condition for existence of E_{VL}, namely $A < B$, both λ_1 and λ_2 will always be negative when E_{VL} exists. Thus, it suffices to show the conditions on λ_3 to prove stability of E_{VL}.

We will discuss conditions necessary for $\lambda_3 < 0$ in three cases: $C > 1$, $C < 1$ and $C = 1$. Before considering the three cases, we should note that $A/(BC)$ is always positive, thus we can simplify our discussion on the conditions for negativity of λ_3 to the sign of $A - B(1 - C)$.

When $C > 1$, it is clear that $1 - C < 0$ and, in turn, $B(1 - C) < 0$. Thus, $A - B(1 - C) > 0$ so long as $C > 1$. Therefore, if E_{VL} exists and $C > 1$, E_{VL} is unstable. It is clear to show all the previous statements hold if $C = 1$, thus we can state in definite terms that if $C \geq 1$, E_{VL} is unstable.

The final case we must consider is when $C < 1$. Here, we notice that $B(1-C) > 0$, therefore to ensure negativity of λ_3, we must have $A < B(1 - C)$, in other words, we must satisfy the inequality $A < A^*$. Therefore, we conclude that if $C \geq 1$, E_{VL} is unstable, with the caveat that at $C = 1$, E_{VL} is non-hyperbolic, and if $C < 1$ then $A < A^*$ is needed for local stability.

Our final discussion on stability will revolve around the coexistence equilibrium, $E_\star$. Although we wish to undergo a similar analytical approach to determining the stability criteria of $E_\star$, the eigenvalues of the Jacobian evaluated at $E_\star$ are not tractable. We can, however, numerically observe that $E_\star$ is stable wherever it exists. That is, $E_\star$ appears to be stable provided $C < 1$ and $A > A^*$. This numerical stability is demonstrated in Figure 3.4. This switch of stability at $A = A^*$ suggests that a transcritical bifurcation occurs between E_{VL} and $E_\star$.
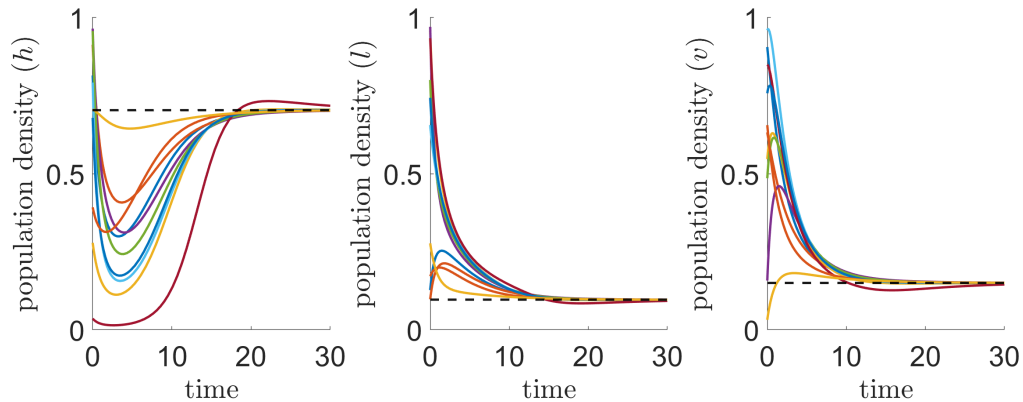
Figure 3.4: Numerical stability of $E_\star$ when $C < 1$ and $A > A^*$. Plotted are numerically integrated results for $h(t)$, $l(t)$, and $v(t)$ using several different initial conditions, all demonstrating that as t tends to infinity the population structure tends towards $E_\star$. Dashed lines in each panel show h^*, l^* and v^* (as given in Equation (3.6)), respectively. Parameters are set at $A = 0.7$, $B = 0.75$, $C = 0.8$, and $p = 0.5$.

We summarize our above discussion in the bifuraction diagrams shown in Figure 3.5,

Proposition 3.3.2 and Conjecture 3.3.3.

Proposition 3.3.2. *Assume all parameters of System (3.2) are strictly positive for biological feasibility. For all biologically feasible parameters, $E_0 = (0,0,0)$ is unstable. If $C > 1$, then $E_H = (1,0,0)$ is globally asymptotically stable, and E_{VL} is unstable; the coexistence equilibrium, $E_\star$ does not exist in this regime. If $C < 1$, E_H is always unstable, and if $A < A^*$, E_{VL} is locally stable and $E_\star$ is unstable.*

Conjecture 3.3.3. *If $C < 1$ and $A > A^*$, then $E_\star$ is locally stable and E_{VL} is unstable. At $A = A^*$ a transcritical bifurcation occurs with a switch of stability from E_{VL} to $E_\star$.*

(a) host equilibrium   (b) lysogen equilibrium   (c) virus equilibrium
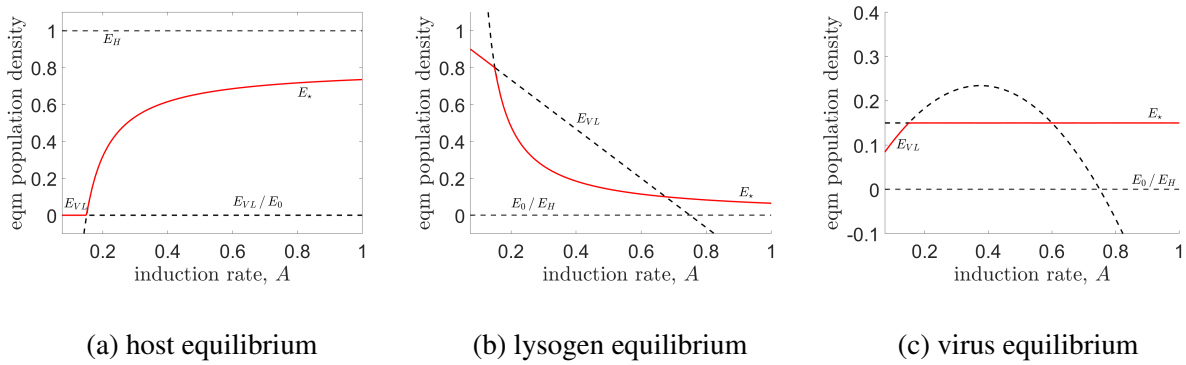
Figure 3.5: Bifurcation diagram for System (3.2) using induction rate, A, as the bifurcation parameter. Solid red lines indicate the stable equilibrium for the system, while black dashed lines indicate an unstable equilibrium. Panel (a) shows the bifurcation diagram for the host population, where E_0 and E_{VL} are indistinguishable. Panel (b) shows the bifurcation diagram for the lysogen population, where E_0 and E_H are indistinguishable. Panel (c) shows the bifurcation diagram for the free virus population, once again where E_0 and E_H are indistinguishable. All diagrams are plotted on the interval $A \in [(1 - p)B(1 - C), 1]$, as all equilibria are biologically feasible at some interval of this range. Parameters are set at $B = 0.75$, $C = 0.8$ and $p = 0.5$.

3.4   Discussion

The mathematical model we propose in this chapter, although simple, presents an interesting

problem when considering the analysis of its equilibria and their associated stability. Here,

we determine a variety of conditions to ensure the existence of all three populations in the

environment. Although four free parameters were revealed to be present when making System

(3.1) dimensionless, two of these are clearly most important on the existence and stability of

equilibria: the viral clearance rate, C and the induction rate, A.

When the viral clearance rate is large, in particular when the dimensionless clearance rate

exceeds unity, the non-dimensional System (3.2) yields three biologically feasible equilibria.

In this same parameter regime, where $C > 1$ and all other parameters are greater than zero, we

show that E_H, the "host only" equilibrium, is not only locally but globally asymptotically stable

(see the proof of Theorem 3.3.1). This makes biological sense, since if the virus is removed at

a fast rate, then viral particles would have little to no chance in infecting the host to produce

lysogens or more free virus.

The induction rate also presents itself as an important parameter in our model. We find

many conditions related to the value of the induction rate to ensure the biological feasibility

of many of the equilibria, as well as their associated stability. We determine that for all four

equilibria of System (3.2) to exist, A must be bounded below by $A^* = B(1 - C)$ and above at

the growth rate of the bacteria, B. Below the lower bound, $E_\star$ becomes negative, losing its

biological feasibility, and above the upper bound, E_{VL} becomes negative.

To summarize our discussion of stability, we expand what is already displayed in Figure

3.3 to now include which equilibria are stable in each regime. This is shown in Figure 3.6. We

notice that A^* not only plays a pivotal role in the existence of the equilibria but also in their

stability. In addition, when $C < 1$, we observe an exchange of stability at $A = A^*$ between

E_{VL} and $E_\star$. For values of A below this bifurcation point, we can analytically show that E_{VL}

is locally stable, which our numerical results confirm. For values above this bifurcation point,

however, we can only show numerically that $E_\star$ is locally stable. We note that this conjecture

regarding the stability of $E_\star$ has been shown to hold [206].

		$C < 1$	$C > 1$
	$A > B$	E_0, E_H, and $E_\star$ exist $E_\star$ stable *(numerical only)*	E_0 and E_H exist E_H stable (GAS)
	$A > A^*$	all 4 equilibria exist $E_\star$ stable *(numerical only)*	
A			E_0, E_H, and E_{VL} exist E_H stable (GAS)
	$A < A^*$	E_0, E_H, and E_{VL} exist E_{VL} stable	

$$C$$
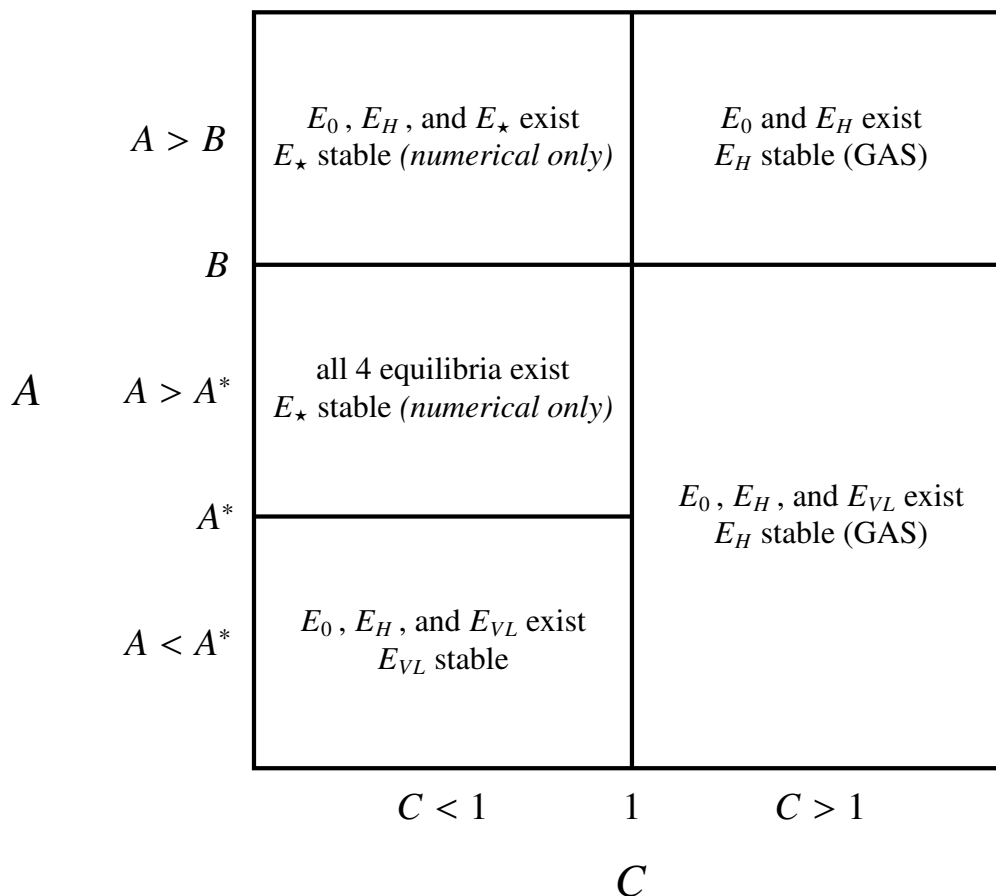
Figure 3.6: An extension of the results shown in Figure 3.3 to now include which equilibrium

is stable in each parameter regime.

Chapter 4

The role of stationary phase in the lysis-lysogeny decision

In this chapter, we extend our previous work to encompass bacteria and phage that exist in a variable environment. Through the development of a chemostat model, we include a time-varying resource compartment to simulate an environment with variable conditions. Through analysis of this model, we reveal the importance of a stationary phase for the bacterial host; stationary phase is a feature of most bacterial populations that may, in fact, play an important role in the lysis-lysogeny decision of an infecting bacteriophage. In this chapter we will demonstrate that, when infecting bacteria with a stationary phase in a time-varying environment, bacteriophage tend to evolve toward complete lysogeny.

4.1   Introduction

As noted many times in previous chapters, bacteriophage have been studied extensively, with their abundance and importance being highlighted as of late [75]. In [183], multiple life-history traits of the host were quantified in an attempt to find correlations between these traits and the prevalence of prophage in the bacterial genome. The strongest correlation was revealed to be between the minimal doubling time of the host and the presence of prophage. Thus, the best predictor of lysogeny seems to be the host's ability to grow quickly when resources are plentiful.

The identified correlation between minimal doubling time and the abundance of prophage seems somewhat counter intuitive. When a bacterial population is experiencing rapid growth, the expected evolutionary payoff for lysogeny would be increased due to fast fission times. The rapid growth of cells would seem to imply, however, an even greater potential payoff for lysis. Why then, across a wide range of measured life-history traits, is the minimal doubling time the strongest empirical predictor of lysogeny? This is the question we explore in this chapter.

Bacteria with smaller minimal doubling times have been identified more commonly in environments with variable conditions [95, 146]. Thus, a correlation between lysogeny and minimal doubling times could arise if lysogeny is also favoured in environments that vary between extreme feast and famine periods. This connection between varying between high and low resource periods due to seasonal variation and the presence of lysogeny was explored within coastal lagoons [122]. Maurice *et al.* found a strong link to lysis and environmental conditions, but found that lysogeny did not seem to be affected to the same degree [122]. This runs counter to the hypothesis posed by Touchon *et al.*, and creates an opportunity for us to identify under

what conditions lysogeny may indeed be favoured [122, 183].

When considering an environment where the bacterial population experiences shifts between feast and famine periods, it would be negligent to ignore the effect of the stationary phase. The bacterial stationary phase, is a phase of bacterial population growth in which the bacteria become non-active and no cellular division occurs (Figure 4.1 shows a general growth time course for a bacterial population with a clear stationary phase) [29, 96, 138]. Often this phase of bacterial growth is triggered by an extreme limitation of resources, and is thought to assist in the long-term survival of the bacteria [59]. Although growth is inhibited while in stationary phase, the bacteria also become immune to most bacteriophage infections [28, 163]. In addition, stationary phase has been suggested to play an important role in bacteria-phage co-evolution, and, as such, should be considered as a possible factor in the lysis-lysogeny decision of temperate bacteriophage [63, 161, 165].

Here, we adapt the methodology of our previous work [190] to investigate the long-term outcome of the evolution of the probability of lysogeny. In particular, we test whether environmental variability might explain the observed correlation between lysogeny and minimal doubling times. Although a multitude of other factors may be present in their usual environment, our results predict that those phage that infect bacterial hosts with a distinct stationary phase are more likely to evolve a high probability of lysogeny in highly variable environments. In fact, as we observed previously, full lysogeny (that is, a phage population that never initiates immediate lysis) is the predicted evolutionary outcome. In contrast, when infecting a bacterial population with no stationary phase in a highly variable environment, the phage will evolve to an intermediary state, where lysis and lysogeny are both able to occur.
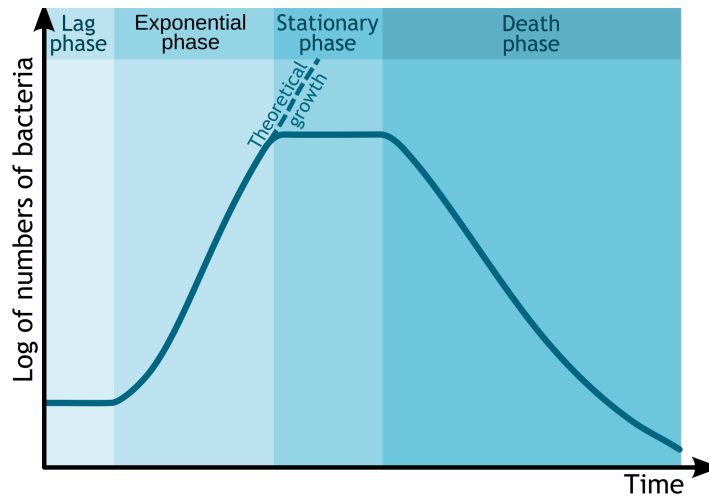
Figure 4.1: General growth curve for a bacterial population. The lag phase is characterised by the bacteria becoming accustomed to their environment, during which they may not replicate but grow in volume. The log phase is when the bacteria exhibit exponential growth, eventually reaching maximal growth, which can be used to estimate the strain's minimal doubling time. The bacteria then enter stationary phase, at which point the bacterial population no longer grows, possibly due to an exhaustion of resources. Finally, the population enters a death phase, during which the number of cells dying exceeds the new progeny produced through cellular fission [193].

4.2 Model

We begin by adopting a chemostat modelling approach, adapted from [176], to analyze the effect of a changing environment, controlled by the forcing of resource in the system (see [172] for a discussion of this approach) and the presence of a bacterial stationary phase. Here, bacterial hosts are able to move between an active state, H, and a stationary state, $\widetilde{H}$. Similarly, we have lysogens (hosts with integrated viral DNA) that can also move between an active, L and stationary, $\widetilde{L}$, state. The active hosts and lysogens grow according to a Michaelis-Menton

resource uptake growth rate, $\psi(R)$, first described by Monod [133] and utilized in a vast number of similar models ([73, 84, 173, 181], for instance). In addition, the hosts and lysogens may become stationary at a rate, $\sigma(R)$, that exponentially decays with the resource concentration, R, meaning that at low resource levels the active cells become stationary more quickly [158]. Stationary cells become active once again at a per capita rate, w. Both hosts and lysogens are lost or cleared at a rate ρ_B, while the lysogens are induced at a rate ξ, producing β new viral particles. Finally, we assume that the lysogens cannot be re-infected, as prophages often confer immunity to superinfection by the same phage [2, 185].

Free virus, V, adsorbs into the active hosts using mass-action dynamics at a rate α. Once the virus has adsorbed, the virus utilises a lysogenic strategy with probability p, and with probability $1 - p$ utilises a lytic strategy, producing β new viral particles, as assumed in System (3.1). The free virus is cleared, lost or denatured at a rate κ. We assume that the viruses are unable to infect host cells in stationary phase.

The resource flows in and out of the system at constant rate ρ. However, the input flow is given by a sinusoidal forcing term, $C(t)$, to simulate a changing environment. In addition, the bacteria and lysogens use up the resource with yield ε.

These assumptions yield a system of ordinary differential equations:

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \psi(R)H - \sigma(R)H - \alpha HV + w\widetilde{H} - \rho_B H$$

$$\frac{\mathrm{d}L}{\mathrm{d}t} = \psi(R)L - \sigma(R)L + \alpha pHV + w\widetilde{L} - (\xi + \rho_B)L$$

$$\frac{\mathrm{d}\widetilde{H}}{\mathrm{d}t} = \sigma(R)H - w\widetilde{H}$$

$$\frac{\mathrm{d}\widetilde{L}}{\mathrm{d}t} = \sigma(R)L - w\widetilde{L} \tag{4.1}$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \xi\beta L + \alpha(1 - p)\beta HV - \kappa V$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \rho(C(t) - R) - \varepsilon\psi(R)(H + L) \ ,$$

where

$$\psi(R) = \frac{mR}{a + R} \ , \quad \sigma(R) = \zeta e^{-\theta R} \quad \text{and} \quad C(t) = R^\star (1 + C_0 \sin(\omega t)) \ .$$

Although it is possible to scale some parameters of a non-autonomous system to be dimensionless, we take a fully numerical approach in the following analysis, and, as such, find it unnecessary to make System (4.1) dimensionless. Dimensional parameters are given in Table A.1 in Appendix A.

Here, we look to compare results with and without the inclusion of the bacterial stationary phase. Often, this phase is neglected in mathematical models of phages and their hosts [176, 190]. Figure 4.2 shows time courses of System (4.1) when stationary phase is "turned on" and "turned off"; where to "turn off" the host's and lysogen's ability to enter the stationary phase, we simply set $\zeta = 0$ in $\sigma(R)$. We see a stark qualitative difference between the two time courses.
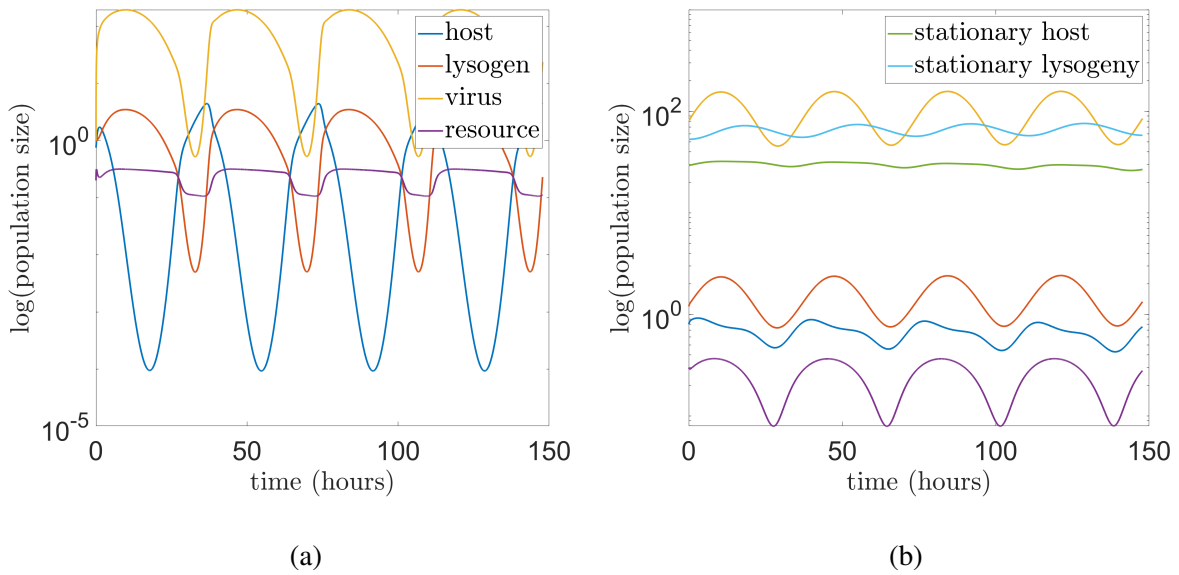
Figure 4.2: Time course of System (4.1). Panel (a) shows the time course without stationary phase (i.e., $\zeta = 0$) and panel (b) shows the time course with stationary phase (i.e., $\zeta = 2.5$). Parameter values are given in Table A.1.

4.3 Evolutionary Invasion Analysis

To determine the long-term trend of the probability of lysogeny in the phage population, we utilise a pairwise invasion analysis. Here, we introduce a rare mutant virus, only differing from the wild-type by the probability of lysogeny. To do so, we add three additional compartments

to System (4.1):

$$\frac{dH}{dt} = \psi(R)H - \sigma(R)H - \alpha H(V_r + V_\mu) + w\widetilde{H} - \rho_B H$$

$$\frac{dL_r}{dt} = \psi(R)L_r - \sigma(R)L_r + \alpha p_r H V_r + w\widetilde{L}_r - (\xi + \rho_B)L_r$$

$$\frac{d\widetilde{H}}{dt} = \sigma(R)H - w\widetilde{H}$$

$$\frac{d\widetilde{L}_r}{dt} = \sigma(R)L_r - w\widetilde{L}_r$$

$$\frac{dV_r}{dt} = \xi\beta L_r + \alpha(1 - p_r)\beta H V_r - \kappa V_r \qquad\qquad (4.2)$$

$$\frac{dR}{dt} = \rho(C(t) - R) - \varepsilon\psi(R)(H + L_r + L_\mu)$$

$$\frac{dL_\mu}{dt} = \psi(R)L_\mu - \sigma(R)L_\mu + \alpha p_\mu H V_\mu + w\widetilde{L}_\mu - (\xi + \rho_B)L_\mu$$

$$\frac{dV_\mu}{dt} = \xi\beta L_\mu + \alpha(1 - p_\mu)\beta H V_\mu - \kappa V_\mu$$

$$\frac{d\widetilde{L}_\mu}{dt} = \sigma(R)L_\mu - w\widetilde{L}_\mu \ ,$$

where $\psi(R)$, $\sigma(R)$, and $C(t)$ are defined as before. Wildtype strains of the phages and lysogens are denoted with a subscript r (to signify it as the resident strain[1]), and mutant strains are denoted by a subscript μ. We assume that the mutant strain only differs from the resident strain by their probability of lysogeny.

In a traditional pairwise invasibility plot (PIP) the determination whether the mutant can invade or not is often completed through a nonlinear invasion analysis, as described in Section 1.4, commonly through the use of an invasion R_0 (see [86] or [142] for a succinct discussion of the topic). In Section 1.6.1, we outline multiple approaches to calculate the basic reproduction number in models with time-periodic parameters, such as System 4.2. Unfortunately, the

---

[1]A subscript w was not used as to avoid confusion with the transition rate from stationary to active phase of the cells.

assumptions of Wesley and Allen [201] do not hold for System (4.2), and therefore we cannot use a time-averaged approach. Through numerous computational experiments, we found that a direct approach to determine invasion (described below) was just as computationally efficient as Posny and Wang's discretization approach, which was determined to be more efficient than the root-finding approach [147, 160]. This is because to ensure accuracy in a numerical discretization approach the necessary matrices become cumbersome [147]. We therefore proceed with our analysis using a direct approach implemented as follows.

For a given wildtype strain, we first numerically integrate System (4.2) with $L_\mu = V_\mu = 0$ until we reach a periodic solution. To ensure that we have indeed reached a periodic solution we check the condition $|H(t_N) - H(t_N + T)| < \gamma$, where t_N is some sufficiently large time, $T = 2\pi/\omega$ (i.e., the period of $C(t)$) and γ is some small threshold. After determining that we have reached a periodic solution, we seed in a mutant virus and lysogen at at ratio of 10^{-10} of the resident virus and lysogen population, respectively. We then numerically integrate the entire system for approximately 25 more periods of time (corresponding to approximately 25 more days). In order to determine if the mutant strain has successfully invaded the wildtype strain, we only consider the virus population; however, the results would be identical if we considered the lysogeny population. We record the average of the virus population over one period when initially seeded into the system, and one period from the end of our numerical integration. We then check the slope created from these two averages. If said slope is positive (negative), the mutant is recorded as successfully (unsuccessfully) invading the wildtype population. If the absolute value of the slope is within some small tolerance, then we record coexistence between the mutant and wildtype strains. Figure 4.3 shows an example of successful and unsuccessful invasion of the mutant strain. We then plot the results of this competition PIP and determine the

critical probability of lysogeny (corresponding to what is known as an "evolutionarily stable strategy").

In the majority of our results, especially when stationary phase is included, we observe an unusual structure in the pairwise invasibility plots; a substantial coexistence region is observed that separates regions wherein the mutant fails to invade (an example PIP is shown in Figure 4.4a, while the corresponding coexistence plot is shown in Figure 4.4b). In this case, we identify an upper and lower bound on the critical probability of lysogeny, between which we could expect complex but constrained dynamics as described more fully in the sections to follow. Figure 4.5 shows the identification of such bounds on the PIP shown in Figure 4.3a.

The interpretation of what constitutes a variable environment is an issue that we address by observing the effects of changing three properties of the input function, $C(t)$. We observe the effect of changing the amplitude, C_0, the frequency, ω, and the average value of the input forcing, $R^\star$. Results for each case are displayed in Figures 4.6, 4.7, and 4.8. It is worth noting that we vary each parameter individually, fixing the other two. A scenario in which more than one parameter is allowed to change is an avenue for future work.

First, we will focus our attention on the effect of changing the amplitude of the forcing function (see Figure 4.6); note that we restrict $C_0 \in (0, 1)$ as any greater value would cause the resource input function to produce negative resource. In this case, we notice a stark difference when stationary phase is included. When no stationary phase is available to the host and lysogen populations, we observe that the critical probability of lysogeny, in fact, decreases as the amplitude increases (shown in Figure 4.6a). This suggests that in a variable environment, where fast-growing bacteria are indeed more prevalent, we should expect less integrated phage DNA, contradictory to the findings of Touchon *et al* [183]. However, when stationary phase is

(a) successful invasion

(b) unsuccessful invasion
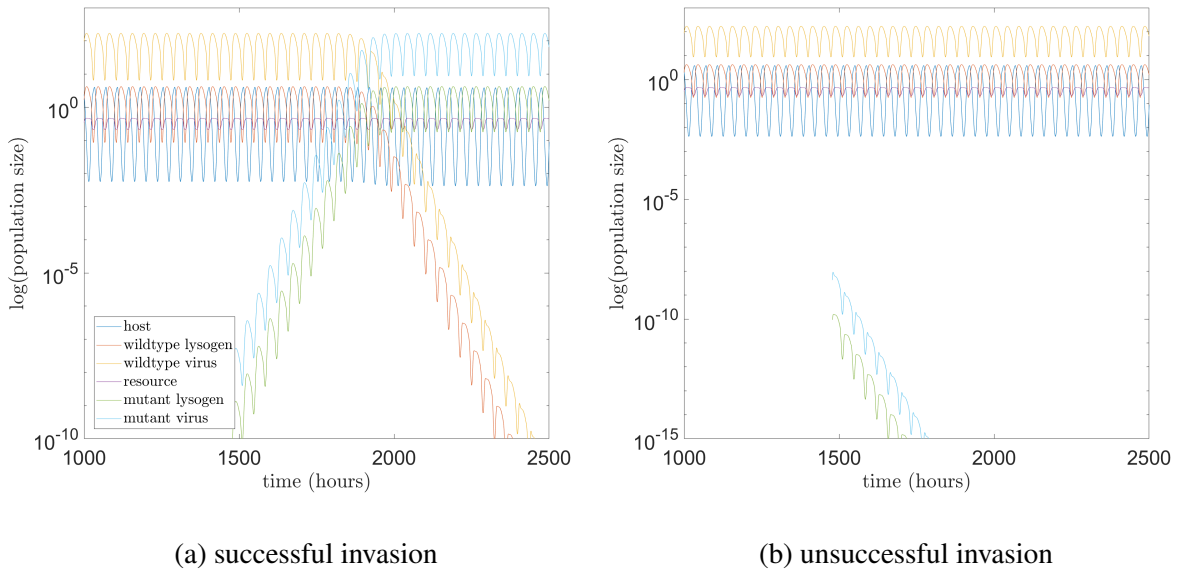
Figure 4.3: Numerical simulation of System (4.2) to demonstrate the competition time courses between a rare mutant and established wild-type phage species. System (4.2) was integrated numerically for 2500 days, with a rare mutant seeded in at $t \approx 1500$ days. The rare mutant lysogen and virus populations were seeded in at a value 10^{-10} times the corresponding wildtype population values at the seeding time. Panel (a) shows a successful mutant invasion, while panel (b) shows a unsuccessful mutant invasion. Parameter values are given in Table A.1 with the exception of p_μ (note that p_r is set at the p value given in the table). In panel (a) $p_\mu = 0.8$, and in panel (b) $p_\mu = 0.2$.

available to the hosts, we see that as the amplitude increases so too does the critical probability of lysogeny. We, in fact, see that at $C_0 \approx 0.65$, the population should evolve to a fully lysogenic population (i.e., $p = 1$). It should be noted that there is an intermediate stage between a population that has both lysis and lysogeny and one that is fully lysogenic. We observe a population where there exists the possibility of two phenotypically different phage coexisting

(a)   (b)
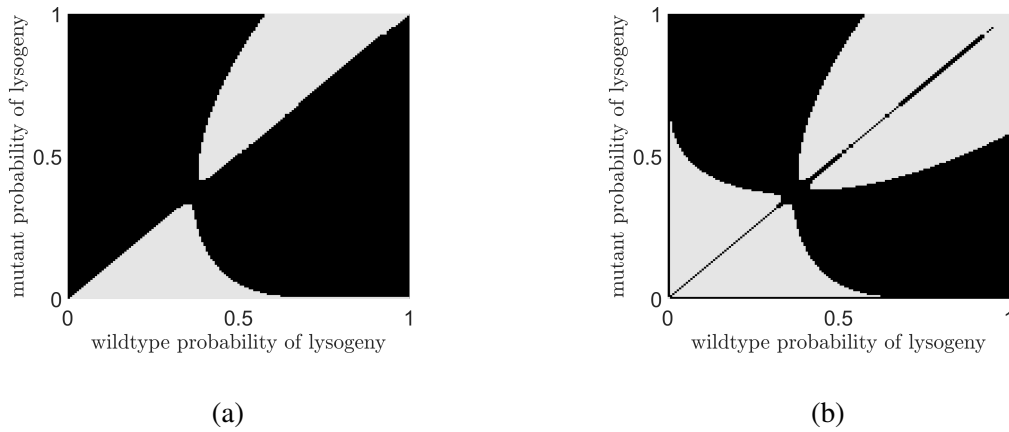
Figure 4.4: Panel (a) shows an example of a pairwise invasibility plot for a resident wildtype phage competing against an invading mutant phage that only differs from the resident in the probability of lysogeny, p. Black identifies competitions where the mutant invades and takes over, while grey identifies competitions where a mutant does not invade. Panel (b) shows the coexistence plot for the PIP in panel (a). Similarly, black represents regions where a polymorphism of resident and mutant can occur, while grey represents competitions with no possibility of coexistence. We note that the area of invasion separating the two non-invasion areas in (a) are indeed an area of coexistence, where we may expect a polymorphism of two phage species to exist.

(for example, a wild-type phage with $p = 0.1$ can coexist with a mutant with $p = 0.5$). This would occur if we allow large evolutionary steps , that is, single mutations that have large effects on the probability of lysogeny, p. If we allow only small evolutionary steps, however, we predict that over the long term the population would evolve to become fully lysogenic, as we see in the right-most region in Figure 4.6b.

Similar results are observed when we turn our focus to the frequency of the input function, ω, and the average of the input function, $R^\star$; Figures 4.7 and 4.8 display these results,
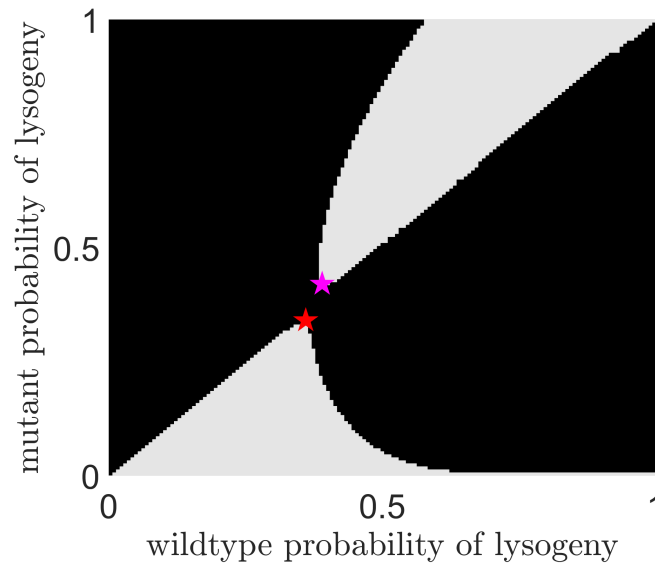
Figure 4.5: Upper and lower bounds for the critical probability of lysogeny shown for the same pairwise invasibility plot shown in Figure 4.4. Lower bound for the critical probability is plotted with a red star, while the upper bound is plotted with a magenta star.

respectively. In contrast to the results from changing the amplitude, we see that when increasing variability through either of these parameters, an increase in the critical probability of lysogeny occurs. This increase is seen when stationary phase is and is not present. In support of our claim of the importance of stationary phase, we do see that the level of variability (either in ω or $R^\star$) at which we expect a fully lysogenic population to singularly evolve decreases when stationary phase is present (note the difference in x-axis in Figures 4.7a and 4.7b).

An interesting feature, not seen in either of the other two cases, occurs when the average value of the input function is increased greatly. After a certain value of $R^\star$, $R_C^\star$, when stationary phase is included, coexistence between the wildtype and any mutant strain is predicted. This suggests that when $R^\star > R_C^\star$, we expect that no matter what level of lysogeny an invading mutant has, coexistence between the mutant and resident is expected. This region is shown in

(a) without stationary phase   (b) with stationary phase

Figure 4.6: The effect of changing the amplitude, C_0, of the forcing term in the resource compartment of System (4.1) on the critical probability of lysogeny. The various forms the pairwise invasibility plots used to determine the critical probability of lysogeny are shown in each region as a cartoon (panel (a) only has one type of PIP). The blue and red lines represent the lower and upper bounds on the critical probability of lysogeny; the purple line (in panel (b)) at $p = 1$ represents a fully lysogenic population.

the shaded panel of Figure 4.8b.

4.4   Discussion and Conclusion

The question regarding temperate bacteriophage, "why be temperate?" has puzzled researchers for some time [176]. Biologists, mathematicians and anything in between seem to be looking to elucidate the reason phage have evolved the ability to become inactive, housed within their "prey" [82, 108, 183, 192, 190]. In this chapter, we examine a novel idea: bacterial stationary phase may play an important role in the evolution of lysogeny. As noted previously, this distinct state of zero growth for the bacterial host is often excluded from models of both bacterial

(a) without stationary phase

(b) with stationary phase

Figure 4.7: The effect of changing the frequency, ω, of the forcing term in the resource compartment of System (4.1) on the critical probability of lysogeny. The various forms of the pairwise invasibility plots used to determine the critical probability of lysogeny are shown in each region as a cartoon. The blue and red lines represent the lower and upper bounds on the critical probability of lysogeny; the purple line at $p = 1$ represents a fully lysogenic population.

population dynamics [149] and bacteria-phage interactions [12, 13, 105, 190, 200].

Here, we develop a model to fill in the gaps of these past studies, where we include a stationary phase for the hosts and lysogens. The inclusion of such a compartment, along with the use of a chemostat model in order to simulate a variable environment has shed new light on the findings by Touchon *et al.* [183], and a possible route for the evolutionary trajectory of bacteriophage.

To simulate a changing environment, we use a simple sinusoidal function as the input to the chemostat; taking inspiration from [84] and [172]. We then allow three parameters of the forcing term to vary, namely, the amplitude, frequency and average of the sinusoidal function. By doing so, we are able to view how each parameter, and in turn variability of some sort, affects the critical or evolutionarily predicted probability of lysogeny of a phage population.

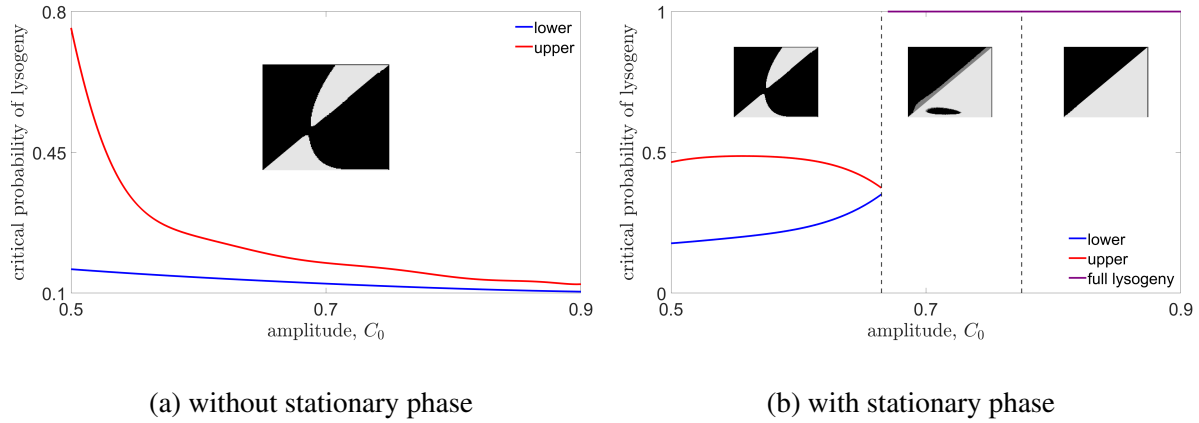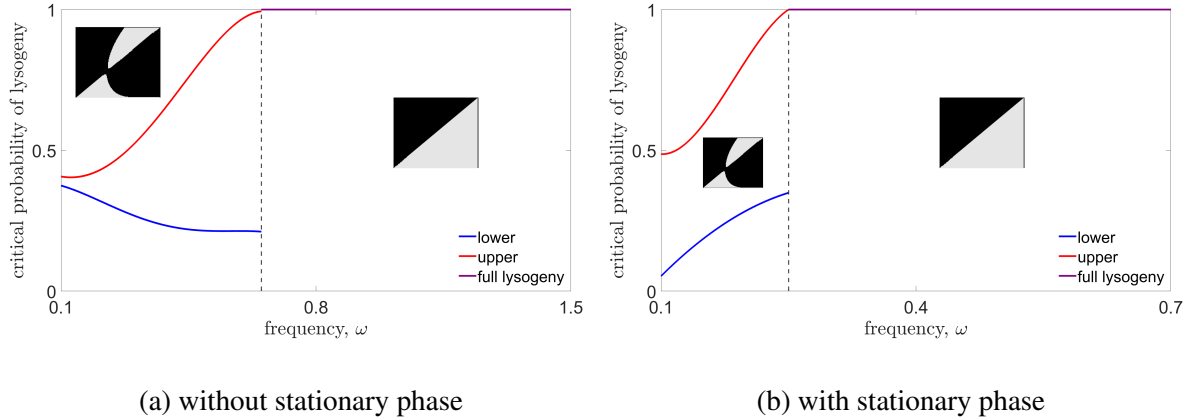(a) without stationary phase　　　(b) with stationary phase
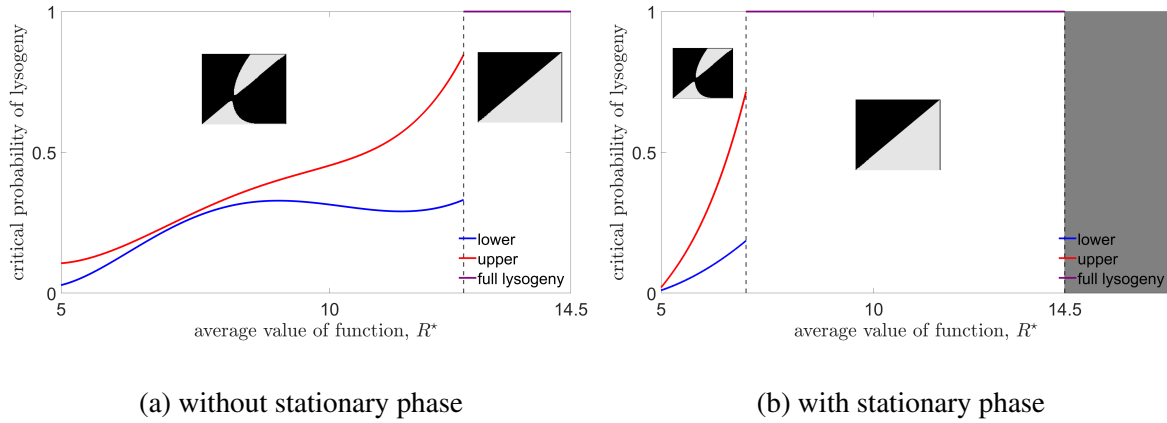
Figure 4.8: The effect of changing the average input, $R^\star$, of the forcing term in the resource compartment of System (4.1) on the critical probability of lysogeny. The various forms of the pairwise invasibility plots used to determine the critical probability of lysogeny are shown in each region as a cartoon. The blue and red lines represent the lower and upper bounds on the critical probability of lysogeny; the purple line at $p = 1$ represents a fully lysogenic population. The shaded, right-most region of Panel (b) indicates an area where a mutant may invade and coexist with the resident regardless of the probability of lysogeny.

To determine such a trajectory for the evolution of the phage, we utilise a pairwise invasion approach, competing a resident phage population against an invading mutant population that only differs in their probability of lyosgeny, p. In our analysis, we often find pairwise invasibility plots (PIPs) that have a distinct coexistence region that separates two regions of invasion (see Figure 4.4). Here, we indicate an upper bound and lower bound for a critical probability, between which we would expect a polymorphism of phage species to be able to coexist. This coincides with an idea presented by Wolkowicz and Zhao [207]. The authors state that in a periodic chemostat with n species, one can expect to oberserve coexistence between multiple species. This result is explained by identifying some species may thrive on high re-

source, while another on low resource (and anything in between). We see this same behaviour in our model; phage species with low probability of lysogeny are better suited for high resource environments, while those with higher probabilities of lysogeny thrive in low resources.

In addition to the common upper/lower bound for coexistence in our results, we also see an interesting feature when we increase the average of the forcing function to a large degree. When the resource is increased greatly, the population does not exhibit a small region of coexistence, but instead we find that any invading mutant may coexist with the established resident population. This result suggests that when resources are overly abundant to the resident, and invading, types the resident population is not affected in any meaningful way by the invading type; competition is completely relaxed. We can therefore conclude that when resources are overly abundant, any two species of varying degrees of probability of lysogeny may coexist with one another.

As variability increases, we see a tendency for the population to evolve toward a fully lysogenic population of temperate bacteriophage when variability is high enough. Biologically speaking, this is not realistic, as the vast majority of temperate phage have the ability to undergo both infection strategies [36, 81, 82]. We posit that this full lysogeny predicted by our model may be biologically consistent with the idea of cryptic prophage. Perhaps, our model suggests that when variability in the environment is high, a phage is more likely to stay integrated than induce and enter into a lytic cycle. Thus, we might expect to observe an abundance of cryptic prophage in fast-growing bacteria, since highly-variable environments favour both fast-growing bacteria and high rates of lysogeny. Although we have not investigated how variability affects induction in this work, this question regarding cryptic prophage is examined more deeply in Chapter 5.

From our results, we see clearly that not only does variability seem to affect the evolutionarily stable strategy for temperate phage and their propensity for lysogeny, but so too does the existence of a distinct stationary phase for the host. This result demonstrates a possible explanation to the findings of Touchon *et al.* [183]. Due to the prevalence of fast-growing bacteria in variable environments [95], we suggest that high levels of variability coupled with the bacterial stationary phase may be the reason for the increased amount of lysogeny found by Touchon *et al.* [183]. Perhaps it is worthwhile in further work on models of bacteria-phage to include the host's stationary phase, as it is clear to us it may indeed play an important role in not only the lysis-lysogeny decision, but the coevolution of host and phage alike.

Chapter 5

The effect of life-history traits on the distribution of prophage length

The results of Chapter 4 regarding a bacteriophage population evolving towards full lysogeny create an additional area of exploration into prophage population dynamics. Classifying bacterial species according to their growth rates, we examine the prophage length distribution in a set of fast- and slow-growing bacterial genomes. We then apply the techniques developed by Khan and Wahl [94] to fit a partial differential equation model to these distributions, in order to elucidate key features of the phage that infect these specific hosts. Our overall goal is to offer explanations for several interesting but unexplained findings of Touchon *et al.* [183].

5.1 Introduction

Prophages can constitute a large portion of a given bacterial genome, on occasion accounting for up to 16% of the host's genome [183]. Recently, investigations into the length distributions

of chromosomal prophages have revealed that these distributions are often multimodal, running contrary to the previous expectation that prophage length distributions would be unimodal with a negative skew [15, 25, 43, 94, 106]. Khan and Wahl [94] examined the underlying lysis-lysogeny processes as an explanation for this multimodal length distribution. Their model predicted large prophages were primarily maintained through lysogeny, while small prophages, that do not contain genes for induction but confer benefits to the host, are primarily maintained through selection [94].

The frequency under which a temperate phage undergoes lysogeny is highly dependent on a multitude of factors, as previously discussed. However, the most common explanation for an increase in lysogeny is that lysogeny may be favoured in conditions where susceptible bacteria populations are at low densities, possibly due to low temperature or resource concentration [42, 66, 123, 127, 148, 167, 190, 203]. In addition, it has been suggested that characteristics of the bacterial host may play a pivotal role in the propensity for a temperate phage to undergo lysogeny, notably the host's growth rate [1, 183] and the pathogenicity of the host [3, 26, 183]. Motivated by this suggestion regarding the host itself playing an important role in the lysis-lysogeny decision, Touchon *et al.* [183] compiled a large data set of hosts and their prophage to reveal any correlations between previously suggested life-history traits of the host and the number of prophages integrated into the genome. As noted in Chapter 4, the most significant association observed was between the frequency of lysogeny and the bacteria's minimal doubling time [183]. As the host's minimal doubling time decreased (i.e., a faster growing host), the frequency of lysogeny increased [183]. In addition, a positive correlation between a host's pathogenicity and the frequency of lysogeny was observed, wherein pathogenic hosts were found to have a much higher frequency of lysogeny [183].

The goal of this chapter is to further investigate the relation between bacterial genetic and life-history traits, specifically host minimal doubling time, and the distribution of prophage lengths. Classifications of bacterial minimal doubling times are taken from the publicly available data of Touchon *et al.* [183], while prophage lengths are gathered using PHAge Search Tool Enhanced Release (PHASTER) [7]. Recall that the results of Chapter 4 suggested an increase in cryptic prophage in fast-growing bacteria. We therefore first examine the fraction of prophages that are intact (functional) or incomplete (cryptic) in fast- and slow-growing hosts. We then apply Khan and Wahl's model [94] in order to elucidate any differences between the strategies of temperate phage infecting fast- and slow-growing bacteria, and how these decisions might influence prophage length distributions.

5.2   Length Distributions

Touchon *et al.* [183] identify a total of 1160 phages in their minimal doubling time data set; splitting the group, they identify 615 found in fast-growing hosts, and 545 in slow-growing hosts. The minimal doubling times (d) used to classify the bacteria were determined through Vieira-Silva and Rocha's experimental work on bacterial species [188]. Bacterial species were classified into fast-growers if their minimal doubling time under optimal conditions was less than 2.5 hours (i.e., $d < 2.5$ h) or slow-growers if $d \geq 2.5$ h [183, 188]. The summary of data for these growth classes is given in Table 5.1. We note that the reduced number of prophages considered in the fast-growing classification is due to a few outliers with length greater than 80 kilobases (kb). These were excluded from our data set as they could not be fit by the model.

We passed the list of accession numbers (unique genome identifiers) for both the fast- and

Bacterial Growth Class	Prophage Number	Min (kb)	Max (kb)	Average (kb)
Fast	609	3.1	72.1	29.883
Slow	545	5.4	72.0	22.869
Total	1154	3.1	72.1	26.571

Table 5.1: Summary of data for growth classes. Data from Touchon *et al.* [183].

slow-growing hosts identified by Touchon *et al.* [183] into the PHASTER web interface [7] to produce results files for each host. We note that not all hosts are readily available through PHASTER, which reduced our data set to 176 genomes for both the fast- and slow-growing data sets. After submitting accession numbers into PHASTER, we download a results file. This file lists the number of prophages in the genome, length of each prophage, the possible proteins present in the prophage and which prophage the sequence is most likely to be [7]. The PHASTER algorithm also identifies if a given prophage is "intact", "questionable", or "incomplete" [7]. These classifications are determined through a completeness score within the search algorithm of PHASTER; a score of greater than 90 is classified as intact, between 70 and 90 is questionable, while less than 70 is incomplete [7]. The data in the sections to follow were extracted by post-processing of the results files obtained through this shell-scripted bioinformatics pipeline. This pipeline is given in Appendix B.

After classifying the prophages according to their bacterial hosts, we were interested in testing our hypothesis regarding an increased number of prophages in fast-growing hosts. Using the completeness score for prophages from PHASTER, we identify the number of intact, ques-

tionable and incomplete prophages a specific genome contains [7]. These results are shown in
Figure 5.1.
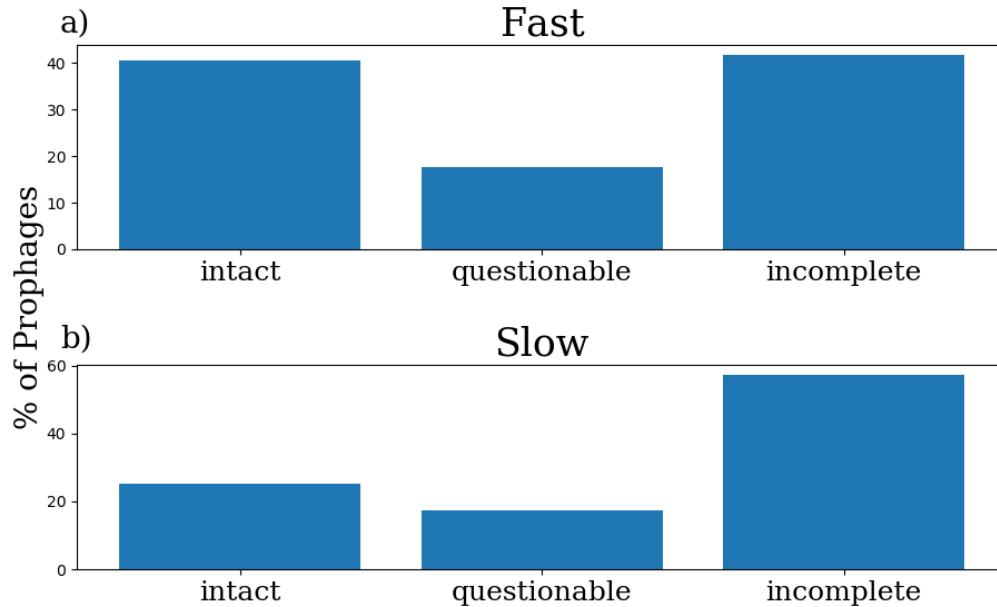
Figure 5.1: Percentage of prophages in each classification (either intact, questionable or in-
complete), according to PHASTER [7]. Panel a) shows the percentage of prophages in each
classification for fast-growing hosts, and panel b) shows the percentage for slow-growing hosts.

These initial data regarding the completeness of prophages are not consistent with the
suggestion that fast-growing bacteria will have more cryptic prophages. We notice that fast-
growing hosts, in fact, have similar fractions of intact and incomplete prophages within their
genomes, while slow-growing hosts have almost 30% more incomplete prophages than intact.
These findings lead us to question what factors may be influencing the lengths of prophages in
fast- and slow-growing hosts.

Figure 5.2 shows the prophage length distributions of prophages in fast- and slow-growing
hosts, as well as the combination of the two data sets. For fast-growing hosts, we see a clear

multimodal distribution, similar to those considered by Khan and Wahl [94]. However, the prophages within slow-growing hosts and, to some extent, the total prophage distribution seem to have a unimodal distribution with positive skewness, in contrast to the previously predicted negatively skewed distribution [15, 94]. This is consistent with the findings shown in Figure 5.1, in that slow-growing hosts tend to have more incomplete prophages within their genomes. This degradation of prophages would account for the more prominent positive skewness of the distribution when compared to the fast-growing host prophage length distribution (see [99] or [128] for a discussion).

5.3　Model

To fit the data shown in Figure 5.2, we adapt the model developed by Khan and Wahl [94]. Following the findings of Khan and Wahl [94], we fit each data set (i.e., fast- and slow-growing distributions) separately using a model comprised of functions describing lysogeny, selection, induction and degradation. For a full description of each function in the model, see [94].

Khan and Wahl [94] outline the derivation of the partial differential equation,

$$\frac{\partial Q(x,t)}{\partial t} = \alpha f(x) + \frac{\partial}{\partial t}(D(x)Q(x,t)) + r_s S(x)Q(x,t) - r_I I(x)Q(x,t), \qquad (5.1)$$

where the model parameters and functions are given in Table 5.2. Equation (5.1) describes the time evolution of the prophage length distributions.

The distribution of interest, $y(x)$, is the steady state solution of $Q(x,t)$, that is

$$y(x) = \lim_{t\to\infty} Q(x,t).$$
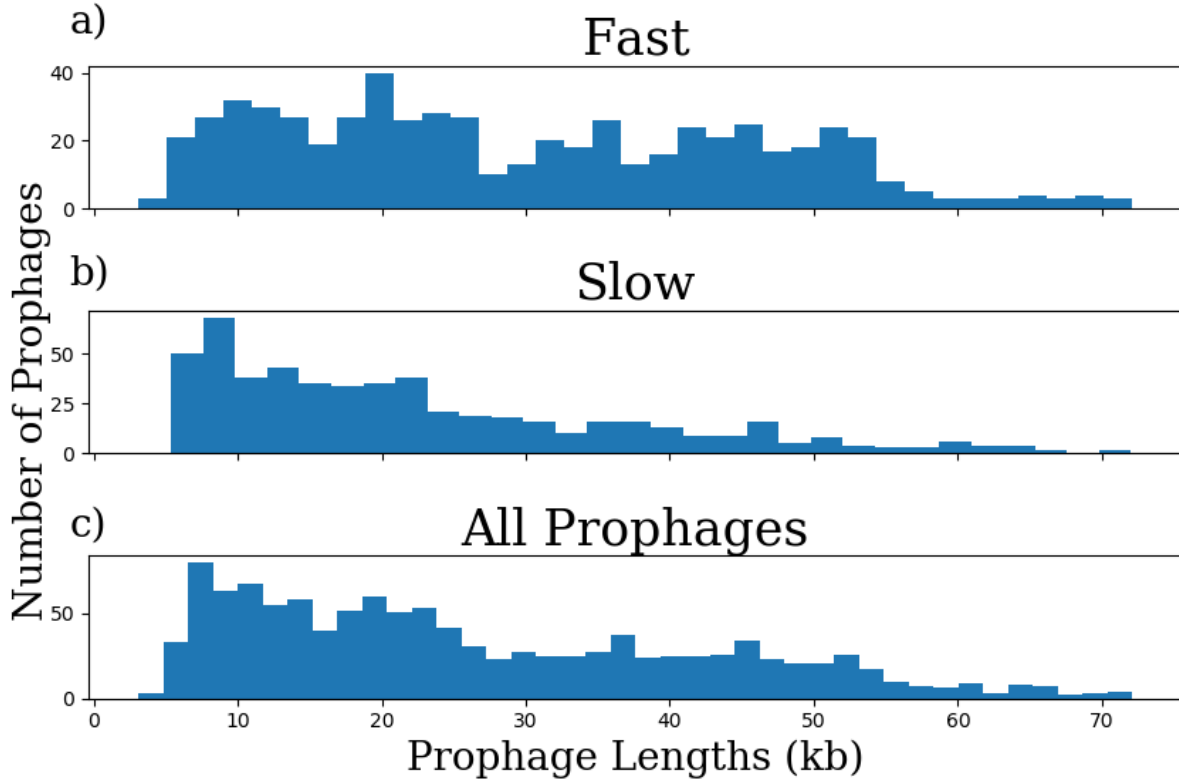
Figure 5.2: Length distributions of prophages found in fast- and slow-growing bacterial hosts. Panel a) shows the histogram of prophage lengths in fast-growing hosts, with a multimodal distribution. Panel b) shows the histogram of prophage lengths in slow-growing hosts, with a positively skewed unimodal distribution. Panel c) shows the total prophage length distribution (i.e., all prophages considered in the following analysis), again with a positively skewed unimodal distribution.

This steady state solution is found to be,

$$y(x) = \frac{-\alpha e^{-\int \mathcal{F}(x)dx}}{r_D x} \int f(x) e^{\int \mathcal{F}(x)dx} dx + \frac{C}{x} e^{-\int \mathcal{F}(x)dx}, \tag{5.2}$$

where C is an arbitrary constant and

$$\mathcal{F}(x) = \frac{r_S S(x)}{r_D x} - \frac{r_I I(x)}{r_D x}. \tag{5.3}$$

$Q(x,t)$	frequency of prophages of length x (kb) at time t
$y(x)$	steady state solution of $Q(x,t)$
$f(x)$	length distribution of prophage sequences entering via lysogeny
$D(x)$	mutational degradation rate
$S(x)$	expected fraction of r_S conferred by prophage of length x
$I(x)$	probability that prophage carries genes required for induction
α	Relative rate of lysogeny
r_D	Relative rate of degradation
r_S	Relative selection coefficient (intact prophage)
r_I	Relative rate of induction
n_I	Number of genes required for induction

Table 5.2: Model functions and parameters adapted from [94]

For brevity, we do not include the full derivation of Equation (5.2), but rather invite the reader

to see the description given by Khan and Wahl [94].

5.4   Results

To fit both the fast- and slow-growing data sets, we used a finite difference scheme to obtain,

numerically, the steady-state solution (Equation (5.2)) and compared this steady-state solution

to the data, optimizing the log-likelihood to identify the best fit parameter values. The log-

likelihood is defined as

$$\log(L) = \sum_{i=1}^{n} \log y(x_i),$$

where x_i are the n observed lengths of prophage genomes in the data set, and $y(x_i)$ is the numerically obtained steady-state solution.

The incoming phage distribution, $f(x)$, is described by a mixed distribution incorporating one to three Gaussian distributions [94]. Dependent on the number of Gaussian distributions selected for $f(x)$, our model could include $k = 9$, $k = 12$, or $k = 15$ free parameters. To select the best model among the candidate models, we used the Akaike Information Criterion (AIC) [5]. A full description of this process is outlined in Appendix C. For both data sets, the best model among the candidate models included two Gaussian distributions in the function $f(x)$, that is,

$$f(x) = \sum_{i=1}^{2} p_i \exp\left(\frac{-(x - (\theta + \mu_i))^2}{\sigma_i^2}\right),$$

where θ gives the length of the smallest autonomous temperate phage. Bobay *et al.* [15] report that the smallest autonomous phage that can infect bacteria is 30 kb; in our following results we, therefore, set $\theta = 30$ kb.

The best fit parameters for the fast- and slow-growing data sets are given in Table 5.3. All the given best fit rate parameters are given relative to the rate of degradation, r_D. This is because the steady state solution does not depend independently on the rates, but only on the ratio of the rates. In particular, as shown in Equations (5.2) and (5.3), the steady state solution depends on the rates α/r_D, r_S/r_D, and r_I/r_D. While in principle we could provide the rates relative to any of the four rate parameters in the model, normalizing by the degradation rate has the additional advantage that the rate of degradation, a mutational rate, is not expected to

differ between the two growth classes, and as such allows for a cleaner comparison [196].

	Parameter	Fast-growing	Slow-growing
α	relative rate of lysogeny	5.9	1.4
r_D	relative rate of degradation	1	1
r_S	relative selection coeff. (intact prophage)	10.774	65.167
r_I	relative rate of induction	44.0	65.99
n_I	number of genes required for induction	5.50	1.50

Table 5.3: Parameter values for the best fits for both bacterial growth groups. All of the given relative rates have been normalized by the degradation rate, r_D.

Figure 5.3 shows the best fit for $y(x)$ for the fast- and slow-growing data sets respectively. Parameters used in the Gaussian distributions underlying $f(x)$ are given in the captions of each figure, while the other best fit parameters are those given in Table 5.3.

Comparing the best-fit parameter values for fast- and slow-growing hosts, we find that the rates of lysogeny, induction and selection all seem substantially different. This, however, leads us to ponder if these values are indeed significantly different from one another. We will discuss the methodology for determining whether the parameters are statistically different in the following section. It is worth noting that the differences in parameter values are not surprising, given the markedly different distributions illustrated in Figure 5.2. We will return to the implications of these results in the Discussion.

(a) fast-growing hosts   (b) slow-growing hosts
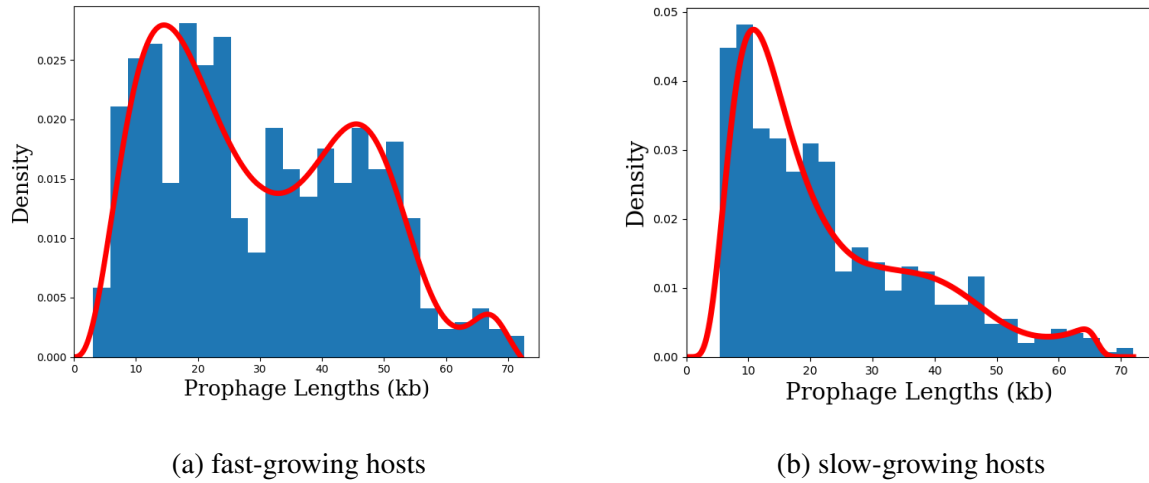
Figure 5.3: The best fit predicted by the model ($y(x)$, red curve) to each data set. Panel (a) shows the best fit model for the fast-growing data set, with $p_1 = 50.487$, $\mu_1 = 20.161$, $\sigma_1 = 8.547$, $p_2 = 20.230$, $\mu_2 = 38.722$ and $\sigma_2 = 3.408$. Panel (b) shows the best fit model for the slow-growing data set, with $p_1 = 13.114$, $\mu_1 = 16.332$, $\sigma_1 = 10.364$, $p_2 = 27.906$, $\mu_2 = 36.001$ and $\sigma_2 = 1.648$.

5.4.1 Statistical Analysis of Parameter Estimates

We want to test which of the parameter estimates for the best fit model, as given in Table 5.3, are statistically different between the data sets, or whether these differences might have arisen by chance. Ideally, one would undertake a bootstrapping approach to identify the variance in the best fit parameters (see [55] for a discussion of this method). However, to ensure the accuracy of the findings of a bootstrap method, we would need to fit, on the order of, 1000 sampled distributions. Due to the computational cost of the optimization algorithm (on the order of one day to complete a single fit), this is not feasible. We therefore turn our attention towards a weighted least squares approach, as described by Landaw and DiStefano [103].

Before our discussion of this method and its results, we note that we fit our data sets by minimizing the log-likelihood. This is in contrast to the fits used in the examples of Landaw and DiStefano, who discuss results and statistical methods for fits that have used a weighted least squares approach [103]. Although the assumptions of these two approaches are quite different, we interpret the results to follow as a rough estimate for the confidence in our best fit parameters. If feasible, we would utilise a bootstrapping methodology to ensure the accuracy of the estimated coefficients of variation.

To determine the coefficients of variation in the best fit parameters of our fits, we construct the variance-covariance matrix (hereafter referred to as the covariance matrix) as described by Landaw and DiStefano [103].

We first denote $\hat{\mathbf{p}}$ as a $P \times 1$ vector whose entries, $\hat{p}_i$, are the best fit parameters. Note that in our best fit models, for both the fast- and slow-growing data sets, $P = 11$. This reduction of free parameters by one is because we scale each rate by r_D; there are only 11 free parameters in the steady-state solution. Then, the $P \times P$ dimensional covariance matrix of $\hat{\mathbf{p}}$ (denoted as $\mathrm{COV}(\hat{\mathbf{p}})$) is a measure of the spread of $\hat{\mathbf{p}}$ about its mean. The ith diagonal entry of $\mathrm{COV}(\hat{\mathbf{p}})$ is the variance of $\hat{p}_i$; therefore, its square root is the standard deviation of $\hat{p}_i$ (denoted as $\mathrm{sd}(\hat{p}_i)$). We then can determine the coefficient of variation for each parameter (denoted as CV_i, where $i = 1 \ldots P$) by

$$\mathrm{CV}_i = \frac{\mathrm{sd}(\hat{p}_i)}{\hat{p}_i} \times 100\% \,. \tag{5.4}$$

We can approximate $\mathrm{COV}(\hat{\mathbf{p}})$ by constructing the complete information matrix for the complete set of data [103]. We first let $\mathbf{s}(x, \hat{\mathbf{p}})$ be a $P \times 1$ vector whose ith component is

$$\frac{\partial Y}{\partial p_i} \,,$$

where $Y(x, \hat{\mathbf{p}})$ is the best fit model of the cumulative distribution function for the prophage lengths. That is,

$$Y(x, \hat{\mathbf{p}}) = \int_0^x y(\tau) \, d\tau,$$

for parameter values $\hat{\mathbf{p}}$. The vector $\mathbf{s}$ can therefore be viewed as the sensitivity of the model output to changes in the parameters [119]. Due to the fact the steady state solution $y(x)$ and the associated $Y(x, \hat{\mathbf{p}})$ are intractable analytically, we approximate the partial derivatives in $\mathbf{s}$ using first principles with a step size of $h = \varepsilon p_i / r_D$, where r_D is the best fit rate of degradation. We then construct the $P \times P$ point information matrix, $M(x_i)$, for each prophage length in our data set, x_i, by

$$M(x_i) = \mathbf{s}(x_i, \hat{\mathbf{p}})[\mathbf{s}(x_i, \hat{\mathbf{p}})]^T. \tag{5.5}$$

Thus, the complete information matrix is given by

$$M = \sum_{i=1}^N M(x_i), \tag{5.6}$$

where N is the number of prophage lengths in each data set. This complete information matrix may then be used to approximate $\mathrm{COV}(\hat{\mathbf{p}})$ as

$$\mathrm{COV}(\hat{\mathbf{p}}) = \sigma^2 M^{-1},$$

where σ^2 is the variance of the error between the data set and its fit. Due σ^2 being unknown, we may use an unbiased approximation

$$\hat{\sigma}^2 = \frac{\mathrm{WRSS}(\hat{\mathbf{p}})}{\mathrm{DF}}, \tag{5.7}$$

where DF is the value of the degrees of freedom, *i.e.,* $N - P$. The function $\mathrm{WRSS}(\hat{\mathbf{p}})$ is the weighted residual sum of squares that is, for a given data set, $z(x_i)$, defined as

$$\mathrm{WRSS}(\hat{\mathbf{p}}) = \sum_{i=1}^N [z(x_i) - Y(x, \hat{\mathbf{p}})]^2. \tag{5.8}$$

In a more general sense, each term in the summation of Equation (5.8) should be scaled by the weight, w_i, given to each data point [103]. However, in our log-likelihood fit all points are equally weighted.

Using COV($\hat{\mathbf{p}}$), we may extract the variance (and standard deviation) from the diagonal entries. We can then use Equation (5.4) to find the coefficient of variation for each parameter. Table 5.4 gives the best fit parameters, as in Table 5.3, with their associated standard deviations and coefficients of variation. We note that the parameters describing lysogeny (α), selection (r_S) and induction (r_I, n_I) all differ significantly between fast- and slow-growing bacteria.

	Value ± sd		Coefficient of Variation (%)	
Parameter	Fast	Slow	Fast	Slow
α/r_D	5.9 ± 0.2	1.4 ± 0.6	3.6334	40.8655
r_S/r_D	10.77 ± 0.04	65.17 ± 0.05	0.3745	0.0841
r_I/r_D	44.0 ± 2.0	65.99 ± 0.04	4.6432	0.0532
n_I	5.50 ± 0.03	1.50 ± 0.01	0.6343	0.8736

Table 5.4: Parameter values for the best fits for both bacterial growth groups with standard deviations. All of the given relative rates have been normalized by the degradation rate, r_D. Standard deviation and coefficient of variation were calculated with $\varepsilon = 0.1$.

One critical piece to note regarding the numerical calculation of the coefficients of variation: through experimentation, we notice that the coefficients of variation for the slow-growing data set are sensitive to the step-size used in the first principle calculation of the partial derivatives. To illustrate this sensitivity, we refer the reader to Appendix D. The coefficients of

variation given in Table 5.4 are calculated with $\varepsilon = 0.1$, meaning the step-size used for each partial derivative is 10% of the best fit parameter value, scaled by the best fit r_D value.

5.5   Discussion

As an extension of the work of Chapter 4, we further investigated how life-history traits of the bacterial host, such as minimal doubling time, affect the fractions of intact and incomplete prophages in a bacterial genome, as well as the distribution of prophage lengths. We developed a bioinformatics pipeline to extract both lengths and completeness classifications for prophages in a set of publicly-available bacterial genomes and implemented the model and numerical scheme developed by Khan and Wahl [94] to fit data sets created via this pipeline in order to test whether prophages in fast- and slow-growing bacteria are subject to different rates of lysogeny, selection or induction.

In the work of Touchon *et al.* [183], the strongest correlation between the host's life-history traits and prevalence of lysogeny is found to be the minimal doubling time. We therefore use their curated list of bacterial genomes to generate the data used in our model fits, in order to determine the possible differences in the forces affecting prophage. The fast-growing bacterial hosts (with minimal doubling time less than 2.5 hours) have a multimodal prophage length distribution, similar to distributions that have been observed in multiple data sets previously [15, 43, 106]. However, the slow-growing hosts exhibit a unimodal distribution with positive skewness, opposite to the previously predicted negatively skewed distribution [15, 94]. This stark difference between the two data sets, which are seen in Figure 5.2, seems to suggest that the forces acting on the prophages that are integrated in fast- and slow-growing hosts may be

significantly different.

Utilising the model of Khan and Wahl [94], we fit each data set using a finite difference scheme and optimized the log-likelihood to determine the best fit parameters. These best fit values are given in Table 5.3 with their associated standard deviations. We note that each of the rates given in this table are scaled by the rate of degradation, r_D. As mentioned, although the scaling by rate of induction would give each rate in units of expected prophage lifetime, we expect the rate of induction to be different between the two hosts. However, it has been found that the rate of degradation is more than likely constant across all bacterial hosts, and as such is the better value to scale by [196].

Our results support the findings of Touchon *et al.* [183]. The rate of lysogeny, α, is much larger for fast-growing hosts when compared to slow-growing hosts. This is congruent with the strong correlation between minimal doubling time and the prevalence of prophage sequences in the observations of Touchon *et al.* [183]. We note that the rates of lysogeny for the two data sets are statistically different, with no overlap between the best fit parameters and their associated standard deviations.

Our best fit parameters suggest several other differences between the forces acting on prophage sequences. The relative rate of induction, r_I, is larger for slow-growing hosts when compared with fast-growing hosts. This difference is consistent with prophages infecting slow-growing hosts favouring lysis, the process that requires induction, while those infecting fast-growing hosts favour lysogeny [183, 190]. The relative selection coefficient, r_S, is also larger for prophages in slow-growing hosts. We predict that although fewer prophages become integrated in slow-growing hosts, many of those that do are degraded to become cryptic prophages, and are then maintained by selection. Thus, an experimentally testable prediction arising from

this work is that prophages in slow-growing bacterial species confer greater selective benefits, on average, than prophages in fast-growing bacteria.

Our findings support those of Touchon *et al.* [183], demonstrating that prophages that infect fast-growing hosts have a higher rate of lysogeny. Although minimal doubling time and lysogeny was the strongest correlation found in Touchon *et al.* [183], the authors note that pathogenicity of the host may also play an important role in the distribution of prophages. An investigation into the forces that act on prophages in pathogenic and non-pathogenic hosts would be a clear avenue for future work.

Chapter 6

Conclusions and Future Work

The interaction between phage and bacterial host has become of increasing interest in scientific research in recent years due to pivotal role both bacteria and phage play in a wide array of areas, including medicine [100, 124], ecosystem health [89, 135] and evolution [12, 21, 34, 38, 60]. The sheer quantity of phage has transformed the views of many about phage-bacteria dynamics from a simple interaction to a complex and important set of interactions [37].

Interest has increased since the 1980s in the evolution of so-called temperate phage [176], bacteriophage that are able to initiate two distinct infection strategies on their bacterial prey [114]. The answer to the long-standing question "why be temperate?" [176] has evaded researchers for some time [12, 108, 190, 199]. Our contribution to this question comes in the form of the results of Chapters 2 and 4. In addition, we investigate the existence and stability of equilibria of a temperate phage model in Chapter 3 and further investigate the unexplained some unexplained experimental results [183] using bioinformatical techniques in Chapter 5. Below we provide a summary of all four projects and the conclusions derived from them.

We first begin by investigating the role that prophage sequences, in the form of a plasmid,

may play in the adaptation of temperate phage in Chapter 2. Here, we develop a life-history model for a free temperate phage and its associated plasmid prophage. We focus soley on temperate phage creating plasmid prophages, as it is an interesting and oft neglected area of prophage studies [34, 37, 79, 144]. The results of our model suggest that *de novo* beneficial mutations are more likely to survive drift if they first occur during plasmid prophage replication, rather than during lytic replication. This suggests that plasmid prophages may have played a disproportionate role in genetic innovation for their associated temperate phage lineages. This work is published, in part, in [192].

In Chapter 3, we develop and analyse a non-linear ordinary differential equation model describing the dynamics of temperate phages, prophages and their hosts. The evolutionary dynamics of this model have been studied, in part, in Wahl *et al.* [190] and Berngruber *et al* [12, 13]. However, no analysis of the equilibria or stability was conducted in this previous work. We identify four distinct equilibria for the associated dimensionless system, and elucidate the precise conditions necessary for stability and existence of each.

We extend the resource-implicit model of Chapter 3 to a resource-explicit model in Chapter 4, to determine the evolutionary trajectory of temperate prophage in variable environments. Our model development includes the novel inclusion of a bacterial stationary phase, which is often neglected in models of temperate phage-bacteria dynamics (for instance in [12], [13], [108], [176] and [190]). Through the utilisation of evolutionary invasion analysis, we observe the importance of the stationary phase in the lysis-lysogeny decision when environments are variable. We find that if bacteria have the ability to enter stationary phase, then when environments vary greatly the propensity for lysogeny increases. Our results suggest a possible explanation for the correlation between minimal doubling time and lysogeny experimentally in

previous work [183].

The results of Chapter 4 regarding a phage population evolving towards full lysogeny created an additional area of exploration that is the focus of Chapter 5. Throughout the utilisation of a bioinformatic pipeline, we create a data set of genomes and their associated prophages from a curated list of publicly-available bacterial genomes [183]. This data set was constructed using PHASTER's web interface [7]. By adapting the model of Khan and Wahl [94], we were able to best fit the distribution of prophage lengths in order to identify any differences in strategies of temperate phage infecting fast- and slow-growing bacteria. Our resulting best fit models support the experimental observation that lysogeny is more prevalent for fast-growing bacteria.

6.1   Future Work

Although the results of Chapter 4 support the claim that variable environments with fast-growing bacteria would have an increased propensity of lysogeny [95, 183], we investigate what constitutes "variability" in a limited number of ways. In our model, the resource influx is controlled by a sinusoidal function $C(t)$, which has parameters that control the frequency, amplitude and the average amount of resource influx. Our results allow each parameter to vary, while holding the other two constant. A clear area for future work would be to allow multiple parameters to vary, in order to observe if the results continue to support our hypothesis.

In addition, the concept of "variability" could be interpreted more broadly. For example, the influx of resource could be controlled stochastically. We could extend this idea further, and build a stochastic model to include all sorts of variability and determine the role stochasticity plays on the lysis-lysogeny decision. Another avenue for future work for this model is an

extension that could be suggested for a variety of phage-host models, the inclusion of spatial dynamics. Since many bacteria exist in complex spatial communities known as biofilms [74], spatial dynamics may be important to phage-host interactions in many ecological settings.

Our results from Chapter 5 support one experimental observation in regards to minimal doubling time and lysogeny, however, a variety of host life-history traits have also been suggested to influence the lysis-lysogeny decision [183]. An avenue for future work would be to investigate how several of these other life-history traits, such as the pathogenicity of the host, influence the distribution of prophage. Using our developed bioinformatics pipeline, the model of Khan and Wahl [94], and the curated list of publicly-available bacterial genomes [183], this would be a relatively straightforward and potentially impactful project to undertake in the future.

Bibliography

[1] S. T. Abedon, editor. *Bacteriophage ecology: population growth, evolution, and impact of bacterial viruses*. Cambridge University Press, Cambridge, June 2008.

[2] S. T. Abedon. Bacteriophage secondary infection. *Virologica Sinica*, 30(1):3–10, 2015.

[3] S. T. Abedon and J. T. LeJeune. Why bacteriophage encode exotoxins and other virulence factors. *Evolutionary Bioinformatics*, 1:97–110, 2005.

[4] M. H. Adams. Bacteriophages. *Bacteriophages*, 1959.

[5] H. Akaike. Likelihood of a model and information criteria. *Journal of Econometrics*, 16(1):3–14, 1981.

[6] J. Arino, S. S. Pilyugin, and G. S. Wolkowicz. Considerations on yield, nutrient uptake, cellular growth, and competition in chemostat models. *Canadian Applied Mathematics Quarterly*, 11(2):107–142, 2003.

[7] D. Arndt, J. R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang, and D. S. Wishart. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic acids research*, 44(W1):W16–W21, 2016.

[8] I. Aviram and A. Rabinovitch. Bactria and lytic phage coexistence in a chemostat with periodic nutrient supply. *Bulletin of Mathematical Biology*, 76(1):225–244, 2014.

[9] D. Axelrod and M. Kimmel. *Branching processes in biology*. Springer-Verlag, 2015.

[10] N. Bacaër. Approximation of the basic reproduction number R_0 for vector-borne diseases with a periodic vector population. *Bulletin of Mathematical Biology*, 69(3):1067–1091, 2007.

[11] E. Beretta and Y. Kuang. Modeling and analysis of a marine bacteriophage infection. *Mathematical Biosciences*, 149(1):57–76, 1998.

[12] T. W. Berngruber, R. Froissart, M. Choisy, and S. Gandon. Evolution of virulence in emerging epidemics. *PLoS Pathogens*, 9(3):e1003209, 2013.

[13] T. W. Berngruber, S. Lion, and S. Gandon. Spatial structure, transmission modes and the evolution of viral exploitation strategies. *PLoS Pathogens*, 11(4):e1004810, 2015.

[14] R. J. Beverton and S. J. Holt. *On the dynamics of exploited fish populations*, volume 11. Springer Science & Business Media, 2012.

[15] L.-M. Bobay, M. Touchon, and E. P. Rocha. Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences*, 111(33):12127–12132, 2014.

[16] J. Bondy-Denomy and A. R. Davidson. When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. *Journal of Microbiology*, 52(3):235–242, 2014.

[17] J. Bondy-Denomy, J. Qian, E. R. Westra, A. Buckling, D. S. Guttman, A. R. Davidson, and K. L. Maxwell. Prophages mediate defense against phage infection through diverse mechanisms. *The ISME Journal*, 10(12):2854–2866, 2016.

[18] S. R. Bordenstein and S. R. Bordenstein. Eukaryotic association module in phage WO genomes from *Wolbachia*. *Nature Communications*, 7(1):1–10, 2016.

[19] R. M. Borysowski. *Phage Therapy*. Caister Academic Press, 2014.

[20] J. D. Bouchard and S. Moineau. Homologous recombination between a lactococcal bacteriophage and the chromosome of its host strain. *Virology*, 270(1):65–75, 2000.

[21] E. F. Boyd, B. M. Davis, and B. Hochhut. Bacteriophage–bacteriophage interactions in the evolution of pathogenic bacteria. *Trends in Microbiology*, 9(3):137–144, 2001.

[22] Å. Brännström, J. Johansson, and N. Von Festenberg. The hitchhiker's guide to adaptive dynamics. *Games*, 4(3):304–328, 2013.

[23] H. Bremer and P. P. Dennis. Modulation of chemical composition and other parameters of the cell by growth rate. Escherichia coli *and* Salmonella*: Cellular and Molecular biology*, 2(2):1553–69, 1996.

[24] N. F. Britton. *Essential mathematical biology*. Springer Science & Business Media, 2012.

[25] A. B. Brueggemann, C. L. Harrold, R. R. Javan, A. J. Van Tonder, A. J. McDonnell, and B. A. Edwards. Pneumococcal prophages are diverse, but not without structure or history. *Scientific Reports*, 7(1):1–13, 2017.

[26] H. Brüssow, C. Canchaya, and W.-D. Hardt. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews*, 68(3):560–602, 2004.

[27] R. Bruynoghe and J. Maisin. Essais de thérapeutique au moyen du bacteriophage. *CR Society of Biology*, 85:1120–1121, 1921.

[28] D. Bryan, A. El-Shibiny, Z. Hobbs, J. Porter, and E. M. Kutter. Bacteriophage T4 infection of stationary phase *E. coli*: life after log from a phage perspective. *Frontiers in Microbiology*, 7:1391, 2016.

[29] R. E. Buchanan and E. D. F. Buchanan. *Bacteriology*. MacMillan, New York, 5th edition, 1951.

[30] J. J. Bull, I. J. Molineux, and W. Rice. Selection of benevolence in a host–parasite system. *Evolution*, 45(4):875–882, 1991.

[31] A. R. Burmeister, R. E. Lenski, and J. R. Meyer. Host coevolution alters the adaptive landscape of a virus. *Proceedings of the Royal Society B: Biological Sciences*, 283(1839):20161528, 2016.

[32] K. Burnham and D. Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York, 3rd edition, 2003.

[33] A. M. Campbell. Cryptic prophages. Escherichia coli *and* Salmonella*: cellular and molecular biology*, 2:2041–2045, 1996.

[34] C. Canchaya, G. Fournous, and H. Brüssow. The impact of prophages on bacterial chromosomes. *Molecular Microbiology*, 53(1):9–18, 2004.

[35] C. Canchaya, C. Proux, G. Fournous, A. Bruttin, and H. Brüssow. Prophage genomics. *Microbiol. Mol. Biol. Rev.*, 67(2):238–276, 2003.

[36] S. Casjens. Prophages and bacterial genomics: what have we learned so far? *Molecular microbiology*, 49(2):277–300, 2003.

[37] S. R. Casjens. Comparative genomics and evolution of the tailed-bacteriophages. *Current opinion in microbiology*, 8(4):451–458, 2005.

[38] S. R. Casjens, E. B. Gilcrease, W. M. Huang, K. L. Bunny, M. L. Pedulla, M. E. Ford, J. M. Houtz, G. F. Hatfull, and R. W. Hendrix. The pKO2 linear plasmid prophage of *Klebsiella oxytoca. Journal of bacteriology*, 186(6):1818–1832, 2004.

[39] H. Caswell. *Matrix Population Models: Construction, Analysis, and Interpretation*. Sinauer Associated, 2nd edition, 2001.

[40] N. Chanishvili and R. Sharp. Bacteriophage therapy: experience from the Eliava Institute, Georgia. *Microbiology Australia*, 29(2):96–101, 2008.

[41] M. R. Clokie, A. D. Millard, A. V. Letarov, and S. Heaphy. Phages in nature. *Bacteriophage*, 1(1):31–45, 2011.

[42] P. K. Cochran and J. H. Paul. Seasonal abundance of lysogenic bacteria in a subtropical estuary. *Applied and Environmental Microbiology*, 64(6):2308–2312, 1998.

[43] J. S. Crispim, R. S. Dias, P. M. P. Vidigal, M. P. de Sousa, C. C. da Silva, M. F. Santana, and S. O. de Paula. Screening and characterization of prophages in *Desulfovibrio* genomes. *Scientific reports*, 8(1):1–10, 2018.

[44] J. M. Cushing and O. Diekmann. The many guises of R_0 (a didactic note). *Journal of Theoretical Biology*, 404:295–302, 2016.

[45] F. Dercole and S. Rinaldi. *Analysis of evolutionary processes: the adaptive dynamics approach and its applications*, volume 3. Princeton University Press, 2008.

[46] U. Dieckmann, M. Doebeli, J. A. Metz, and D. Tautz. *Adaptive speciation*. Cambridge University Press, 2004.

[47] O. Diekmann. A beginners guide to adaptive dynamics. *Summer School on Mathematical Biology*, pages 63–100, 2002.

[48] O. Diekmann, J. Heesterbeek, and M. G. Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, 7(47):873–885, 2009.

[49] O. Diekmann and J. A. P. Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, volume 5. John Wiley & Sons, 2000.

[50] I. B. Dodd, B. Kalionis, and J. B. Egan. Control of gene expression in the temperate coliphage 186: VIII. control of lysis and lysogeny by a transcriptional switch involving face-to-face promoters. *Journal of Molecular Biology*, 214(1):27–37, 1990.

[51] I. B. Dodd, K. E. Shearwin, and J. B. Egan. Revisited gene regulation in bacteriophage λ. *Current opinion in genetics & development*, 15(2):145–152, 2005.

[52] H. M. Doekes, G. A. Mulder, and R. Hermsen. How repeated outbreaks drive the evolution of bacteriophage communication: Insights from a mathematical model. *bioRxiv*, 2020.

[53] M. K. Donnelly-Wu, W. R. Jacobs Jr, and G. F. Hatfull. Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. *Molecular microbiology*, 7(3):407–417, 1993.

[54] G. Ebersbach and K. Gerdes. Plasmid segregation mechanisms. *Annu. Rev. Genet.*, 39:453–479, 2005.

[55] B. Efron and R. Tibshirani. The bootstrap method for assessing statistical accuracy. *Behaviormetrika*, 12(17):1–35, 1985.

[56] Z. Erez, I. Steinberger-Levy, M. Shamir, S. Doron, A. Stokar-Avihail, Y. Peleg, S. Melamed, A. Leavitt, A. Savidor, and S. Albeck. Communication between viruses guides lysis–lysogeny decisions. *Nature*, 541(7638):488–493, 2017.

[57] T. Evans, R. G. Bowers, and M. Mortimer. Adaptive dynamics of temperate phages. *Evolutionary Ecology Research*, 12(4):413–434, 2010.

[58] W. J. Ewens. *Population genetics*. Springer Science & Business Media, 2013.

[59] S. E. Finkel. Long-term survival during stationary phase: evolution and the gasp phenotype. *Nature Reviews Microbiology*, 4(2):113–120, 2006.

[60] L.-C. Fortier and O. Sekulovic. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, 4(5):354–365, 2013.

[61] D. E. Fouts. *Phage_Finder*: automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic acids research*, 34(20):5839–5851, 2006.

[62] C. Fraley, J. Kim, M. McCann, and A. Matin. The *Escherichia coli* starvation gene *cst*C is involved in amino acid catabolism. *Journal of bacteriology*, 180(16):4287–4290, 1998.

[63] R. Gallet, T. Lenormand, and I.-N. Wang. Phenotypic stochasticity protects lytic bacteriophage populations from extinction during the bacterial stationary phase. *Evolution: International Journal of Organic Evolution*, 66(11):3485–3494, 2012.

[64] S. Gandon. Why be temperate: lessons from bacteriophage λ. *Trends in microbiology*, 24(5):356–365, 2016.

[65] S. A. Geritz, G. Mesze, and J. A. Metz. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary ecology*, 12(1):35–57, 1998.

[66] D. Ghosh, K. Roy, K. E. Williamson, D. C. White, K. E. Wommack, K. L. Sublette, and M. Radosevich. Prevalence of lysogeny among soil bacteria and presence of 16S rRNA and trzN genes in viral-community DNA. *Applied and Environmental Microbiology*, 74(2):495–502, 2008.

[67] B. Gibson, D. J. Wilson, E. Feil, and A. Eyre-Walker. The distribution of bacterial doubling times in the wild. *Proceedings of the Royal Society B: Biological Sciences*, 285(1880):20180789, 2018.

[68] G. S. Gordon and A. Wright. DNA segregation in bacteria. *Annual Reviews in Microbiology*, 54(1):681–708, 2000.

[69] A. Górski, J. Borysowski, R. Międzybrodzki, and B. Weber-Dąbrowska. *Bacteriophages in medicine*. Caister Academic Press Norfolk, 2007.

[70] D. Greenhalgh and I. Moneim. SIRS epidemic model and simulations using different types of seasonal contact rate. *Systems Analysis Modelling Simulation*, 43(5):573–600, 2003.

[71] B. Guttman, R. Raya, and E. Kutter. Basic phage biology. *Bacteriophages: Biology and applications*, 4, 2005.

[72] J. B. Haldane. A mathematical theory of natural and artificial selection, part V: selection and mutation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 23, pages 838–844. Cambridge University Press, 1927.

[73] S. P. Hansen and S. P. Hubbell. Single-nutrient microbial competition: qualitative agreement between experimental and theoretically forecast outcomes. *Science*, 207(4438):1491–1493, 1980.

[74] D. R. Harper, H. M. Parracho, J. Walker, R. Sharp, G. Hughes, M. Werthén, S. Lehman, and S. Morales. Bacteriophages and biofilms. *Antibiotics*, 3(3):270–284, 2014.

[75] G. F. Hatfull. Bacteriophage genomics. *Current opinion in microbiology*, 11(5):447–453, 2008.

[76] F. Hayes. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science*, 301(5639):1496–1499, 2003.

[77] J. M. Heffernan, R. J. Smith, and L. M. Wahl. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2(4):281–293, 2005.

[78] R. W. Hendrix. *Lambda II*. Cold Spring Harbor Laboratory, 1983.

[79] R. W. Hendrix, G. F. Hatfull, M. E. Ford, M. C. Smith, and R. N. Burns. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. In *Horizontal Gene Transfer*, pages 133–VI. Elsevier, 2002.

[80] F. Herelle. An invisible microbe that is antagonistic to the dysentery bacillus *Cozzes rendus*. In *Academy of Sciences*, volume 165, pages 373–375, 1917.

[81] Z. Hobbs and S. T. Abedon. Diversity of phage infection types and associated terminology: the problem with 'lytic or lysogenic'. *FEMS microbiology letters*, 363(7), 2016.

[82] C. Howard-Varona, K. R. Hargreaves, S. T. Abedon, and M. B. Sullivan. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *The ISME journal*, 11(7):1511–1520, 2017.

[83] N. M. Høyland-Kroghsbo, R. B. Mærkedahl, and S. L. Svenningsen. A quorum-sensing-induced bacteriophage defense mechanism. *mBio*, 4(1), 2013.

[84] S.-B. Hsu, S. Hubbell, and P. Waltman. A mathematical theory for single-nutrient competition in continuous cultures of micro-organisms. *SIAM Journal on Applied Mathematics*, 32(2):366–383, 1977.

[85] J. E. Hubbarde, G. Wild, and L. M. Wahl. Fixation probabilities when generation times are variable: The burst–death model. *Genetics*, 176(3):1703–1712, 2007.

[86] A. Hurford, D. Cownden, and T. Day. Next-generation tools for evolutionary invasion analyses. *Journal of the Royal Society Interface*, 7(45):561–571, 2010.

[87] P. Hyman and S. T. Abedon. Bacteriophage host range and bacterial resistance. In *Advances in applied microbiology*, volume 70, pages 217–248. Elsevier, 2010.

[88] P. Jagers. The growth and stabilization of populations. *Statistical Science*, 6(3):269–274, 1991.

[89] I. Jasser, I. Kostrzewska-Szlakowska, J. Ejsmont-Karabin, K. Kalinowska, and T. Węgleńska. Autotrophic versus heterotrophic production and components of trophic chain in humic lakes: the role of microbial communities. *Pol. J. Ecol*, 57(3):423–439, 2009.

[90] S. Jiang and J. Paul. Significance of lysogeny in the marine environment: studies with isolates and a model of lysogenic phage production. *Microbial ecology*, 35(3-4):235–243, 1998.

[91] D. R. Kadavy, J. J. Shaffer, S. E. Lott, T. A. Wolf, C. E. Bolton, W. H. Gallimore, E. L. Martin, K. W. Nickerson, and T. A. Kokjohn. Influence of infected cell growth state on bacteriophage reactivation levels. *Appl. Environ. Microbiol.*, 66(12):5206–5212, 2000.

[92] A. D. Kaiser. Mutations in a temperate bacteriophage affecting its ability to lysogenize *Escherichia coli*. *Virology*, 3(1):42–61, 1957.

[93] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of London. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[94] A. Khan and L. M. Wahl. Quantifying the forces that maintain prophages in bacterial genomes. *Theoretical Population Biology*, 2019.

[95] A. L. Koch. Oligotrophs versus copiotrophs. *Bioessays*, 23(7):657–661, 2001.

[96] R. Kolter, D. A. Siegele, and A. Tormo. The stationary phase of the bacterial life cycle. *Annual review of microbiology*, 47(1):855–874, 1993.

[97] E. V. Koonin and V. V. Dolja. A virocentric perspective on the evolution of life. *Current opinion in virology*, 3(5):546–557, 2013.

[98] P. Kourilsky. Lysogenization by bacteriophage lambda. *Molecular and General Genetics MGG*, 122(2):183–195, 1973.

[99] C.-H. Kuo and H. Ochman. Deletional bias across the three domains of life. *Genome biology and evolution*, 1:145–152, 2009.

[100] E. Kutter, D. De Vos, G. Gvasalia, Z. Alavidze, L. Gogokhia, S. Kuhl, and S. T. Abedon. Phage therapy in clinical practice: treatment of human infections. *Current pharmaceutical biotechnology*, 11(1):69–86, 2010.

[101] S. J. Labrie and S. Moineau. Abortive infection mechanisms and prophage sequences significantly influence the genetic makeup of emerging lytic lactococcal phages. *Journal of bacteriology*, 189(4):1482–1487, 2007.

[102] R. C. Lacy. Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation biology*, 1(2):143–158, 1987.

[103] E. Landaw and J. DiStefano 3rd. Multiexponential, multicompartmental, and noncompartmental modeling. II. Data analysis and statistical considerations. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 246(5):R665–R677, 1984.

[104] H. Lehnherr, E. Maguin, S. Jafri, and M. B. Yarmolinsky. Plasmid addiction genes of bacteriophage P1: *doc*, which causes cell death on curing of prophage, and *phd*, which prevents host death when prophage is retained. *Journal of molecular biology*, 233(3):414–428, 1993.

[105] R. E. Lenski and B. R. Levin. Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. *The American Naturalist*, 125(4):585–602, 1985.

[106] R. Leplae, G. Lima-Mendez, and A. Toussaint. ACLAME: a CLassification of Mobile genetic Elements, update 2010. *Nucleic acids research*, 38(suppl_1):D57–D61, 2010.

[107] T. J. Lerner and P. Model. The "steady state" of coliphage f1: DNA synthesis late in infection. *Virology*, 115(2):282–294, 1981.

[108] G. Li, M. H. Cortez, and J. S. Weitz. When be temperate: On the fitness benefits of lysis vs. lysogeny. *bioRxiv*, page 709758, 2019.

[109] M. Lieb. Studies on lysogenization in *Escherichia coli*. In *Cold Spring Harbor symposia on quantitative biology*, volume 18, pages 71–73. Cold Spring Harbor Laboratory Press, 1953.

[110] D. M. Lin, B. Koskella, and H. C. Lin. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. *World journal of gastrointestinal pharmacology and therapeutics*, 8(3):162, 2017.

[111] J. W. Little. Lysogeny, prophage induction, and lysogenic conversion. In *Phages*, pages 37–54. American Society of Microbiology, 2005.

[112] M. B. Łobocka, D. J. Rose, G. Plunkett, M. Rusin, A. Samojedny, H. Lehnherr, M. B. Yarmolinsky, and F. R. Blattner. Genome of bacteriophage P1. *Journal of bacteriology*, 186(21):7032–7068, 2004.

[113] A. J. Lotka. Théorie analytique des associations biologiques. *Hermann*, 1939.

[114] A. Lwoff. Lysogeny. *Bacteriological Reviews*, 17(4):269, 1953.

[115] K. H. Lynch, P. Stothard, and J. J. Dennis. Genomic analysis and relatedness of P2-like phages of the Burkholderia cepacia complex. *BMC genomics*, 11(1):599, 2010.

[116] J. Ma and Z. Ma. Epidemic threshold conditions for seasonally forced SEIR models. *Mathematical Biosciences & Engineering*, 3(1):161, 2006.

[117] A. Mai-Prochnow, J. G. K. Hui, S. Kjelleberg, J. Rakonjac, D. McDougald, and S. A. Rice. Big things in small packages: the genetics of filamentous phage and effects on fitness of their host. *FEMS Microbiology Reviews*, 39(4):465–487, 2015.

[118] D. D. Majewski, L. J. Worrall, and N. C. Strynadka. Secretins revealed: structural insights into the giant gated outer membrane portals of bacteria. *Current opinion in structural biology*, 51:61–72, 2018.

[119] C. R. Markussen and J. Distefano III. Evaluation of four methods for computing parameter sensitivities in dynamic system models. *Mathematical Biosciences*, 61(1):135–148, 1982.

[120] D. Marvin and B. Hohn. Filamentous bacterial viruses. *Bacteriological reviews*, 33(2):172, 1969.

[121] R. C. Matos, N. Lapaque, L. Rigottier-Gois, L. Debarbieux, T. Meylheuc, B. Gonzalez-Zorn, F. Repoila, M. de Fatima Lopes, and P. Serror. *Enterococcus faecalis* prophage dynamics and contributions to pathogenic traits. *PLoS genetics*, 9(6):e1003539, 2013.

[122] C. Maurice, C. d. Bouvier, R. De Wit, and T. Bouvier. Linking the lytic and lysogenic bacteriophage cycles to environmental conditions, host physiology and their variability in coastal lagoons. *Environmental microbiology*, 15(9):2463–2475, 2013.

[123] L. McDaniel and J. Paul. Effect of nutrient addition and environmental factors on prophage induction in natural populations of marine *Synechococcus* species. *Applied and environmental microbiology*, 71(2):842–850, 2005.

[124] R. McNerney. TB: the return of the phage. A review of fifty years of mycobacteriophage research. *The International Journal of Tuberculosis and Lung Disease*, 3(3):179–184, 1999.

[125] J. Metz. Adaptive dynamics. 2012.

[126] J. R. Meyer, D. T. Dobias, J. S. Weitz, J. E. Barrick, R. T. Quick, and R. E. Lenski. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science*, 335(6067):428–432, 2012.

[127] M. Middelboe. Bacterial growth rate and marine virus–host dynamics. *Microbial Ecology*, 40(2):114–124, 2000.

[128] A. Mira, H. Ochman, and N. A. Moran. Deletional bias and the evolution of bacterial genomes. *Trends in Genetics*, 17(10):589–596, 2001.

[129] J. E. Mittler. Evolution of the genetic switch in temperate bacteriophage. I. Basic theory. *Journal of theoretical biology*, 179(2):161–172, 1996.

[130] K. Moebus. Marine bacteriophage reproduction under nutrient-limited growth of host bacteria. I. Investigations with six phage-host systems. *Marine Ecology Progress Series*, 144:1–12, 1996.

[131] S. Moineau, S. Pandian, and T. R. Klaenhammer. Evolution of a lytic bacteriophage via dna acquisition from the *Lactococcus lactis* chromosome. *Appl. Environ. Microbiol.*, 60(6):1832–1841, 1994.

[132] I. Moneim. Seasonally varying epidemics with and without latent period: a comparative

simulation study. *Mathematical medicine and biology: a journal of the IMA*, 24(1):1–15, 2007.

[133] J. Monod. The growth of bacterial cultures. *Annual review of microbiology*, 3(1):371–394, 1949.

[134] G. J. Morgan, G. F. Hatfull, S. Casjens, and R. W. Hendrix. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. *Journal of molecular biology*, 317(3):337–359, 2002.

[135] A. M. Motlagh, A. S. Bhattacharjee, F. H. Coutinho, B. E. Dutilh, S. R. Casjens, and R. K. Goel. Insights of phage-host interaction in hypersaline ecosystem through metagenomics analyses. *Frontiers in microbiology*, 8:352, 2017.

[136] A. Nadeem and L. M. Wahl. Prophage as a genetic reservoir: promoting diversity and driving innovation in the host community. *Evolution*, 71(8):2080–2089, 2017.

[137] K. Nordström and S. J. Austin. Mechanisms that contribute to the stable segregation of plasmids. *Annual review of genetics*, 23(1):37–69, 1989.

[138] A. Novick. Growth of bacteria. *Annual Reviews in Microbiology*, 9(1):97–110, 1955.

[139] A. Novick and L. Szilard. Description of the chemostat. *Science*, 112(2920):715–716, 1950.

[140] N. Obeng, A. A. Pratama, and J. D. van Elsas. The significance of mutualistic phages for bacterial ecology and evolution. *Trends in microbiology*, 24(6):440–449, 2016.

[141] A. B. Oppenheim, O. Kobiler, J. Stavans, D. L. Court, and S. Adhya. Switches in bacteriophage lambda development. *Annual Review of Genetics*, 39:409–429, 2005.

[142] S. P. Otto and T. Day. *A biologist's guide to mathematical modeling in ecology and evolution*. Princeton University Press, 2011.

[143] Z. Patwa and L. Wahl. Fixation probability for lytic viruses: the attachment-lysis model. *Genetics*, 180(1):459–470, 2008.

[144] J. H. Paul. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *The ISME journal*, 2(6):579–589, 2008.

[145] J. P. Payet and C. A. Suttle. To kill or not to kill: the balance between lytic and lysogenic viral infection is driven by trophic status. *Limnology and Oceanography*, 58(2):465–474, 2013.

[146] J. S. Poindexter. Oligotrophy. In *Advances in microbial ecology*, pages 63–89. Springer, 1981.

[147] D. Posny and J. Wang. Computing the basic reproductive numbers for epidemiological models in nonhomogeneous environments. *Applied Mathematics and Computation*, 242:473–490, 2014.

[148] A. Pradeep Ram and T. Sime-Ngando. Resources drive trade-off between viral lifestyles in the plankton: evidence from freshwater microbial microcosms. *Environmental microbiology*, 12(2):467–479, 2010.

[149] K. M. Pruitt and D. N. Kamau. Mathematical models of bacterial growth, inhibition

and death under combined stress conditions. *Journal of Industrial Microbiology*, 12(3-5):221–231, 1993.

[150] M. Ptashne. A genetic switch: phage and higher organisms. *Cell Press & Blackwell Scientific Publications*, 1992.

[151] M. Ptashne. A genetic switch: phage lambda revisited. *Cold Harbor Spring*, 421, 2004.

[152] N. Ravin and D. Lane. Partition of the linear plasmid N15: interactions of N15 partition functions with the *sop* locus of the F plasmid. *Journal of bacteriology*, 181(22):6898–6906, 1999.

[153] N. V. Ravin. Mechanisms of replication and telomere resolution of the linear plasmid prophage N15. *FEMS microbiology letters*, 221(1):1–6, 2003.

[154] V. Ravin and M. Shulga. Evidence for extrachromosomal location of prophage N15. *Virology*, 40(4):800–807, 1970.

[155] D. Refardt and P. B. Rainey. Tuning a genetic switch: experimental evolution and natural variation of prophage induction. *Evolution*, 64(4):1086–1097, 2010.

[156] W. E. Ricker. Stock and recruitment. *Journal of the Fisheries Board of Canada*, 11(5):559–623, 1954.

[157] F. Rodriguez-Valera, A.-B. Martin-Cuadrado, B. Rodriguez-Brito, L. Pašić, T. F. Thingstad, F. Rohwer, and A. Mira. Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology*, 7(11):828, 2009.

[158] D. Roszak and R. Colwell. Survival strategies of bacteria in the natural environment. *Microbiological reviews*, 51(3):365, 1987.

[159] V. N. Rybchin and A. N. Svarchevsky. The plasmid prophage N15: a linear dna with covalently closed ends. *Molecular microbiology*, 33(5):895–903, 1999.

[160] M. A. Safi, M. Imran, and A. B. Gumel. Threshold dynamics of a non-autonomous SEIRS model with quarantine and isolation. *Theory in Biosciences*, 131(1):19–30, 2012.

[161] J. E. Samson, A. H. Magadán, M. Sabri, and S. Moineau. Revenge of the phages: defeating bacterial defences. *Nature Reviews Microbiology*, 11(10):675–687, 2013.

[162] S. E. Schoustra, A. M. McCollum, K. N. Garner, and B. R. Levin. The population dynamics and ecology of temperate phage: a killing and recruiting allelopath, 2004.

[163] H. Schrader, J. Schrader, J. Walker, N. Bruggeman, J. Vanderloop, J. Shaffer, and T. Kokjohn. Effects of host starvation on bacteriophage dynamics. *Bacteria in oligotrophic environments*, 1997.

[164] H. S. Schrader, J. O. Schrader, J. J. Walker, T. A. Wolf, K. W. Nickerson, and T. A. Kokjohn. Bacteriophage infection and multiplication occur in *Pseudomonas aeruginosa* starved for 5 years. *Canadian journal of microbiology*, 43(12):1157–1163, 1997.

[165] S. Schrag and J. Mittler. Host-parasite coexistence: the role of spatial refuges in stabilizing bacteria-phage interactions. *The American Naturalist*, 148(2):348–377, 1996.

[166] M. Sengupta and S. Austin. Prevalence and significance of plasmid maintenance

functions in the virulence plasmids of pathogenic bacteria. *Infection and immunity*, 79(7):2502–2509, 2011.

[167] J. Shan, S. Korbsrisate, P. Withatanung, N. L. Adler, M. R. Clokie, and E. E. Galyov. Temperature dependent bacteriophages of a tropical bacterial pathogen. *Frontiers in microbiology*, 5:599, 2014.

[168] J. W. Shapiro and P. E. Turner. Evolution of mutualism from parasitism in experimental virus populations. *Evolution*, 72(3):707–712, 2018.

[169] J. W. Shapiro, E. S. Williams, and P. E. Turner. Evolution of parasitism and mutualism between filamentous phage M13 and *Escherichia coli*. *PeerJ*, 4:e2060, 2016.

[170] F. R. Sharpe and A. J. Lotka. A problem in age-distribution. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 21(124):435–438, 1911.

[171] V. Sinha, A. Goyal, S. L. Svenningsen, S. Semsey, and S. Krishna. In silico evolution of lysis-lysogeny strategies reproduces observed lysogeny propensities in temperate bacteriophages. *Frontiers in microbiology*, 8:1386, 2017.

[172] H. L. Smith. Competitive coexistence in an oscillating chemostat. *SIAM Journal on Applied Mathematics*, 40(3):498–522, 1981.

[173] H. L. Smith and P. Waltman. *The theory of the chemostat: dynamics of microbial competition*, volume 13. Cambridge university press, 1995.

[174] J. M. Smith. *Evolution and the Theory of Games*. Cambridge university press, 1982.

[175] H. Spencer. *The Principles of Biology*, volume 2. Williams and Norgate, 1864.

[176] F. M. Stewart and B. R. Levin. The population biology of bacterial viruses: why be temperate. *Theoretical population biology*, 26(1):93–117, 1984.

[177] W. J. Stewart. *Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling*. Princeton university press, 2009.

[178] J. M. Sturino and T. R. Klaenhammer. Engineered bacteriophage-defence systems in bioprocessing. *Nature Reviews Microbiology*, 4(5):395–404, 2006.

[179] S. Sun and X. Zhang. Asymptotic behavior of a stochastic delayed chemostat model with nonmonotone uptake function. *Physica A: Statistical Mechanics and its Applications*, 512:38–56, 2018.

[180] A. Szczepankowska. Role of CRISPR/Cas system in the development of bacteriophage resistance. In *Advances in virus research*, volume 82, pages 289–338. Elsevier, 2012.

[181] B. Tang and G. S. Wolkowicz. Mathematical models of microbial growth and competition in the chemostat regulated by cell-bound extracellular enzymes. *Journal of Mathematical Biology*, 31(1):1–23, 1992.

[182] T. F. Thingstad. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography*, 45(6):1320–1328, 2000.

[183] M. Touchon, A. Bernheim, and E. P. Rocha. Genetic and life-history traits associated with the distribution of prophages in bacteria. *The ISME journal*, 10(11):2744, 2016.

[184] A. Towle. *Modern Biology*. Holt, Rinehart and Winston Inc, New York, 1989.

[185] P. E. Turner and S. Duffy. Evolutionary ecology of multiple phage adsorption and in-
      fection. *Bacteriophage ecology: population growth, evolution, and impact of bacterial
      viruses*, pages 195–216, 2008.

[186] F. W. Twort. An investigation on the nature of ultra-microscopic viruses. *Lancet*,
      2:1241–1243, 1915.

[187] P. Van den Driessche and J. Watmough. Reproduction numbers and sub-threshold en-
      demic equilibria for compartmental models of disease transmission. *Mathematical bio-
      sciences*, 180(1-2):29–48, 2002.

[188] S. Vieira-Silva and E. P. Rocha. The systemic imprint of growth and its uses in ecological
      (meta) genomics. *PLoS Genetics*, 6(1):e1000808, 2010.

[189] P. L. Wagner and M. K. Waldor. Bacteriophage control of bacterial virulence. *Infection
      and immunity*, 70(8):3985–3993, 2002.

[190] L. M. Wahl, M. I. Betti, D. W. Dick, T. Pattenden, and A. J. Puccini. Evolutionary
      stability of the lysis-lysogeny decision: Why be virulent? *Evolution*, 73(1):92–98,
      2019.

[191] L. M. Wahl and A. Dai Zhu. Survival probability of beneficial mutations in bacterial
      batch culture. *Genetics*, 200(1):309–320, 2015.

[192] L. M. Wahl and T. Pattenden. Prophage provide a safe haven for adaptive exploration in
      temperate viruses. *Genetics*, 206(1):407–416, 2017.

[193] L. Wang, D. Fan, W. Chen, and E. M. Terentjev. Bacterial growth, detachment and cell size control on polyethylene terephthalate surfaces. *Scientific reports*, 5(1):1–11, 2015.

[194] W. Wang and X.-Q. Zhao. Threshold dynamics for compartmental epidemic models in periodic environments. *Journal of Dynamics and Differential Equations*, 20(3):699–717, 2008.

[195] X. Wang and N. P. Higgins. 'Muprints' of the *lac* operon demonstrate physiological control over the randomness of *in vivo* transposition. *Molecular microbiology*, 12(4):665–677, 1994.

[196] X. Wang, Y. Kim, Q. Ma, S. H. Hong, K. Pokusaeva, J. M. Sturino, and T. K. Wood. Cryptic prophages help bacteria cope with adverse environments. *Nature communications*, 1:147, 2010.

[197] X. Wang, Y. Kim, and T. K. Wood. Control and benefits of CP4-57 prophage excision in *Escherichia coli* biofilms. *The ISME journal*, 3(10):1164, 2009.

[198] J. Weitz. *Quantitative viral ecology: dynamics of viruses and their microbial hosts*. Princeton University Press, 2016.

[199] J. S. Weitz, S. J. Beckett, J. R. Brum, B. Cael, and J. Dushoff. Lysis, lysogeny and virus–microbe ratios. *Nature*, 549(7672):E1–E3, 2017.

[200] J. S. Weitz, T. Poisot, J. R. Meyer, C. O. Flores, S. Valverde, M. B. Sullivan, and M. E. Hochberg. Phage–bacteria infection networks. *Trends in microbiology*, 21(2):82–91, 2013.

[201] C. L. Wesley and L. J. Allen. The basic reproduction number in epidemic models with periodic demographics. *Journal of biological dynamics*, 3(2-3):116–129, 2009.

[202] S. W. Wilhelm and C. A. Suttle. Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. *Bioscience*, 49(10):781–788, 1999.

[203] S. Williamson, L. Houchin, L. McDaniel, and J. Paul. Seasonal variation in lysogeny as depicted by prophage induction in Tampa Bay, Florida. *Applied and environmental microbiology*, 68(9):4307–4314, 2002.

[204] S. Williamson and J. H. Paul. Nutrient stimulation of lytic phage production in bacterial populations of the Gulf of Mexico. *Aquatic Microbial Ecology*, 36(1):9–17, 2004.

[205] C. Winter, J. A. Garcia, M. G. Weinbauer, M. S. DuBow, and G. J. Herndl. Comparison of deep-water viromes from the Atlantic Ocean and the Mediterranean Sea. *PloS one*, 9(6):e100600, 2014.

[206] G. S. Wolkowicz. personal communication.

[207] G. S. Wolkowicz, X.-Q. Zhao, et al. *n*-species competition in a periodic chemostat. *Differential and Integral Equations*, 11(3):465–491, 1998.

[208] M. A. Woody and D. O. Cliver. Effects of temperature and host cell growth phase on replication of F-specific RNA coliphage Q beta. *Appl. Environ. Microbiol.*, 61(4):1520–1526, 1995.

[209] A. Wright, C. Hawkins, E. Änggård, and D. Harper. A controlled clinical trial of

a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa*; a preliminary report of efficacy. *Clinical otolaryngology*, 34(4):349–357, 2009.

[210] M. B. Yarmolinsky. Programmed cell death in bacterial populations. *Science*, 267(5199):836–838, 1995.

[211] L. You, P. F. Suthers, and J. Yin. Effects of *Escherichia coli* physiology on growth of phage T7 *in vivo* and *in silico*. *Journal of bacteriology*, 184(7):1888–1894, 2002.

[212] B. Youngren, L. Radnedge, P. Hu, E. Garcia, and S. Austin. A plasmid partition system of the P1-p7*par* family from the pMT1 virulence plasmid of *Yersinia pestis*. *Journal of bacteriology*, 182(14):3924–3928, 2000.

[213] U. Zielenkiewicz and P. Cegłowski. Mechanisms of plasmid stable maintenance with special focus on plasmid addiction systems. *Acta Biochimica Polonica*, 48(4):1003–1023, 2001.

Appendix A

Parameter values for System (4.1)

Parameter	Value	Parameter	Value
α	0.015 virus^{-1}day^{-1}	$R^{\star}$	6 resource·litre^{-1}
β	100 viruses	C_0	0.8
ξ	1 day^{-1}	ω	0.17 day^{-1}
ρ_B	1 day^{-1}	m	12.5 day^{-1}
κ	4 day^{-1}	a	1 resource·litre^{-1}
p	0.5	ζ	2.5 day^{-1}
ρ	1 day^{-1}	θ	0.5 litre·resource^{-1}
ε	1 resource· (cell·litre)$^{-1}$	w	0.05 day^{-1}

Table A.1: Parameter values used for numerical simulations of System (4.1), unless stated otherwise.

Appendix B

Shell-scripted bioinformatics pipeline

Here, we present the shell-scripted bioinformatics pipeline that accesses PHASTER's [7] web interface to download a Results file for each bacterial genome.

```bash
#!/bin/bash
while read accnum; do

  wget "http://phaster.ca/phaster_api?acc=${accnum}" -O tmp/tmp.out;
  awk  '{gsub(/\\n/,"\n")}1' tmp/tmp.out > tmp/${accnum}.txt;
  awk '{print $3 "," $4}' tmp/${accnum}.txt | grep -E ...
       "intact\(|incomplete\(|questionable\(" > tmp/tmp${accnum}.txt
  for score in intact questionable incomplete; do
    grep -E "${score}" tmp/tmp${accnum}.txt | wc -l >> ...
         output/iqi_${accnum}.txt
  for protein in terminase portal head injection tail protease ...
       transposase integrase lysis plate capsid lysin flippase; do
```

```
10        grep -E "${score}\S+${protein}" tmp/tmp${accnum}.txt | wc -l >> ...

              output/output_${accnum}.txt

11    done

12    done

13  done < accnums.txt

14  exit 0
```

The following short shell script searches the corresponding Results file for each prophage

length and stores it in a separate file that we read into Matlab to plot the histograms shown in

Chapter 5.

```
1  while read accnum; do

2

3  grep -E "Kb" tmp/${accnum}.txt | awk '{print $2 }' | grep -E "Kb" | ...

        awk '{gsub("Kb"," ")}1' >> output/lengths_${accnum}.txt

4

5  done < accnums_sm_slow.txt

6

7  exit 0
```

Appendix C

Results from model selection

The Akaike Information Criterion (AIC) [5] estimates the quality of a model and protects against overfitting. The lowers AIC is therefore the optimal model fit. To select the best model among the candidate models, we used the AIC defined as,

$$\text{AIC} = 2k - 2\log\left(\hat{L}\right) \tag{C.1}$$

where k is the number of free parameters, and $\log\left(\hat{L}\right)$ is the maximum log-likelihood. Equation (C.1) was first derived by Akaike [5].

The AIC may, however, insufficiently punish extra parameters if the sample size comparted to the number of parameters is small. To alleviate this potential problem, Burnham and Anderson [32] developed an improvement on Equation (C.1) called the corrected Akaike Information Criterion (AICc). Their correct version is given by

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n - (k+1)} \tag{C.2}$$

where k is once again the number of free parameters and n is the sample size.

While the lowest AIC (or AICc) value corresponds to the best fit, it is possible that several candidate models offer equivalently good fits. This would then correspond to models that cannot be rejected statistically. To address this issue, we compute the relative probability. Defining AIC_{min} as the lowest AIC of the candidate models, the relative probability [32] is defined for each candidate model as

$$R = \exp\left(\frac{AIC_{min} - AIC}{2}\right). \tag{C.3}$$

The best fit model will then have relative probability of one.

In Tables C.1 and C.2. we provide both AIC and AICc values, the log-likelihood and compute the relative probabilities using the AICc values. The difference in the number of parameters reflects how many Gaussian components were included in the incoming prophage distribution, $f(x)$; we define the number of Gaussian distributions used as g, where $g = 1, 2,$ or 3.

For each data set, the best fit model was found to be a mixed distribution of two Gaussian distributions in $f(x)$, meaning $k = 12$ free parameters were used. We note that in all testing of the fits, we did not test the effect that the inclusion of HGT may have on the best fit. This exclusion of the HGT process is done due to the findings of Khan and Wahl [94], where the rate of HGT was found to be extremely small in comparison to the other rates.

	Parameters	AIC	AICc	Log-likelihood	Relative Probability
1	12	4957.9482	4958.0877	−2466.9741	1
2	15	4958.9073	4959.7153	−2464.4533	0.6191
3	9	4965.0218	4965.1613	−2473.5109	0.0291

Table C.1: Number of parameters, AIC, AICc, log-likelihood and the corresponding relative probabilities for the fast-growing data set. The best fit model includes a mixed distribution to describe autonomous temperate phages ($g = 2$), degradation, induction and selection. No model was tested with HGT included, as this was shown in [94] to have little effect on the results.

	Parameters	AIC	AICc	Log-likelihood	Relative Probability
1	12	4201.6520	4202.2373	−2088.8260	1
2	9	4203.8107	4204.1466	−2092.9054	0.3398
3	15	4205.7795	4206.6869	−2087.8898	0.1270

Table C.2: Number of parameters, AIC, AICc, log-likelihood and the corresponding relative probabilities for the slow-growing data set. The best fit model includes a mixed distribution to describe autonomous temperate phages ($g = 2$), degradation, induction and selection. No model was tested with HGT included, as this was shown in [94] to have little effect on the results.

Appendix D

Sensitivity of the coefficients of variation

As stated in Chapter 5, the intractability of the steady state $y(x)$ and the associated $Y(x, \hat{\mathbf{p}})$ analytically means we need to approximate the partial derivatives of $\mathbf{s}(x_i, \hat{\mathbf{p}})$. To do so, we use first principles with a step-size of $h = \varepsilon p_i$, where p_i is the best fit parameter as given in Table 5.3. The results given in Table D.1 use $\varepsilon = 0.12$, in comparison to the results of Table 5.4 that use $\varepsilon = 0.1$.

	Coefficient of Variation(%)	
Parameter	Fast	Slow
α/r_D	3.2319	11.926
r_S/r_D	0.9238	0.0062
r_I/r_D	4.0269	0.0234
n_I	0.3853	0.2785

Table D.1: Coefficients of variation for best fit parameters of fast- and slow-growing data sets with $\varepsilon = 0.12$.

Tyler Pattenden

Personal info

Full Name: Tyler Christopher Pattenden

Education

University of Western Ontario 2015 - in progress
PhD, Applied Mathematics
- Expected graduation in August 2020
- Direct entry from Master's enrollment
- Supervisor: Dr. Lindi Wahl

University of Ontario Institute of Technology 2014 - 2015
Bachelor of Education
- Qualified to instruct Ontario grades 7 - 12 in mathematics, and science

University of Ontario Institute of Technology 2010 - 2014
Bachelor of Science (Hons.), Applied and Industrial Mathematics
- Honours Thesis Project: "*Small solutions of differential delay equations*" (supervised by Dr. P-L Buono)

Teaching

Instructor
- AM1201b – Calculus and Probability with Biological Applications (Winter 2019/2020)

Course Coordinator
- AM1201b – Calculus and Probability with Biological Applications (Winter 2017/18)

Graduate Student Teaching Assistant (UWO)
- AM3815a – Partial Differential Equations (Fall 2018)
- AM1411a – Linear Algebra for Engineers (Fall 2017)
- AM2270a – Applied Mathematics for Engineers II (Fall 2016)
- AM1201b – Calculus and Probability with Biological Applications (Winter 2016)
- CALC1000a – Calculus I (Fall 2015)

Undergraduate Student Teaching Assistant (UOIT)
- MATH1020 – Calculus II (Winter 2014)
- MATH1850 – Linear Algebra for Engineers (Fall 2013)

Publications

Books
1. Pattenden, T. (2019) "AM1201b Course Manual (2nd Edition)", UWO Bookstore.
   - Previously released a 1st edition version with Dr. Lindi Wahl in 2018

Papers in Refereed Journals and Collections
2. Wahl, L.M., Betti, M.I., Dick, D.W., Pattenden, T., and Puccini, A.J. (2019) "Evolutionary stability of the lysis-lysogeny decision: Why be virulent?", Evolution, 73(1): 92-98.

1. Wahl, L.M. and Pattenden, T. (2017) "Prophage provide a safe haven for adaptive exploration in temperate viruses", Genetics, 206: 407-416.

Presentations

Invited Talks

2. Pattenden, T. (2018) My life in the world of mathematical biology. UWO Undergraduate Society of Applied Mathematics Undergraduate Research Conference, London, Ontario.

1. Wahl, L.M. & Pattenden, T. (2017) Prophage provide a safe haven for adaptive exploration in temperate viruses. Undergraduate Society of Applied Mathematics Seminar Talk, London, Ontario.

Talks and Posters

6. Pattenden, T., & Wahl, L.M. (2020) The role of stationary phase in the lysis-lysogeny decision. The Second Joint Meeting of SIAM/CAIMS, Toronto, Ontario.
   **Note: Accepted to give lecture, but not delivered due to COVID-19 Pandemic conference cancellation.

5. Pattenden, T. & Wahl, L.M. (2019) In search of an optimal probability of lysogeny in a variable environment: the role of stationary phase. Annual Meeting of the Society of Mathematical Biology, Montreal, Quebec.

4. Pattenden, T. (2018) The effects of the environment on bacteriophage virulence. Annual Meeting of CAIMS, Toronto, Ontario.

3. Pattenden, T. (2018) Do variable environments promote virulence? A bacteriophage model. BioMath Days, Ottawa, Ontario.

2. Pattenden, T. & Wahl, L.M. (2017) How the environment affects the lysis-lysogeny decision. The Fourth AMMCS International Conference, Waterloo, Ontario.

1. Pattenden, T. & Wahl, L.M. (2017) The effect of environmental variability on the lysis-lysogeny decision. Annual Meeting of CAIMS, Halifax, Nova Scotia.

Conferences Attended

- Canadian Mathematics Education Study Group Annual Conference (2019), St. Francis Xavier Univeristy, Antigonish, NS.
  - Attended multiple sessions on mathematical pedagogy including a plenary on the inclusion of indigenous teachings in post-secondary mathematics
  - Group memeber of the "Problem-Based Learning in the Post-Secondary Classroom" working group led by Dr. Jeff Hooper (Acadia)
- Canadian Mathematics Education Study Group Annual Conference (2018), Quest University, Squamish, BC.
  - Group member of the "Developing the 21st Century Curriculum" working group led by Dr. Peter Taylor (Queens)
- First Year Mathematics Across Canada (2018), Fields Institute, Toronto, ON.
  - Assisted in the devleopment of a database of first year mathematics and statistics courses across Canada's post-secondary institutions

Awards & Honours

Graduate Student Teaching Assistant Award, UWO (2017)
Student Leadership Award - Respect, UOIT (2014)
Student Leadership Award - General, UOIT (2013)

Research Interests

- Introductory post-secondary mathematics pedagogy, with interests in how to build an effective transition into post-secondary education

- Active learning and problem-based learning techniques for large classes (> 250 students)

- Population genetics, including experimental evolution models

- Modeling the evolution of temperate bacteriophage, and the search for an optimal probability of lysogeny

- Bacteriophage "social" dynamics, including the interactions between phage and phage, and their interactions with host bacterial cells

Technical skills

Languages	English (native)
Programming Languages	Matlab, Maple, SageMath
Software	Microsoft Office, LaTeX, OWL, GradeScope, CrowdMark