The Role of Virus-Host Interactions in the Evolutionary Dynamics of Bacteriophage Populations

Michael Thomson Hunter

Supervisor: Dr. Diana Fusco

Department of Physics
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

Churchill College June 2022

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This thesis is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This thesis does not to exceed 60,000 words, including summary/abstract, tables, footnotes and appendices, in accordance with the Physics and Chemistry Degree Committee.

Michael Thomson Hunter
June 2022

Abstract

Bacteriophage (phage) are viruses which infect and replicate within bacteria cells. Phages replicate by finding a host bacteria cell, adsorbing to it, injecting their DNA, hijacking the host machinery to produce more phage, and finally lysing the cell to release the new phage. The evolutionary dynamics of phage populations are therefore strongly tied to their interactions with the host bacteria. In a spatial context, I was able to show that unavoidable physical interactions between the virus and host hinder viral dispersal, leading to a phage population in which stochastic fluctuations are much weaker. This allows beneficial mutations to establish more easily and deleterious ones to be purged, making the phage population more adaptable to changes in the environment. To facilitate better comparisons between theory and experiment in spatial contexts, I also developed a variety of new experimental techniques. These were aimed at addressing shortcomings of existing techniques, which are almost exclusively based on population averages and are carried out in well-mixed liquid cultures, making their applicability to spatial settings challenging. Finally, phages are also able to interact with *each other* via their host, by encoding superinfection-exclusion mechanisms which, after initial infection, prevent subsequent phage from successfully infecting that host. Using stochastic simulations, I show that in the long term such mechanisms limit the adaptive potential of the phage, by limiting the genetic diversity in the population and reducing the efficiency of selection. In the short term however, having such a mechanism provides a very large advantage over other phage in the population which lack such a mechanism, possibly explaining the ubiquity of such mechanisms in nature.

The Role of Virus-Host Interactions in the Evolutionary Dynamics of Bacteriophage Populations

Michael Hunter

Acknowledgements

I have it from multiple sources that this is the only part of my thesis that anyone, save a select few such as my supervisor, my examiners, and my parents are likely to read. My more cynical friends and colleagues caution me that even that may be expecting too much! I guess then, that I should really do my best to make sure I don't miss anyone that deserves thanks for helping me one way or another throughout this process.

First and foremost, I would like to thank my supervisor Diana. I think being the first student in a new lab brings with it its own set of challenges, as, I have no doubt, does being the person in charge of starting that lab. Despite this, or perhaps because of it, as I reach the end of my PhD I'm very happy and grateful to say that I genuinely could not have asked for a better supervisor. From walking around the grounds of Girton College during her maternity leave, newborn in tow, to the everyday willingness to respond to my questions and concerns, Diana has always been there for me. For this, I will be eternally grateful.

To my lab and office mates, as well as the wider BSS community past and present, thank you for your help and support throughout. From the scientific discussions in group meetings and offices, to the social chats during lunch and over drinks, you have helped me further my research and genuinely enjoy coming to work. Special thanks must go to Racha, who was there from (almost) the outset to brighten our very small group meetings, and offer me her support when it became clear that, despite embarking upon a PhD in viral evolution, I knew very little about biology.

I would also like to take a moment to acknowledge and thank my collaborators, i.e. the people that 'did stuff' that appears somewhere in this thesis. Their contributions will be noted explicitly throughout in detail, but in summary: Racha for her help in bacterial transformations, and for her images of sectoring during plaque growth; Nikhil for his stochastic simulations; Wolfram for his help preparing my first manuscript; Chris for his early work on phage adsorption protocols; Somenath for the use of his lab and resources for single-cell imaging; and Charlie for actually collecting those images.

Of the people who were less directly involved in this work, the biggest thanks of course goes to my parents for their unwavering support over the years. I am also grateful to the friends I have made in Cambridge, from those that I met in Churchill in my first year to all those that I have met along the way. Moving to a new place and starting anew is always difficult, but I was immediately made to feel welcome. Special thanks to Ang, Jannat, Ben, Arthur, and the members of One Man's Tree, who kept me going when things were tough. As we part ways and head off to whatever comes next, I'm sure these friendships will remain.

I must also finish by extending my thanks to EPSRC for supplying the funds to carry out this research, along with Churchill College and the Department of Physics whose support also made this work possible.

Table of contents

List of figures

But Mousie, thou art no thy-lane,
In proving foresight may be vain:
The best laid schemes o' Mice an' Men
Gang aft agley,
An' lea'e us nought but grief an' pain,
For promis'd joy!

from To A Mouse, by Robert Burns

Chapter 1

Introduction

Starting from some unknown initial life form, the process of evolution has created all of the biological diversity we see in the world today. At a basic level, evolution is the change of heritable characteristics or traits (such as morphology, physiology and behaviour) over time. The extent of such changes and the timescales involved can vary enormously, from the rapid increase in bacteria strains resistant to antibiotics, to the many million year evolution of mammals [1].

The traits of an individual are controlled by genes, which can come in different variants, known as 'alleles'. Each gene corresponds to a section of the genome, which consists of a nucleotide sequence of DNA.[1] During cell division the DNA is replicated, but this is an imperfect process, and errors can be introduced. These errors, or 'mutations,' could be a change to a single nucleotide, or could be the insertion, removal, or swapping of longer stretches of DNA. Crucially, these mutations are heritable, i.e. a cell with a mutated gene will pass this version of the gene on to its offspring.

The impact of mutations on the observable characteristics, or 'phenotype,' of an organism varies significantly. Some will be lethal, some will have a negative (or 'deleterious') effect on the ability to survive and reproduce, some will have a positive (or 'beneficial') effect, while most will be neutral, i.e. they do not affect one's ability to survive and reproduce (known as 'fitness') [2].

The effect of a mutation will often be dependent on the environment in which the individual carrying it lives. As much as physicists might wish it so, organisms do not live in a vacuum.[2] They live in an environment that contains different quantities and types of resources, which they must utilise to survive and reproduce. If, for instance,

[1] or RNA, in the case of RNA viruses, but for now we will ignore this distinction.
[2] nor are they spherical ...

a mutation occurs which allows the individual to make use of a previously untapped resource, that mutation would likely[3] be beneficial in an environment where that resource is abundant, but not in an environment where that resource was absent. Not only that, but the environment almost always contains other organisms that must use the same resources to ensure their own survival, leading to competition between organisms.

The way organisms interact with each other and with their environment will therefore shape their evolution, and indeed evolution will shape the way in which they interact. This is perhaps most true of viruses, which lack their own metabolism and so *require* a host to be able to reproduce. While many might argue that viruses are therefore not "alive," like organisms they replicate, posses genes, and undergo evolution.

Exploring how viral evolution is tied to the interactions the virus has with its host will be the overarching theme of this thesis. In particular, this thesis will focus on bacteriophage (phage), which are viruses that infect bacteria. Phage, alongside bacteria, offer a great model system to observe evolutionary dynamics at work in real time. A continuous four week experiment of phage growth will see the evolution of over 2000 generations of phage, owing to their reproduction time of approximately 20 minutes. To give this a human context, assuming an average age of 20 years to birth of first offspring, 2000 generations would take us far back enough to see the extinction of the Neanderthals [3].

In addition to the experimental and theoretical insights gained, this work also has important practical implications. For instance, anti-microbial resistance is acknowledged as a major clinical threat to humans, with deaths due to drug-resistant bacterial infections estimated to reach 10 million each year by 2050 [4]. Phage therapies offer a potential route out of this crisis, although one clear difference between phages and antibiotics is that phages are replicating, *evolving* entities. The safe and effective use of phages in treating bacterial infections will therefore require an understanding of how they will evolve in different settings that might be more representative of clinical environments.

[3] I say likely, because in principle the ability to utilise this new resource might come at some other cost, like the ability to utilise the original resources. If the original resource was even more abundant, or nutrient rich, then the net effect could be negative.

1.1 Key Concepts in Evolution

To understand the findings discussed throughout this thesis, it is first necessary to understand some key, fundamental concepts in evolution. As mentioned previously, at its core, biological evolution is the change in the heritable characteristics of a population over time. Such changes initially occur within a single individual, and the long term shift in frequency of that change is then (in asexual reproducing populations) governed by two main evolutionary processes: natural selection and genetic drift.

1.1.1 Natural Selection

The theory of natural selection was famously set out in detail by Charles Darwin in his book *On the Origin of Species* in the mid 19th century [5].[4] As was outlined earlier, broadly Darwin argues that some of the genetic variation that arises in a population will give rise to phenotypic differences that alter an individual's ability to survive and reproduce (known as fitness).[5] In each generation, successful parents may therefore differ from the population as a whole. He also posits that there is likely to be a heritable component to this, so that the offspring's phenotype matches its parents (i.e. heritable fitness). As this process is repeated over many generations, the characteristics of the population will gradually change, with the frequency of beneficial characteristics (i.e. those that increase one's ability to survive and reproduce) increasing, and the frequency of deleterious characteristics (i.e. those that decrease one's ability to survive and reproduce) decreasing. This is the basis of natural selection.

1.1.2 Genetic Drift

Genetic drift is the evolutionary process whereby the frequency of a given allele changes over time through random sampling (i.e. chance). To study genetic drift we often rely on idealised models that capture some biological aspects such as reproduction and death, while simplifying many of the specific details [6], for instance death could occur for various reasons that may be distinct in different demographics.

One of the most widely used models of genetic drift is known as the Wright-Fisher model. The Wright-Fisher model consists of discrete, non-overlapping generations,

[4] More completely, the work is titled *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*

[5] There are many other aspects to fitness, for instance, as outlined previously, the fitness benefit of a given allele is not absolute, but depends on the environmental conditions.

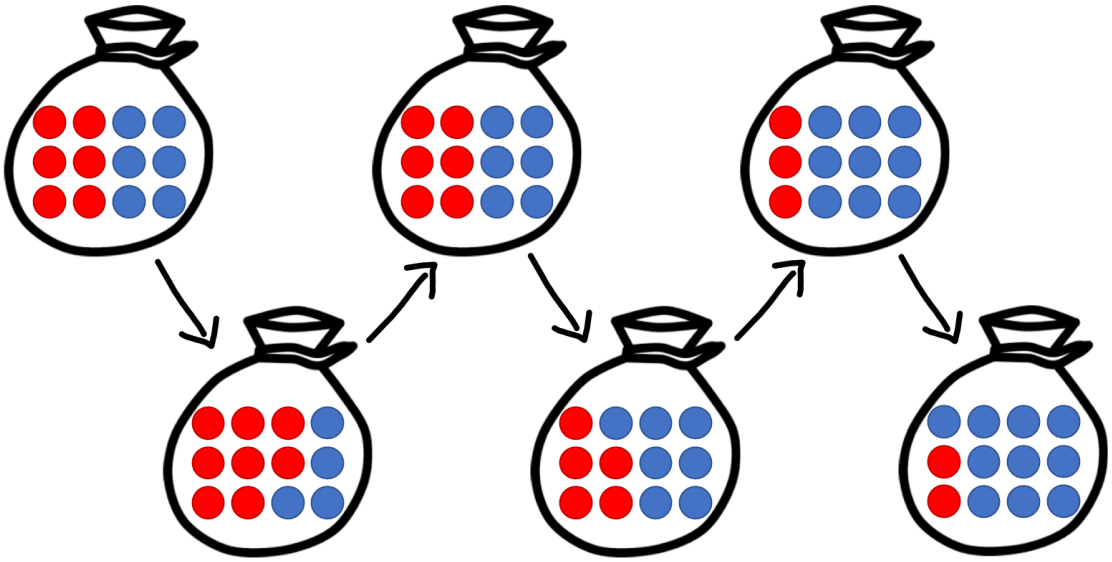


Fig. 1.1 An illustration of genetic drift. The balls represent organisms in a population, with the colours representing 2 selectively neutral alleles. Each bag represents successive generations of the population, with the frequency of the alleles fluctuating due to genetic drift. The process depicted here, and described in the main text is an example of the Wright-Fisher model of genetic drift [6].

where reproduction occurs via random sampling of the parent generation [7–9]. This concept is often illustrated by analogy with a bag of balls (Fig. 1.1). In this analogy, there is a bag which contains 12 balls, representing $N = 12$ haploid (where the cells contain one set of chromosomes) organisms in a population.[6] Initially, in the starting generation, half of the balls are red, and half are blue, representing two different alleles in the population. These alleles are selectively neutral with respect to each other, i.e. being red or blue does not impact that individual's chance of reproducing. In each new generation, the organisms reproduce at random. To represent this process, a ball is chosen at random from the original bag, and a ball of the same colour is deposited into a new bag (i.e., the ball is chosen with replacement). The new ball is meant to represent the offspring of the original organism. This process is then repeated until there are 12 balls in the second bag, representing the second generation of the population. Unless exactly 6 red balls and 6 blue balls were chosen, a random shift will have occurred in the frequency of the two variants. More formally, if we say that the frequency of one allele (red) is p and the frequency of the other allele (blue) is $1 - p$, then in a population of N individuals the probability that k copies of a given allele (in

[6]For most of this discussion we will focus on haploid organisms as it should be simpler to explain, and will be more relevant to the phage studied throughout this thesis.

this case red) survive until the next generation can be determined from the binomial distribution

$$\frac{N!}{k!(N-k)!}p^k(1-p)^{N-k}. \tag{1.1}$$

Alternative models of genetic drift exist, most notably the Moran model [10]. Unlike the Wright-Fisher model, the Moran model has overlapping generations. In the Moran model, in each timestep one individual is chosen to reproduce and one individual is chosen to die, thereby creating a population with constant size in which the number of copies of a given allele (like red or blue balls) can go up or down by one, or stay the same. For now, we will continue with the Wright-Fisher model as our illustrative example as it is computationally faster, since each new generation requires only one computational timestep, rather than many in the Moran model. It is worth noting, however, that in practice both models produce very similar results [6–10].

As the process of random sampling is repeated over many successive generations, the frequency of each allele (colour) will fluctuate, with these fluctuations being the essence of genetic drift. Multiple realisations of this process illustrate that the fluctuations are much higher in smaller populations (Fig. 1.2). It can also be seen that, particularly when fluctuations are high, it is possible that in one generation either no blue or red balls are selected to reproduce. In this instance, one of the alleles has gone extinct, and cannot be present in future generations, whereas the other allele is said to have reached fixation. This can happen very quickly in small populations because a small difference in the number of offspring corresponds to a large change in frequency. If no new mutation occurs, genetic drift will always eventually lead to the fixation of one allele and the extinction of the others.

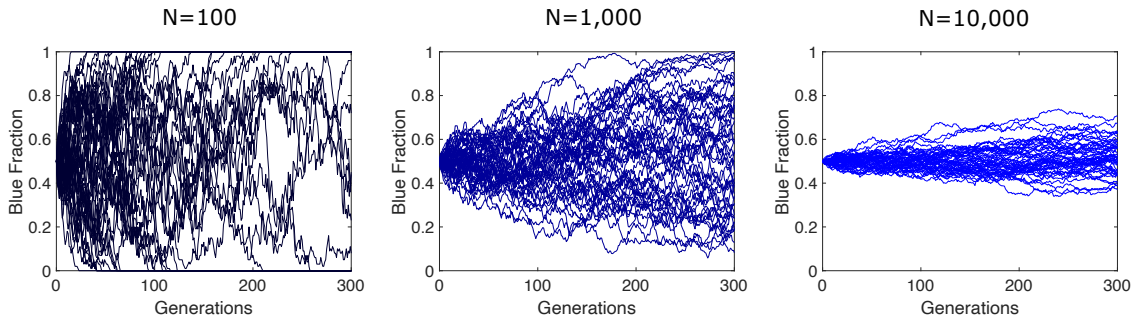


Fig. 1.2 An illustration of different realisations of the situation shown in Fig. 1.1, where the fraction of blue balls is plotted as a function of time. Each line represents one independent realisation of the algorithm described. Each panel represents a different population size N. It can be clearly seen that fluctuations are reduced as the population size increases.

A quantitative method of characterising the genetic diversity in such a population is through the heterozygosity (Fig. 1.3). In a biallelic population (i.e. a population with two alleles) like we have here, the heterozygosity H is defined as [6–10]

$$H = 2\langle f(1-f)\rangle, \tag{1.2}$$

where f and $(1-f)$ represent the frequencies of the two neutral alleles in the population (i.e. the frequency of the red and blue balls respectively), and $\langle \dots \rangle$ indicates the average over independent simulations. $H(t)$ can be understood to be the time-dependent probability that two balls chosen from the population have different colours.

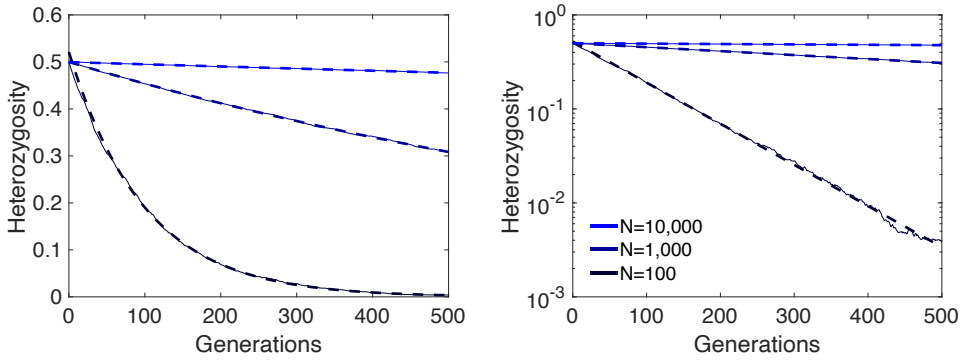


Fig. 1.3 Plots of the average heterozygosity H as a function of time for three different population sizes N. It can be seen that the diversity declines faster when the population size N is smaller. Data is taken as the average of 500 individual simulations, and dashed lines represent exponential decay with decay rate $1/N_e$. In this instance $N = N_e$.

We can then ask the question, how do we expect $H(t)$ to behave? In our set up we have a haploid population of N individuals, and say at time t we have a heterozygosity H_t. Assuming there are no mutations, what do we expect H_{t+1} to be? If we randomly sample two individuals with replacement in generation $t+1$, the probability that they will have the same parent will be $1/N$, and the probability they will have the different parents will be $(1-1/N)$. In the case where their parents are different, what then is the probability that those two parents will carry different alleles? This is simply given by the heterozygosity at time t, i.e. H_t. The heterozygosity in generation $t+1$ will then be given by

$$H_{t+1} = \frac{1}{N} \times 0 + \left(1 - \frac{1}{N}\right) \times H_t. \tag{1.3}$$

It follows from this that, if at time $t = 0$ the heterozygosity is H_0, we can write that

$$H(t) = \left(1 - \frac{1}{N}\right)^t H_0. \tag{1.4}$$

If we then assume that our population is large, i.e. $1/N \ll 1$, at long times we can approximate this decay as

$$H(t) = H_0 e^{-t/N}. \tag{1.5}$$

It can be seen in Fig. 1.3 that the heterozygosity H decays exponentially at long times due to genetic drift so that $H(t) = H_0 e^{-t/N_e}$ as expected, with the decay rate in units of generations being expressed in terms of an effective population size $1/N_e$ [6–10]. The effective population size can be considered to be the number of individuals in the population that on average are able to contribute to the next generation. In the Wright-Fisher model the effective population size is equal to the size of the actual population $N = N_e$, although this doesn't need to be the case. Indeed, in nature the effective population size is often much smaller than the size of the actual population [7, 8, 11]. An alternative way of thinking about the effective population size is therefore that it describes the size of an idealised population (e.g. a Wright-Fisher type population) in which the genetic drift occurs at the same rate as in the real population under consideration.

There are various reasons that the effective population size is often smaller than the actual population size, such as individuals in the population which are unable to reproduce, or individuals whose progeny are unable to survive. To illustrate how this might be the case with a perhaps unrealistic but more concrete example, imagine a population of $N = 1000$ deer, where there is $N_f = 1$ female and $N_m = 999$ males.[7] In this scenario, the number of deer that contribute to the next generation, i.e. the effective population size N_e, will clearly be smaller than 1000, as the female cannot carry the offspring of all 999 of the males.

In addition to heterozygosity, there are other properties and statistics that can be used to monitor the genetic diversity in populations. One of the more common is the site frequency spectrum, which summarises the distribution of allele frequencies of a given set of loci in the population. This approach has been employed to experimentally study the spatial growth of yeast colonies [12], and to theoretically study the genealogical

[7] I appreciate that sexual reproduction adds an extra layer of complexity compared to previous analogies, and changes the mathematics of genetic drift slightly, but I think it is an easier example to illustrate why the effective population size might be smaller than the actual population size.

structure of pulled and pushed waves [13]. In this thesis we will focus on heterozygosity as it is a simpler summary statistic, which will be useful for building a baseline understanding of viral evolutionary dynamics. It is also a useful measure to consider for simple experimental setups, particularly in a spatial context. As we shall see in Chapter. 3, we can label two strains with different fluorescent markers (representing two different alleles) that can be tracked using fluorescence microscopy to measure the heterozygosity, without the need to sequence the full genome.

1.1.3 Spatial Range Expansions

In biology, the spatial range of a population refers to that population's geographic, or spatial, limits. When a population expands and colonises new areas it is therefore said to undergo a spatial range expansion. This is a concept which has been studied extensively within the context of ecology, from the expansion of invasive plant species, to the range shifts of many organisms to higher altitudes and latitudes due to climate change [14–21]. However, spatial range expansions can also have a profound impact on the *evolution* of populations.

Individuals in a population undergoing a range expansion experience different conditions depending on their location, whereas individuals in a "well-mixed" setting (such as bacteria in liquid culture) experience the same conditions regardless of position. This is because, in a range expansion where dispersal is limited, pioneering individuals near the expansion front benefit from access to new territory, where there is a higher concentration of resources and lack of competition relative to the region behind the front [22]. This same benefit is then afforded to the progeny of these individuals who, by virtue of their initial location, are more likely to be the pioneers of the next generation. As a result, the population's genetic diversity declines at the front more rapidly than in an equivalent well-mixed population, as the gene pool at the front is dominated by the few individuals that happened to be there first, in what is known as the founder effect [22, 23].

For similar reasons, the genes of the pioneering individuals are also more likely to be carried along in the expansion front, in a phenomenon known as gene surfing [24]. The legacy of such effects can be observed in the world today, where the historical migration of humans out of the African continent is reflected in the approximately linear decrease in genetic diversity, as distance from the continent is increased [18, 20, 21]. Similarly, an observed latitudinal gradient in genetic diversity of multiple species

in the northern hemisphere is thought to reflect migration events prompted by repeated ice ages [16, 17].

The most widely used model to describe a species' colonisation of new habitats is the generalised Fisher-Kolmogorov (FKPP) equation [25–27]:

$$\frac{\partial n}{\partial t} = \frac{\partial}{\partial t}\left(D(n)\frac{\partial n}{\partial x}\right) + nr(n) + \sqrt{\gamma_n(n)\, n}\,\eta(x,t) \tag{1.6}$$

where $n(x,t)$ is the population density, $D(n)$ is the density-dependant rate of dispersal and $r(n)$ is the density-dependent growth rate. The final term accounts for demographic fluctuations, with $\eta(x,t)$ being a Gaussian noise term, and $\gamma_n(n)$ controlling the strength of the fluctuations.[8]

In travelling population waves, the dynamic of the colonisation front has generally been split into two distinct classes: "pulled" waves and "pushed" waves (see Fig. 1.4) [29] . Pulled waves occur when the growth rate $r(n)$ is maximal at small n, such as when growth is logistic [29, 30]:

$$r(n) = r_0\left(1 - \frac{n}{N}\right) \tag{1.7}$$

In the above, r_0 is the growth rate at low densities, and N is the carrying capacity, which defines the population density behind the front. Such waves are characterised by high growth rates at the leading edge of the front which *pull* the front forward. This is because the leading edge of the front is where the population density is lowest, and so individuals have better access to resources and unobstructed access to virgin territory. In this case, the speed of the front is independent of the form of $r(n)$, and is given by the Fisher velocity $v_F = 2\sqrt{D(0)r(0)}$, which is defined for any model where $r(0) > 0$ and $D(0) > 0$, even those which are not pulled [25–27].

It should be noted that, when considering a *discrete* population of N_{dis} individuals, rather than the mean-field population density n, the speed of the front v is lower than v_F, even in pulled waves, and approaches v_F as $N_{dis} \to \infty$ (i.e., as the mean-field limit is approached) [31]. The speed correction is proportional to $1/\log^2 N_{dis}$.

Contrasting with pulled waves, in pushed waves, maximal growth occurs at intermediate population densities, meaning it occurs behind the leading edge of the

[8] In much of the cited literature (e.g., [25, 28]) the n-dependence of the strength term is specified twice because they discuss demographic fluctuations γ_n and frequency fluctuations γ_f (i.e., genetic drift), both of which depend on the population density n, hence $\gamma_n(n)$ and $\gamma_f(n)$. While $\gamma_f(n)$ does not appear here, I opt to keep both n dependencies to minimise confusion for the reader if transitioning to the cited literature.

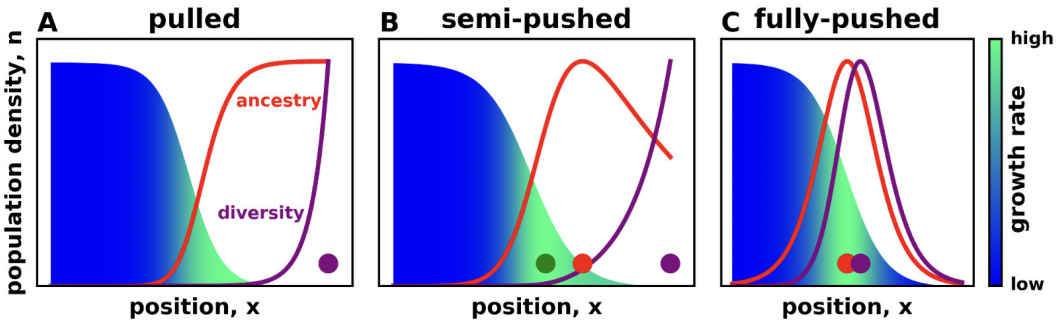


Fig. 1.4 This figure and its caption are reproduced from the Supplemental Information of Ref. [25]. "The three panels compare the spatial distribution of growth, ancestry, and diversity among pulled, semi- pushed, and fully-pushed waves. The color gradient shows how the per-capita growth rate changes along the wave front. The ancestry curve shows the distribution of the most recent common ancestor of the entire population at the front, which is the same as the probability of a neutral mutation arising at a particular location and then reaching fixation. The diversity curve shows the spatial distribution of the most recent common ancestor of two individuals sampled randomly from the front. In other words, this curve shows the probability that two ancestral lineages coalesce at a specific location at the front. Thus, the maximum of the diversity curve corresponds to the location that contributes most to the rate of diversity loss. The colored dots show the positions of foci of growth (green), ancestry (red), and diversity (purple). In both pulled and pushed waves these foci are colocalized, but they are spatially separated in semi-pushed waves. As a result of this, semi-pushed waves posses characteristics of both pulled and pushed waves."

front, and *pushes* the front forward. Pushed waves often result from a cooperative benefit at intermediate population densities, known as Allee effect [32], which can occur for a variety of reasons, such as anti-predator vigilance, reduction of inbreeding or co-operative feeding [33, 34].

The Allee effect that causes pushed waves is usually accounted for by modifying the equation for $r(n)$ to include a cooperative term [25]

$$r(n) = r_0 \left(1 - \frac{n}{N}\right) \left(1 + B\frac{n}{N}\right), \tag{1.8}$$

where B quantifies the strength of the cooperativity. For this model, exact solutions are known for the expansion speed and front profile [25]. Expansions are pulled for $B \leq 2$, and the speed is given by v_F, i.e. it is independent of B. For $B > 2$, expansions

are pushed, with the speed being given by [25]:

$$v = \sqrt{\frac{r_0 DB}{2}} \left(1 + \frac{2}{B}\right). \tag{1.9}$$

For $B > 2$, this expression is always greater than v_F, and increases with increasing B. No matter the form of $r(n)$, it is always true that pushed waves have a speed greater than v_F. It is therefore worth explicitly highlighting that cooperativity does not *always* increase the expansion speed (i.e. lead to pushed waves), even though it will lead to increased growth at high densities. It has recently been shown that not only co-operative growth, but also co-operative dispersal can lead to a density-dependent behaviour that results in the transition from pulled to pushed waves [28].

Because the dynamics of pulled waves are dominated by the very tip of the expansion, where the number of individuals is low, they are very susceptible to stochastic fluctuations (i.e. genetic drift). In pushed waves, however, genetic drift is reduced, due to the larger effective population that contributes to the dynamics [35]. Recent work has also identified the existence of two sub-classes within the pushed regime: "semi-pushed" and "fully pushed" (see Fig. 1.4) [25]. In semi-pushed waves, both the bulk of the wave and the leading edge contribute to the dynamics, resulting in a wave that has the kinetics of a fully pushed wave, but is highly susceptible to fluctuations, similar to pulled waves. The transitions between pulled, semi-pushed and fully pushed have been found to occur at specific wave speeds. Pulled expansions spread with a speed equal to the speed of the corresponding linearised system c_{linear} - the speed determined solely by the linear dynamics at the tip of the front - while pushed expansions spread faster [29]. The transition between semi-pushed and fully pushed occurs at a speed of $3/2\sqrt{2} \times c_{linear}$, with waves below this speed being semi-pushed, and above this speed being fully pushed [25]. These thresholds have been shown to be robust to the details of the population dynamic, suggesting that they can be used as universal conditions to distinguish between the different wave classes.

The existence and properties of pushed waves have been well studied experimentally with regard to single species expansions. In nature, Allee effects have been shown to explain key aspects observed in the invasions of both Eurasian gypsy moths and house finches in North America [36, 37], and spatial sorting, where traits that enhance dispersal accumulate at the front, resulted in a five-fold increase in invasion speed of cane toads in Australia [38, 39]. In the laboratory, studies have shown that the expansion dynamics in populations of the budding yeast *Saccharomyces cerevisae*

transition from pulled to pushed as growth becomes more cooperative [40], with a corresponding preservation of genetic diversity [41].

1.2 Bacteriophage

Now that we have introduced the general evolutionary principles which will govern our population, we need to introduce *what* our population is. As outlined at the start of this Chapter, this thesis will focus on bacteriophage (phage) - viruses that infect and replicate within bacteria. Phage are as ubiquitous as bacteria, and are thought to be the most plentiful and diverse organisms on Earth [42]. The discovery of phages is usually attributed to independent work carried out by Frederick W. Twort in 1915, and Félix d'Herelle in 1917 [43, 44]. Twort observed an agent that infected and killed bacteria, but it was only after the work by d'Herelle that Twort's observations were recognised as dealing with phage.

Since their discovery, phage have been considered as a potential method of treating bacterial infections in humans. Early applications included the treatment of acute intestinal diseases and skin infections [45, 46]. Interest in phage generally declined in the western world with the advent of antibiotics, although this interest has now been revived due to the growing concern surrounding drug resistant bacteria, with renewed focus on human phage therapies, alongside other applications in areas such as food production [47, 48].

Effectively utilising phage in any application is challenging, however, because an understanding of how phage grow and evolve in settings with explicit spatial structure (such as the human gut, or the surface of plants) is currently lacking. As has been highlighted in the previous section, the differing conditions experienced by individuals in a spatially structured population result in unique selective pressures on the population, which can drastically change the evolutionary dynamics of the population.

This thesis is therefore generally motivated by a desire to better understand how spatial structure alters the interactions between the phage and its host, and in turn how these interactions shape their evolution. As will be outlined in greater detail later in this Chapter, that means that there will be particular focus in this thesis on the growth of phage in plaques (clearings formed in lawns of bacteria by repeated cycles of phage infection). Later, we will also look at the impact of multiple phage infecting the *same* host cell. Such a process is likely common during spatial growth, due to the high local concentrations of phage relative to the number of available hosts, and

is particularly interesting in this context since phage in plaques are more likely to be in competition with clonal phage in their local vicinity than they would be in a well-mixed population where competition is global.

1.2.1 Replication

Phage typically replicate either through the lytic life cycle or the lysogenic life cycle. The lytic life cycle (Fig. 1.5) can be split into several stages. Initially, phage particles diffuse freely and are not yet attached to a host cell. The length of time phage spend in this state is determined by several environmental factors, such as host cell density, the rate of phage diffusion and the rate of phage adsorption [49]. Next, the phage adsorbs to the surface of the host cell, and injects its genetic material into the host. This process is highly specific, with phage only binding to certain receptor sites on the host's surface, such as oligosaccharides, lipopolysaccharides, or proteins [50].

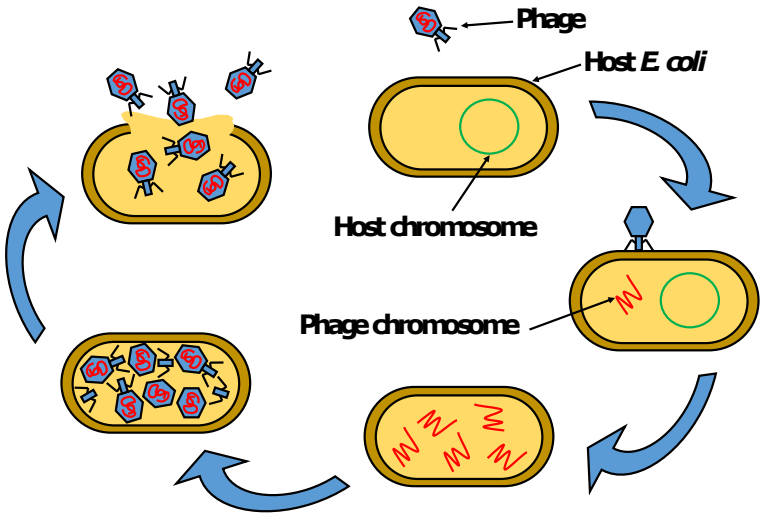


Fig. 1.5 The lytic life cycle. Initially, the phage is not attached to a host cell and is diffusing freely. The phage then adsorbs to the surface of the host cell and injects its genetic material, which then hijacks the internal machinery of the host cell. New phage are then produced and assembled inside the cell. The new phage are then released in the final step when the cell bursts, known as lysis.

Then follows the eclipse period, where infection has occurred but no progeny have yet been produced, as the phage is in the process of hijacking the host's cellular machinery. At the end of the eclipse period, new phage production starts inside the cell, with the production having been shown experimentally in an array of isogenic λ-phage

to be linear in time [51]. While the phage produced inside are fully constructed, they remain inside the cell until the final stage of the cycle, known as lysis.

Lysis is controlled by two components: a lysin (an enzyme capable of cleaving a key bond in the peptidoglycan matrix that forms the cell wall) and a holin (a protein which forms pores in the inner membrane to facilitate the lysin's access to the peptidoglycan matrix) [50]. Lysis occurs when the cell wall is breached, releasing the new phage progeny which are now free to diffuse onward to new hosts. Thus, the cycle is complete, and the infection spreads.

In the lysogenic cycle, instead of the host's cellular machinery being hijacked, the phage nucleic acid either forms an extra-chromosomal plasmid, or is integrated into the host's genome [50]. This phage genome (known as a prophage) remains dormant, and allows the host bacteria to live and replicate as normal. The prophage can then be passed on to future bacterial generations, where certain environmental stressors such as UV light or low nutrient conditions can cause it to be released, resulting in continued infection through the lytic cycle [50]. It should thus be noted, that phage which are capable of undergoing the lysogenic life cycle, known as temperate phage, remain capable of undergoing the lytic cycle. For this reason, the common equating of *lytic* phage with *virulent* phage should be avoided. This work instead follows the convention laid out by S. Abedon [42], which opts to refer to phage *only* capable of undergoing replication through the lytic life cycle as obligately lytic.

There also exists another type of phage known as filamentous phage. These phage are named for their long, thin filamentous shape. Unlike most phage which are pathogens, and kill their host bacteria, filamentous phage live in more co-operative relationships with their host bacteria [52]. Filamentous phage impose little burden on the host cell, and replicate without the need to lyse the host, instead being extruded out through the membrane. Filamentous phage can also offer benefits to their hosts, such as aiding the creation of biofilms, or modifying behaviours to provide motile activity to their hosts [52].

1.2.2 Bacteriophage T7

In this work, we will focus on phage T7 and its host *Escherichia coli* as a model system. Bacteriophage are classified by the International Committee on Taxonomy of Viruses (ICTV) according to morphology and nucleic acid. The phage used in this study is bacteriophage T7, a double stranded DNA virus of the order *Caudovirales* (tailed bacteriophage), and the family *Podoviridae* (short, non-contractile tail). Phage T7

consists of an icosahedral head structure, known as the capsid, which is connected to a short non-contractile tail. Contained within the capsid is the almost 40 kbp viral genome, which codes for 55 proteins [53]. Phage T7 is an obligately lytic phage.

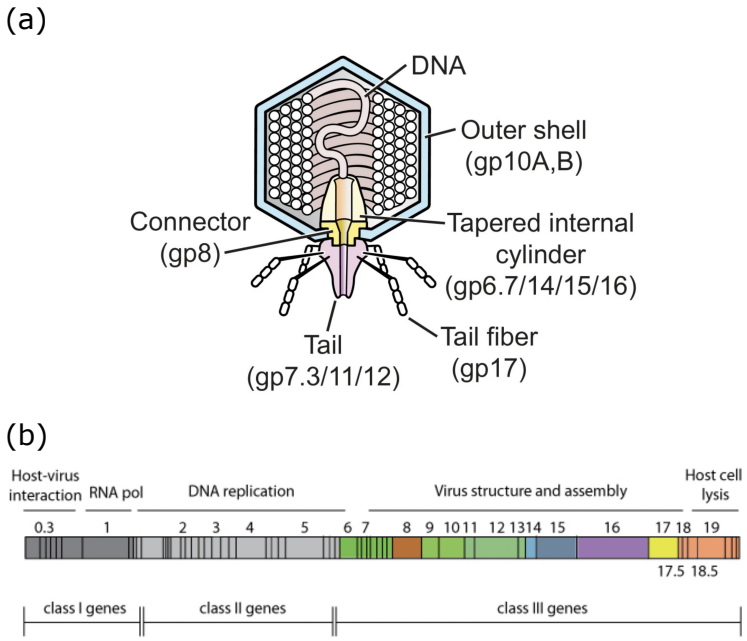


Fig. 1.6 (a): A schematic of the structure of bacteriophage T7, taken from Ref. [54]. (b): An annotated bacteriophage T7 genome, taken from *https://viralzone.expasy.org/518?outline=all_by_species*.

1.2.3 Plaques

It is possible to observe the spread of phage infection through susceptible host bacteria with the naked eye. When a small quantity of phage is inoculated onto a bacterial lawn (a mat of bacteria grown on an agar plate), the process of infection, reproduction and diffusion described previously results in a growing clearing of lysed (dead) cells. This clearing is known as a plaque (Fig. 1.7). Plaque based assays have long been used as a reliable and accurate method for determining the concentration of infective units in a viral suspension [55], as well as more recently a measure of viral fitness [56].

Alongside virus characterisation, the study of plaques also provides an interesting medium to study the growth (and evolution) of viruses. Previous work on plaques has typically centred on how virus-host interactions affect the speed at which the plaque grows. Many models of plaque growth exist, the oldest and simplest of which was proposed by Koch [57], who estimated that the speed of the spread was proportional

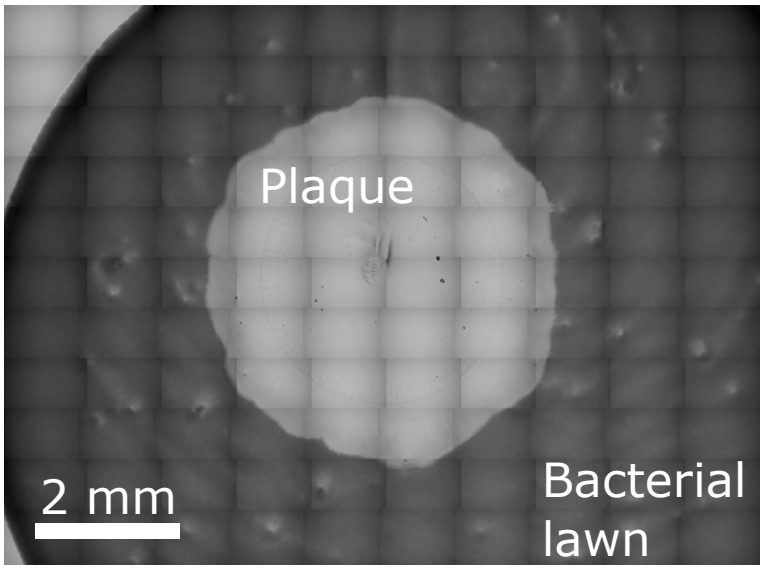


Fig. 1.7 An image of a T7 plaque. The plaque is visible as a clearing in the middle of a bacterial lawn, on an agar plate.

to $\sqrt{D/\tau}$, where D is the diffusion coefficient of the phage and τ is the latent period (the time in which phage are immobile inside the cell).

Yin and McCaskill Model

Later, Yin and McCaskill [58] constructed a reaction-diffusion system for the spread of phage T7 through host *E. coli*, and obtained the speed from travelling wave solutions. This system is laid out in Eq. 1.10:

$$\frac{\partial[B]}{\partial t} = -\underbrace{k_1[V][B]}_{\text{adsorption}} + \underbrace{k_{-1}[I]}_{\text{desorption}} \tag{1.10a}$$

$$\frac{\partial[I]}{\partial t} = \underbrace{k_1[V][B]}_{\text{adsorption}} - \underbrace{k_{-1}[I]}_{\text{desorption}} - \underbrace{k_2[I]}_{\text{lysis}} \tag{1.10b}$$

$$\frac{\partial[V]}{\partial t} = \underbrace{D\frac{\partial^2[V]}{\partial r^2}}_{\text{diffusion}} - \underbrace{k_1[V][B]}_{\text{adsorption}} + \underbrace{k_{-1}[I]}_{\text{desorption}} + \underbrace{Yk_2[I]}_{\text{lysis}} \tag{1.10c}$$

where $[B]$, $[I]$ and $[V]$ denote the concentrations of uninfected host bacteria, infected bacteria and phage at the leading edge of the front respectively. Rate constants of phage adsorption, desorption and host lysis are given by the parameters k_1, k_{-1} and k_2 respectively, while D represents the diffusion coefficient of the phage, and Y represents

the burst size (i.e. the quantity of new phage released at each lysis event). These equations are based on the assumptions that the host bacteria are immobilised in the agar, and that adsorption of phage to infected hosts is negligible.

It is worth taking a brief pause here to discuss the later of these assumptions. Some bacteriophage do indeed have mechanisms that prevent adsorption to already infected cells, usually by blocking receptor sites post-infection [59]. For instance, Bacteriophage T5 produces a lipoprotein (Llp) that is expressed at the beginning of infection, preventing superinfection by blocking its own receptor site (FhuA protein), and protecting newly produced phage from inactivation by binding to free receptors released by lysed cells [60, 61]. This mechanism is not universal however. Indeed, bacteriophage T7 - used both in this thesis and in Ref. [58] - has no such mechanism. The justification for this assumption must instead come from the fact that the speed estimates obtained from the model come from linearising the system of reaction-diffusion equations about the front of the expansion, where the number of infected cells tends to zero.[9] A further discussion on the use of the solution to the linearised model will come later in Sec. 1.4.

Density-Dependent Diffusion Coefficient

Yin and McCaskill also attempted to correct the free diffusion coefficient D, to take account of the hindrance that is posed by the bacterial cells. This manifests in the model as the free diffusion coefficient being replaced with an effective diffusion coefficient D_{eff} given by [58]:

$$D_{eff} = \frac{1-f}{1+\frac{f}{x}}D \tag{1.11}$$

where $f = B_0/B_{max}$ is the density of bacteria relative to the density at which diffusion is completely prohibited. The parameter x is a shape factor of the bacteria cells, which Yin and McCaskill assumed to equal 2 (i.e. spherical) [58], and Fort and Méndez later calculated from theory, based on the size of the cells along their axes of symmetry, to be equal to 1.67 [62]. The free diffusion coefficient D in the absence of bacteria is taken to be approximately the same for phage T7 as it is for phage P22, as they are of similar size and shape, and so takes a value of $D = 4\ \mu\text{m}^2/\text{s}$ in 1% agar [63, 64].

[9] No justification for this assumption is actually provided in Ref. [58], however this is the only sensible justification I can think of.

Equation 1.11 is based on Fricke's equation [65], which describes the diffusion of a solute through a suspension of diffusionally impermeable spheroids.

In Equation 1.11, it can be seen that when $f = 0$, i.e. when there is no bacteria, we have the simple case that $D_{eff} = D$. As the bacterial fraction is increased, D_{eff} is reduced until it reaches a value of 0 when $f = 1$, i.e. when all of the available space is occupied by bacteria and so the phage is unable to diffuse. Despite some observations that the diffusion coefficient *is* reduced in the presence of bacteria [66], the effective diffusion coefficient has yet to be robustly demonstrated experimentally, as noted by Abedon and Culler in a review of the topic [67].

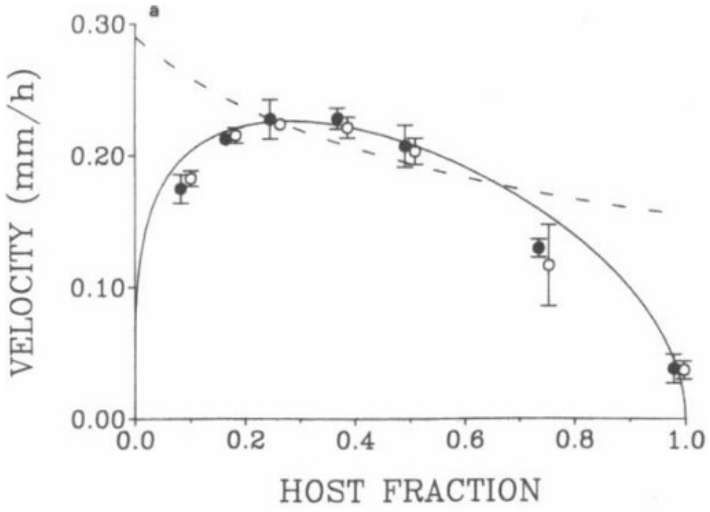


Fig. 1.8 A comparison of Koch's model (dashed line) and Yin and McCaskill's model (solid line) to experimental data. Yin and McCaskill's model accounts for hindered diffusion while Koch's does not. In addition, both models are fit to the data using two additional fitting parameters. Figure adapted from Ref. [58].

Nevertheless, Yin and McCaskill found that the inclusion of this density-dependent diffusion coefficient was necessary to qualitatively capture an observed reduction in the plaque expansion speed at high density (Fig. 1.8) [58]. It was also found that, in the limit where adsorption and desorption processes are fast relative to the death rate of infected cells, the effect of desorption on the speed is negligible. For this reason, future models do not include the terms in Eq. 1.10 which are proportional to k_{-1}.

You and Yin Model

You and Yin [68] subsequently supported the claim of travelling wave solutions through numerical simulations, although the speeds obtained by both this work and the work

by Yin and McCaskill were much larger than those observed experimentally [58]. It is important to note that in this study, the linearisation assumption was shown to break down in the limit of fast adsorption (again, this will be discussed further in Sec. 1.4).

Fort and Méndez Model

Subsequent work by Fort and Méndez [62] tried to resolve the discrepancy between prediction and experiment by taking account of the latent period, i.e. the "delay time between the moment when a virus adsorbs into a cell and that in which the cell dies and the new generation of viruses begins to spread [62]." This was achieved by modifying Eq. 1.10c to its hyperbolic generalisation given by:

$$\frac{\partial[V]}{\partial t} + \frac{\tau}{2}\frac{\partial^2[V]}{\partial t^2} = D\frac{\partial^2[V]}{\partial r^2} + F_V + \frac{\tau}{2}\frac{\partial F_V}{\partial t}. \tag{1.12}$$

This modification attempts to account for the latent period by delaying the diffusion of phage after lysis. F_V is the virus growth function given by:

$$F_V = -k_1[V][B] + Yk_2[I]\left(1 - \frac{[I]}{[I]_{max}}\right) \tag{1.13}$$

It should be noted here that another key difference between this model and that of Yin and McCaskill is the use of the logistic growth term to describe the growth in the absence of uninfected cells - the final term in Eq. 1.13. In this model, the equivalent term in the equation governing the concentration of infected cells is modified in the same way:

$$\frac{\partial[I]}{\partial t} = k_1[V][B] - k_2[I]\left(1 - \frac{[I]}{[I]_{max}}\right). \tag{1.14}$$

The intuition behind the use of the hyperbolic generalisation is perhaps better understood in the context in which it was originally implemented, namely the neolithic transition from hunter-gatherer to agricultural economies in Europe [69]. Based on archaeological data, it had been concluded that European farming originated in the Near East, and then spread across Europe in a wave of expansion driven by human migration. In Ref. [69], it is assumed that there is a well-defined time scale between two successive migrations. In other words, they assume that families will have travelled to a new location, which may have taken some days or weeks, and then remained in that location for a period of "residence," which the authors assume to be about the

length of one generation, before moving again. In this context, the delay period τ should be understood to be approximately the time of residence.

While this description seems readily applicable to human migrations, I think it is perhaps less relevant to viral expansions. By modelling the delay in this way, the authors are equating the time which phage are incubated inside immobilised bacteria with the residence period between migrations. A key difference between the two scenarios, however, is that while humans are in residence at a given location they can and do reproduce (increase in population density), whereas while incubated inside bacteria, phage should not be capable of infecting new bacteria. The description of plaque growth in this model seems to imply that some host bacteria lyse immediately following infection, releasing new phage, and while those phage are unable to diffuse away from that location for a time τ, they are capable of immediately infecting other cells at that location, which is unrealistic.

Jones Model

Later, work by Jones *et al.* [70] adopted a different approach to this issue, and described a perfect delay model where all of the cells lyse at exactly the same time after infection. This was achieved by replacing the $k_2[I]$ term in the Yin and McCaskill model with $k_1[V](t-\tau)[B](t-\tau)$, where the concentrations are considered at time $(t-\tau)$. This is perhaps a more intuitive description of the delay, in that phage and uninfected bacteria are converted into infected cells, and after the incubation period τ, these infected cells are converted back into free phage that are able to diffuse onward and infect new hosts. In this description, there is no *delay* to the diffusion of the phage, because during incubation there effectively are *no phage*, only infected cells which by definition do not move.

de Rioja Model

Most recently work by de Rioja *et al.* [71] proposed a new model which combines both of the two previously discussed aspects: firstly, the model takes into account the higher order corrections proposed by Fort and Méndez, and adopted by several subsequent models (terms proportional to τ in Eq. 1.12) [62, 72, 73]. Secondly, a description of the cell death process, that takes into account the concentrations at time $t-\tau$, instead of through the use of logistic growth functions is used. This results in the terms $k_2[I]$ in Equation 1.10 being replaced with the term $k_2[I](t-\tau)$. Irrespective of my misgivings

about the approach taken by Fort and Méndez [62] compared to Jones *et al.* [70] when it comes to modelling the delay time τ, the adoption of *both* approaches in the *same* model is surely redundant. This would seem to imply that phage and uninfected bacteria will be converted to infected cells, there will be a delay time τ before the infected cells are converted to free phage, and then for a further period τ those phage will be unable to diffuse.

1.2.4 Multiple Infections

When phage are in a setting such that there are relative few bacteria compared to the number of phage, competition for viable host can lead to different strains or even species of phage superinfecting or co-infecting the same bacterial cell, ultimately resulting in the production of more than one type of phage [74–76]. In the following, we define infection terminology in line with Turner & Duffy [77], such that co-infection occurs when two or more phage have successfully infected a single bacteria, and superinfection occurs when there is a delay between infection by the first and second phage. Therefore, all cells which have been successfully superinfected can be said to be co-infected [77]. To account for different usages throughout the literature and across fields, we also refer to multiple infections, to indicate any case where multiple viruses exist within a single host simultaneously.

Interestingly, several phages have evolved mechanisms that prevent superinfection, known as superinfection-exclusion (Fig. 1.9). This can be achieved at the early stage of infection, by preventing further adsorption of phage, or at a later stage, by preventing the successful injection of subsequent phage DNA [59, 78]. For instance, as mentioned earlier bacteriophage T5 encodes a lipoprotein (Llp) that is synthesised by the host at the start of infection and prevents further adsorption events by blocking the outer membrane receptor site (FhuA protein) [60, 61]. Bacteriophage T4 encodes two proteins, Imm and Sp, that prevent superinfection by other T-even phages (including T4) by inhibiting the degradation of bacterial peptidoglycan, whose presence hinders the DNA transfer across the membrane [79, 80]. Bacteriophage T7 also possesses a superinfection exclusion mechanism, which acts to exclude superinfection by either T7 or T3 phage after adsorption, but before the genome becomes available for gene expression or replication [81, 82].

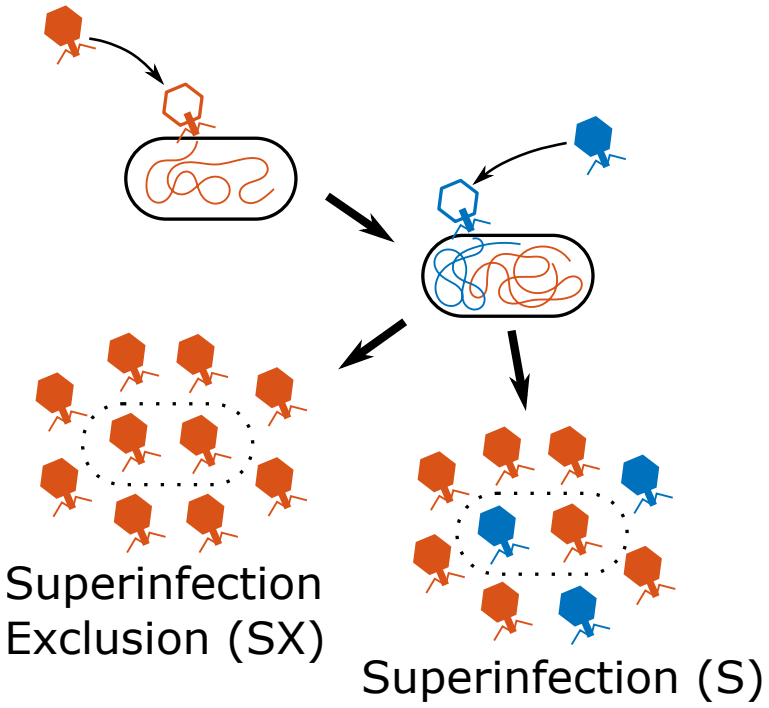


Fig. 1.9 An illustration of the difference between superinfection and superinfection-exclusion. When superinfection-exclusion occurs, on the first phage to infect is produced, while if superinfection occurs, both types of phage are produced when the host lyses.

1.3 The Impact of Virus-Host Interactions on Evolution

1.3.1 Life-History Parameters

The question of how virus-host interactions affect the evolution of bacteriophage has often been studied in the context of how changes to these interactions affect the fitness of the phage. In particular, much attention has been focused on the timing of lysis. Intuition might suggest that shorter lysis times would be favourable, as phage could more rapidly continue on to a new host and grow the population. This is borne out by experiments which have shown that in well-mixed liquid cultures, if the bacteria density is high, phage populations evolve to have shorter lysis times [83].

When bacteria density is low, however, it has been shown that phage can evolve to have *longer* lysis times [84, 85]. The intuition behind this stems from the fact that there is a positive linear relationship between lysis time and burst size [51], meaning that a reduction in lysis time generally results in the production of fewer phage. In the high bacteria density case, when fresh hosts are very abundant, this reduction in phage output is not significant, as the few phage that are produced are able to quickly find

new hosts to continue the infection. However, lysing hosts faster is not particularly useful if the phage are released into an environment where they will struggle to find new hosts, as in the case of low bacteria density cultures. In this scenario, it is more useful to release lots of phage, thereby increasing the chances that at least one of them finds a host quickly, even if it means taking longer to lyse the hosts. In essence, there's no point in lysing a host quickly if the new phage can't find a host to continue the infection. This tradeoff between lysis time and burst size has been shown to lead to an 'optimum' lysis time which depends on host density and other environmental conditions [86].

As with the well-mixed case, the linear correlation between lysis time and burst size has been shown to impose an optimum selection criteria during plaque growth [87]. This optimum is not always entirely predictable however. It has been shown, for instance, that when wild-type T7 are grown on a lawn of *E. coli* which expresses an essential early enzyme of the phage infection (T7 RNA polymerase), phage populations evolved with large deletions from the genome (up to 11%), including the gene for their own RNA polymerase [88]. These phage were able to outcompete the wild-type, and were shown to have both shorter lysis times *and* larger burst sizes, as they were able to more efficiently use the host's resources without the need to produce their own RNA polymerase.

Similar to the well-mixed case, in the spatial context of plaques intuition suggests that shorter lysis times, larger burst sizes, and faster adsorption rates should all be beneficial [67]. Spatial structure, however, can impose selection criteria that may initially seem counter-intuitive. For instance, spatial growth of phage can actually favour *lower* adsorption rates, due to a tradeoff between adsorption and dispersal i.e., the longer a phage goes without adsorbing to a host the further it is able to spread [89]. Such differences are perhaps not surprising given that spatial and well-mixed settings fundamentally select for different things: in simple terms, growth in well-mixed settings selects for greater ability to increase in number, whereas growth in plaques selects for greater ability to spread through space [67].

Even in more 'artificial' spatial settings with imposed migration schemes similar results can occur, albeit for different reasons. Kerr *et al.* employed a setup where phage T4 and its host *E. coli* were embedded in a 96-well plate, which imposed a meta-population structure that was perpetuated by serial transfer every 12 hours [90]. At each transfer, migration also occurred between wells, with this migration either being 'restricted,' meaning that migration could only occur between neighbouring

wells, or 'unrestricted,' meaning that migration could occur between all wells. Kerr *et al.* found that the two migration patterns selected for different viral strategies, termed 'prudent' and 'rapacious.' Rapacious phage are able to out-compete their prudent counterparts when in competition, but prudent phage are more productive in isolation. Restricted migration selects for prudent phage, while rapacious phage dominate when migration is unrestricted. The intuition behind this is in the fact that in the unrestricted case, all spatial structure is essentially destroyed every 12 hours. As a result, the rapacious type can limit the access of other phage to hosts by infecting and lysing them quickly, meaning prudent phage do not have a chance to grow. In the restricted case competition and resource availability is local, and so prudent use leads to higher productivity. For instance if phage increase their lysis time, the uninfected hosts have more time to replicate, thereby increasing availability of the resource and ultimately phage growth in the long run.

1.3.2 Intracellular Interactions

The question of how phage compete for hosts, and the evolutionary consequences of this competition, is of course also extremely relevant when multiple phage attempt to infect the *same* host, i.e. superinfection. Given that populations which allow and prevent superinfection both exist in the wild, it is natural to wonder what impact either strategy has on the evolution of viral populations.

This question has been studied in various viral systems (including those beyond bacteriophage) from the perspective of intracellular interactions and competition [91–99]. Multiple infections allow for the exchange of genetic material between viruses through recombination, which can increase diversity, but may also decrease fitness by promoting the presence of deleterious mutants at low frequencies [100–102]. Additionally, in RNA viruses with segmented genomes, multiple infections can lead to hybrid offspring containing re-assorted mixtures of the parental segments (reassortment). This mechanism can in principle improve selection efficiency (i.e., deleterious mutations are more likely to be purged, and beneficial ones are more likely to reach fixation), as re-assorted segments may generate highly deleterious variants that will be easily out-competed by the rest of the population [103]. Multiple infections can also lead to viral complementation, where defective viruses can benefit from superior products generated by ordinary viruses inside the host [103–107]. This process increases the diversity of the population, but also allows cheating individuals to persist in the viral population for long times [103, 104].

The likelihood of multiple infections occurring increases with the number of free phage available per viable host - multiplicity of infection (MOI) - and several experimental systems have been used to study the impact of MOI on viral dynamics [105, 106, 108–112]. For instance, high MOI in RNA phage $\phi 6$ has been shown to result in a behaviour conforming to the Prisoner's Dilemma strategy in game theory, and a reduction in viral diversity [108–111, 113]. Theoretically, the same question has been investigated in different scenarios [114], in particular in the context of human immunodeficiency virus (HIV) infections [100, 101, 115–120]. These studies have focused on determining whether multiple infections preferentially occur simultaneously or sequentially, in an effort to explain experimental data, and on the role of recombination in the acquisition of drug resistance, showing that its impact depends on the effective population size. The role of MOI has also been studied in terms of diversity and evolution of the viral population [100, 101, 117, 121–126], with theoretical predictions suggesting that multiple infection favours increased virulence, and that within-host interactions can lead to a more diverse population.

1.4 Open Questions

While much work has been done on the role of virus-host interactions, both on properties like the rate of plaque growth, as well as on the fitness and consequent evolution of phage populations, there remain several key, open questions. In the remainder of this section, I will outline some of these questions, which will form the basis of the work presented in the rest of this thesis.

1.4.1 Are Plaque Fronts Pulled or Pushed?

It is crucial to appreciate that, as mentioned previously, regardless of the differences between each of the plaque models described in Sec. 1.2.3, the speed estimates obtained in each case come from linearising the system of reaction-diffusion equations. As described in Sec. 1.1.3, using the solution to the linearised system equates to assuming that the viral dynamic is correctly described by a pulled wave. No justification is given for this assumption, and indeed it was demonstrated by You and Yin to be invalid in the limit of fast adsorption [68]. The failure of the linearised solution however, occurred

at rates of adsorption far higher than those found in the original measurements made with phage T7 [58], and so this line of inquiry was not pursued further.[10]

In addition to this observation, other key aspects of the current models could potentially challenge this assumption. For instance, the models assume that the diffusion coefficient of the phage is reduced depending on the concentration of bacteria according to Eq. 1.11. However, because the models are generally only concerned with the solution to the linearised model, they implement a *constant* diffusion coefficient that is set only by the concentration of bacteria *at the expansion front*. This therefore does not take into account the fact that the phage behind the expansion front, in the centre of the plaque, will diffuse faster relative to those phage at the tip of the expansion, due to the lack of bacteria in the plaque centre. It seems natural to think that this relative increase in diffusion could afford phage behind the front a better ability to catch up to the front and influence the expansion dynamics. By only using the bacterial concentration at the front to set a constant diffusion coefficient, however, it is impossible to tell if this density-dependent reduction in phage diffusion could lead to pushed waves [28].

1.4.2 How To Compare Plaque Models to Experiments?

Primarily, each of the models discussed in Sec. 1.2.3 was created and examined with the goal of accurately reproducing plaque speeds observed in experiments. To achieve this comparison the authors make use of independent measurements of the system's properties (e.g. bacteria density, phage burst size, lysis time, etc ...). A comparison of various models, reproduced from Ref. [71], can be seen in Fig. 1.10.

Briefly for context, the model by Gourley and Kuang shown in Fig. 1.10 is very similar to the model by Jones *et al.*, but additionally assumes a natural death process for only infected cells at rate μ_I that is unrelated to virus infection and decreases the number of infected cells after time τ by a factor of $e^{-\mu_I \tau}$ [70, 71, 127]. No biological reason is given for the inclusion of an additional death process that affects infected and not uninfected cells. Additionally, the reason that this model was not discussed in more detail above is that it is not *actually* a plaque model, but a model of phage infection in marine environments. The full model presented in Ref. [127] also includes the diffusion of both infected and uninfected hosts, bacterial growth and the decay of

[10] The nature of these measurements themselves will also be discussed shortly.

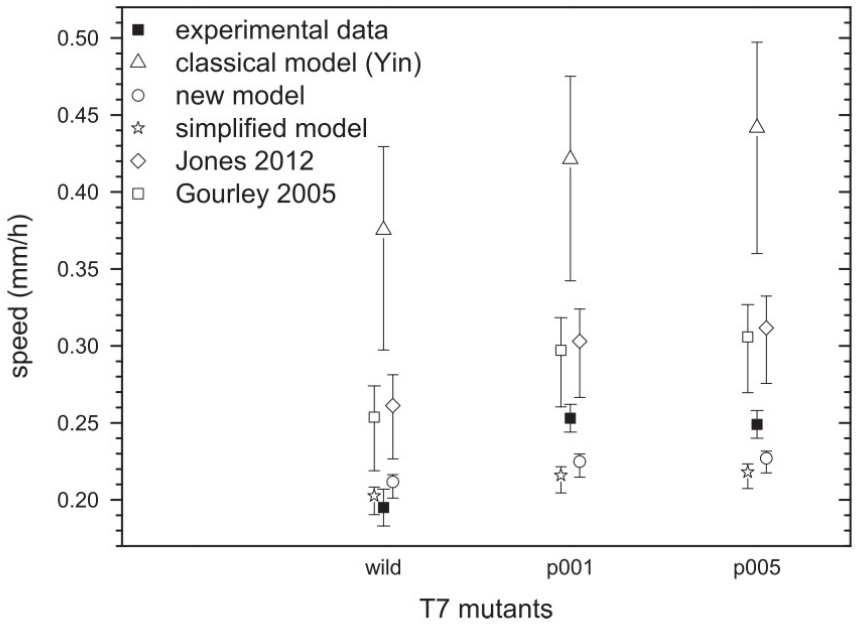


Fig. 1.10 Front propagation speeds for three T7 mutants (wild, p001 and p005). The mutants vary in their lysis time and burst size, and their adsorption rate is assumed to be unchanged [71, 88]. Black squares refer to experimental data obtained in Ref. [88]. White symbols refer to various theoretical models: triangles for the classical Yin and McCaskill model [58], circles for the "new model" by de Rioja *et al.*, stars for a simplified version of this model [71], rhombuses for the model by Jones *et al.* [70], and white squares for a model by Gourley and Kuang [127]. This figure is reproduced without adaptation from Ref. [71].

free phage. All of these processes, and their corresponding mathematical descriptions, are neglected when calculating the speeds shown in Fig. 1.10 [71, 127].

Secondly, the "simplified model" referred to is a version of the model by de Rioja *et al.* with fewer terms. The author's removed terms from their model one at a time to evaluate their contribution to the speed calculated. Through this process, it was found that terms relating to the adsorption of virus to host (i.e. $k_1[V][B]$) could be removed in the viral equation (but not in the equations describing infected and uninfected bacteria) with only a 4% alteration to the speeds calculated. Additionally, many of the higher-order terms could be removed without a significant alteration to the speeds obtained. It's worth quickly reiterating here that transitions to fully pushed waves occur at $3/2\sqrt{2} \times c_{linear}$, and so can be as little as 6% faster than pulled waves. While removing terms may not significantly impact the speed, they may significantly impact the population dynamic.

With that aside, it can be seen in Fig. 1.10 that the classic model by Yin and McCaskill, which does not take into account the incubation period, results in speeds significantly higher than those observed experimentally. The other models shown here show better agreement with the experiments, although the Jones and Gourley models appear to overestimate the speeds, while the de Rioja model is more inclined to underestimate the speeds (with the exception of the wild type).

While much effort has been devoted to examining the impact of all of the many possible terms in the model, essentially *no* attention has been given to the suitability of the actual parameter values used in them. I think there are various issues with this. For one thing, in experiments originally performed in Yin and McCaskill's work, the host fraction $f = B/B_{max}$ was assumed to be proportional to the concentration of nutrient broth used in the experiments [58]. The basis for this assumption is not entirely clear, and it appears no effort was made to test its veracity. The authors go on to say that this "implies that a critical nutrient concentration ($\sim$ 50 g/l in this work) corresponding to a host fraction of unity exists." Again, how this value was reached is not discussed. de Rioja *et al.* use these assertions to conclude that the experiments carried out by Yin in Ref. [88], which they use as the point of comparison for their model, must be at a density of $f = 0.2$, based on the fact that the nutrient concentration was 10 g/l [71].[11]

While it does seem likely that the final bacteria density will in some way be related to the nutrient concentration, I think that to say that they are directly proportional across a wide range of concentrations is a very big assumption. It seems perfectly possible that the final density of bacteria is in part determined by the density of the agar or other factors, and that the nutrient concentration impacts the time taken for the bacteria to reach this density, particularly at very high or low nutrient concentrations. Also implicit in this is the assumption that varying nutrient concentration has *no* effect on the phage life-history parameters. I think it is very possible, if not likely, that the host's access to nutrients will impact the production of new phage. For instance, varied availability of nutrients may promote alterations to burst size [128] and lysis time [84].

[11] They also go on to state that $B_{max} = 10^7$ ml^{-1} and therefore $B_0 = 2 \times 10^6$ ml^{-1}. Ref. [58] is cited as their source for the value that $B_{max} = 10^7$ ml^{-1}. Again, this is simply not what Ref. [58] says. In fact, Ref. [58] states that bacteria were plated with initial concentrations of 10^7 ml^{-1} *and* 10^8 ml^{-1}, and were then left to grow for 12 h before any plaque measurements were taken. This is so that the bacteria had time to grow to the relevant concentration B_0, which is not known precisely, but is assumed to be proportional to the nutrient concentration. Common sense however, would suggest that it will clearly be larger than 2×10^6 ml^{-1}.

Another, arguably more significant issue, is that the phage adsorption rate, burst size, and lysis time are all determined by experiments carried out in a well-mixed liquid cultures of exponentially growing bacteria, despite the fact that plaques by definition occur in bacteria immobilised in solid agar [62, 71, 88].[12] It is perfectly possible that these parameters could take significantly different values when the infection occurs in spatially structured environments. Moreover, the plaque growth measurements are taken over the course of 24 hrs, over which time the bacteria will almost certainly enter stationary phase, which could also significantly impact these parameters [88, 129]. Any such change would also be difficult to observe in the data underlying Fig. 1.10, given that the speed of each plaque was determined from size measurements at only 3 timepoints (13, 18 and 23 hours) [88].

Having said all of this, I do think it is important to attempt to compare predictions from a model with experimental observations to determine if the model is physically realistic. From that point of view, all of the models which in some way incorporate a time delay show some level of quantitative agreement with the experimental observations (i.e. the models predict speeds that are within 25% of the experimental values), and the models show good qualitative agreement with the behaviours observed (e.g. speeds are reduced at high bacteria density, increased at large burst sizes, reduced at large lysis times etc). Beyond this however, I think that saying model X is better than model Y because model X predicts a speed 10% lower while model Y predicts a speed 15% higher is perhaps not very useful at present given the uncertainty surrounding the underlying parameters used to make the predictions.

1.4.3 What is the Impact of Multiple Infections Absent of Intracellular Interactions?

Despite the active work in the area, several fundamental questions on the role of superinfection exclusion on viral dynamics remain unanswered. First, while decreasing MOI in viral populations that allow superinfection decreases the *likelihood* of superinfection, it does not introduce a superinfection exclusion mechanism that prevents superinfection altogether, making it difficult to draw conclusions about the (dis)advantages of this

[12] de Rioja *et al.* claim that the burst size data "have been obtained for cells in agar-immobilized microcolonies containing many cells [71]," however this is not accurate. This can be verified by reading the method described in Ref. [88], which is the stated source of their data. I believe the confusion arises from the fact that to measure the concentration of phage, it *is* necessary to plate an aliquot of the phage with host bacteria immobilised in agar, but this is simply the measurement technique, not the environment of the actual experiment.

viral strategy. Second, little is known about how the occurrence of superinfection alone, before even accounting for the additional effects of any intracellular interactions, impacts the evolution of viral populations, particularly when it comes to fundamental evolutionary outcomes such as mutant fixation probabilities. A quantitative understanding of this baseline behaviour is necessary to evaluate the impact of the many additional intracellular interactions that can occur (recombination, defective viruses, etc.). The limited work in this area has shown that in the absence of intracellular interactions, high MOI in superinfecting viral populations can promote the presence of disadvantageous mutants in the "short term," and obstruct it in the "long term" [130, 131], but how the evolutionary outcomes in each case depend on the parameters describing the viral life-cycle (adsorption rate, lysis time and burst size) and the (dis)advantages of either strategy remain unclear. Addressing these knowledge gaps in a well-mixed setting must also occur before we can draw any firm conclusions about whether either strategy becomes more or less useful in a spatial context.

1.5 Thesis Outline

In Chapter 2 I will address the question of whether plaque fronts are pulled or pushed. In particular, I will (i) experimentally determine whether and how the rate of phage diffusion depends on the density of surrounding bacteria, (ii) under what conditions, if any, transitions to semi-pushed and fully pushed expansions can occur, and (iii) what role density-dependent diffusion plays in this.

I will follow this up in Chapter 3 by developing an image analysis pipeline for the identification of large monoclonal sectors that occur during phage range expansions, which can be used to measure the strength of genetic drift as a function of bacterial density and other parameters.

In Chapter 4 I will present a variety of novel experimental techniques for working with phage in settings with spatial structure, with a view to facilitating better comparison between theory and experiment. This includes protocols to measure phage adsorption rate and lysis time on solid media rather than in liquid cultures, as well as design and demonstration of agar plates with subsurface macrofluidic channels for the controlled delivery of nutrients.

In Chapter 5 I will present a computational exploration of the evolutionary impact of superinfection absent of any intracellular interactions. This will particularly focus on the (dis)advantages of the choice to allow or exclude superinfection. This will include a

discussion of the impact on both the neutral dynamics and the fixation probabilities of non-neutral mutants, as well as looking at the benefits of either allowing or preventing superinfection when both strategies are competed directly.

This will be followed in Chapter 6 by some initial observations of phage infection at a single-cell level in microfluidic mother machine devices. This aims to characterise the level of cell-to-cell variability when it comes to infection parameters such as the lysis time, with the ultimate goal of understanding the impact of phenotypic heterogeneity on viral evolution.

Finally in Chapter 7 I will conclude by summarising the key findings presented in this thesis, as well as discussing the many directions future research in this area may take, and my opinion on the promises and challenges ahead.

'There is no safety in numbers, or in anything else.'

James Thurber

'Defer no time, delays have dangerous ends.'

William Shakespeare

Chapter 2

Density-dependent diffusion alters dynamics of viral range expansions

This chapter is based on results presented in

M. Hunter, N. Krishnan, T. Liu, W. Möbius, and D. Fusco, *Virus-Host Interactions Shape Viral Dispersal Giving Rise to Distinct Classes of Traveling Waves in Spatial Expansions,* Phys. Rev. X **11**, 21066, (2021)

I performed all experiments, analysis and interpretation of results, with the exception of the stochastic simulations used to measure the rate of decay of genetic diversity, which were performed by Dr. Nikhil Krishnan.

2.1 Introduction

Spatial range expansions are ubiquitous in nature, from the expansion of invasive plant species, through the migration of ancient human populations, to the range shifts of many organisms to higher altitudes and latitudes due to climate change [14–21]. One of the hallmarks of spatial expansions is the rapid loss of genetic diversity due to the enhanced fluctuations at the front [22, 24]. This effect can, however, be significantly mitigated in the presence of density-dependent growth [29, 35], such as an Allee effect [32], or density-dependant dispersal, where individuals in highly dense patches tend to disperse more quickly [28]. In particular, it has recently been shown theoretically that the ratio between the deterministic speed of the front and that of its linearised approximation is sufficient to classify the expansions in three distinct types of travelling

waves, nominally pulled, semi-pushed and fully-pushed, which respectively exhibit qualitatively distinct behaviors in the decay of heterozygosity, the stochastic wandering of the front position, and the probability distribution of the most recent common ancestor [25, 28].

Because density-dependent growth can play such a crucial role in the evolutionary dynamic of a population, it has been extensively investigated in both naturally occurring range expansions in animals, such as the invasions of both Eurasian gypsy moths and house finches in North America [36, 37], and in laboratory microbial model systems, where the expansion dynamics in populations of the budding yeast *Saccharomyces cerevisae* transition from pulled to pushed as growth becomes more cooperative [40], with a corresponding preservation of genetic diversity [41]. In comparison, relatively little is known about the population dynamic of experimental systems that exhibit density-dependant dispersal, even if it has been documented in several natural populations [132] and the transition to pushed waves has been theoretically predicted [28].

One laboratory system that has been hypothesized to undergo density-dependent dispersal is bacteriophage expanding in a bacterial lawn. The crowded bacterial environment is thought to hinder phage diffusion because of steric interactions, resulting in a density-dependent diffusion coefficient due to the coupling between the host and the viral population densities [58]. Direct experimental quantification of this density-dependent diffusion is, however, limited [66], and its consequence on the front population dynamic mostly unknown.

Here, we address the open questions of (i) whether and how the rate of phage diffusion depends on the density of surrounding bacteria, (ii) under what conditions transitions to semi-pushed and fully-pushed expansions can occur, and (iii) what role density-dependent diffusion plays in this. We first design an experimental protocol to measure the effect of steric interactions between phage and the surrounding bacteria on phage dispersal. We then construct a system of reaction-diffusion equations to determine the phage front speed, demonstrating that transitions to both semi-pushed and fully-pushed waves can occur. We find that the presence and location of these transitions are controlled by two independent effects that alter the density-dependent diffusion of the virus: the first is associated with the excluded-volume interactions with the surrounding bacteria, while the second spontaneously emerges from the viral infection dynamic, which prevents a viral particle from diffusing during infection of the host. Using stochastic agent-based simulations, we show that even the second effect

alone, which applies to viruses beyond phage, can lead to a significant reduction in the rate of diversity loss in the viral population.

Taken together, our results identify bacteriophages as a controllable laboratory model system to investigate the role of density-dependent dispersal in evolution and provide a quantitative explanation of the physical mechanisms that control the phage population dynamic during a range expansion. Going beyond phages, our findings suggest that a broad range of viruses may expand via pushed travelling waves and, consequently, may be much more adaptable then previously thought.

2.2 Experimental Measurements of Density Dependent Dispersal in Coliphage T7

Starting with Yin and McCaskill [58], it is usually recognized that bacteria can act as a barrier to phage diffusion, resulting in a diffusion coefficient that depends on the bacterial density. This dependence is indeed necessary in phage expansion models to correctly reproduce the non-monotonicity of front speed observed as a function of bacterial density [58, 62, 68, 72]. While it has been shown that phage diffuses faster in the bulk of a plaque than at the edge [66], the dependence of the phage diffusion coefficient on bacterial density has never been quantitatively measured.

To address this need, we designed an experimental setup where (i) host density can be quantified and maintained uniform in space and constant in time, and (ii) the effect of steric interactions is decoupled from the viral infection dynamic. To this end, we moved away from classic plaque-in-agar assay, which exhibit a fragmented host distribution at the microscopic level, and built a uniform bacterial lawn by directly pouring an *E. coli* liquid culture of known density on top of 2% agar LB plates containing chloramphenicol (background bacteria in Fig. 2.1b). These bacteria are susceptible to the antibiotic, which prevents their growth ensuring a constant host density during the experiment, and are engineered to prevent phage adsorption [133, 134], so as to serve as passive barrier to phage dispersal. Phage droplets were then inoculated across the lawn (grey in Fig. 2.1b) at different distances from droplets of a second *E. coli* strain, susceptible to phage and resistant to chloramphenicol (black in Fig. 2.1b). The time Δt required by the phage to travel the distance r between a viral droplet and a close-by susceptible bacteria droplet was monitored *in vivo* by tracking the appearance of clearings in the susceptible droplets (Fig. 2.1a and b).

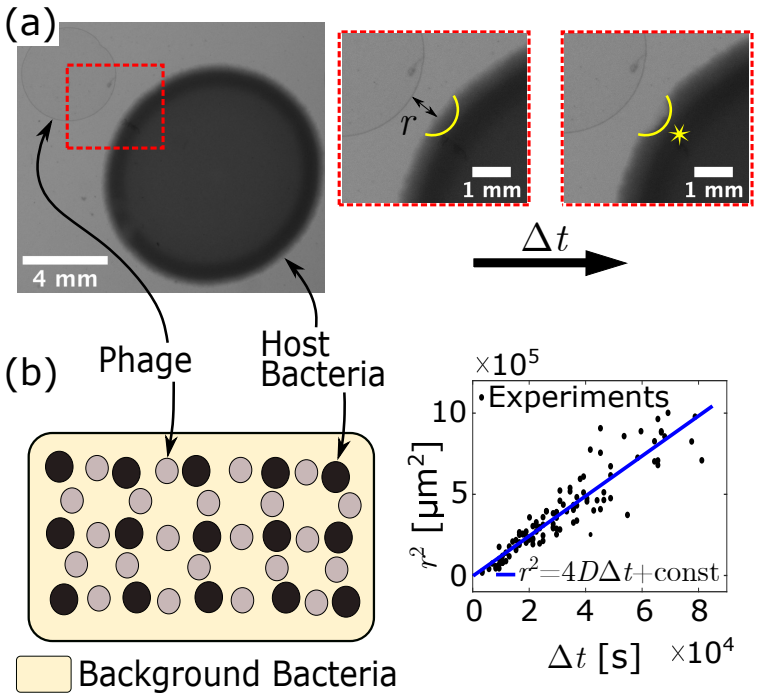


Fig. 2.1 (a): The basis of the experimental set-up, consisting of a droplet of phage and host bacteria, separated by a distance r. After time Δt, a plaque begins to form in the host bacterial droplet (starred region). (b): The full experimental set-up, consisting of many phage-host droplet pairs on top of a lawn of phage resistant bacteria of variable density. The presence of chloramphenicol in the plate media ensures that the background bacterial density is constant over the course of the experiment (see Sec. 2.8.7). An example plot of r^2 against Δt data, with linear fit, for a resistant bacteria density of 0.36 μm^{-2} is shown.

By gathering statistics over many droplet-droplet pairs, we were able to first confirm that the relationship between distance travelled and mean first passage time is consistent with diffusive behavior for the whole range of background densities tested (Fig. 2.1b), and then calculate the rate of phage diffusion D as a function of background bacterial density (Fig. 2.2). Additionally, in Sec. 2.2.1 we apply a correction to these measured diffusion rates that arises due to the geometry and number of phage used in our set-up, however in the remainder of this Section we report the experimental results without that correction. Additional tests were also performed to confirm that phage did not significantly diffuse out of the plane and into the agar during the course of the experiment (Sec. 2.2.2).

Building on previous efforts to account for density-dependence in plaque models [58], we fit our data using Fricke's Law [62, 65], which describes the diffusion of a solute

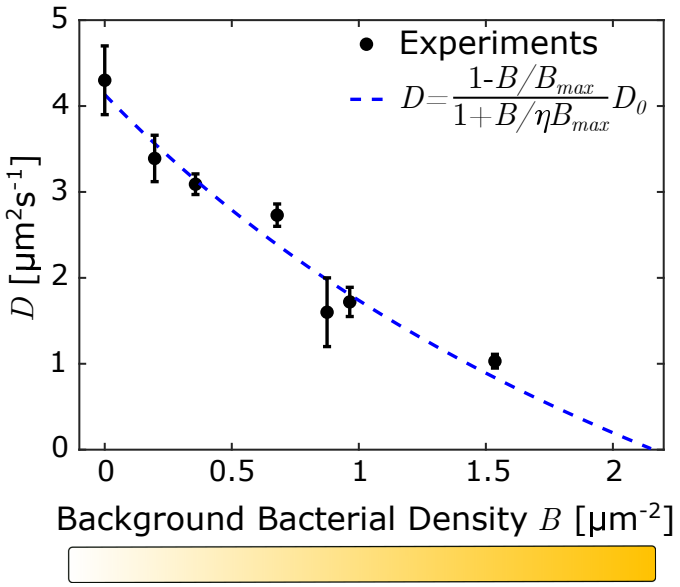


Fig. 2.2 Experiments show how rate of phage diffusion is reduced by surrounding bacteria. The diffusion rate obtained as a function of resistant bacteria density (i.e. for several instances shown in Fig. 2.1(b)), fit with Fricke's Law (Eq. 2.1).

through a suspension of spheroids [135]:

$$D = \frac{1-b}{1+\frac{b}{\eta}}D_0 \quad ; \quad b = \frac{B}{B_{max}}, \tag{2.1}$$

where b indicates the fraction of bacteria B relative to a maximum value B_{max} and η accounts for the shape of the cells: spherical cells correspond to η=2, while *E. coli* cells have previously been determined to correspond to η=1.67 [62]. Our experimental data allow for the first time to estimate the two fitting parameters required by Frickes's law in this context: the free diffusion coefficient D_0 (i.e. the diffusion coefficient in the absence of surrounding bacteria), and the bacterial density B_{max} at which diffusion is expected to be completely halted. We estimate $D_0 = 4.13 \pm 0.19 \mu\text{m}^2/\text{s}$, which is in good agreement with the rate of 4 $\mu\text{m}^2/\text{s}$ previously determined by Ouchterlony double immunodiffusion[1] in 10 g/l agar of phage P22 (similar size and shape of T7) [63, 64]; and $B_{max} = 2.16 \pm 0.19\ \mu\text{m}^{-2}$, which is consistent with the typical dimensions of an *E.*

[1]This technique involves cutting wells into an agar plate. One well contains a sample of interest (in this case phage), and while the other well contains purified antibodies. Antigens from the sample and the antibodies then diffuse outward through the agar from their respective wells, eventually meeting, and binding to form an immune complex which precipitates, forming a band in the gel which can be seen. The geometry of the wells and the band can then be used to determine the diffusion rate.

coli cell (assuming *E. coli* cells are approximately 0.5 × 2 μm, we would expect a 1 μm^2 cross section to contain between 1 and 4 closely packed cells, depending on their orientation and deformation). Note that while we will use the expression in Eq. 2.1 to account for the effect of steric interactions on phage dispersal in a uniform bacterial lawn, this relationship can be applied to any scenario in which the bacterial density distribution is known, even if it is non-uniform.

2.2.1 Correction to Measurements of Phage Diffusion

Given that there is more than one phage in the droplets in our experimental set-up, we consider that there may be a correction to our calculated diffusion rate that depends on the initial number of phage. As the number of phage increases, the time taken for *one* of them to reach the bacteria droplet should decrease, thereby increasing the diffusion coefficient calculated. On the other hand, the bacteria do not form a ring around the initial droplet of phage, so we are unable to truly measure the first time a phage travels a distance r, rather, we are measuring the first time for a phage to travel a distance r in a *specific* direction. Additionally, the phage are not all initially located in precisely the same spot, but are spread randomly throughout a circular droplet that is large relative to diffusion scales. If the phage which first reaches the bacteria droplet does not originate at the edge of the droplet closest to the bacteria, then we would be underestimating the distance travelled, and therefore the diffusion rate.

To determine how these aspects alter our diffusion rate measurements, we perform a 2D simulation of the diffusive process in our experimental set-up (see Sec. 2.8.10). Through these simulations, we are able to determine the 'true' diffusion rate D_{true} of the phage from their mean-squared displacement $\overline{r^2}$ as a function of time t (Fig. 2.3a), along with the 'experimental' diffusion rate $D_{experiment}$ that we would determine using our experimental approach (Fig. 2.3b).

Fig. 2.3c shows the comparison of D_{true} with $D_{experiment}$. It can be seen that our experimental procedure systematically overestimates the diffusion rate of the phage. Heuristically, we fit a straight line to D_{true} as a function of $D_{experiment}$ (blue dotted line in Fig. 2.3c), which yields a gradient of 0.65 ± 0.04 and an intercept of -0.2 ± 0.2. This indicates that on average the 'true' diffusion rate is $\sim$2/3 of the rate measured in our experiments, at least in the range of diffusion rates observed.

Using the heuristic fit from Fig. 2.3c, we adjust the data collected through our experimental procedure (Fig. 2.3d). It can be seen that while the bacterial density B_{max} at which diffusion is expected to be completely halted does not change substantially -

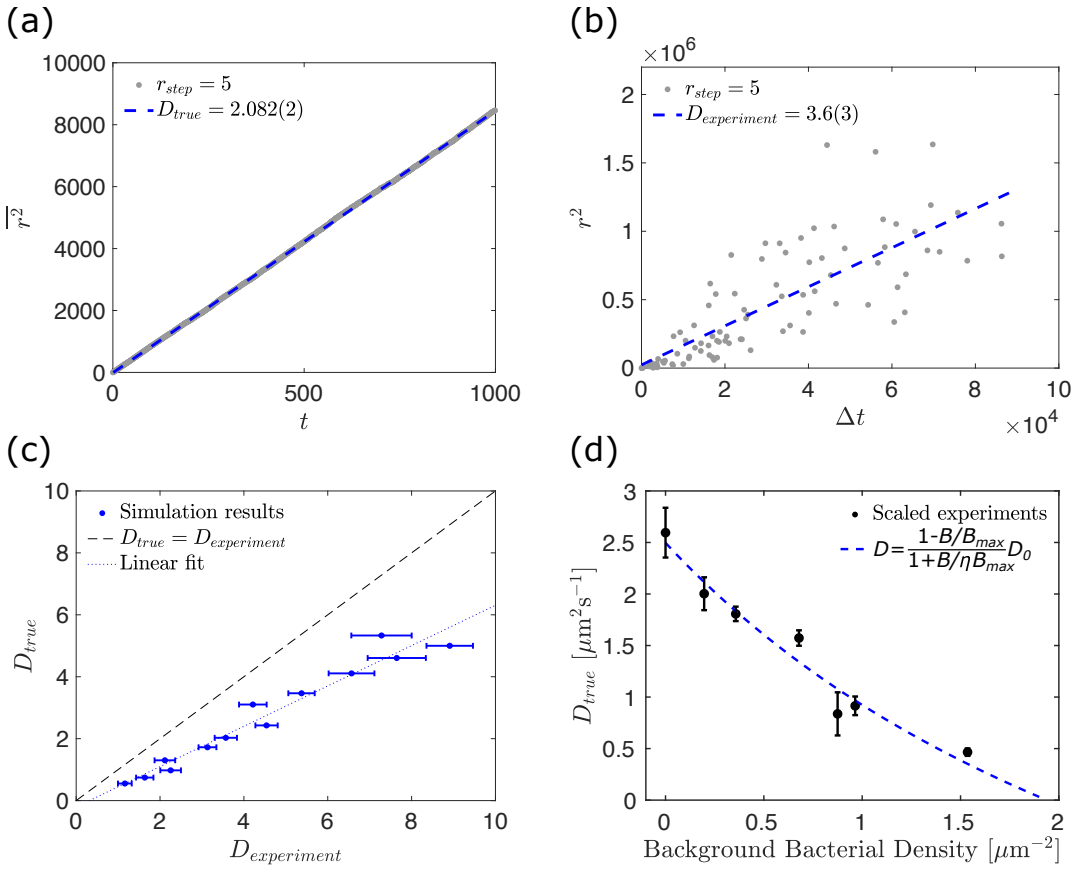


Fig. 2.3 (a) The mean-squared displacement $\overline{r^2}$ of 1000 tracers as a function of time t. A linear fit yields a 'true' diffusion rate of $D_{true} = 2.082 \pm 0.002$. (b) Simulated experimental data of for the first time Δt that one of the tracers reaches the target region, and the distance r separating the initial and target region, as would be measured in the experiment. A linear fit yields an 'experimental' diffusion rate of $D_{experiment} = 3.6 \pm 0.3$. (c) D_{true} as a function of $D_{experiment}$. It can be seen that $D_{true} < D_{experiment}$. A heuristic linear fit to the data yields a gradient of 0.65 ± 0.04 and an intercept of -0.2 ± 0.2. (d) The experimental data presented in Fig. 2.2 re-scaled according to the relationship determined from (c). Our data now yields $D_0 = 2.50 \pm 0.12$ μm^{2}s^{-1} and $B_{max} = 1.94 \pm 0.16$ μm^{-2}.

1.94 ± 0.16 μm^{-2} now vs 2.16 ± 0.19 μm^{-2} before - the free diffusion rate D_0 drops significantly from 4.13 ± 0.19 μm^{2}s^{-1} to 2.50 ± 0.12 μm^{2}s^{-1}. It should be noted, however, that this change has no impact on our ability to use the expression in Eq. 2.1 to account for the effect of steric interactions on phage dispersal in a uniform bacterial lawn.

2.2.2 Phage Remain On Agar Surface

To verify that phage are not diffusing out of the approximately 2D plane and into the agar during the course of the diffusion experiments, additional tests were carried out. If significant diffusion into the agar was occurring, we would expect the number of phage at the surface to be reduced over time. As with the measurements of diffusion coefficient (see Sec. 2.8.7), 35 ml omni-plates of 20 g/l (2%) agar, with LB and 15 μg/ml chloramphenicol were prepared. Similarly, overnight liquid cultures of susceptible host (*E. coli* eMTH43) were grown from single colonies at 37 °C in LB with 15 μg/ml chloramphenicol.

Then, 1.5 ml of stock bacteriophage T7 diluted in LB was spread across the plate with glass beads, such that the plate should contain a countable (order of tens) number of phage.

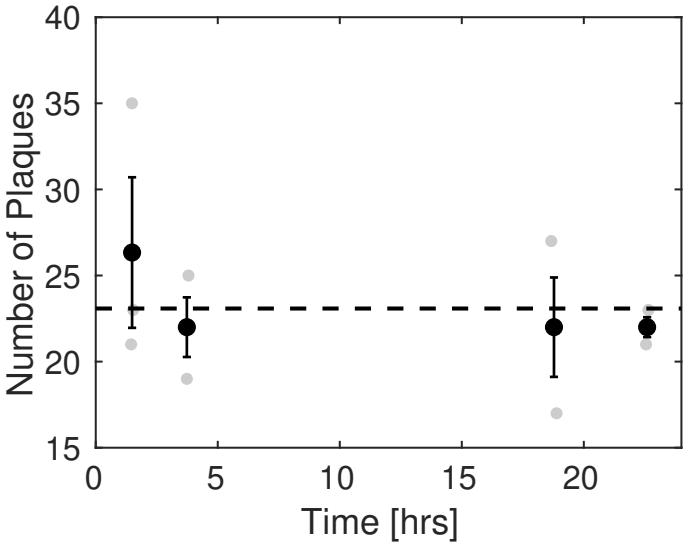


Fig. 2.4 Plaques counted on the surface of a 2% agar plate as a function of time after the phage were originally added to the plate. Data for the four time points measured (black) and their corresponding errors are calculated from three replicates (grey). The dashed line indicates the mean number of plaques counted across time-points. It can be seen that the data is consistent with a constant number of phage being recovered over time, indicating that phage diffusing into the agar is very limited.

After a set time period (approximately 1.5, 4, 19 or 23 hours), 100 μl of overnight eMTH43 was mixed with 10 ml of molten 7 g/l (0.7%) agar and poured over the surface of the plate and left on a lab bench overnight. Any phage that were located on the surface of the 2% agar at the time that the 0.7% agar and susceptible eMTH43 was

poured on top is expected to be able to infect the susceptible host and result in a plaque. By counting the number of plaques visible in the 0.7% agar the following morning, we are able to determine whether a significant amount of phage diffuses into the 2% agar plate over a 24 hour period as these phage would not be able to form a plaque.

The results from this test (Fig. 2.4) clearly show that there is no significant reduction in the number of phage recovered from the surface of the plate over a roughly 24 hour period (the period over which diffusion measurements were gathered). We believe that this is because the pore size of our 2% agar is small enough to significantly limit diffusion. Indeed, similar 2% agarose substrates have been shown to have a pore size of under 80 nm [136], which is comparable with the size of T7.

2.3 Modelling Plaque Growth: Density-Dependent Diffusion and Adsorption to Infected Cells

To investigate whether the phage expansion on a bacterial lawn occurs as a pulled or a pushed wave, and to uncover the role of host density-dependence, we compare the actual front speed with the speed c_F of the corresponding linearised system, as their ratio has been shown to be sufficient to determine the wave class in single species range expansions [25]. To this end, we develop a mathematical model that accommodates the density-dependent diffusion we have experimentally measured.

We model the spatial dynamics of bacteriophage plaque growth by considering the interactions between three populations: viruses (phage) V, uninfected host bacteria B and infected host bacteria I, similar to [58, 62, 68, 70–73, 137]. The process may be summarised as

$$V + B \xrightarrow[\alpha]{\text{rate}} I \xrightarrow[\tau]{\text{delay}} \beta V, \tag{2.2}$$

where β is the burst size, α is the rate of adsorption, and τ is the lysis time.

As the model is deterministic, without loss of generality, we describe these populations with a set of reaction-diffusion equations in 1D, similar to those examined by Jones *et al.* [70]:

$$\frac{\partial B}{\partial t} = -\alpha V B, \tag{2.3a}$$

$$\frac{\partial I}{\partial t} = \alpha V B - \alpha V_{t-\tau} B_{t-\tau}, \tag{2.3b}$$

$$\frac{\partial V}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial V}{\partial x}\right) - \alpha V B - \alpha^* V I + \beta\alpha V_{t-\tau} B_{t-\tau}, \tag{2.3c}$$

where V, B and I indicate the concentration of the population as a function of space and time. The subscript is used to indicate that those terms are evaluated at time $t-\tau$. D is the density-dependent diffusion coefficient of the phage, determined from fitting Fricke's law to experimental results in the previous section (Eq. 2.1). $\alpha^* = \alpha$ or $\alpha^* = 0$, depending on whether adsorption to previously infected hosts is allowed or prevented, respectively. We assume that the host bacteria are motionless and that adsorption to uninfected hosts always leads to successful infection while neglecting desorption.

Our model introduces two ingredients that are biologically and physically relevant, and that are expected to affect the front dynamic. First, in contrast to previous work [58, 62, 68, 71–73, 137], where the diffusion coefficient D only depends on the initial bacterial density B_0 ($b = \frac{B_0}{B_{max}}$ in Eq. 2.1), we allow D to vary in time and space according to the local bacterial density ($b = \frac{B+I}{B_{max}}$ in Eq. 2.1), resulting in faster diffusion inside the phage clearing (Fig. 2.5). Secondly, we allow for the possibility that phage can adsorb to previously infected cells ($-\alpha^* V I$ term in Eq. 2.3c), as is the case for phage T7. The presence or absence of these two effects generates four model variants that are summarized in Fig. 2.5: Uniform vs. Variable Diffusion model (UDM vs. VDM), and adsorption vs. non-adsorption to infected cells (+ vs. -).

In line with previous studies, we cast the equations using dimensionless variables. We measure concentrations in terms of the initial bacterial density B_0, time in units of τ, and length in units of $L = \sqrt{D(B_0)\tau}$ (spatial scale of diffusion at the front within the lysis time). This results in the following set of dimensionless variables: $\overline{B} \equiv B/B_0$, $\overline{I} \equiv I/B_0$, $\overline{V} \equiv V/((\beta-1)B_0)$, $\overline{t} \equiv t/\tau$, $\overline{x} \equiv x/L$ and $K \equiv \alpha\tau B_0$. Consequently, $\overline{c} = c\sqrt{\tau/D}$, where $\overline{c}$ and c are the dimensionless and dimensional speed of the expansion front, respectively (Fig. 2.5).

In these units, the UDMs are characterized by a constant dimensionless diffusion coefficient $\overline{D} = D\tau/L^2 = D/D(B_0) = 1$ by definition, while the VDMs exhibit a dimensionless density-dependent diffusion coefficient of the form:

$$\overline{D} = \frac{D}{D(B_0)} = \frac{1 - f(\overline{B}+\overline{I})}{1 + f(\overline{B}+\overline{I})/\eta} . \frac{1 + f/\eta}{1 - f}, \tag{2.4}$$

where $f = B_0/B_{max}$ is the initial fraction of bacteria. Note that $\overline{D}$ corresponds to the phage diffusion coefficient relative to the diffusion coefficient at the very front of the

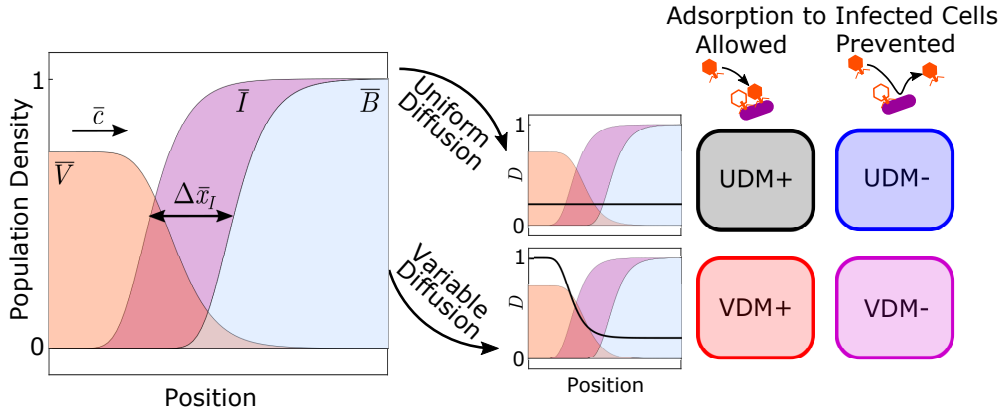


Fig. 2.5 A sketch of the population concentrations $\overline{B}$, $\overline{I}$ and $\overline{V}$ as a function of location at the expansion front (the precise location is not important here, as the qualitative shape of the fronts remain constant during the expansion). The front is propagating with dimensionless speed $\bar{c}$ to the right. The dimensionless width $\Delta\overline{x}_I$, characterising the width of the infected region is given by the difference in position of the uninfected ($\overline{B}$) and infected fronts ($\overline{B}+\overline{I}$). The differing diffusion and adsorption behaviours explored lead to four different model variants in this work. Variants either have a Uniform or Variable diffusion rate (UDM or VDM respectively, black line), and adsorption to previously infected cells either does (+), or does not (-) occur, leading to the four model variants (UDM+, UDM-, VDM+ and VDM-).

expansion, where host density is maximal. As a consequence, $\overline{D}$ from Eq. 2.4 is always greater than or equal to 1, and can therefore be interpreted as a "boost" in diffusion that the VDMs exhibit in the bulk of the plaque in comparison to the corresponding UDMs. This boost mathematically describes the decrease in steric interactions between phage and bacteria due to the lysis of the host as the viral infection proceeds (black line in Fig. 2.5).

In terms of these variables, our model (Eqs. 2.3) becomes:

$$\frac{\partial\overline{B}}{\partial\bar{t}} = -K(\beta-1)\overline{V}\,\overline{B}, \tag{2.5a}$$

$$\frac{\partial\overline{I}}{\partial\bar{t}} = K(\beta-1)\overline{V}\,\overline{B} - K(\beta-1)\overline{V}_{\bar{t}-1}\overline{B}_{\bar{t}-1}, \tag{2.5b}$$

$$\frac{\partial\overline{V}}{\partial\bar{t}} = \frac{\partial}{\partial\overline{x}}\left(\overline{D}\frac{\partial\overline{V}}{\partial\overline{x}}\right) - K\overline{V}\,\overline{B} - K^*\overline{V}\,\overline{I} + \beta K\overline{V}_{\bar{t}-1}\overline{B}_{\bar{t}-1}, \tag{2.5c}$$

where $K=\alpha\tau B_0$ and $K^*=\alpha^*\tau B_0$.

As our goal is to determine whether the travelling waves are either pulled or pushed, we will require the solution to the linearised approximation of the model. To achieve

this we expand the model (Eqs. 2.5) about the tip of the front where ($\overline{V}$, $\overline{B}$, $\overline{I}$) $\approx$ (0, 1, 0), keeping only linear terms. This results in the following set of equations:

$$\frac{\partial \overline{B}}{\partial \overline{t}} = -K(\beta - 1)\overline{V}, \tag{2.6a}$$

$$\frac{\partial \overline{I}}{\partial \overline{t}} = K(\beta - 1)\overline{V} - K(\beta - 1)\overline{V}_{\overline{t}-1}, \tag{2.6b}$$

$$\frac{\partial \overline{V}}{\partial \overline{t}} = \frac{\partial}{\partial \overline{x}}\left(\overline{D}\frac{\partial \overline{V}}{\partial \overline{x}}\right) - K\overline{V} + \beta K\overline{V}_{\overline{t}-1}. \tag{2.6c}$$

From Eqs. 2.5 three natural parameters emerge: the dimensionless adsorption coefficient $K = \alpha\tau B_0$, the burst size β and the dimensionless diffusion coefficient $\overline{D}$. In the UDMs, $\overline{D} = 1$, leaving K and β as the only two parameters of the model. By contrast, in the VDMs, $\overline{D}$ is a function of B_0 (Eq. 2.4), which entangles the effect of initial bacterial density on K and $\overline{D}$. To decouple adsorption and diffusion, we define a set of three new independent parameters that we will use in the following to analyse the model variants: the initial fraction of bacteria $f = B_0/B_{max}$, the maximum dimensionless adsorption coefficient $K_{max} = \alpha\tau B_{max}$ ($K = fK_{max}$), and the burst size β. In the linearised approximation (Eqs. 2.6) $\overline{c}_F$ is the same for all four model variants and, therefore, depends only on the dimensionless adsorption coefficient $K = fK_{max}$ and the burst size β - see Sec. 2.8.12.

2.4 From Pulled, to Semi-Pushed to Fully Pushed

By numerically solving the PDE system in Eqs. 2.5, we obtain the front speed $\overline{c}$ and compare it with the speed $\overline{c}_F$ of the linearised approximation of the model (Eqs. 2.6, see Sec. 2.8.11 and Sec. 2.8.12 for details). In addition to front speed, we also determine the characteristic width of the infection region $\Delta\overline{x}_I$ (Fig. 2.5), which we will discuss later on.

The transitions between different types of travelling wave are then determined from the ratio $\frac{\overline{c}}{\overline{c}_F}$ according to Ref. [25]: (i) pulled waves for $\frac{\overline{c}}{\overline{c}_F} = 1$, (ii) semi-pushed wave for $1 < \frac{\overline{c}}{\overline{c}_F} < \frac{3}{2\sqrt{2}}$, (iii) fully pushed waves for $\frac{\overline{c}}{\overline{c}_F} \geq \frac{3}{2\sqrt{2}}$ - see Sec. 2.8.11. We point out here that the transition between semi-pushed and fully pushed waves has been uncovered and investigated only for single species range expansions, so far. The viral model presented here is more complex because of the coupling between the dynamics of different populations (bacterial and viral) and because of the presence of a time delay.

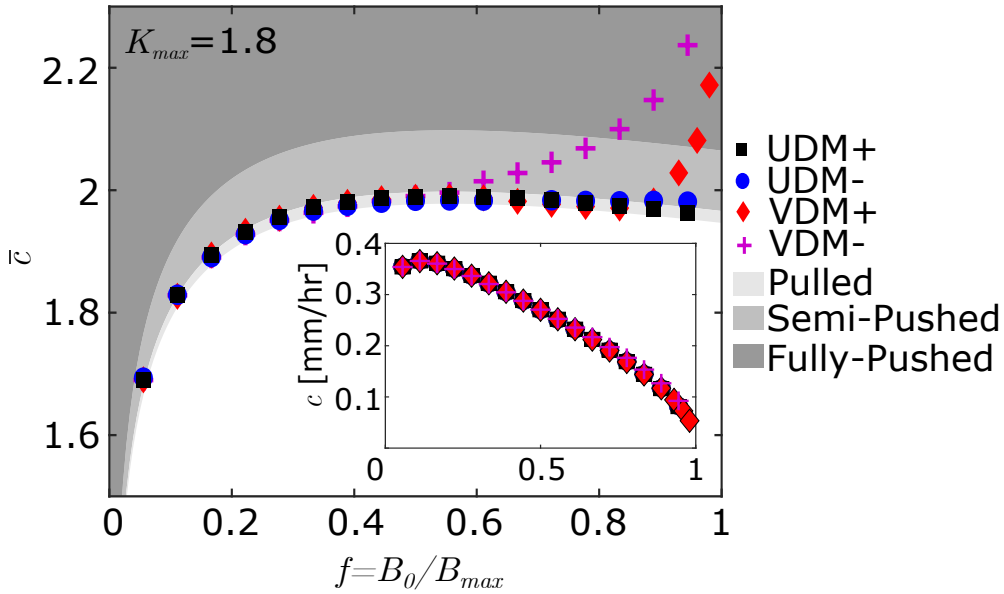


Fig. 2.6 Dimensionless front speed $\bar{c}$ as a function of bacteria fraction f, with shaded regions indicating the different expansion types. Error bars on the speeds are smaller than the symbols. Inset also shows the dimensional speed c. Model parameters are chosen to represent typical T7 expansions with $\beta = 50$, τ=18 mins, and αB_{max}=0.1 min^{-1} [58, 70, 71], corresponding to $K_{max} = 1.8$ in our model.

As a result, the demographic noise in our model may differ from that in Ref. [25]. While this does not affect the transition between pulled and pushed waves, the distinction between semi-pushed and fully pushed waves might, in principle, be different.

The location of these transitions in the different model variants for a set of infection parameters typical of T7 is shown in Fig. 2.6. Under these conditions, we observe that the UDM+ exhibits a pulled wave for the full range of initial bacterial fraction, while the UDM-, the VDM+ and the VDM- waves become increasingly more pushed as f increases. In terms of dimensional speed, the difference between the model variants is minimal (inset in Fig. 2.6), justifying why these effects have gone unnoticed in past theoretical work that aimed at predicting experimental phage front speeds.

2.4.1 Wave Transitions are Very Sensitive to Virus-Host Interactions

To generalise our findings and fully characterise the origin and nature of the transitions in front dynamic for the different model variants, we extend our investigation to a broader range of parameter values, by varying K_{max} (and β, see Sec. 2.5.1) about the parameters used in Fig. 2.6.

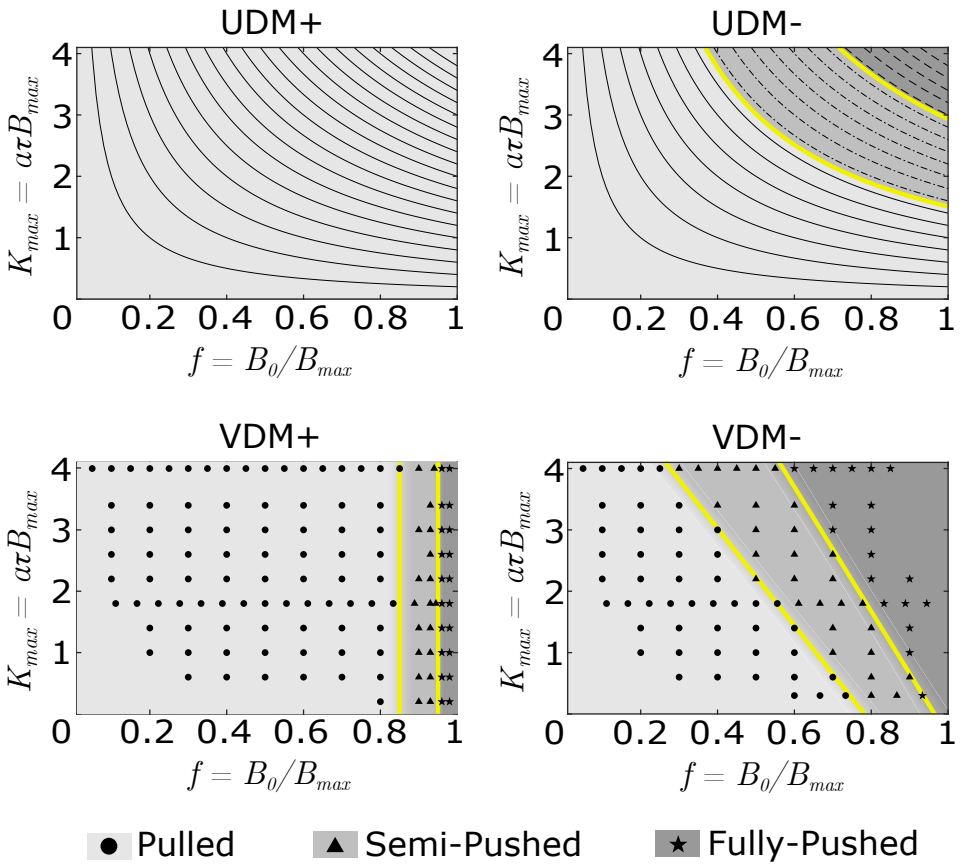


Fig. 2.7 Phase diagrams showing the expansion types for the four model variants as a function of bacterial density f and maximum dimensionless adsorption coefficient K_{max} - burst size β=50 throughout. Lines in the UDMs, and data points in the VDMs indicate the parameter combinations for which numerical integration was performed, and speeds calculated. These values are interpolated to estimate the transition boundaries between different classes of travelling waves (yellow lines). In the UDM+, we do not observe pushed transitions, while in the VDM+ transitions occur at approximately constant bacteria fractions. In the UDM-, as K and β are the only free parameters, transitions occur at specific values of K. In the VDM-, the transitions are heuristically approximated as linear relationships with gradient m and intercept a ($f = mK_{max} + a$).

Fig. 2.7 shows the type of expansion that occurs in each of the models as a function of f and K_{max}. The results clearly indicate that the presence or absence of density-dependent diffusion and adsorption to infected cells can dramatically alter the type of travelling wave undergone by phage, with the UDM+ being the only model resulting in a pulled population wave for the whole range of parameters explored.

In the following, we will provide a physical interpretation for these observations, by identifying two independent mechanisms that alter phage dispersal in a density-

dependent fashion. The first (Sec. 2.4.2), which we name the 'explicit' effect, is caused by steric interactions between phage and the bacterial host, and represents the effect measured in our experiments (Fig. 2.2). The second (Sec. 2.4.3), which we name the 'implicit' effect, arises spontaneously from the infection dynamic due to the fact that during incubation, phage are *trapped* inside the host cells, unable to diffuse, thereby resulting in a density-dependent effect on phage diffusion (Fig. 2.8).

2.4.2 Decreased Steric Effects due to Host Lysis Promote the Transition to Pushed Waves at High Bacterial Densities

The effect of virus-host steric interactions can be best appreciated by comparing the phase diagram of the UDM+ to that of the VDM+, and is a direct consequence of the variable diffusion coefficient that our model *explicitly* introduces in Eq. 2.4.

In the VDM+, transitions to semi-pushed and fully pushed waves occur at high values of f with very weak dependence on K_{max}. This results from the boost in phage diffusion that occurs in the bulk of the plaque as host cells lyse and steric effects decrease. Because the boost increases with increasing difference in bacterial density between the front ($\overline{B} = B/B_0 = 1$) and the back ($\overline{B} = 0$), higher initial bacterial density f will lead to a stronger boost. We find empirically that, beyond a given point controlled exclusively by f, the phage behind the propagating front will disperse sufficiently fast to be able to catch up with the front and generate a semi-pushed or even a fully-pushed wave. For what follows, it is useful to name this explicit boost to diffusion $\overline{D}_{exp}$, which is mathematically identical to the dimensionless diffusion coefficient in Eq. 2.4, and reaches its maximum in the bulk of the plaque where no bacteria are left (Fig. 2.5 and dashed red line in Fig. 2.9).

2.4.3 A Second "Implicit" Density-Dependent Diffusion Emerges from the Viral Infection Dynamics

Since the UDMs lack the explicit density-dependent diffusion, the appearance of transitions to pushed regimes in the UDM- may seem surprising (Fig. 2.7). To understand the origin of these transitions, it is helpful to consider the effects of the parameter $K = \alpha\tau B_0$ that controls the transition. Adsorption and incubation (quantified by the parameter K) are not only key for the effective growth rate of the phage population, but also for the effective dispersal rate of the phage, as they control

the time and the probability that phage particles are “trapped" in a host cell, unable to disperse. As K increases, either *more* phage adsorb to host cells per unit of time (increased adsorption rate), or they are trapped in the host for *longer* (increased lysis time), resulting in a hampered dispersal of the phage (Fig. 2.8). The strength of this effect, by which phage is kept *prisoner* by the host cell, has to depend on the number of host available to infect, and thus be the strongest at the edge of the expansion, where there is plenty of uninfected host, and the weakest in the bulk, where all the host has been removed. Beyond a certain point, we therefore expect phage diffusion to be sufficiently hindered at the front to allow the phage in the back to catch up and generate a pushed wave.

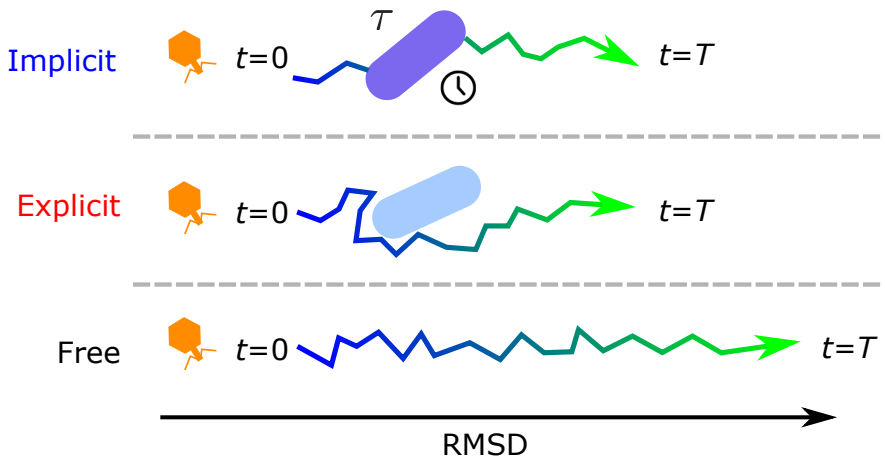


Fig. 2.8 An illustration of both explicit and implicit effects to phage diffusion. Due to the explicit effect, phage diffusion is hindered by steric interactions with bacterial hosts, while the implicit effect hinders phage diffusion by trapping the virus for a period τ during which it cannot disperse. Colour gradients on the phage trajectories indicate the passage of time.

To quantify the reduced dispersal resulting from viral infection, we consider a system of point-like phage particles diffusing across a field of completely permeable “sticky” obstacles, mimicking host bacteria that trap phage for a time τ. A simple mean-field analytical argument (see Sec. 2.8.13), demonstrates that the particles in this system exhibit a hindered diffusion D compared to their free diffusion $D_0 = D(B = 0)$, such that

$$\frac{D}{D_0} = \hat{D}_{imp} = \frac{1}{1 + bK_{max}}, \tag{2.7}$$

where b is the local density of host that can be infected by phage relative to the density B_{max} at which diffusion is completely prevented. We note here that, by definition, $\hat{D}$

is the phage diffusion coefficient relative to the bulk of the expansion, in parallel to $\overline{D} = D/D(B_0)$, defined earlier, which is the phage diffusion coefficient relative to the front of the expansion.

When adsorption to infected cells occurs (UDM+), infected cells *trap* phage as much as uninfected cells ($b = (B + I)/B_{max}$ in Eq. 2.7) resulting in a diffusion coefficient in the dimensionless model of the form

$$\overline{D}_{imp+} = \frac{\hat{D}_{imp+}}{\hat{D}_{imp+}(B_0)} = \frac{1+K}{1+(\overline{B}+\overline{I})K}. \tag{2.8}$$

Because of the shape of the bacterial density profile during the expansion, phage diffusion will then be highest in the bulk and slowest at the front (black line in Fig. 2.9). Yet, fast diffusing phage appear too far from the front to contribute to the expansion, resulting in pulled waves across parameter space (Fig. 2.7).

By contrast, when adsorption to infected cells is prevented (UDM-), phage can no longer become trapped in the infected region behind the front ($b = B/B_{max}$ in Eq. 2.7), so that

$$\overline{D}_{imp-} = \frac{\hat{D}_{imp-}}{\hat{D}_{imp-}(B_0)} = \frac{1+K}{1+\overline{B}K}, \tag{2.9}$$

and fast diffusing phage emerge much closer to the expansion front (blue vs. black lines in Fig. 2.9, Sec. 2.8.15). Preventing adsorption to infected cells is therefore equivalent to a boost in implicit diffusion in the infected region just behind the front, which can be approximated to

$$\frac{\overline{D}_{imp-}}{\overline{D}_{imp+}} = 1 + \psi \approx 1 + \frac{\overline{I}K}{1+\overline{B}K}, \tag{2.10}$$

(blue dashed line in Fig. 2.9). This boost is sufficient to shift the fast diffusing phage closer to the expanding front and, if K is sufficiently large, to generate a transition to pushed waves.

It is important to point out that this implicit density-dependent diffusion emerges spontaneously from the viral infection dynamics (common to many viruses), where infecting viruses trapped in the host cannot contribute to the advancement of the front until they are released from the host. As a consequence, unlike the explicit density-dependent diffusion, this effect cannot be easily accommodated into the diffusion coefficient of our model, as it does not act independently of the infection and growth processes. Indeed, an alternative interpretation for this mechanism can be provided by a density-dependent death rate. In our model, adsorption to previously infected hosts

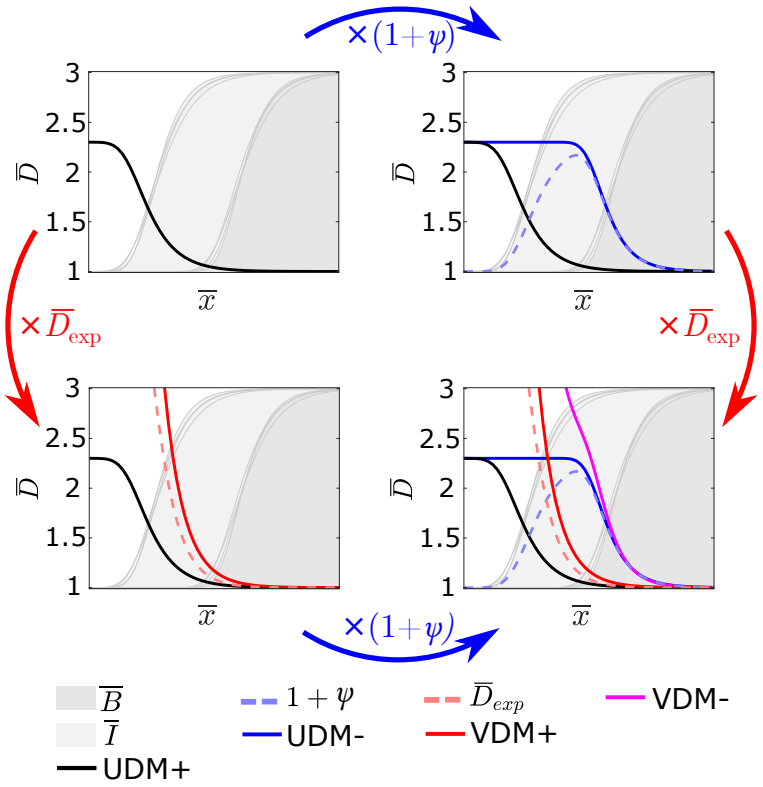


Fig. 2.9 Proxies for the diffusive behaviour in each of the model variants plotted as a function of position across the expansion front. The base diffusion rate $\overline{D}_{imp+}$ (Eq. 2.8) in the UDM+ (black solid line) is modified either by the term $1+\psi$ (blue arrow and blue dashed line) in the UDM- (blue solid line), which accounts for the now unhindered diffusion in the region of infected cells, or by an additional term $\overline{D}_{exp}$ (red arrow and red dashed line) in the VDM+ (red solid line) which accounts for the hindrance due to steric effects. Both modifications occur in the VDM- (magenta line). Faint grey lines indicate the different front profiles from the model variants used to calculate the diffusion rate profiles (see Sec. 2.8.16).

is equivalent to phage 'death', as it results in the permanent loss of these phage. Going from the case where adsorption to infected cells occurs to the case where it does not (from + to - models) will then lead to an increase in net growth rate in the region of infected cells, which lies at intermediate viral density (Fig. 2.5). The result is a higher net growth rate at intermediate population densities similar to what an Allee effect would generate in a mono-species expansion [32, 35].

2.4.4 Implicit and Explicit Density-Dependent Diffusions Act Independently with Multiplicative Effects

Because the implicit and explicit boosts to diffusion discussed above have different physical origins and are controlled by different parameters (K and f, respectively), they play significant roles in different regions of parameter space. The implicit boost that results from a lack of adsorption to infected cells, encoded in $(1+\psi)$, is stronger at large K_{max}, where more phages are trapped by hosts for a longer period of time. Instead, the explicit boost caused by steric interactions, encoded in $\overline{D}_{exp}$, is dominant at low K_{max}. The ratio of the two effects over parameter space is shown in Fig. 2.10.

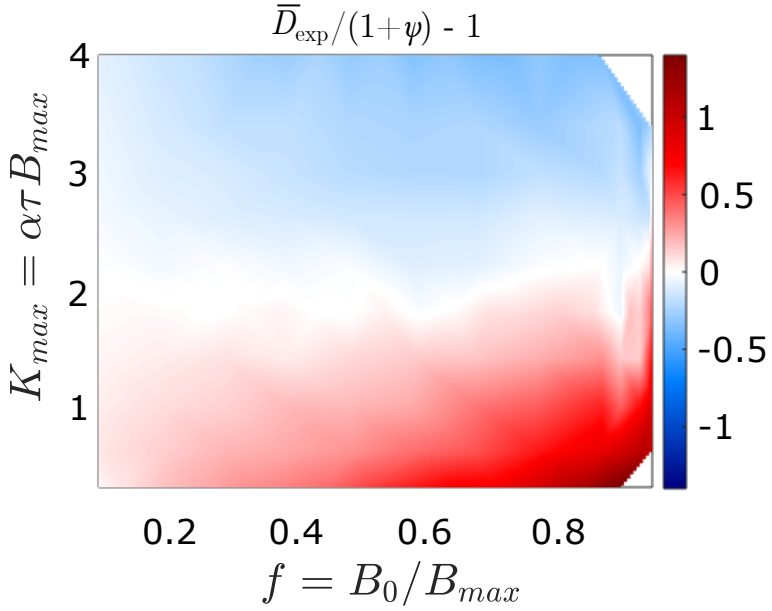


Fig. 2.10 The ratio of the explicit boost to diffusion $\overline{D}_{exp}$ to the implicit boost to diffusion $(1+\psi)$, as a function of f and K_{max}. This is obtained by determining the strength of each effect at the front position where the phage population is $3/4$ times the steady state population $\overline{V}_{max}$. It can be seen that the implicit boost is dominant at large K_{max}, while the explicit boost dominates at low K_{max}.

Extending the analytical argument with which we defined the implicit boost to diffusion, we can show that, to a first approximation, explicit and implicit effects act independently over a basal diffusion coefficient (see Sec. 2.8.14). As a consequence, preventing adsorption to infected cells corresponds to multiplying the diffusion coefficient by $1+\psi$ (from + to - models, blue arrows in Fig. 2.9 and Fig. 2.11). Similarly, including steric effects corresponds to multiplying the diffusion coefficient by $\overline{D}_{exp}$ (from UD to VD models, red arrows in Fig. 2.9 and Fig. 2.11). As a result, we can

write the dimensionless diffusion coefficient of the VDM-, which exhibits both effects, as

$$\overline{D}_{VDM-} = (1+\psi)\overline{D}_{exp}\overline{D}_{imp+}, \tag{2.11}$$

where all the terms are calculated with respect to $\overline{B}$ and $\overline{I}$ from the VDM- simulations.

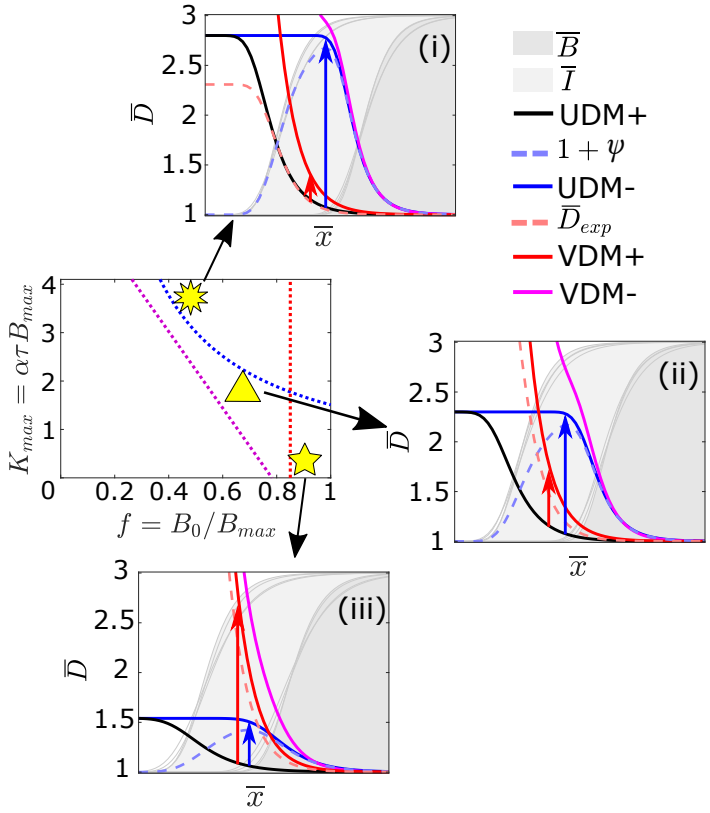


Fig. 2.11 Diffusion profiles for three representative cases are shown, all highlighted in comparison to the semi-pushed transition boundaries for the models shown in Fig. 2.7 (dotted lines in central panel). Red and blue arrows highlight the shift from the UDM+ to the VDM+ and UDM- respectively at the position where the viral profile is approximately 3/4 times its steady-state. (i) A region of parameter space with high K_{max} and low f, where the UDM- and VDM- are pushed (note that the blue arrow is much larger than the red arrow, indicating a much greater boost to diffusion from the implicit effect). (ii) A region of parameter space with intermediate f and K_{max}, where only the VDM- is pushed (note similar sizes of both blue and red arrows) (iii) A region of parameter space with low K_{max} and high f, where the VDM+ and VDM- are pushed (note that the red arrow is much larger than the blue arrow, indicating a much greater boost to diffusion from the explicit effect).

In contrast to the other models, this function depends non-trivially on K_{max} and f, making it challenging to find a simple parameter combination that controls the transitions to pushed waves. Nonetheless, we see that the diffusion coefficient

determined at 3/4 times the steady state phage population is able to qualitatively capture the behavior of the transition lines in all models (Fig. 2.12) and it explains why the transition lines in the VDM- approaches the transition lines in the UDM- and the VDM+ at high and low K_{max}, respectively, where either effects dominate (Fig. 2.7 and Fig. 2.11). While the phage diffusion at a specific population density is, in principle, insufficient to predict whether the expansion is pushed, which by definition depends on the whole wave dynamic, Fig. 2.12 illustrates that regions in parameter space with similar effective diffusion within a model correspond to similar types of expansions. This supports the idea that the density-dependent diffusion, whether implicit or explicit, is the key ingredient that leads to transitions to pushed waves.

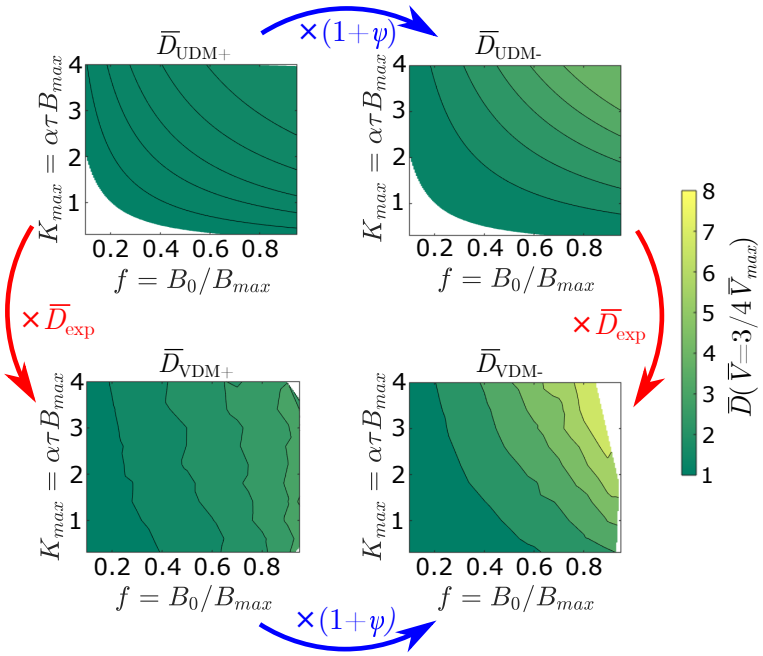


Fig. 2.12 Dimensionless diffusion rates in each of the models determined at the front position where the phage population is 3/4 times the steady state population $\overline{V}_{max}$, plotted as a function of both f and K_{max}. Contour lines indicate levels of constant $\overline{D}$. The behaviour of the contours qualitatively matches that of the transition boundaries in Fig. 2.7.

2.5 Transitions are Robust to Other Model Factors

2.5.1 Burst Size

Fig. 2.7 illustrates the transition to pushed expansions for a fixed burst size β. Because burst size affects the growth rate of the phage and, consequently, its expansion speed, it is natural to wonder whether it also has any effect on the transitions between expansion types. In Fig. 2.13 we present phase diagrams such as in Fig. 2.7, but for a burst size β=20 instead of β=50. It can be seen by comparing both Figures that the qualitative behaviour of the transitions remains unchanged: in the UDM-, transitions are characterised by constant values of K_s and K_p; in the VDM-, transitions are approximately straight lines characterised by gradients m_s and m_p, and intercepts a_s and a_p ($f = m_{s,p}K_{max} + a_{s,p}$); in the VDM+, transitions are largely independent of K_{max}, and occur at critical values f_s and f_p.

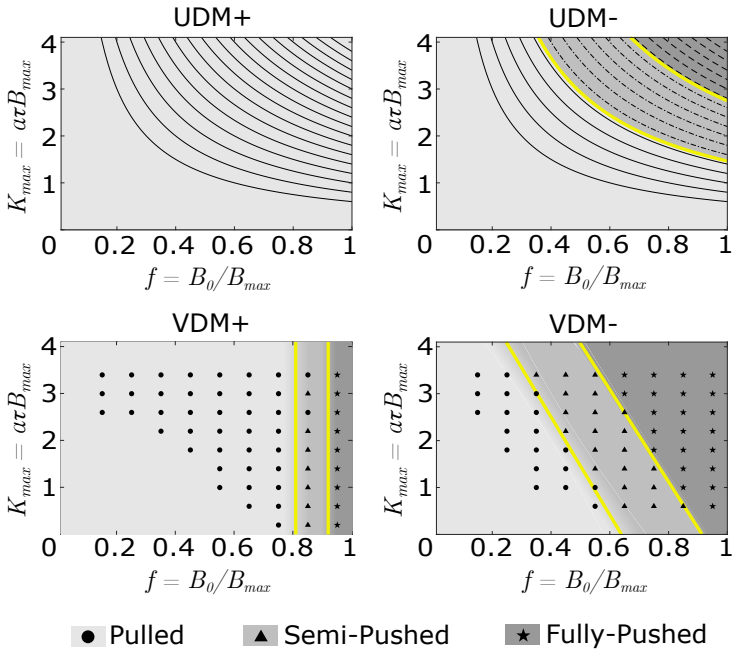


Fig. 2.13 Phase diagrams showing the expansion types for the four model variants as a function of f and K_{max}, β=20 throughout. As can be seen by comparison to Fig. 2.7, the qualitative behaviour of the transitions in each of the models remains the same, and can be characterised using the same parameters: $K_{s,p}$ for the UDM-; $f_{s,p}$ for the VDM+; $m_{s,p}$ and $a_{s,p}$ for the VDM-. As before, lines in the UDMs, and data points in the VDMs indicate the parameter combinations for which numerical integration was performed, and speeds calculated. Transition boundaries (yellow lines) are inferred from the data points calculated.

As it seems the behaviour can be characterised in the same manner when burst size is changed, we simplify our examination by focusing only on how the burst size alters these specific characteristic parameters. Rather than attempting to produce the whole phase diagram for each of the models at various β, as this is very computationally intensive when β is either small or large, we instead choose a specific K_{max} value, and for various values of β, we vary f at this K_{max} value. From this, the parameters $K_{s,p}$ and $f_{s,p}$ as a function of β can be easily obtained (i.e. the parameters describing the UDM- and VDM+ transitions respectively).

To simplify our investigation of the behaviour in the VDM-, and limit the number of computationally intensive calculations required, we assume that $m_{s,p}$ are constant with burst size, and using data from one specific K_{max} value we calculate how $a_{s,p}$ vary as a result. In the two cases where the full phase diagrams were computed, the transition gradients were calculated as $m_s = -0.125(5), -0.091(11)$ and $m_p = -0.097(5), -0.101(2)$ for $\beta = 50, 20$, indicating agreement to within 2σ and 1σ respectively.

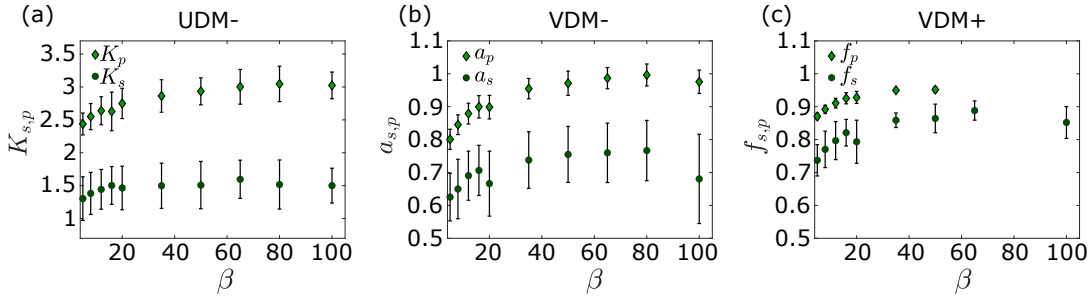


Fig. 2.14 Behavior of the critical parameters describing the location of the transitions (Fig. 2.7 and Fig. 2.13) as a function of burst size β. (a): Critical values K_s and K_p in the UDM-. (b): Critical values a_s and a_p in the VDM-, calculated from transition locations at K_{max}=2.2, assuming the gradients of the transitions m_s and m_p are approximately constant when varying β, to maintain computational feasibility. (c): Critical values f_s and f_p in the VDM+, similarly calculated from transition locations at K_{max}=2.2. For $\beta > 50$ we did not observe transitions to fully pushed waves in the parameter regime of $f \leq 0.95$ investigated. Extending the parameter regime was computationally unfeasible. In each case, to determine the error bars, we assume a 1% error in the model speeds as before, and shift the speeds accordingly, to determine the resultant shift in the transition parameters.

We find that while the general shape of the transitions for the model variants does not depend on β (Fig. 2.13), the exact location of the transitions are affected (Fig. 2.14). The dependence of all of the transition parameters on β is similar across models. Above $\beta \approx 40$, the parameters exhibit only a weak dependence on burst size,

whereas when β decreases below this value, the transition parameters also decrease, increasing the parameter range of K_{max} and f in which we observe a pushed wave.

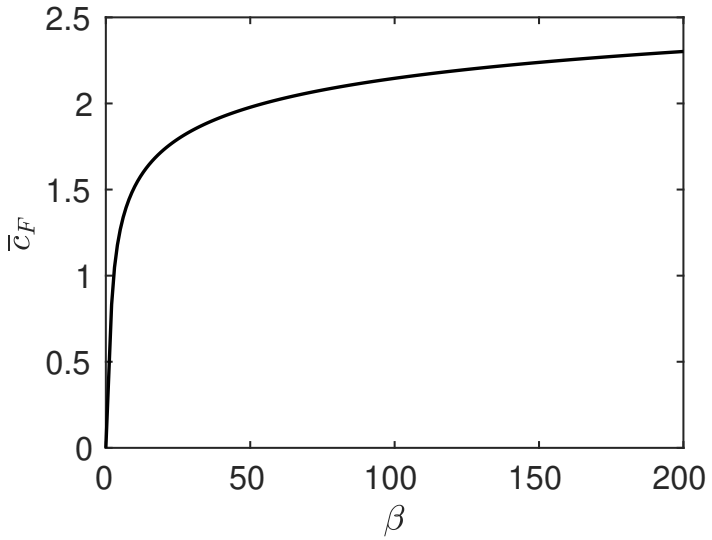


Fig. 2.15 Spreading speed of the linearised model for $K = 1.0$ as a function of burst size β (in the linearised model, K and β are the only independent parameters).

This behaviour qualitatively matches the dependence of the spreading speed of the linearised model $\overline{c}_F$ on burst size (Fig. 2.15). While f and K_{max} determine the ability of phage in the bulk to catch up to the front and contribute to the dynamics, either due to explicit or implicit hindrance to diffusion, β only contributes to the phage growth rate and, as a result, the speed of the front. At lower values of β, the spreading speed is greatly reduced as the limited number of phage released at the tip after each lysis event struggle to clear the host cells around them, allowing the phage in the back to catch up more easily, regardless of the mechanism, and contribute to the expansion. As burst size is increased however, the opposite is true, although the speed gains that come with increased β become increasingly marginal, as the uninfected host within the vicinity of recently lysed cells become saturated with newly released phage.

2.5.2 Bacterial Growth

Our model thus far assumes that the host bacteria are not growing. While this certainly can be the case, it is by no means always true, and so it is natural to ask how our results are affect by a growing host population. To this end, we modify our reaction-diffusion

model to include a logistic growth term:

$$\frac{\partial B}{\partial t} = -\alpha V B + \mathbf{r_0 B}\left(\mathbf{1} - \frac{\mathbf{B}+\mathbf{I}}{\mathbf{B_0}}\right), \tag{2.12a}$$

$$\frac{\partial I}{\partial t} = \alpha V B - \alpha V_{t-\tau} B_{t-\tau}, \tag{2.12b}$$

$$\frac{\partial V}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial V}{\partial x}\right) - \alpha V B - \alpha^* V I + \beta\alpha V_{t-\tau} B_{t-\tau}, \tag{2.12c}$$

where terms in bold indicate new terms, and r_0 is the growth rate of the bacteria at low densities. We estimate this growth rate by assuming a doubling time of bacteria equal to 30 mins (typical of T7 host *E. coli*), resulting in $r_0 = \ln(2)/30$ min^{-1}. Note, that in each case the carrying capacity is equal to the initial bacterial density B_0, meaning that growth only occurs across the front region. This represents the situation that would occur in natural environments where bacteria far from the front have grown to a stationary density (set by B_0) by the time the front arrives. Additionally, we assume that the replication of infected hosts is negligible due to the burden caused by producing new phage.

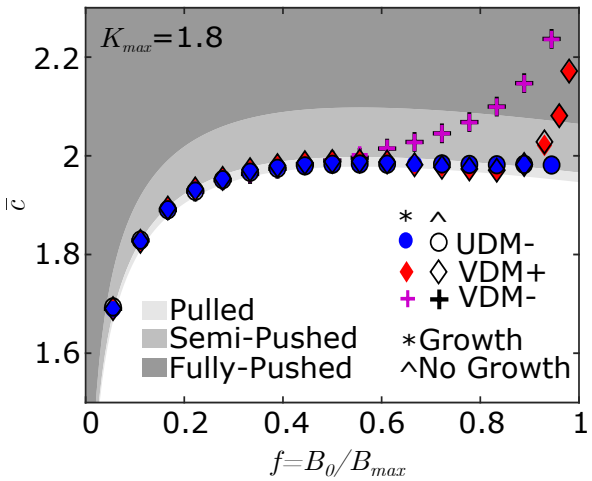


Fig. 2.16 Dimensionless front speed $\bar{c}$ as a function of bacteria fraction f, with shaded regions indicating the different expansion types, when bacteria grow logistically as described in Eqs. 2.12. As in Fig. 2.6, parameters are chosen to represent typical T7 expansions with $\beta = 50$, τ=18 mins, and αB_{max}=0.1 min^{-1} [58, 70, 71], corresponding to $K_{max} = 1.8$ in our model. Additionally, a doubling time of 30 mins has been assumed to calculate the growth rate of the host bacteria. By comparison to the results obtained in the absence of bacterial growth (as in Fig. 2.6), it can be seen that bacterial growth has no effect on the transition to pushed expansions.

It can be seen in Fig. 2.16 that the introduction of this bacterial growth term has no discernible effect on the type of expansion that occurs in the parameter space describing typical T7 expansion through host *E. coli*.

2.5.3 Cell Debris

Another aspect of the system which our model thus far neglects is the possibility that some of the debris from lysed cells could trigger phage adsorption. If this were the case, it would likely result in a reduction of the parameter space corresponding to pushed waves, as there would be fewer phage in the bulk that were able to catch up to the front. To explore this prospect further, we modify our reaction-diffusion model to account for this possibility:

$$\frac{\partial B}{\partial t} = -\alpha V B, \tag{2.13a}$$

$$\frac{\partial I}{\partial t} = \alpha V B - \alpha V_{t-\tau} B_{t-\tau}, \tag{2.13b}$$

$$\frac{\partial V}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial V}{\partial x}\right) - \alpha V(B+I) + \beta\alpha V_{t-\tau}B_{t-\tau} - \alpha\mathbf{V}(\mathbf{B_0} - (\mathbf{B}+\mathbf{I}))\mathbf{d_f}, \tag{2.13c}$$

where terms in bold indicate new terms, and d_f controls the fraction of lysed cells that are capable of adsorbing phage (or alternatively, how able the cell debris is to adsorb phage in comparison to the cells which generated it). Note that we have limited the model to the case where adsorption to infected hosts occurs, as it seems unlikely that phage would be unable to adsorb to infected hosts but would be able to adsorb to the debris from those hosts (VDM+).

It can be seen in Fig. 2.17a that while the introduction of adsorbing cell debris does not qualitatively change the behaviour of the model (i.e. that transitions to pushed waves occur at high bacterial fractions in the VDM+), it does have a slight impact on the location of these transitions (Fig. 2.17b). As the fraction of cell debris capable of adsorbing phage d_f increases, the critical bacteria fractions f_s and f_p where transitions occur also increase. This result confirms our intuition: a larger d_f will cause a greater reduction in free phage in the bulk of the plaque, thereby reducing the number of phage available to catch up to the front and contribute to the expansion dynamics.

While increasing the rate at which cell debris adsorbs phage does result in 'less pushed' expansions, we can see in Fig. 2.18 that the resulting depletion of free phage in the bulk of the expansion is even more significant. Indeed, after an expansion of

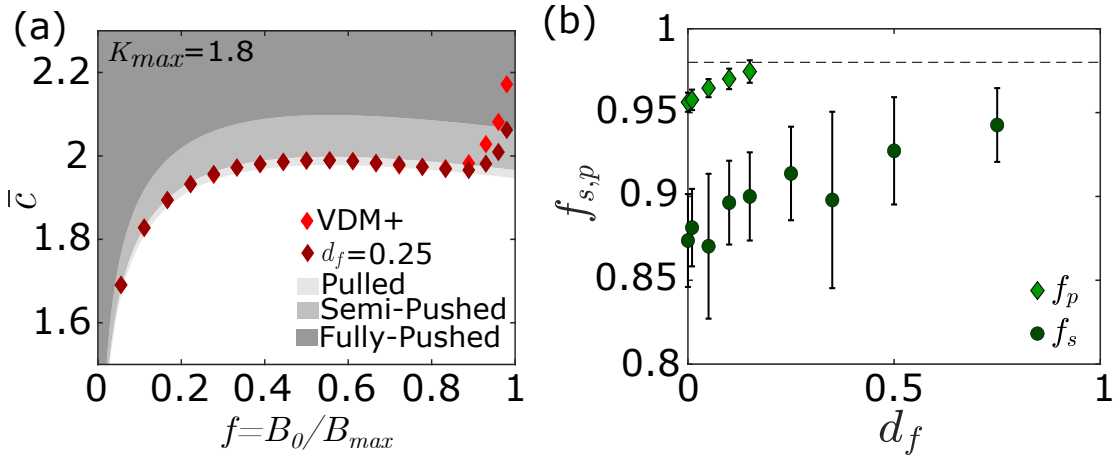


Fig. 2.17 (a): Dimensionless front speed $\bar{c}$ of the VDM+ as a function of bacteria fraction f and debris fraction d_f, with shaded regions indicating the different expansion types. As in Fig. 2.6, parameters are chosen to represent typical T7 expansions with $\beta = 50$, τ=18 mins, and αB_{max}=0.1 min^{-1} [58, 70, 71], corresponding to $K_{max} = 1.8$ in our model. (b): How the critical bacteria fractions f_s and f_p where transitions occur change with d_f. The dashed line indicates the upper bound of f for which results were computed. We expect fully pushed transitions to continue to occur above this threshold for some time as d_f is increased.

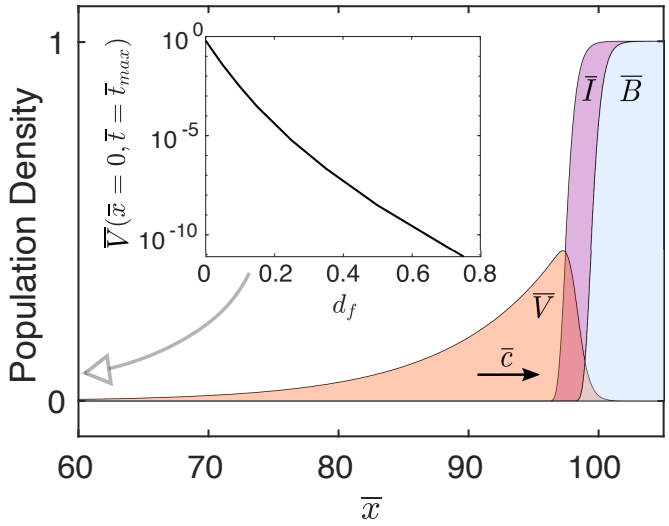


Fig. 2.18 Population concentrations $\overline{B}$, $\overline{I}$ and $\overline{V}$ at time $\bar{t} = \bar{t}_{max}$ for a debris fraction d_f=0.25. The front is propagating with dimensionless speed $\bar{c}$ to the right. It can be seen that behind the expansion front the density of the viral population is quite low. Inset shows the density of the viral population at $\overline{x}$=0 and $\bar{t} = \bar{t}_{max}$ as a function of d_f (i.e. it indicates the viral density in the centre of the plaque).

50 lysis times (15 hours if τ=18 mins) we find that there are approximately 10^5 fewer phage in the bulk than would be found at the expansion front when d_f=0.25, and approximately 10^{10} when d_f=0.6. These values are clearly unrealistic for T7, as in the laboratory phage can be easily recovered from the centre of a plaque by simply stabbing with a needle. Therefore, if we assume that the steady state phage population at the front is approximately βB_0=100 μm^{-2} (β=100, B_0=1 μm^{-2}), and then *very* conservatively assume that easy recovery would require at least a phage density of 10^3 mm^{-2}, then we can conclude that any d_f leading to a central density reduction of more than 10^5 is physically unrealistic (i.e. $d_f \geq 0.25$).

Given that transitions to both semi-pushed and fully-pushed waves occur in our model up to a debris fraction d_f=0.25 (which we believe to be the *very* upper bound of physically realistic behaviour) with only a minor shift in the bacterial fractions where these transitions occur, we can conclude that this effect, should it occur, has only a minor impact on our results.

2.6 Rate of Genetic Diversity Loss

The simulations in this section were performed by Dr. Nikhil Krishnan. Single-species populations expanding via pushed waves have been theoretically and experimentally shown to retain genetic diversity much longer than their pulled counterparts [22, 25, 35, 41]. To test whether this property is maintained in viral populations and determine the effect of virus-host interactions on the rate of diversity loss, we developed an agent-based stochastic implementation of one of our numerical models and tracked the viral heterozygosity H as a function of time (Fig. 2.19, Sec. 2.8.18). The heterozygosity H in a viral biallelic population is given by

$$H = \frac{1}{M} \sum_{i}^{M} 2 f_i (1 - f_i), \tag{2.14}$$

where M is the total number of demes in the simulation box, and the fraction of the two alleles in deme i are f_i and $1 - f_i$. We focus on the UDM- as it is the simplest of our models that exhibits pushed waves, and it is also relevant for viruses beyond phage T7.

Analogously to previous studies on single-species, we find that the heterozygosity decays exponentially over time, so that we can define an effective population size N_e as the inverse of the decay rate in units of generations (Fig. 2.19b, see Sec. 2.8.18

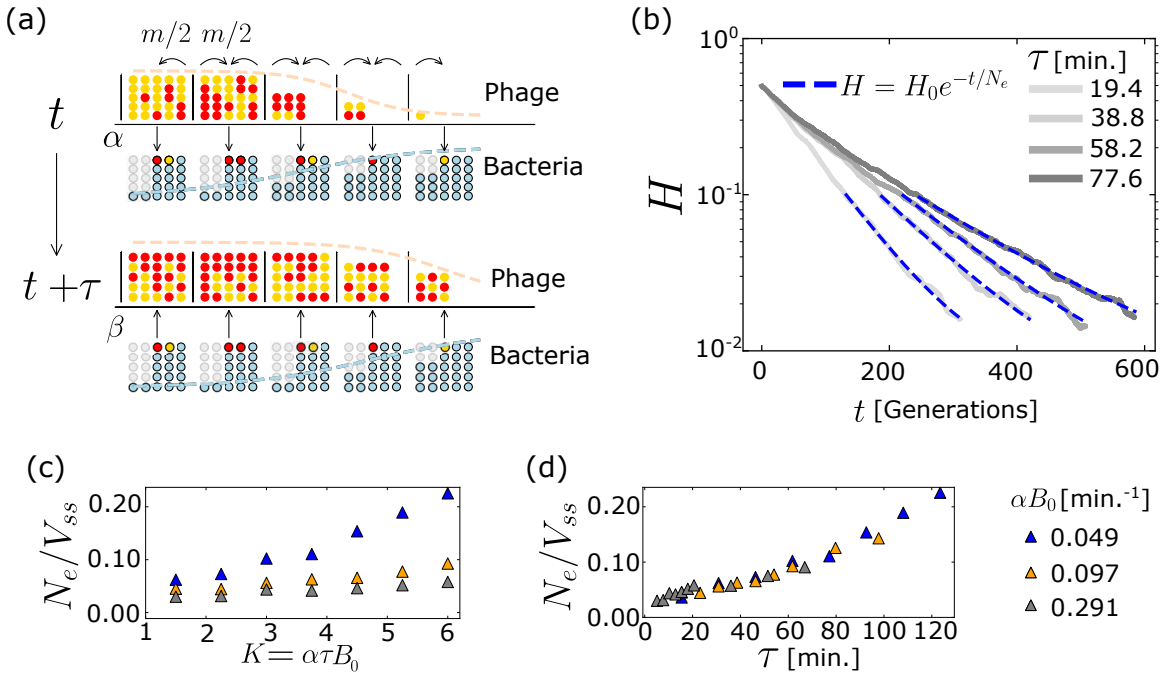


Fig. 2.19 (a): Simulation algorithm of stochastic simulations of plaque growth. Each deme contains two labelled phage populations (yellow and red circles) and bacteria that can be dead (grey), uninfected (light blue) or infected with one type of phage (red/yellow). At each step, phage can migrate to neighboring demes with probability m/2 (right/left curved arrows), infect an uninfected cell with probability α (downwards arrow), and after τ time steps, introduce β phage into the deme upon cell lysis (upwards arrow). Dashed lines indicate the analogous density profile, similar to those shown in Fig. 2.5. (b): Example of linear fit to log transformed heterozygosity data, shown for $\alpha B_0 = 0.077$ min^{-1}, $B_0 = 100$ cells/deme, and a range of τ values. Heterozygosity data represents the average of ∼1000 simulations. Slope of fit, $1/N_e$, is the calculated decay rate for the given parameter values. (c,d): Effective population size N_e for $B_0 = 75$, normalized by the steady state population in the bulk of the population V_{ss} as a function of K (c) and τ (d), over a range that could be expected in various viral populations. Colors indicate specific values of αB_0. Errors in (c) and (d) due to linear fit of heterozygosity decay over time are negligibly small and not shown.

for details). To account for the fact that adsorption rate, lysis time and bacterial density also change the density profile of the viral population, we normalize the effective population size by the steady-state value of the viral population in the bulk of the wave V_{ss}. This normalization aims at providing a direct comparison between our system and previous theoretical studies where the carrying capacity of the population was held constant (see Fig. 2.20 for data without normalization). It should be noted, however, the parameters in those studies (including carrying capacity) could be independently

controlled, whereas in our system the steady state population size is an emergent property that depends both on the phage infection parameters, and the given aspects of the model variant in question. Therefore the phage population size behind the expansion front will not always be directly comparable to the carrying capacity as used in these previous studies - take for example the case where phage can adsorb to cell debris (Sec. 2.5.3), resulting in the depletion of phage behind the front over time.

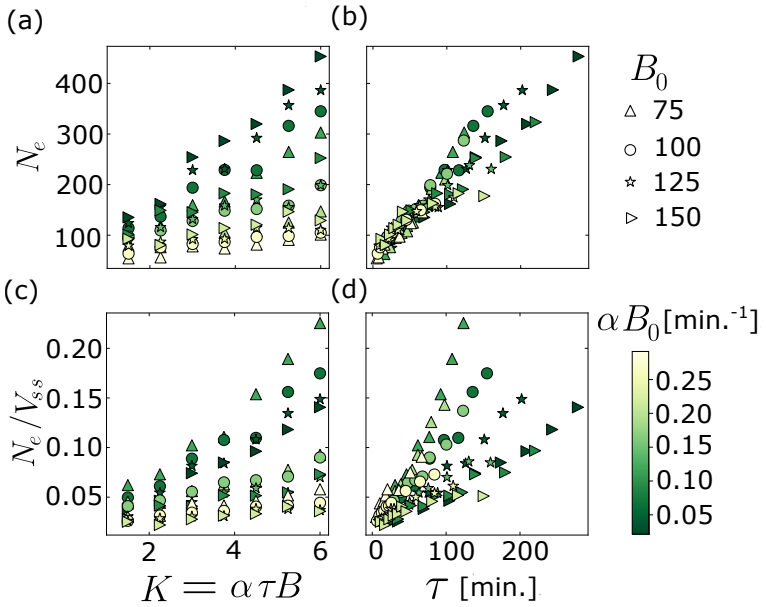


Fig. 2.20 Effective population size N_e over a range of B_0 values shown: (a) unnormalised and as a function of K; (b) unnormalised and as a function of τ; (c) normalised by V_{ss} and as a function of K; (d) normalized by V_{ss} and as a function of τ. Color indicates αB_0 in min^{-1}. B_0 and αB_0 were varried as they are the quantities one would measure independently in experiments. Effective population size is mainly controlled by τ as in Fig. 2.19d, over the range parameters examined. Errors due to linear fit of heterozygosity over time curve are negligibly small and not shown.

Our results show that the level of *pushedness* of the wave, controlled by K in the UDM-, can signficantly increase the normalized effective population size (two-fold increase between a pulled wave, $K = 1.5$, and just above the pushed transition, $K = 3$, Fig. 2.19c). However, we also find that, not surprisingly, K alone is not sufficient to determine the value of the effective population size of the expansion. Similar observations have been made before in single-species expansions, where distinct cooperativity models, all displaying transitions between pulled and pushed waves, were found to be characterised by different values of N_e [28]. Remarkably, we find that an

excellent predictor for the value of N_e in the UDM- is the lysis time τ (collapse of datasets in Fig. 2.19d and Fig. 2.20). A possible reason behind this observation is the different effects that lysis time and adsorption rate have on the steady-state viral density profile (one broadens it, while the other narrows it), which, in turn, impacts N_e [22, 25, 28]. Further analyses are necessary to pin-point the exact mechanisms that link virus-host interactions and viral diversity, urging for future theoretical work to investigate viral genetic diversity in spatial settings.

2.7 Discussion

In this work, we first experimentally quantify how the diffusion of phage in a bacterial lawn is hindered by steric interactions with the host bacterial cells, resulting in a density-dependent diffusion coefficient. Going beyond current descriptions of plaque growth, which have considered host density-dependence only for setting a constant diffusion coefficient parameter, we construct a reaction-diffusion model of the phage-bacteria system that explicitly incorporates a diffusion coefficient that depends on local host density, and therefore varies in time and space. We show that, in contrast to current thinking which assumes that viral expansions are always pulled, this 'explicit' effect can lead to a transition from pulled to pushed waves at high host densities. We also show that a second, independent density-dependence in diffusion emerges implicitly from the underlying viral dynamics, whereby phage are unable to disperse during replication within the host. We find that when adsorption to infected host cells is prevented, this 'implicit' effect can also lead to the transition to pushed waves. Together, this indicates that bacteriophage offer an excellent experimental system to study the effect of density-dependent diffusion on expansion dynamics.

The transition from a pulled wave to a pushed wave has traditionally been associated with increased co-operativity between individuals, quantified by density-dependent growth, or more recently, density-dependent dispersal [25, 28, 32]. By analogy, the density-dependence in phage diffusion can be interpreted as an emergent co-operativity, which stems from the fact that as phage work together towards cell lysis, they remove bacterial obstacles, indirectly favouring the dispersal of neighboring phage. The fact that the diffusion is dynamically changed as phage replicate could lead to interesting ecological feedback. Ecological feedback on diffusion has been theoretically shown in other contexts to lead to pattern formation, and in some cases help maintain genetic diversity and mitigate the risk of extinctions [138]. Indeed, density-dependent dispersal

has been identified as a key ingredient in a generic route to pattern formation in bacterial populations [139].

We find that the transition to a pushed wave can occur due to two separate effects: an explicit density-dependent diffusion coefficient, caused by steric interactions between the phage and the host bacteria, which is dominant in crowded host environments, and an implicit hindrance to the diffusion of the phage population at the front caused by the viral infection dynamics. We therefore expect that the pushed dynamics will be strongest in populations that experience both effects, and where adsorption to infected host is absent (VDM-). Some bacteriophage have mechanisms that prevent adsorption to already infected cells, usually by blocking receptor sites post-infection [59]. Bacteriophage T5 produces a lipoprotein (Llp) that is expressed at the beginning of infection, preventing superinfection by blocking its own receptor site (FhuA protein), and protecting newly produced phage from inactivation by binding to free receptors released by lysed cells [60, 61]. Similar mechanisms are also well documented in several temperate phage. Phage ΦV10 possesses an O-acetyltransferase that modifies the specific ΦV10 receptor site (the O-antigen of *E. coli* O157:H7) to block adsorption [140]. Similarly, *Pseudomonas aeruginosa* prophage D3 modifies the O-antigen of LPS on the host surface to prevent adsorption of the many phage that bind to the O-antigen [141]. This is similar to other *Pseudomonas* prophage which encode for twitching-inhibitory protein (TiP) that modifies the type IV pilus on the *P. aeruginosa*, preventing further adsorption [78, 142].

Mechanisms that prevent superinfection by forbidding adsorption to infected cells have been observed in viruses beyond bacteriophage [143]. For instance, cells recently infected with Vaccinia virus VacV (the live vaccine used to eradicate smallpox) express two proteins that repel super-infecting virions, resulting in plaques that grow four-fold faster than predicted by replication kinetics alone [144]. Our results show that even in the absence of explicit steric effects, pushed expansions can occur if adsorption to infected hosts is prevented (UDM-), as is the case for VacV, simply due to the fact that viruses are unable to disperse during incubation, suggesting that pushed waves might be far more widespread than previously thought among different viral systems. A further example can be found in plant viruses, which spread through a host population that is fixed in place, and so are unable to disperse whilst inside infected cells. Additionally, a wide range of plant viruses are known to prevent superinfection, making them good candidates for pushed expansions, although the precise nature of the exclusionary mechanism is often poorly understood at present [145–147].

Pushed dynamics in range expansions have been shown to have significant consequences for the evolution of the population. In pulled expansions, the high susceptibility to stochastic fluctuations results in inefficient selection, as beneficial or deleterious mutations can effectively behave as neutral due to the small number of individuals contributing to the dynamics [148–150], and leading, for instance, to the accumulation of deleterious mutations, known as expansion load [151, 152]. In fully pushed waves, stochastic fluctuations are much weaker as more individuals contribute to the advancement of the population, allowing beneficial mutations to establish more easily and deleterious mutations to be purged [28, 35]. Our stochastic simulations show that even the implicit density-dependent diffusion alone can slow down the rate of diversity loss up to 5-fold under reasonable phage infection parameters (Fig. 2.19). Remarkably, we find that the rate of diversity loss strongly depends on the lysis time, but only weakly on adsorption (αB_0), even if the two parameters are expected to contribute equally to the level of "pushedness" of the wave. This observation reveals a rich and non-trivial evolutionary dynamic displayed by our viral model that distinguishes it from classic mathematical descriptions of pushed waves, where dispersal, growth and cooperativity are controlled by independent parameters.

Going forward, three clear avenues emerge as a result of our work. Firstly, the complex dependence of the expansion dynamics on the infection parameters that we observe indicates that viral expansions offer a currently untapped ground for further theoretical studies. Our model provides a framework to investigate the evolutionary dynamics of an expanding viral population in terms of the realistic processes that occur therein. Within this framework, future work is required to fully characterise the complex interplay that each of the infection processes exhibit, and ultimately determine what impact they have on the viral evolutionary dynamics.

Secondly, while this work provides theoretical predictions and physical insights regarding the transition from pulled to pushed waves in viral expansions, it also points at phage plaques as a well-controlled model system to experimentally investigate these theories in a laboratory setting. We have shown (Fig. 2.6 and Fig. 2.7) that pushed waves can occur during plaque growth in conditions easily achievable in the laboratory. While it is challenging in a laboratory setting to fully replicate the complex environments found in nature, we believe that plaques offer an alternative and possibly more realistic environment to study these dynamics than the environments typically used thus far, such as cultures in a 96-well plate, where dispersal is achieved by artificial migration schemes [40, 41, 90].

Lastly, our experiments have shown that the rate of phage diffusion strongly depends on the host environment and, in particular, can dramatically differ from liquid culture measurements, where even at high overnight densities of $\sim 10^9$ cells/ml, the volume fraction occupied by the cells is ~ 0.001, and so diffusion is effectively unhindered. This realisation raises the more general question of whether other phage life history parameters also depend strongly on the surrounding host environment. This question will be addressed further in Chapter. 4.

2.8 Methods

2.8.1 Bacterial Strains

Five strains of *E. coli* were involved in this work. The first strain, *E. coli* BW25113 (CGSC# 7636), is susceptible to phage infection. This strain was transformed previously with a plasmid expressing venus YFP to create the second strain, *E. coli* eWM43 [133]. This strain was further transformed (see Sec. 2.8.5) with plasmid pAK501 (Addgene# 48107) [153], which confers resistance to chloramphenicol, to create the third strain, *E. coli* eMTH43. The fourth strain used, *E. coli* $\Delta waaC$ (JW3596-1, CGSC# 11805), is resistant to phage infection through deletion of the *waaC* gene, the product of which is involved in the production of lipopolysaccharide, the recognition of which is essential for the adsorption of phage [134]. This strain was transformed previously with a plasmid expressing mCherry to yield the final strain *E. coli* eWM44 [133]. A summary of these details can be found in Appendix A.

2.8.2 Bacteriophage T7

The phage used in the study is the obligately lytic bacteriophage T7. The wild-type phage was originally obtained as an aliquot from the wild-type stock of the Richardson lab (Harvard Medical School, Boston, MA).

2.8.3 Media Preparation

The *E. coli* media used throughout this work is 25 g/l Luria Broth (henceforth LB), prepared in distilled water from powder (Invitrogen - 10 g/l NaCl) and autoclaved. For growth solid media, LB-agar was prepared in concentrations of 7 g/l and 20 g/l agar (VWR Chemicals), henceforth referred to as 0.7% and 2% agar. If required, 15

μg/ml chloramphenicol (Sigma-Aldrich) or 100 μg/ml ampicillin (Sigma-Aldrich) were added to either the LB or agar after autoclaving.

2.8.4 Culturing

E. coli were streaked from freezer stocks on 2% agar plates and incubated overnight at 37 °C. These plates were then stored at 4 °C for up to 2 weeks.

Liquid cultures of bacteria were prepared by picking single colonies from the streaked plates and depositing in a culture tube with 2-3 ml of LB. These cultures were then incubated overnight under shaking at 37 °C to allow the bacteria to grow - henceforth referred to as an overnight culture.

To prepare phage cultures and stocks, 10 μl of an overnight culture of host *E. coli* (most often BW25113) was diluted in 990 μl of LB and incubated under shaking at 37 °C for approximately 2 hours to generate an exponentially growing culture. Phage were then added to this culture, and re-incubated until clear (usually approximately 1 hour) to create the phage lysate. To create phage stocks, this lysate was then mixed with NaCl to a final concentration of 1.4 M, and centrifuged to remove any cell debris. The resulting supernatant was stored at 4 °C and used as stock phage.

2.8.5 Creating Strain eMTH43

To create strain eMTH43, the previous strain eWM43 was transformed with plasmid pAK501 to confer resistance to chloramphenicol. Chloramphenicol is a bacteriostatic antibiotic, that inhibits protein synthesis through interactions with the 50S subunit of the ribosome [154]. Resistance is most commonly conferred by the enzyme chloramphenicol transacetylase, that acetylates hydroxyl groups of chloramphenicol, rendering it unable to bind to the 50S subunit [154].

To achieve the transformation, first a stock of competent eWM43 cells was produced. The bacteria were grown overnight in LB under shaking at 37 °C to generate an overnight culture. The following day, this culture was diluted in fresh LB to achieve a density of OD_{600}=0.01-0.05. This diluted culture was then incubated at 37 °C with shaking until it reaches exponential phase (OD_{600}=0.6-0.8). This culture was then centrifuged at 13,000 rpm for 10 mins at 4 °C. From this point onward the bacteria were not allowed to warm up. The supernatant was removed, and the bacteria resuspended in cold TSS buffer (LB, 10% PEG MW8000, 20 mM $MgSO_4$, 5% DMSO, 0.22 μm filter sterilised) in a 1:10 ratio. The cells were then split into 100 μl aliquots and stored at -80 °C.

The competent cells were then transformed by thermal shock. One of the 100 μl aliquots was allowed to thaw on ice. Also on ice, an eppendorf tube was filled with 70 μl of distilled water, 10 μl of plasmid pAK501 and 20 μl of KCM buffer 5$\times$ (0.5 M KCl, 0.15 M $CaCl_2$, 0.25 M $MgCl_2$, 0.22 μm filter sterilised). The aliquot of competent cells were then added, and mixed gently by pipetting up and down. They were then incubated in the ice for 20 mins, and then at 37 °C for 5 mins. 750 μl of fresh LB was then added, and the cells were incubated at 37 °C for a further 45 mins. 100 μl of this culture was then plated using glass beads on plates containing 2% LB-agar and 15 μg/ml chloramphenicol. The remainder of the culture was then spun down at 13,000 rpm for 1 minute, and the cells resuspended in 100 μl LB, which was then plated in a similar fashion. The plates were then incubated overnight at 37 °C. The following day, colonies that had grown on the plate must have been successfully transformed with the plasmid pAK501, as otherwise they would not have resistance to chloramphenicol. These colonies were picked, and grown in LB plus 15 μg/ml chloramphenicol, and used to create freezer stocks.

2.8.6 Determining Phage Concentration

The oldest but still most commonly used technique for the quantification of viral concentration relies on the ability of phage to form plaques in a lawn of bacteria [74]. When a single phage encounters a susceptible host bacteria, it will infect and later lyse this cell, releasing many new phage progeny. When this process occurs in some kind of lawn of bacteria, repeated cycles of infection, lysis and diffusion will result in a growing clearing in the lawn, known as a plaque.

To utilise this feature to determine the concentration of phage in a culture, a *plating assay* is performed. To do this, the phage culture is first ten-fold serial diluted in LB. Next, 20 μl of each serial dilution is mixed with 100 μl of overnight bacteria and 3-5 ml of molten 0.7% agar, gently vortexed, plated, and left on a bench-top overnight. In the plates where the serial dilutions contained no phage, the bacteria will have grown in the 0.7% agar, generating a slightly opaque layer of agar. In the plates where the serial dilutions contained many phage ($\geq$ 1000), the bacteria will have all been lysed, resulting in a transparent layer of agar. In plates where the serial dilutions contained approximately 1-300 phage, each of the phage will have resulted in a visible, distinct plaque in the otherwise slightly opaque bacterial layer. By counting these plaques, the concentration of phage in the original culture can be estimated.

It should be noted that each phage that gives rise to a plaque is known as plaque forming unit (PFU). This procedure therefore estimates the phage concentration in terms of PFUs, which gives a count of the viable phage (i.e. those capable of infecting hosts), but not necessarily the absolute number of phage.

It should also be noted that the number of phage in each dilution will be Poisson distributed [50], and so to a first order approximation, the error in the number of plaques n counted in each plated will be given by $\sqrt{n}$. It is therefore preferable to estimate the phage concentration from plates containing a larger number of plaques, as the relative error scales with $n^{-0.5}$.

An alternative method for estimating phage concentration is known as the *spotting assay*. In this assay, 100 μl of overnight bacteria and 3-5 ml of molten 0.7% agar are mixed, gently vortexed, and plated with no phage. Serial dilutions of the phage culture are prepared as before, and 2 μl droplets of each dilution are placed separately on the surface of the plate. Again, the plates are then left overnight on the bench-top, and an order of magnitude estimate for the phage concentration can be drawn from the presence or absence of a plaque at each dilution. This method is less accurate than the plating assay, as only an order of magnitude estimate is gained, but it is simpler to perform, and less time consuming. Despite the lower accuracy, it can still be useful if the precise concentration is not needed, and it can be particularly useful if a rough estimate of the phage concentration is not known beforehand.

2.8.7 Diffusion Experiment Sample Preparation

To measure the diffusion coefficient of phage, 96 well plate sized omni-plates, containing 35 ml of 20 g/l agar (VWR Chemicals), with LB (Invitrogen) - NaCl concentration 10 g/l - and 15 μg/ml chloramphenicol (Sigma-Aldrich) were prepared and kept at room temperature for 2 days. The presence of chloramphenicol in the plate prevents growth of the background strain eWM44, so to maintain its density constant over the course of the experiment. Plates were then refrigerated if they were to be used at a later date. Overnight liquid cultures of *E. coli* were grown from single colonies at 37 °C in LB with either 100 μg/ml ampicillin (Sigma-Aldrich) or 15 μg/ml chloramphenicol (Sigma-Aldrich) for eWM44 and eMTH43 respectively.

To create the background lawn of bacteria, the optical density at 600 nm (OD_{600}) of the eWM44 culture was measured, and diluted into LB to obtain the desired density (calculated on the basis that $OD_{600} = 0.1$ equates to $\sim 10^8$ cells/ml [155]). A 500 μl droplet of this culture was then spread with glass beads (radius 4 mm) across the

surface of the agar until dry. This process (a 500 μl droplet spread with fresh beads) was repeated a further two times to achieve as uniform a distribution as possible. The plate was then left for a further 10 minutes before proceeding to the next step.

10 ml of eMTH43 overnight culture was spun down and re-suspended in fresh LB, so as to give an OD_{600} reading of 0.50 if diluted hundredfold. 2 μl droplets of the concentrated culture were then pipetted onto the lawn of eWM44 in a grid like pattern, spaced approximately 1 cm apart, and left to dry (approximately 15 mins after the last droplet was pipetted). Each plate contained approximately 60 host droplets. The plate was then incubated at 37 °C for 1 hour.

10 μl of stock T7 phage (10^7 pfu/ml) was diluted in 100 μl of black food dye. 1 μl droplets of this dilution were pipetted onto the surface of the agar, in the gaps between the previously pipetted droplets of eMTH43, and left to dry. A schematic of the resulting set-up can be seen in Fig. 2.1.

2.8.8 Data Acquisition

The plates were imaged using a Zeiss Axio Zoom.V16 stereo microscope equipped with a Zeiss PlanApo Z 0.5x/0.125 FWD 114 mm objective. Images of the sample were taken every 20 minutes for a period of 24 hours. During the imaging period, the sample was kept with its lid on at 37 °C using an ibidi Multi-Well Plate Heating System.

2.8.9 Data Analysis

All of the images were analysed using Fiji (v1.52h), an open source distribution of ImageJ focused on scientific image analysis [156, 157].

The time Δt necessary for a clearing to appear in the droplets of eMTH43, and the separation r between the point at which a clearing forms and the edge of the nearest phage droplet was recorded (Fig. 2.1). By measuring r and Δt over many droplet-droplet pairs, we can calculate the rate of phage diffusion by fitting the relationship [158]:

$$r^2 = 4D\Delta t + \text{constant}, \tag{2.15}$$

where D is the diffusion coefficient, and the constant arises due to the delay between the arrival of the phage and the formation of a visible plaque. The diffusion rates determined from this method are then corrected in Sec. 2.2.1 to account for the geometry and number of phage used in our experimental set-ups (i.e., to account for

the fact that phage may not originate at the edge of the droplet, and the fact that we are not measuring the first passage time for a *single* phage).

2.8.10 Simulations of Phage Diffusion

Firstly, 1000 point-like tracers[2] i (representing phage) are generated with random co-ordinates $(x_i,\ y_i)$ within a circle of radius r_V (representing the initial droplet of phage) on a 2D plane (representing the agar surface). The tracers then diffuse in discrete steps, by moving in each timestep to a new random co-ordinate within a circular region of radius r_{step} around their current positions. The simulation proceeds in this fashion for 10^3 steps, with the mean-squared displacement $\overline{r^2}$ of all of the tracers from their initial co-ordinates being recorded at each step. A simple linear fit to $\overline{r^2}$ as a function of t is then used to determine the 'true' diffusion coefficient D_{true} (the gradient of the fit line is equal to $4D_{true}$). Different values of D_{true} can be achieved by altering r_{step}.

Next, we determine the diffusion coefficient that would be found using our experimental set-up. In our simulations the division of space and time are made such that $\Delta x = \Delta y = 1 \equiv 1\,\mu$m and $\Delta t = 1 \equiv 1\,$s. In addition to the 1000 tracers, a circular target region with radius r_B (representing the initial droplet of bacteria) is defined in the 2D plane, with an edge-to-edge distance $r_{separation}$ from the initial bounding circle of tracers. To represent our experimental set-up, we set $r_V = 2000$ and $r_B = 4000$, and $r_{separation}$ is chosen randomly in the range [0, 1000].

The tracers are then allowed to diffuse as before, until one of them crosses the boundary of the target region. At this point the simulation ends, and the time elapsed Δt is recorded. In addition, the distance r between the point at which the tracer entered the target region and the closest point on the edge of the initial bounding circle of tracers is recorded. These two measurements represent the measurements that were made using our experimental set-up. As in our experiments, if no tracer has entered the target region after a time $t = 8.64 \times 10^4 \equiv 24\,$hours then the simulation ends and no data is recorded.

This process is then repeated 100 times (to represent the ∼100 droplet-droplet pairs we have in each experiment) with a new value of $r_{separation}$ chosen each time. The data generated is comparable to the data collected through our experiments, and allows us

[2] i.e., the same number of phage as in the droplets in our experimental set-up.

to determine the 'experimental' diffusion rate $D_{experiment}$ in the same fashion as in the actual experiments.

2.8.11 Numerical Solutions of Reaction-Diffusion Model

We consider the scaled reaction-diffusion equations describing free phage V, uninfected bacteria B and infected bacteria I

$$\frac{\partial \overline{B}}{\partial \overline{t}} = -K(\beta - 1)\overline{V}\,\overline{B}, \tag{2.16a}$$

$$\frac{\partial \overline{I}}{\partial \overline{t}} = K(\beta - 1)\overline{V}\,\overline{B} - K(\beta - 1)\overline{V}_{\overline{t}-1}\overline{B}_{\overline{t}-1}, \tag{2.16b}$$

$$\frac{\partial \overline{V}}{\partial \overline{t}} = \frac{\partial}{\partial \overline{x}}\left(\overline{D}\frac{\partial \overline{V}}{\partial \overline{x}}\right) - K\overline{V}\,\overline{B} - K^*\overline{V}\,\overline{I} + \beta K\overline{V}_{\overline{t}-1}\overline{B}_{\overline{t}-1}, \tag{2.16c}$$

on a finite interval of length L_D with homogeneous Neumann boundary conditions. Throughout we used $L_D = 120$ and a maximum possible time of $\overline{t}_{max} = 50$. Initially, we set $\overline{B} = 1$ over the whole interval, $\overline{V} = 1$ for $\overline{x} \leq 2$, and $\overline{V} = 0$ elsewhere. There are initially no infected bacteria ($\overline{I} = 0$). Solutions are determined on a mesh of uniform space and time, with divisions of $d\overline{t} = 0.1$, and $d\overline{x} = 0.1$ or $d\overline{x} = 0.2$ to give the best balance between precision and compute time for a given parameter set.

A sketch of the fronts during the expansion can be seen in Fig. 2.5. The dimensionless spreading speed $\overline{c}$ of the front is determined by tracking the midpoint of the bacterial wave (i.e. $\overline{B} = 0.5$) over time. For pulled fronts, the spreading speed is known to demonstrate a power law convergence to an asymptotic value [159]. In the case where a steady spreading speed was not reached, the spreading speed was given by the asymptotic value which produced the best power law fit to the data.

The transitions between pulled, semi-pushed and fully pushed have been found to occur at specific wave speeds with respect to the linearised (Fisher) speed c_F - the speed determined solely by the linear dynamics at the tip of the front [25]. Pulled expansions spread with a speed equal to the speed of the linearised model $c = c_F$, while pushed expansions spread faster [29]. The transition between semi-pushed and fully pushed occurs at a speed of $\frac{c}{c_F} = \frac{3}{2\sqrt{2}}$, with waves below this speed being semi-pushed, and above this speed being fully pushed [25]. These thresholds have been shown to be robust to the details of the population dynamics [28], though their appropriateness for multi-species expansions requires further investigations. We here use the same values for illustration purposes.

Even though strictly speaking pulled expansions occur when the spreading speed is equal to the speed of the linearised model ($c = c_F$), due to errors in determining the speed over a finite time (i.e. errors due to the power law fit when determining the asymptotic speed), we conservatively consider speeds within 1% of the linearised speed as corresponding to pulled expansions.

A length scale characterising the width of the infected region $\Delta\overline{x}_I$ was also computed by tracking the separation between the midpoint on the wave of uninfected bacteria ($\overline{B} = 0.5$), and the midpoint on the wave of infected bacteria ($\overline{B} + \overline{I} = 0.5$) over time. An average was taken over the final 20 time points for the reported value of $\Delta\overline{x}_I$.

2.8.12 Linearised Solution of Reaction-Diffusion Model

To determine the transition between pulled, semi-pushed and fully pushed regimes, the solution to the linearised model is required. To achieve this, we first look for travelling wave solutions to Eq. 2.16 in the co-moving co-ordinate $\overline{z} \equiv \overline{x} - \overline{c}\overline{t}$ where $\overline{c}$ is the dimensionless front speed.

As the components approach their limiting concentrations at the leading edge of the front, the linearised form of the model becomes valid, and so, following previous work [58, 62], we assume the concentrations take the form $\overline{V} = a_1\exp(-\overline{\lambda}\overline{z})$, $\overline{B} = 1 - a_2\exp(-\overline{\lambda}\overline{z})$ and $\overline{I} = a_3\exp(-\overline{\lambda}\overline{z})$ where $\overline{\lambda}$ is a dimensionless width parameter, and a_1, a_2 and a_3 are positive constants.

Substituting into the linearised version of the model

$$\frac{\partial\overline{B}}{\partial\overline{t}} = -K(\beta-1)\overline{V}, \tag{2.17a}$$

$$\frac{\partial\overline{I}}{\partial\overline{t}} = K(\beta-1)\overline{V} - K(\beta-1)\overline{V}_{\overline{t}-1}, \tag{2.17b}$$

$$\frac{\partial\overline{V}}{\partial\overline{t}} = \frac{\partial}{\partial\overline{x}}\left(\overline{D}\frac{\partial\overline{V}}{\partial\overline{x}}\right) - K\overline{V} + \beta K\overline{V}_{\overline{t}-1}, \tag{2.17c}$$

and writing in matrix notation yields

$$\begin{pmatrix} K(\beta-1) & -\overline{\lambda}\overline{c} & 0 \\ K(\beta-1)(1-\mathrm{e}^{-\overline{\lambda}\overline{c}}) & 0 & -\overline{\lambda}\overline{c} \\ \overline{\lambda}^2 - \overline{\lambda}\overline{c} + K(\beta\mathrm{e}^{-\overline{\lambda}\overline{c}} - 1) & 0 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = 0. \tag{2.18}$$

To find non-trivial solutions the determinant of the matrix is set to zero, leading to the characteristic equation:

$$\overline{\lambda}^2 - \overline{\lambda}\overline{c} + K(\beta \mathrm{e}^{-\overline{\lambda}\overline{c}} - 1) = 0. \tag{2.19}$$

As we are assuming here that the front is pulled, and that the initial conditions are sufficiently steep[3], the front propagates with the minimum possible speed [159]:

$$\overline{c} = \min_{\overline{\lambda}>0}[\overline{c}(\overline{\lambda})]. \tag{2.20}$$

By implicitly differentiating Eq. 2.19 with respect to $\overline{\lambda}$, and setting $d\overline{c}/d\overline{\lambda} = 0$ according to Eq. 2.20, this leads to the second characteristic equation:

$$2\overline{\lambda} - \overline{c} - K\beta\overline{c}\mathrm{e}^{-\overline{\lambda}\overline{c}} = 0. \tag{2.21}$$

The dimensionless spreading speed $\overline{c}$ is given as the unique solution to both Eq. 2.19 and Eq. 2.21 which we solved numerically for each set of parameters.

2.8.13 Analytical Model of "Implicit" Density Dependence

To develop a simple mean-field analytical model to describe the effect of the underlying viral dynamics on the phage diffusion, we imagine phage diffusing across a lawn of "sticky" penetrable disks. These disks are used to represent host bacteria cells that are able to adsorb phage for a period equivalent to the lysis time, after which the phage desorb and continue to diffuse. The disks do not pose a hindrance to the phage through steric interactions.

In this set-up, phage diffuse through a series of discrete steps, where phage move a certain distance with each step. In any given step, the probability that a phage will become adsorbed to one of the host bacteria is $p_\alpha\phi$, where p_α represents the probability of adsorbing when a phage encounters a host, and ϕ represents the fraction of all space occupied by the host bacteria. This is analogous to a Poisson point process, where events (adsorption) occur continuously and independently. Therefore, the number of

[3]The *sufficiently steep* initial condition is $\lim_{x\to\infty} \phi(x,0)e^{\lambda^* x} = 0$, where $\phi(x,t)$ denotes a generic spreading field. If the initial conditions are such that the front is not *sufficiently steep* ($\lambda > \lambda^*$), then it will propagate with a speed greater than the minimum spreading speed [159]. As noted in Ref. [29], however, while in such fronts one can intuitively describe the expansion tip as 'pulling' the front along, to distinguish them from the pulled front solutions that emerge from sufficiently steep initial conditions it is better to refer to these dynamics as 'leading edge dominated' dynamics.

steps that a phage takes before becoming adsorbed to a host is exponentially distributed with mean $t_{ads} = \frac{1}{p_\alpha \phi}$. Consequently, over the period of time $T = t_{ads} + \tau_s$, where τ_s is the lysis time (in steps), the phage will only have on average actually moved for t_{ads} of that time.

For long times (over many adsorption/desorption events), this process can be thought of as a hindered diffusion process with relative diffusion coefficient equal to

$$\frac{D}{D_0} = \hat{D}_{imp} = \frac{t_{ads}}{t_{ads} + \tau_s} = \frac{1}{1 + p_\alpha \tau_s \phi}, \tag{2.22}$$

where D_0 is the free diffusion coefficient, and $\hat{D}_{imp}$ is the relative density-dependent diffusion coefficient resulting from the hindrance posed by the underlying viral dynamics, which we have termed the "implicit" density-dependence.

This can be re-written in terms of the parameters used in the main text as

$$\frac{D}{D_0} = \hat{D}_{imp} = \frac{1}{1 + p_\alpha \tau_s \phi} = \frac{1}{1 + AbK_{max}}, \tag{2.23}$$

where A is a scaling parameter given by:

$$A = \frac{p_\alpha \tau_s \phi}{\alpha \tau B_{max} b}. \tag{2.24}$$

To compare the parameters used in the two descriptions, we can consider that in our mean-field model $\phi = b$. We can then use the fact that the term αB_{max} determines the rate at which phage are adsorbed when in contact with bacteria (as all space is filled with bacteria at B_{max}), and so, like $p_\alpha \tau_s$, the term $\alpha B_{max} \tau$ measures the total probability that phage will be adsorbed over a lysis time, assuming that the phage are always in contact with bacteria (i.e. $p_\alpha \tau_s = \alpha B_{max} \tau$). This leads to a value for A of:

$$A = \frac{p_\alpha \tau_s \phi}{\alpha \tau B_{max} b} = \frac{p_\alpha \tau_s}{\alpha \tau B_{max}} = 1. \tag{2.25}$$

Consequently, the implicit density dependence can be written equivalently in terms of either parameters as:

$$\frac{D}{D_0} = \hat{D}_{imp} = \frac{1}{1 + p_\alpha \tau_s \phi} = \frac{1}{1 + bK_{max}}. \tag{2.26}$$

2.8.14 Analytical Model Predicts Multiplicative Effects of Steric Interactions and Infection Dynamic

The model introduced above can be modified to account for the presence of steric effects. In the absence of adsorption i.e., if excluded-volume interactions were the only hindrance to diffusion, the average fraction of steps successfully "jumped" by phage compared to the total attempted would be $1-\phi$, and we can define a relative diffusion coefficient $\hat{D}_{exp} = 1-\phi$. Although this is an approximation as it does not take into account the fact that jumps may be correlated, which is why it deviates from the more precise Fricke's equation, it helps extending the analytical model in the previous section.

If we now introduce adsorption, as explained in the previous section, the average number of steps taken by phage before adsorbing to an obstacle will be $t_{ads} = \frac{1}{p_\alpha \phi}$. In the presence of steric interactions, only a fraction $1-\phi$ of these steps will be successful, so that $t_{succ} = \frac{1-\phi}{p_\alpha \phi}$. Thus, on average, the success rate of jumping if both adsorption and steric interactions are taken into account will be:

$$\frac{D}{D_0} = \hat{D}_{exp+imp} = \frac{t_{succ}}{t_{ads} + \tau_s} = \frac{1-\phi}{1 + p_\alpha \tau_s \phi}, \tag{2.27}$$

which is equivalent to the product $\hat{D}_{exp}\hat{D}_{imp}$, indicating that when both implicit and explicit effects are present, the total behaviour can be expressed as the product of both effects individually. Because in our model the functional form is explicitly input into the PDE system, the same argument holds if we replace the simplified $\hat{D}_{exp} = 1-\phi$ with the more precise Fricke's law parameterised by our experiments, resulting in Eq. 2.11.

2.8.15 Implicit 'Boost' to Diffusion

This section will derive how the implicit diffusion coefficient in the UDM- can be thought of as a boost to the implicit diffusion coefficient in the UDM+. Here, let the implicit slow-down to diffusion in the UDM+ and UDM- be denoted by $\hat{D}_{imp+}$ and $\hat{D}_{imp-}$, respectively. Following our previous derivations, these rates are given by:

$$\hat{D}_{imp^+_-} = \frac{1}{1 + b_{\substack{+\\-}} K_{max}}, \tag{2.28}$$

where $b_+ = \frac{B+I}{B_{max}}$ and $b_- = \frac{B}{B_{max}}$, owing to the fact that infected cells are either adsorbing or non-adsorbing in the two models.

So as to compare to the dimensionless set of parameters used in the model (where $\overline{D} = 1$ at the expansion front), we re-scale the implicit coefficients as:

$$\overline{D}_{imp^+_-} = \frac{\hat{D}_{imp^+_-}}{\hat{D}_{imp^+_-}(f)} = \frac{1 + fK_{max}}{1 + \rho_+ fK_{max}}, \tag{2.29}$$

where $\rho_+ = \overline{B} + \overline{I}$ and $\rho_- = \overline{B}$.

If we then compare the ratio of the two coefficients, it can be seen that

$$\frac{\overline{D}_{imp-}}{\overline{D}_{imp+}} \approx \frac{1 + fK_{max}}{1 + \overline{B}fK_{max}} \frac{1 + (\overline{B} + \overline{I})fK_{max}}{1 + fK_{max}} \tag{2.30}$$

$$= \frac{1 + (\overline{B} + \overline{I})fK_{max}}{1 + \overline{B}fK_{max}} \tag{2.31}$$

$$= 1 + \frac{\overline{I}fK_{max}}{1 + \overline{B}fK_{max}}. \tag{2.32}$$

The approximation arises from the assumption that the bacterial curves $\overline{B}$ and $\overline{I}$ are the same in both expansions. This is supported by the observation that the density profiles for $\overline{B}$ and $\overline{I}$ are very similar when compared across models (Fig. 2.9). Therefore we can see that we can write $\overline{D}_{imp-}$ in terms of $\overline{D}_{imp+}$ as

$$\overline{D}_{imp-} \approx (1 + \psi)\overline{D}_{imp+} \; ; \; \psi = \frac{\overline{I}fK_{max}}{1 + \overline{B}fK_{max}}. \tag{2.33}$$

2.8.16 Diffusion Profiles

Fig. 2.9, shows the proxy diffusion coefficients of each of the model variants as a function of position across the expansion front. To generate this, the population profiles $\overline{V}$, $\overline{B}$ and $\overline{I}$ of each of the model variants were taken at the final time step of the numerical solution (so as to be as close to the steady state as possible), and then aligned so that the half max points of the density profiles of uninfected bacteria ($\overline{B} = 0.5$) coincide. These population curves were then used to determine the proxy dimensionless diffusion coefficients:

$$\overline{D}_{UDM+} = \overline{D}_{imp+}, \tag{2.34a}$$

$$\overline{D}_{UDM-} = (1 + \psi)\overline{D}_{imp+}, \tag{2.34b}$$

$$\overline{D}_{VDM+} = \overline{D}_{exp}\overline{D}_{imp+}, \tag{2.34c}$$

$$\overline{D}_{VDM-} = (1+\psi)\overline{D}_{exp}\overline{D}_{imp+}. \tag{2.34d}$$

2.8.17 Stochastic Simulations

Our simulation algorithm is carried out on a one-dimensional lattice (Fig. 2.19a). A finite number of lattice sites (demes), denoted by i, are distributed along a line, with each containing a fixed number of bacteria $B_i = B_0$. Each deme is also initialized with $V_i = 100$ phage. In each time step, there is: a migration step in which a proportion of phage from each deme i, binomially sampled with V_i trials and probability $m/2$ ($m = 0.25$), are exchanged with each of its neighbors; an adsorption step in which the number of adsorbing phage is sampled in each deme from a binomial distribution with B_iV_i number of trials, with success probability α; and a lysis step in which each infected bacteria's state is advanced by one, and bacteria with state τ are labeled as lysed, and β new phage are inserted into the deme. The simulation box is periodically shifted with uninfected bacteria placed ahead of the population and demes with a steady state number of phage omitted and recorded. In this way the simulation box stays in the co-moving frame of the population.

When the traveling wave is established, verified by convergence of the expansion speed, all of the free and adsorbed phage are randomly labeled with one of two neutrals labels. The proportion of the population with each allele selected during the migration and adsorption step is found by binomial sampling with probability equal to the current allele fractions and the total number of events as described above. Upon lysis, all of the new phage released are labelled with the same marker as the phage which infected that bacteria.

2.8.18 Decay in Heterozygosity

The average heterozygosity H in the simulation box is given by

$$H = \frac{1}{M}\sum_{i}^{M} 2f_i(1-f_i), \tag{2.35}$$

where M is the total number of demes in the simulation box, and the fraction of the two alleles in deme i are f_i and $1-f_i$. Timesteps in our simulation are converted to

generation time T by noting that $T = 1/\alpha B_0 + \tau$ timesteps is equal to the average time for an individual virus to be adsorbed and the infected bacterial cell to lyse.

It is expected that heterozygosity decays due to genetic drift in our simulations [25, 28]. We expect that heterozygosity the $H(t)$, within a certain range of $t \in (t_s, t_f)$, will approximately satisfy the relation $H(t) = Ae^{-\Lambda(t+B)} + C$, where A, B, and C are constants, and Λ is the decay rate. With variable transient periods, A and B are unknown, but we assume C to be 0 (the heterozygosity will always decay to 0 as one of the alleles fixes). To estimate Λ, we simply take the natural log of our data, which we expect to be approximated by $\ln H(t) = \ln A - \Lambda(t + B)$. Combining constant terms, we can find Λ by simply performing a linear fit to $\ln H(t)$:

$$\ln H(t) = -\Lambda t + \text{const.} \tag{2.36}$$

We can alternatively express Λ in terms of an effective population size, N_e, where $N_e \equiv 1/\Lambda$ [22]. Following [25, 28], the fit was performed for the average of 1000 simulations, with t_i chosen such that $H(t_i)$ was as close to 0.1 as possible, and t_f was chosen such that at least 5% of simulations had non-zero values of $H(t_f)$ (Fig. 2.19b). Calculated N_e was normalized by the measured average steady state viral population per deme V_{ss} in the bulk of the established traveling waves to account for variable carrying capacity.

2.8.19 Time- and Length-Scales in Stochastic Simulations

To express the τ and αB_0 in our stochastic simulations in terms of real time units, we defined a spatial scale L in our model such that $L^2 = B_0^s/B_0^r = 115\frac{\mu\text{m}^2}{\text{deme}}$, where B_0^s and B_0^r is the initial bacterial density in the simulations and in our experimental data, respectively. With the spatial scale fixed, we can find the equivalence between simulation timesteps and minutes T using the viral diffusion constant in our simulations and our experimental data, D^s and D^r respectively. We first note that $D^s = m/2$ which is 0.125 in all our simulations, and D^r is function of $B_0^r = B_0^s/L^2$, in accordance to our fitted values in (Fig. 2.2). We then have that

$$T = \frac{D^s L^2}{D^r} = \frac{mL^2}{120 D^r (B_0^s/L^2)} \frac{\text{min}}{\text{timestep}}, \tag{2.37}$$

where D^r is in units of $\frac{\mu m^2}{s}$. Given a value of B_0^s, we can use this equivalence to convert αB_0 and τ to minutes.

'Not everything that can be counted counts, and not everything that counts can be counted.'

William Bruce Cameron (and maybe Albert Einstein)

Chapter 3

Identification of Monoclonal Sectors in Phage Range Expansions

All of the analysis presented in this Chapter is my own. The images of sectoring in phage range expansions analysed were collected by Dr. Racha Majed.

3.1 Introduction

The impact of spatial growth on the evolution of an expanding population has been discussed at length in Chapter 1, and in Chapter 2 a reaction-diffusion model was used to make theoretical predictions about the transition from pulled to pushed waves in viral expansions. Our model predicts that as bacterial density is increased, genetic drift and the rate of decline of genetic diversity should be reduced. A natural next step then is to attempt to measure the strength of genetic drift during such expansions experimentally.

In microbial expansions, one previously adopted approach to this problem is through the quantification of the properties of large monoclonal sectors that form during the growth of colonies [160–162]. An example of this can be seen for *E. coli* colonies in Fig. 3.1. Here we can see that cells in colonies initially inoculated with a mixture of YFP (green) and CFP (red) labelled cells quickly segregate into single coloured regions. This segregation occurs due to random number fluctuations in a thin band of 'pioneers' at the expansion front. Due to the spatial structure of the colony, only the cells which are in this band are able to pass on their genes to the outwardly growing cells. Additionally, because this band is so thin (estimated to be $\sim$30 μm in the expansions

shown in Fig. 3.1 [160]), the number of cells in it is very small, meaning that alleles can quickly reach local fixation due to random number fluctuations.

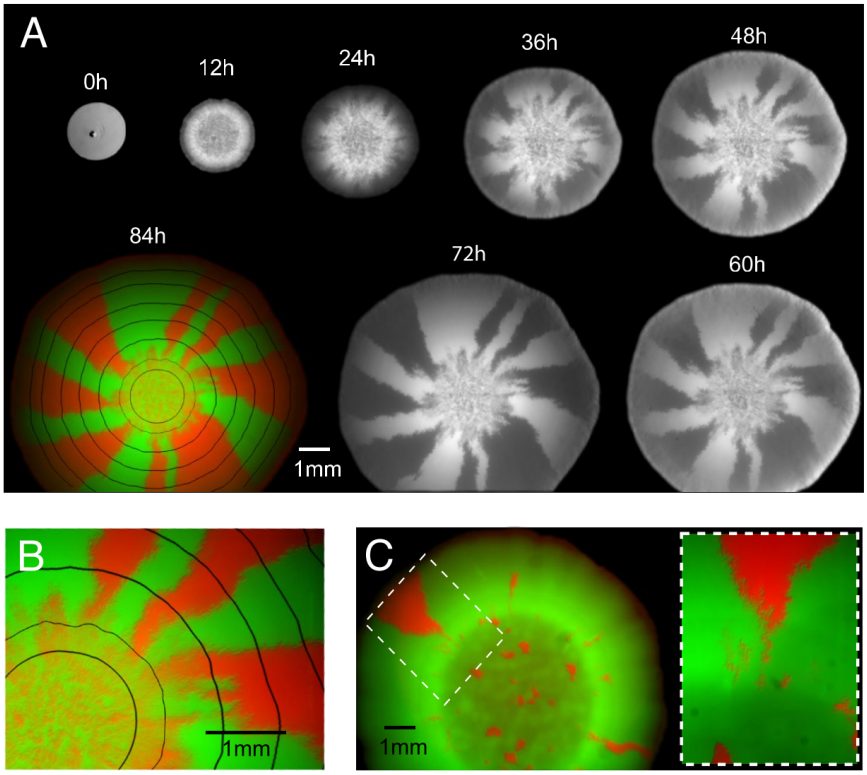


Fig. 3.1 An example of the spatial segregation of alleles into large monoclonal sectors during the expansion of *E. coli*. Colonies were inoculated with a mixture of YFP and CFP labelled cells, false coloured as green and red respectively. Figure taken from Hallatschek *et al.* (2007) [160].

It has been shown previously that, at long times, the number of sectors will reach a steady asymptotic value n_∞ because the random wandering of the sector boundaries is negligible in comparison to the increasing separation between sector boundaries, caused by the colony expansion due to radial inflation. Under the assumption that the two alleles are neutral with respect to each other, the asymptotic number of sectors is given by [161, 162]

$$n_\infty = \frac{2\pi H_0 v_\parallel}{D_g} + H_0 \sqrt{\frac{2\pi R_0 v_\parallel}{D_s}}. \tag{3.1}$$

In the above, R_0 represents the initial radius of the inoculum, $v_\parallel$ indicates the radial expansion velocity of the colony, D_s is the spatial diffusion coefficient of the sector boundary, D_g controls the strength of genetic drift and H_0 represents the initial

heterozygosity, given by $H_0 = 2f_0(1 - f_0)$ where f_0 and $(1 - f_0)$ represent the initial fraction of the two alleles [161, 162].

In bacteria, this relationship has been used to measure the strength of genetic drift in expanding populations. Since a plot of n_∞ against $\sqrt{R_0}$ will yield a straight line with gradient $H_0\sqrt{\frac{2\pi v_\parallel}{D_s}}$ and y-intercept $\frac{2\pi H_0 v_\parallel}{D_g}$, values for D_s and D_g can be extracted by measuring $n_\infty(R_0)$.

Here, we can adopt a similar approach to our viral expansions, with the previously stated goal of exploring the impact of various properties of the system (such bacterial density and lysis time) on the strength of genetic drift, in an attempt to confirm our predictions of Chapter 2. Since D_g controls the strength of genetic drift (i.e. $D_g \propto 1/N_e$ [162]), we would expect that as expansions shift from pulled to pushed, such as when bacteria density or lysis time are increased, we should see a reduction in D_g (large increase in N_e).

Our approach will different slightly from the approach taken for bacteria however, since we are not able to observe the full sectors like those shown in Fig. 3.1. This is because, if a fluorescently labelled phage was used in a similar fashion to Fig. 3.1 the continued diffusion of the phage in the plaque centre would likely result in a significant blurring of the sector boundaries. Instead, we opt to image a time-lapse of what the sectors looks like at the very edge of the plaque, using phage which are genetically engineered to induce the expression of fluorescent proteins in infected bacteria cells. By inoculating a lawn of bacteria with a mixture of two such phage, each inducing the expression of distinguishable fluorescent proteins, we can in principle measure the number of sectors during the course of an expansion, based on how many fluorescent regions there are on the plaque boundary.

The actual experiments were performed by Dr. Racha Majed, and the methods are described in Sec. 3.2. To be clear at the outset, my contribution to this project was only to develop an automated image analysis pipeline that is able to quickly and reliably identify the number of sectors in such images. To illustrate why this is important, imagine that in one such experiment, you can fit 12 individual plaques on a single agar plate at one time (or in the wells of a 12-well plate for instance). Each of these plaques will then be imaged for ≈7 hours, with images being taken every 20 mins. This means that in a single experiment one can expect to capture 12×7×3=252 images (not including multiple channels). This means that if you were to run such an experiment four times in a week, you would have generated over 1000 images. Even if it took only a couple of minutes to open each image and count the number of sectors, that

would correspond to approximately 33.5 hours a week, which is not far off a standard 37.5 hour working week. Counting the number of sectors in these images by hand is therefore at best laborious, if not completely infeasible.

3.2 Experimental Methods

As stated at the outset of this Chapter, these experiments were performed by Dr. Racha Majed.

3.2.1 Bacteriophage T7 Modification

We use phage modified to induce infected cells to produce fluorescent proteins prior to lysis, in this case mEYFP or mCherry. The procedure for producing these phage follows that set out in Ref. [163]. Briefly, recombinant T7 phage were engineered using the T7select415-1 phage display system (Millipore) with standard molecular biology techniques. The fluorescent gene was cloned under control of the strong T7 phi10 promoter downstream of the T7select415-1 10B capsid gene (the inserted DNA contains a copy of the promoter). Infected host cells produce the fluorescent protein rapidly enough to clearly fluoresce in the relevant channel before subsequently lysing and releasing a cohort of viable progeny phages. Henceforth, these two strains inducing expression of mEYFP and mCherry will be referred to as $T7_{\mathrm{mEYFP}}$ and $T7_{\mathrm{mCherry}}$ respectively.

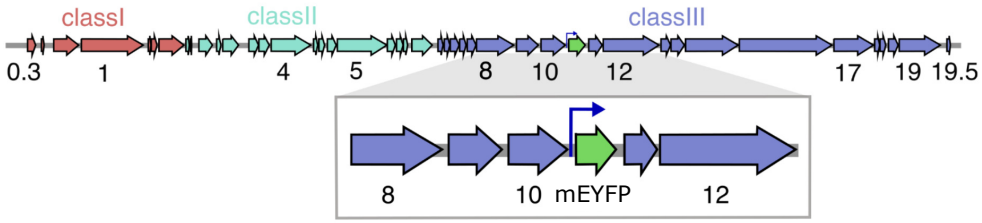


Fig. 3.2 Insertion of the fluorescent mEYFP gene under control of the strong phi10 promoter downstream of the T7select415-1 10B capsid gene. Figure adapted from Ref. [163].

3.2.2 Imaging Sectoring During Plaque Growth

To image bacteriophage sectoring, 96 well plate sized omni-plates (a plate with one single, large well), containing 35 ml of 7 g/l agar (VWR Chemicals), with LB (Invitrogen)

- NaCl concentration 10 g/l - and 100 μg/ml trimethoprim (Sigma-Aldrich) were prepared. The presence of trimethoprim in the plate prevents growth of the bacteria, so to maintain its density constant over the course of the experiment. Plates were then refrigerated if they were to be used at a later date.

Overnight liquid cultures of *E. coli* BW25113 were grown from single colonies at 37 °C in LB. To create the bacterial lawns through which the phage would expand, 12 separate 100 μl droplets of overnight culture were pipetted onto the agar surface and left to dry. Next, 1 μl of $T7_{\text{mEYFP}}$ was mixed with 1 μl of $T7_{\text{mCherry}}$, both at a concentration of 10^3 pfu/μl, and this 2 μl droplet was then pipetted onto the surface of the bacterial droplets. To investigate the impact of initial droplet radius, the volume of the phage droplet was varied over a range from 0.5 - 7 μl.

The plates were imaged using a Zeiss Axio Zoom.V16 stereo microscope equipped with a Zeiss PlanNeoFluar Z 2.3x/0.57 FWD 10.6 mm objective objective. Images of the sample were taken every 20 minutes for a period of 7 hours. During the imaging period, the sample was kept with its lid on at 37 °C using an ibidi Multi-Well Plate Heating System.

3.3 Image Analysis Pipeline

An example of the type of images collected during the sectoring experiments are shown in Fig. 3.3. One can see that there are several bright patches at the boundary of the plaque in the YFP channel. These bright patches correspond to regions where cells are currently infected with $T7_{\text{mEYFP}}$ which is inducing expression of YFP. These patches therefore correspond to $T7_{\text{mEYFP}}$ sectors. In the mCherry channel, the whole plaque appears fluorescent, meaning it is not possible to identify sectors of $T7_{\text{mCherry}}$ from these images.[1]

At its most basic level, the pipeline that I have developed has two key components: (i) finding the fluorescent intensity profile along the boundary of the plaque and (ii) identifying sectors from this intensity profile. This is obviously an incredibly course overview, although I think for the more detailed description that follows it will be

[1] I'm not entirely sure why this happens. It is possible that, owing to the relatively long maturation time of mCherry, it fails to mature before the cell lyses. However, I would expect this to result in a very week, diffuse signal, rather than what *appears* to be a strong signal throughout all of the plaque. Further controls would be required to accurately assess this, however, as fluorescence intensity is not easily compared across channels.

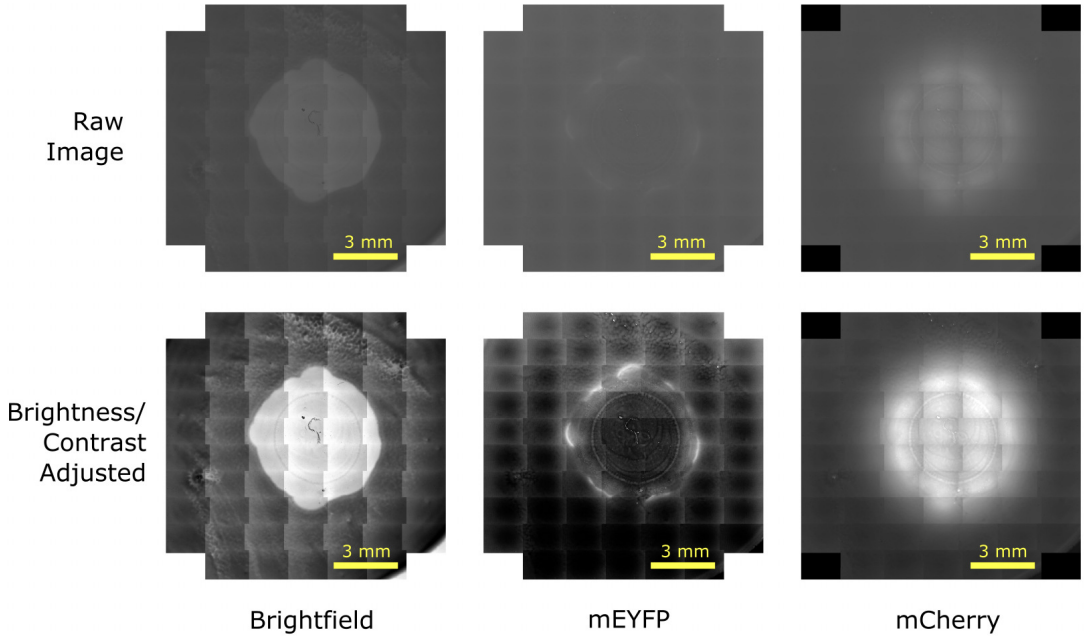


Fig. 3.3 An example set of images collected showing sector patterns during phage plaque growth. On the left are the brightfield images, which shows the plaque in the bacterial lawn. In the middle are the fluorescent images in the YFP channel, in which several bright patches can be seen at the boundary of the plaque. These bright patches show sectors where bacteria are infected by $T7_{mEYFP}$. On the right are the fluorescent images in the mCherry channel, in which the whole plaque appears quite fluorescent, meaning it cannot easily be used to find sectors. The top row shows the 'raw' images, whereas the bottom row shows images where the brightness and contrast have been manually adjusted to help the reader.

helpful to have in mind that this is ultimately the direction that I have taken. All of the analysis that follows is conducted in MATLAB.

Before we start the analysis, we first have to do some image prep. You will notice in Fig. 3.3, that each of the images has the four corner tiles 'missing,' i.e. they are completely white or black. The first step of the pipeline is to fill-in these tiles with the average pixel value of the image (Fig. 3.4).

This is because the next step is to automatically threshold the brightfield image, such that the plaque is completely white, and almost everything else is completely black (Fig. 3.4). This is performed by the MATLAB function *imbinarize* using Otsu's method [164], which "chooses the threshold value to minimize the intraclass variance of the thresholded black and white pixels." At this point the centre of the plaque is estimated by calculating the centre of mass of all of the white pixels. Additionally an

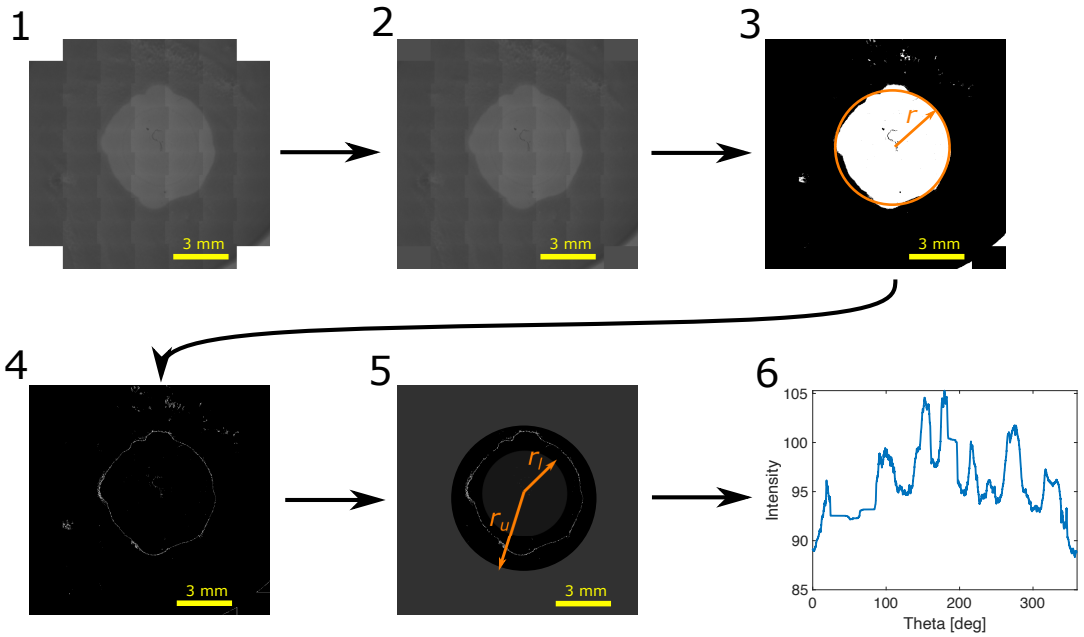


Fig. 3.4 An illustration of the first steps of the image analysis pipeline, which show how the boundary of the plaque is identified. (1) The original brightfield image. (2) The white tiles in the corner of the image are filled in. (3) An automatic threshold is set on the image. At this point, the centre of the plaque is estimated using the centre of mass of all of the white pixels, and a radius r is calculated assuming that all white pixels are contained in a circle. (4) Edges are detected in the image. (5) Edge points that are identified with a distance $< r_l$ or $> r_u$ from the centre of the plaque, where $r_l = 0.8r$ and $r_u = 1.2r$, are removed. (6) The fluorescent intensity at the remaining edge points is plotted, representing the fluorescent profile around the boundary of the plaque.

average plaque radius r is calculated by assuming that all white pixels are distributed in a perfect circle.

The next step is to identify the edges between white and black pixels in this thresholded brightfield image, which will indicate the boundary of our plaque (Fig. 3.4). This is done using the MATLAB function *edge*, which uses the Sobel edge detection method. This works by finding "edges at those points where the gradient of the image I is maximum, using the Sobel approximation to the derivative." Due to various imaging artefacts, this process will inevitably find some edge points that do not correspond to the boundary of the plaque. Examples of this can be seen in Fig. 3.4 point (4), where irregularities in the agar surface as well as the boundary of the bacteria droplet in the corner of the image are identified as edge points. A very basic filtering method is

therefore applied where any edge points with a distance from the centre of the plaque outside the range $[r_l, r_u]$ are removed ($r_l = 0.8r$, $r_u = 1.2r$).

The edge points are then sorted by angle (according to some reference direction). In the fluorescent image, the intensity is then determined at these edge points, and a moving mean is taken with an angular size of 1 degree, so as to reduce the impact of noise in the values of individual pixels. We now have the fluorescent intensity profile around the boundary of the plaque.

The final step in the analysis is to determine the number of sectors from this intensity profile. This is done using the signal processing function *findpeaks* which finds local maxima in the input signal vector (i.e. fluorescent intensity as a function of angle in our case). What is considered a "peak" is controlled by a set of optional inputs including maximum number of peaks, minimum peak height, minimum height difference between neighbouring peaks, minimum peak separation, minimum peak width, maximum peak width and minimum peak prominence, amongst others.

The options that I have used as they seem to give reasonable results (more on this in the next section) are: (i) minimum peak distance p_d, which sets a minimum angular separation between sectors (ii) minimum peak width p_w, which sets a minimum angular sector width and (iii) minimum peak prominence p_p, which measures how much a peak stands out relative to its surrounding landscape (Fig. 3.5). An example of the output of *findpeaks* can be seen in Fig. 3.6 for a given set of input parameters, where the sectors identified by the function are shown in relation to the sectors in the actual fluorescent image.

3.4 Optimising Sector Detection

As you might imagine, the images used in the previous section to illustrate the pipeline represent a relatively cherry-picked case, that results in a nice fluorescent profile where peaks can easily be seen, found by the code, and mapped to sectors in the original image (Fig. 3.6). Often, the images and the results are not so clean (Fig. 3.7). In some cases, the images contain artefacts such as bubbles in the agar, or the lighting is non-uniform, which can make it difficult to accurately identify the plaque boundary. In others, the fluorescent signal is either weak or diffuse, which makes it difficult to get an accurate count of the number of sectors even by hand. Finally, the input parameters to *findpeaks* that work well for one image may not work well for another, and might result in over/under counting the number of sectors.

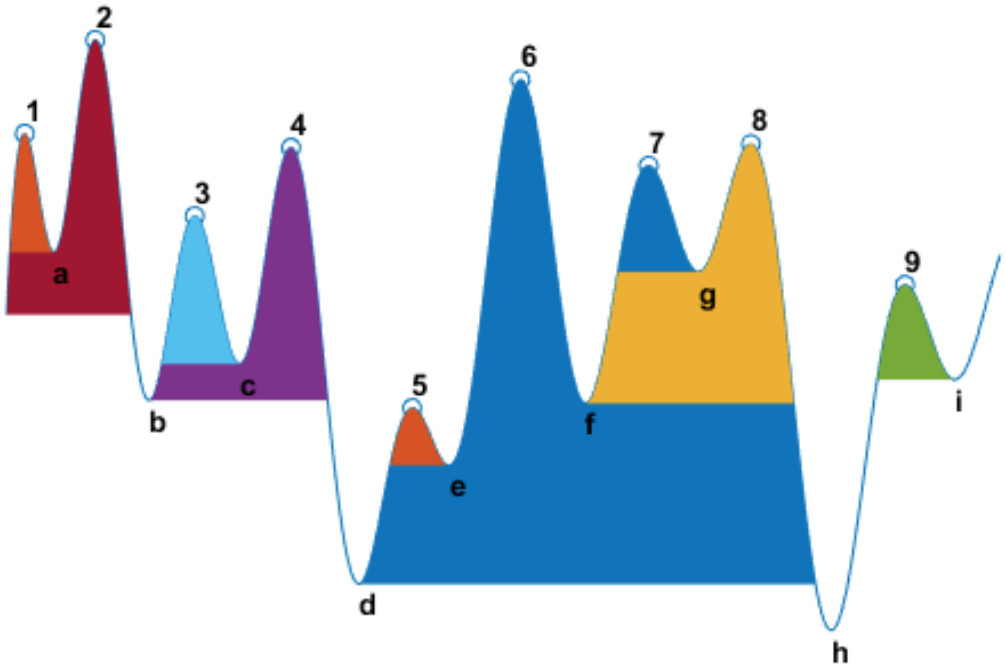


Fig. 3.5 An illustration of peak prominence. Each of the peaks 1-9 are shown, with their prominence height indicated by the coloured shading below that peak. To measure the peak prominence: (i) place a marker on the peak; (ii) extend a horizontal line to the left and right of the peak until it either crosses the signal because there is a higher peak or reaches the end of the signal; (iii) find the minimum of the signal in the two intervals just identified; (iv) determine which of these minima is higher to define a reference level. The prominence of the peak is then its height above that reference level. Figure and measurement procedure taken from *https://uk.mathworks.com/help/signal/ref/findpeaks.html*.

To try to find a combination of settings that worked well across a broad range of images, I opted to compare 100 randomly selected images across a range of timepoints and initial conditions with manual counts of the number of sectors in those same images. These 100 images were then split up between myself, my supervisor Dr. Diana Fusco and my fellow PhD student Dr. Nikhil Krishnan to count the number of sectors by hand. The reason that multiple people were used was partly to try to account for the fact that what I consider to be a sector may be slightly different to what someone else considers to be a sector. As it is difficult to tell the number of sectors in many of the images very precisely, for every image we provided a 'best estimate' for the number of sectors n_{person}, along with an upper and a lower bound of the number of sectors we thought could be in the image (n_u and n_l respectively).

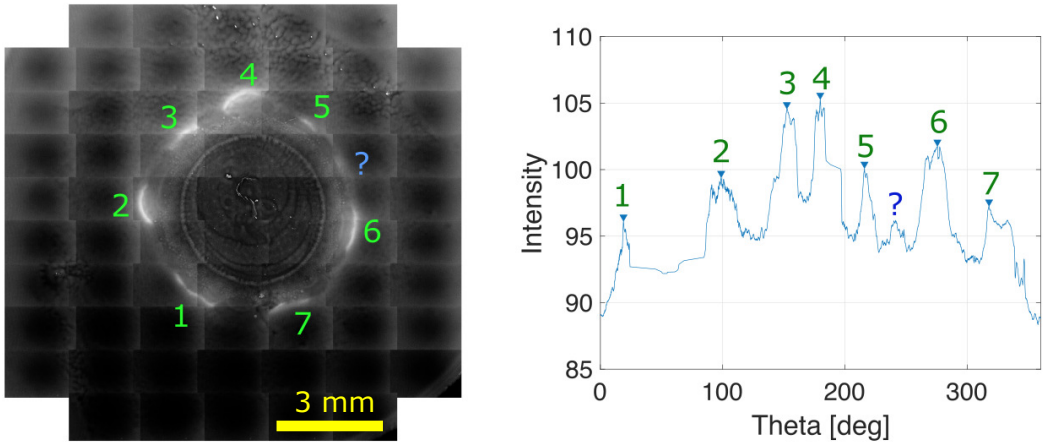


Fig. 3.6 A "good" example of a fluorescent intensity profile that relatively clearly maps to sectors that can be seen in the actual fluorescent image. For a given set of options, *findpeaks* outputs that there are 7 sectors in the image. These are indicated by the numbers 1-7 in both the image and the profile. A smaller, potential peak/sector is also indicated with a question mark (?). If different options had been used this could also have been counted as a sector. Parameters used were $p_d = 10$, $p_w = 9$ and $p_p = 2$.

The code was then run on all of the test images with a given set of input parameters, and compared to the manual measurements. An example of this is shown in Fig. 3.8a and Fig. 3.8b, which show histograms of the difference between the code and the manual measurement. In the initial run ($p_d = 5$, $p_w = 3$ and $p_p = 1$) it can be seen that the distributions are skewed significantly to the right, indicating that the code on average counts more sectors than the person identified in the images. More quantitatively, the code counts on average $\Delta n = 5.0$ more sectors than the best manual measurements. We also found an average $\delta = 4.2$, where δ indicates the difference between the best estimate of the code n_{code} and the lower or upper bound (n_l, n_u) identified by the person (if $n_l \leq n_{code} \leq n_u$ then $\delta = 0$).

Ideally, we would like these histograms to be centred on 0 in both cases, and be as narrow as possible. The process of refining the input parameters to better achieve this involved looking at some of the images where the level of disagreement was largest. The input parameters were then changed by trial and error so that the code showed better agreement with the manual measurements. For instance, starting from Fig. 3.8a and Fig. 3.8b, the main issue is simply that the code very often considers too many parts of the signal to be a sector, when in fact it is just noise, at least in the mind of the person that looked at the image. As a result, parameters were changed so that the

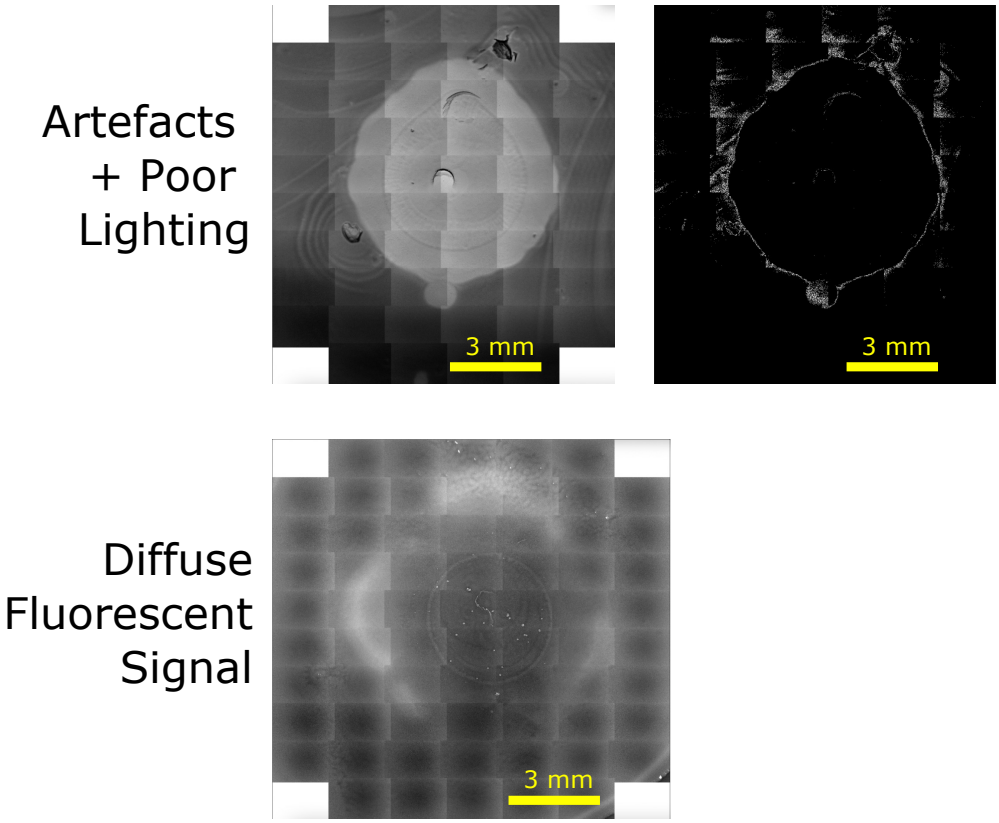


Fig. 3.7 An example of two issues that commonly arise in the sectoring images. The first is imaging artefacts and poor or uneven lighting which results in many incorrect edge points being detected from the brightfield image. The second issue is that the fluorescent signal becomes very diffuse at long times, so sectors are less clearly defined. This second example actually represents the same plaque as the 'good' example (Fig. 3.6), just later in time.

threshold for counting as a sector was higher, i.e. by increasing the minimum height or width of a peak that would be counted. Using these new parameters, the code was then run again on all 100 of the images, and the histograms shown in Fig. 3.8 were plotted. If the histograms were better, i.e. more narrowly centred on 0, then the above process of looking at outlier images and tuning the parameters was repeated. If the histograms were worse than before, then different changes were made to the input parameters to try to address this. For the results presented here this process was done manually, although in principle could be automated.

Through multiple cycles of this refinement process, it was determined that the parameters that were the most useful in altering the shape of the distribution were minimum peak separation p_d, minimum peak width p_w, and minimum peak prominence p_p. Changing just these three parameters eventually resulted in the histograms that can

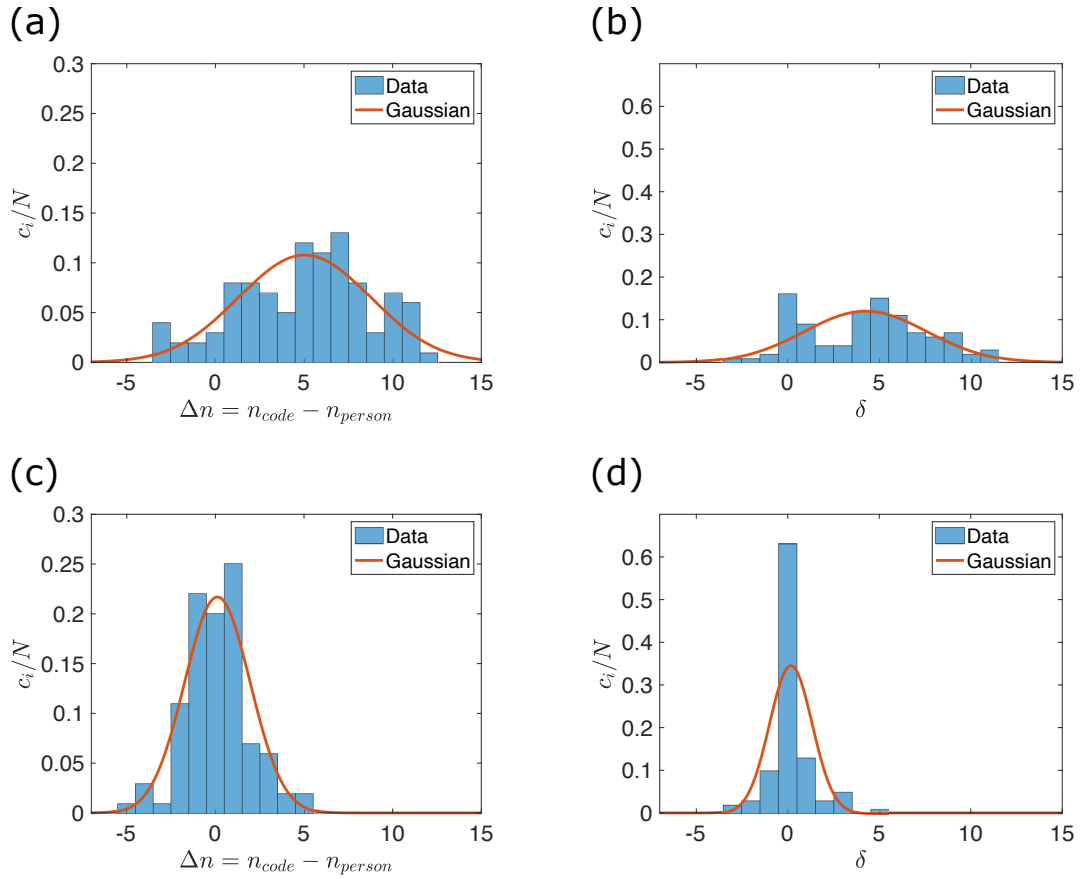


Fig. 3.8 The performance of the code at identifying sectors in comparison to the number identified by hand. For an initial attempt at choosing input parameters, (a) shows a histogram of the difference Δn between the best estimate of the code n_{code} and the best estimate of the person n_{person}, while (b) shows a histogram of the difference δ between the best estimate of the code and the range of values identified using the upper and lower bounds (n_u and n_l respectively). If $n_l \leq n_{code} \leq n_u$ then $\delta = 0$. (c),(d) show the same as (a),(b) but for an improved set of input parameters. The number of images in each bin c_i is normalised by the total number of images $N = 100$. In (a), (b) $p_d = 5$, $p_w = 3$ and $p_p = 1$, and in (c),(d) $p_d = 10$, $p_w = 9$ and $p_p = 2$.

be seen in Fig. 3.8c and Fig. 3.8d. It can immediately be seen that these distributions are much better, with an average $\Delta n = 0.1$ and $\delta = 0.2$. It can also be seen that the distributions are much narrower, in particular the distribution of δ where approximately $2/3$ of the estimates of the code fall within the lower and upper bounds identified by the person in each case. The parameters used here were $p_d = 10$, $p_w = 9$ and $p_p = 2$.

3.5 Discussion

This level of agreement reached may seem underwhelming, as there are still 1/3 of the images which the code produces an estimate that is outside the bounds defined by hand. A useful comparison however, is how 'well' a person does in comparison to other people. For this comparison, I went through all of Diana's and Nikhil's images by hand and gave a best estimate of the number of sectors in the image. I can then compare my best estimate with theirs, and the bounds they set, in the same manner that I have done with the code. The results of this are shown in Fig. 3.9. It can clearly be seen from this that the distribution of differences between my estimates and their estimates are is almost identical to the distribution of differences between the code's estimates and their estimates. This highlights the earlier point that in these images it is often simply quite difficult to judge the number of sectors even by hand, resulting in each person making a different judgement. In this regard then, the code performs just as well as we might expect a person to, i.e. it does not systematically count more or fewer sectors, and the differences are on average about the same level as the person-to-person differences.

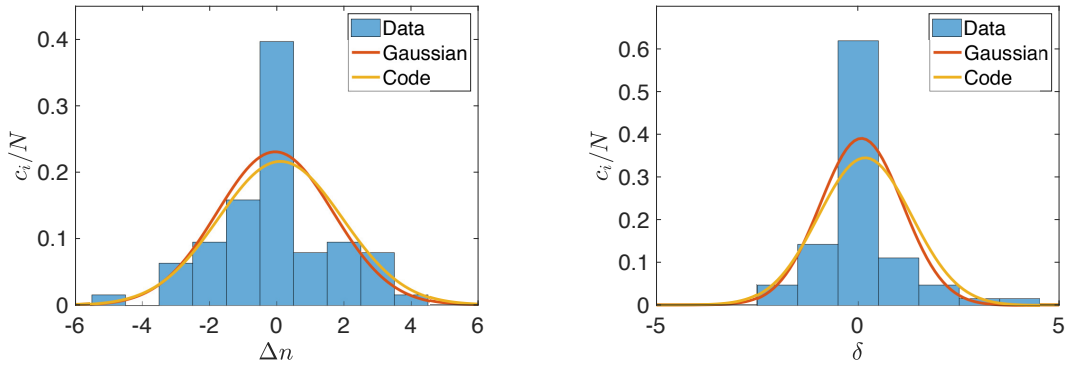


Fig. 3.9 My 'performance' at identifying sectors in comparison to the number identified by Diana and Nikhil. This is shown both in terms of the difference Δn between the my best estimate and the best estimate of the person n_{person}, and in terms of the difference δ between my best and the range of values identified using the upper and lower bounds (n_u and n_l respectively). If $n_l \leq n_{code} \leq n_u$ then $\delta = 0$. This is also shown in comparison to the best estimate of the code on the same images. Parameters used were $p_d = 10$, $p_w = 9$ and $p_p = 2$.

You might reasonably ask whether the code and I perform similarly when presented with Nikhil's and Diana's images simply because *I* designed the code, and so it is more likely to agree with my assessment of the images. This does not appear to be

the case however. Fig. 3.10 shows the performance of the code on only *my* images, in comparison to its performance on all of the images. It can clearly be seen that the code does not agree with me any more than it agrees with Nikhil and Diana. Indeed, the code actually performs slightly worse on my images than it does on Nikhil's and Diana's. This further emphasises the earlier conclusion: the images are difficult to analyse, but the code does as well as a person would.

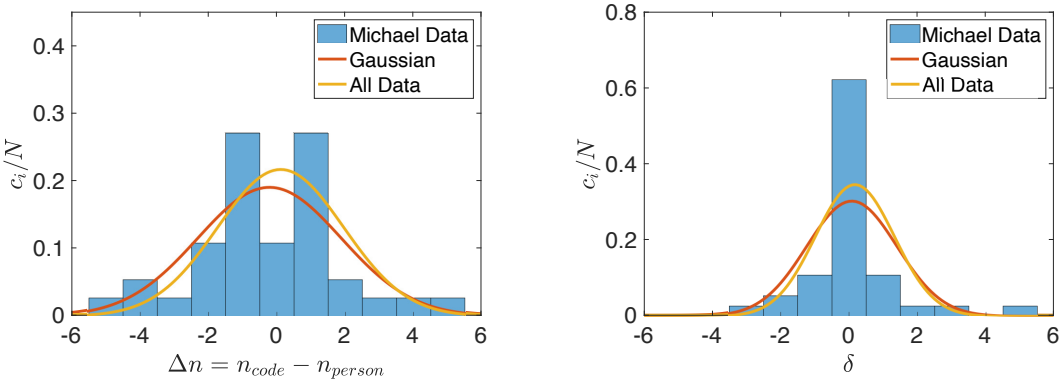


Fig. 3.10 The performance of the code at identifying sectors in the images which I looked at, compared to its performance on all of the images. This is shown both in terms of the difference Δn between the my best estimate and the best estimate of the person n_{person}, and in terms of the difference δ between my best and the range of values identified using the upper and lower bounds (n_u and n_l respectively). If $n_l \leq n_{code} \leq n_u$ then $\delta = 0$.

There obviously remains several ways in which this approach could be improved going forward. For one thing, although it would be very computationally expensive, automating the search for optimum parameters might reasonably be expected to eventually yield results which are better than the current manual search. Perhaps more significantly, modifications could be made to correlate the number and location of sectors in a plaque as a function of time. This would likely improve accuracy, both of the script and of the by hand measurements, as borderline cases (i.e. ones in which you think "that might be a sector") could potentially be ruled in or out more clearly based on whether a sector could be seen in that location in the previous/future frames. Similarly, if both fluorescent channels were reliably available, then the sectors in one channel could be matched with the sectors in the other channel to measure the degree of agreement, and potentially improve counts.

Regardless of these potential improvements, the main point is that we have now developed a tool which can be used to identify sectors during phage range expansions.

This will now allow researchers to address the original questions which motivated the work: how is the genetic drift affected by bacteria density and lysis time, and does this match our expectations from Chapter 2.

'A bad workman always blames his tools.'

English proverb

Chapter 4

Novel Experimental Techniques for Plaque Studies

All of the work presented in this Chapter is my own, except the evolutionary experiment discussed as a motivating example, which was performed by Dr. Diana Fusco and Dr. Wolfram Möbius

4.1 Introduction

As discussed in Chapter 2 and Chapter 3, our results highlight that phage plaques have the potential to be used as a well-controlled model system to experimentally investigate the impact of density-dependent dispersal, and the transition from pulled to pushed waves in a laboratory setting. At present however, this poses several technical challenges.

Firstly, if we are to use theoretical and computational models to understand and interpret experimental observations, it is crucial that we have accurate and appropriate measurements of the phage life-history parameters that characterise the system. In Chapter 2, it was demonstrated that the rate of phage diffusion strongly depends on the host environment and, in particular, can dramatically differ from liquid culture measurements. This raises the more general question of whether other phage life-history parameters also depend strongly on the surrounding host environment. As was noted in Chapter 1, existing methods for the experimental determination of these phenotypic parameters are almost exclusively performed in well mixed environments (see [62, 88] for instance). It is perfectly possible that these parameters could take significantly different values when the infection occurs in different spatially structured environments

and varying metabolic states of the host cells [129]. Moreover, it is possible that these parameters are not only different on solid media when compared to liquid, but may also vary across the expansion in a similar fashion to the diffusion coefficient. For instance, upon lysis, release of cytoplasmic fluids could affect the infection or lysis of neighbouring cells, resulting in life-history parameters that vary with cell density. To be able to accurately compare model predictions to experiments it is therefore critical to develop assays to measure the phage life-history parameters in the environment where the range expansions occur i.e., on agar plates.

In addition to the ability to accurately map experiments onto computational models or theories, there are other practical challenges. For instance, studying the spatial growth of plaques often requires observation over a 24 hour period, or more [88]. This is because, if one wants to accurately sample from both the initial population and the population at the front after a period of expansion, for instance, these two regions need to be significantly separated in space (e.g. at least 3 mm apart). In the case of wild-type T7 under standard conditions, plaques grow at a rate of $\sim$100-200 μm/hr, meaning such a separation is only achieved after $\sim$15-30 hours [58]. Over this period of time, the level of nutrients available to the host bacteria (and consequently the phage) will vary, and the bacteria will almost certainly enter stationary phase, which could also significantly impact the phenotypic parameters [88, 129]. It would be preferable therefore, to be able to maintain a constant supply of nutrients to the host bacteria during the course of the expansion. This would have the additional benefit of increasing the length of experiments that could be performed, as the limiting factor would simply become the physical size of the plate, rather than the time until nutrients are exhausted.

In this Chapter, I will present various experimental techniques intended to address each of the points discussed. This includes methods to measure the phage adsorption rate (Sec. 4.4) and lysis time (Sec. 4.5) on agar plates, and a technique for constantly supplying nutrients to an agar plate (Sec. 4.6).

4.2 Existing Phenotypic Assays

4.2.1 Measuring Adsorption Rate in Liquid

Many previous studies have shown that in the first minutes post exposure to phage, reaction kinetics between free phage (V), bacteria (B) and bacteria-phage complexes,

i.e. infected bacteria (I), can be described by a simple reaction model [165]

$$V + B \xrightarrow[\alpha]{\text{rate}} I, \tag{4.1}$$

with adsorption rate α. This leads to three different ways of measuring phage adsorption rate: as a function of time, one can measure the number of (i) free phage, (ii) adsorbed phage or (iii) uninfected bacteria [74]. The first of these approaches is the most commonly used, and it is the one we shall focus on here.

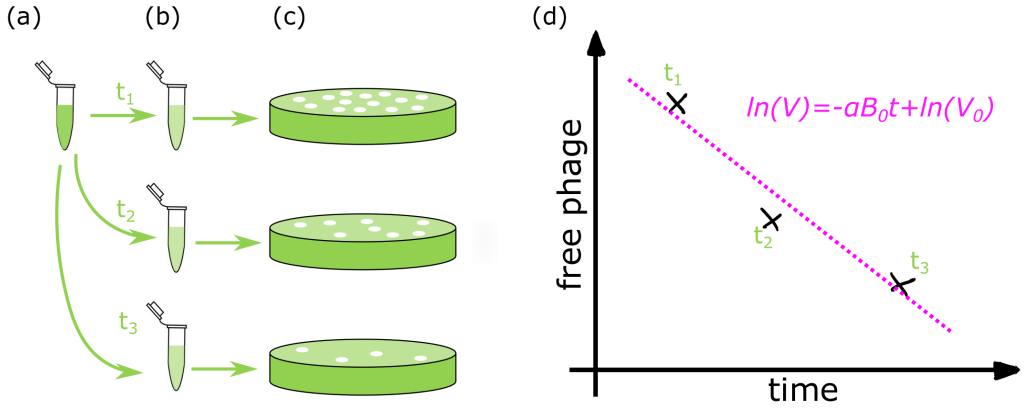


Fig. 4.1 A cartoon to illustrate how adsorption rate is measured in liquid. (a) Phage are added to a bacterial culture in log-phase. (b) Samples are taken periodically, and diluted into an LB/chloroform mixture. (c) Plating assays are performed on these samples. (d) Free phage are plotted as a function of time, and the adsorption rate is extracted.

The procedure is as follows: phage are added to a bacterial culture in log-phase, and kept at 37 °C. At predetermined time-points, samples of this culture are then removed, and diluted into an LB/chloroform mixture. The chloroform kills both infected and uninfected host cells, but has no effect on any remaining free phage. Plating assays (Sec. 2.8.6) are then performed on these samples to determine the number of free phage as a function of time. It must be noted that the samples must be taken before the end of the eclipse period, otherwise fully constructed phage (i.e. those with a complete capsid and tail, and the viral DNA packaged into the capsid) inside infected hosts may be counted in the plating assays, skewing the results.

Assuming that the concentration of available host remains relatively constant over the several minute long period adsorption is usually measured (up to ∼5 mins), this is

modelled as a first order reaction resulting in an exponential decay in free phage

$$\frac{dV(t)}{dt} = -\alpha B_0 V(t), \tag{4.2}$$

$$\ln(V(t)) = -\alpha B_0 t + \ln(V(0)). \tag{4.3}$$

A cartoon of this process is shown in Fig. 4.1.

4.2.2 Measuring Lysis Time and Burst Size in Liquid

The so called *single-burst curve*, or *lysis curve*, is used for experimentally measuring the average lysis time and burst size of a phage population. The technique used here is adapted from that used originally by Ellis and Delbrück [166]. The procedure is as follows: initially, 10 μl of susceptible host overnight culture is diluted in 990 μl of LB and incubated at 37 °C for approximately 2 hours to generate an exponentially growing culture. At the start of the experiment (t=0), 10 μl of phage (approximately 10^7-10^8 phage) is added to this culture and incubated for one minute to allow infection. The culture is then diluted 10^5 and 10^6 in LB to obtain a countable number of phage, and kept incubated at 37 °C. Every 2-3 mins for 30-40 mins (the length depends on the phage's typical lysis time), 20 μl of each dilution is plated with 100 μl of host overnight and 3-5 ml of molten 0.7% agar. As in the plating assay (Sec. 2.8.6), these plates are left on a bench-top overnight, and plaques counted the following day. This should result in a single step growth curve for the phage, from which the lysis time and burst size can be measured. Cartoon data that could result from such measurement with the parameters highlighted are shown in Fig. 4.2.

For fitting lysis curves, multiple different functions have been considered, including piece-wise [167], logistic-like [62, 66] and Gaussian functions [168]. Previous analysis has concluded that the use of logistic-like and Gaussian functions yield very similar results, and that both are preferable to the use of piece-wise functions [169]. Here, we use a Gaussian function, and so the expected number of PFUs per initial phage n_p at time t will be given by [169]

$$n_p = \frac{\beta}{2}\left[1 + \mathrm{erf}\left(\frac{t-\tau}{\sigma_\tau\sqrt{2}}\right)\right] + 1, \tag{4.4}$$

where β is the burst size (i.e. the yield of new phage produced at a lysis event), τ is the lysis time and σ_t is the standard deviation of the lysis times. It's worth noting here that while σ_t could be interpreted as a measure of the stochasticity in lysis time, it is not

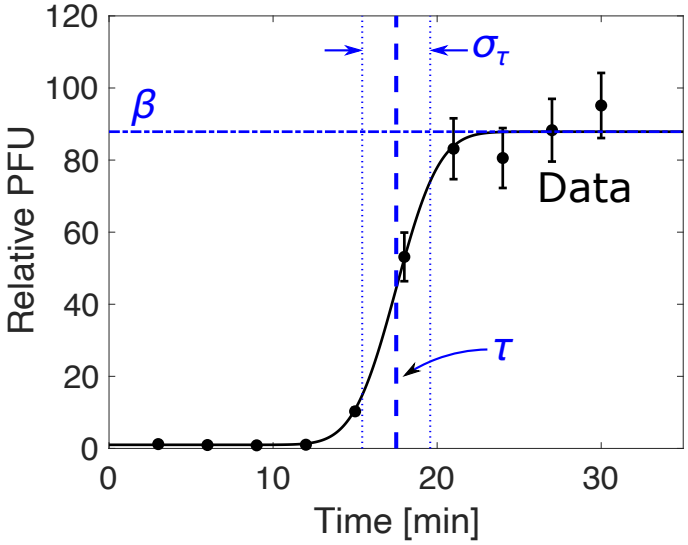


Fig. 4.2 Cartoon data to illustrate the type of results obtained during a lysis curve. The burst size β, lysis time τ and the standard deviation of lysis times σ_τ are shown.

clear exactly what its mechanistic origin is, i.e. this variability could represent variation amongst the infecting phages *or* variation amongst the cells they infect.[1] The number of initial phage is usually taken to be the average of the first three measurements, although this can be adjusted where appropriate, as long as the points are taken before it appears that any lysis has occurred. The final +1 in Eq. 4.4 is to adjust for the fact that, by the nature of calculating the ratio of PFUs at time t to those at the outset, the distribution should be equal to 1 at time $t = 0$. $\mathrm{erf}(x)$ is the error function given by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy. \tag{4.5}$$

4.3 Motivating Example

The importance of being able to answer the questions laid out in Sec. 4.1 is perhaps best illustrated with a more concrete example. As stated at the beginning of this Chapter, the work presented in this section is not my own, and represents unpublished results of Dr. Diana Fusco and Dr. Wolfram Möbius. Briefly, in this work bacteriophage T7 was inoculated onto a lawn of bacteria on an agar plate and allowed to form a plaque and grow for 17 hours (Fig. 4.3). At this point, phage from the very edge of the plaque were

[1] This idea of biological stochasticity will be discussed further in Chapter 6.

picked with a needle or a razor blade, and used to inoculate a fresh lawn of bacteria. After another 17 hours of plaque growth, phage were again picked from the edge of the plaque and inoculated onto a fresh lawn, with this process being repeated 14 times.

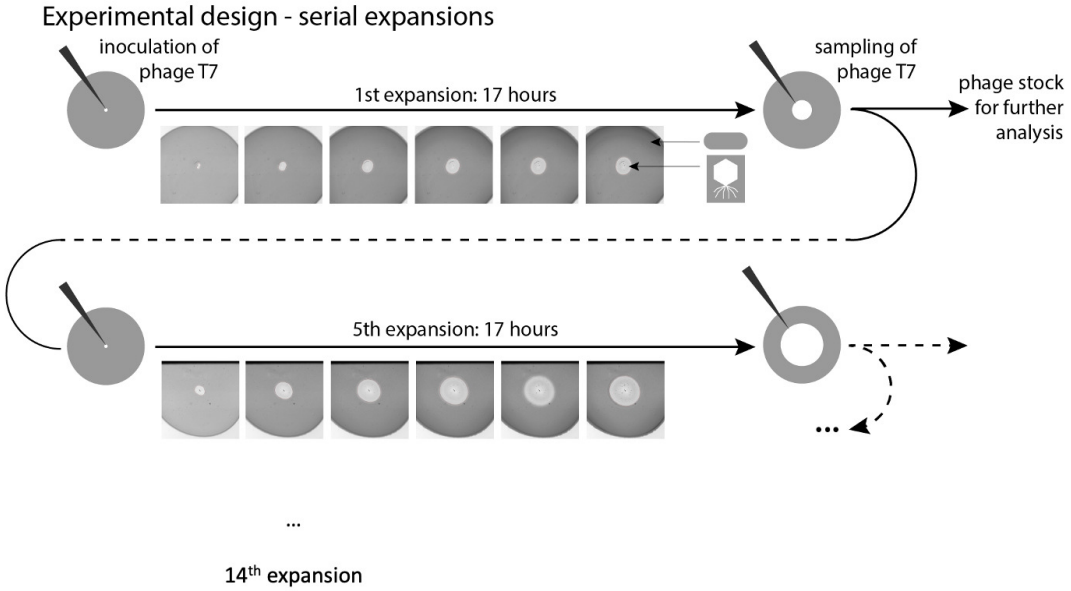


Fig. 4.3 The setup of a bacteriophage evolutionary experiment performed in bacterial lawns of agar plates. Phage are inoculated onto a lawn of bacteria, allowed to grow for 17 hours, at which point phage are picked from the edge of the plaque and used inoculated onto a new plate. This process was repeated 14 times. Figure provided by Dr. Diana Fusco.

While there are various findings from this evolutionary experiment which I will not detail here, the key finding for the purposes of this Chapter is that the phage consistently evolved to form faster and faster growing plaques (Fig. 4.4). This increase in speed was observed across replicates, and the speed after 14 cycles could be as much as 4× larger than the original inoculate (i.e. the *ancestor*).

If we attempt to explain this in terms of the reaction-diffusion model presented in Chapter 2, such a speed increase would require very significant changes to the phage life-history parameters, such as a 10-fold reduction in adsorption rate α *and* a 2-fold reduction in lysis time τ (Fig. 4.5). Changes to burst size β are unlikely to be able to explain such an increase in speed given that, as shown in Fig. 2.15, changes to burst size β result in little change to speed for $\beta > 50$. When existing techniques (i.e. methods that rely on bulk liquid culture measurements, described in Sec. 4.2) are used to measure these parameters in both the ancestor and evolved strains however, they show no consistent or significant difference (Fig. 4.5).

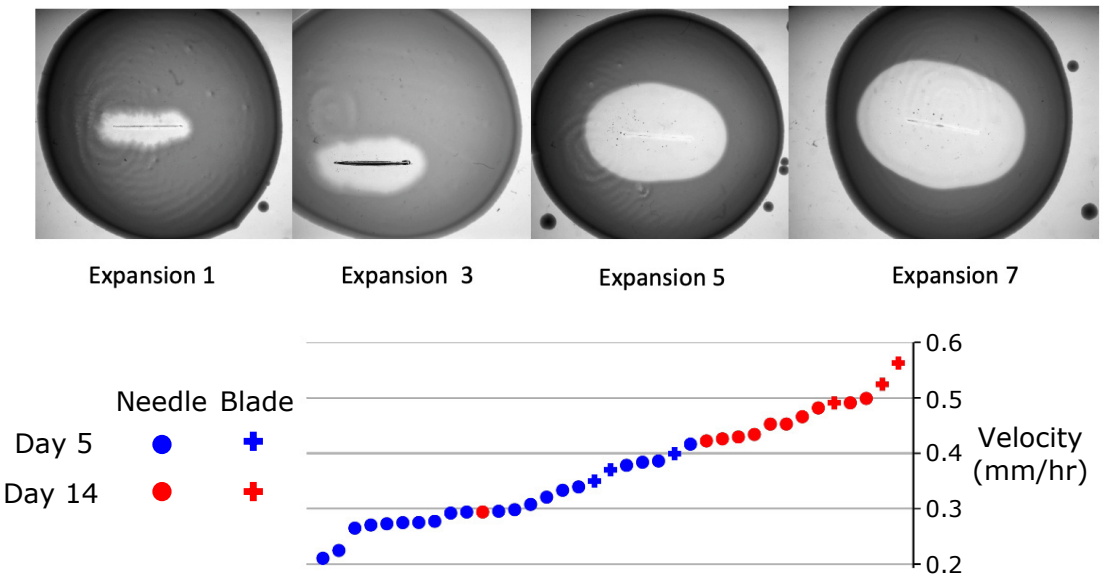


Fig. 4.4 Each of the top images are taken the same time after inoculation of the phage. The phage consistently evolved to form faster and faster plaques, as can be seen both in the top images, and in the data provided across replicates in the bottom plot. Figure provided by Dr. Diana Fusco.

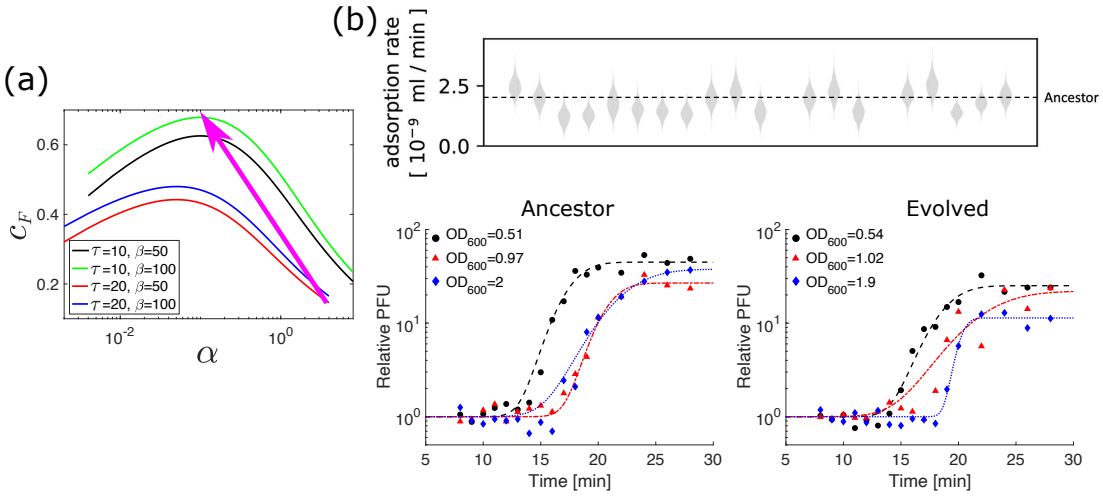


Fig. 4.5 (a) The reaction-diffusion model presented in Chapter 2 indicates that a 4-fold increase in lysis time ($\sim c_F$ in the model) would require large changes to the phage life-history parameters. (b) Phenotypic assays in liquid culture reveal that there is little difference in the adsorption rate α (top), or the lysis time τ (bottom) of the ancestor and evolved strains of the evolutionary experiment. Lysis time data collected by Lucy Witherall of the Möbius Lab. The lysis curves were performed in partially 'spent' media, i.e. media which had previously been used to grow cells to a density specified by the OD_{600} reading. Adsorption rate figure provided by Dr. Diana Fusco.

Each of the issues highlighted in the introduction could potentially play a role in addressing this discrepancy. While the phenotypic assays do not reveal a significant difference between ancestor and evolved strains, these assay were performed in a well-mixed liquid culture, where the bacteria are *typically* in exponential phase with ample access to nutrients.[2] By contrast, the evolutionary experiments themselves were performed on agar plates, where the metabolic state of the host would change during the course of the expansion, and so could significantly impact the phage life-history parameters.

4.3.1 Diffusion of Evolved Phage

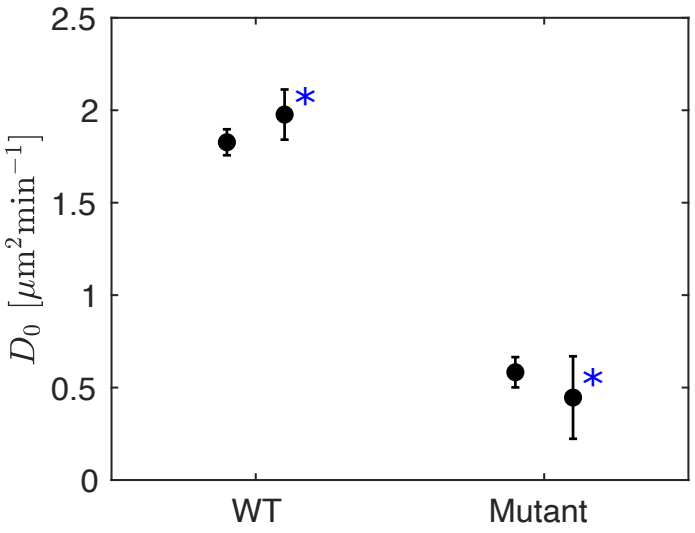


Fig. 4.6 Free diffusion rates measured for standard wild-type (WT) T7 compared to those measured for one of the evolved mutants discussed in Sec. 4.3. The measurement procedure is described in Chapter 2, including correction due to geometry and number of phage (Sec. 2.2.1). The blue asterisk (*) indicates that these data points were collected from the same plate, i.e. half of the phage droplets were WT and half were mutant.

While it may seem unlikely, in principle one other characteristic of the phage that could have changed is the diffusion rate D, which I presented a technique for experimentally measuring in Chapter 2. I therefore present here results of the measured

[2]The lysis curves in Fig. 4.5 were also performed in liquid cultures where the nutrients were partially depleted. This was achieved by growing a culture of bacteria to a pre-determined OD_{600} (density), spinning down the culture and removing the cells to leave partially 'spent' media.

free diffusion rate D_0 (i.e. without any background bacteria) for both the wild-type (WT) T7 and one of the evolved mutants - note that the data has been corrected due to geometry and number of phage as shown in Sec. 2.2.1. It can be seen in Fig. 4.6 that surprisingly the measured diffusion rate is quite different for the mutant. Even more surprisingly, it is lower than the measured diffusion rate for the WT phage. This is exactly the opposite of the change we would expect to see if the increased plaque speed was explained by a change in diffusion rate - lower diffusion rates should result in a slower plaque speed. This effect is seen across replicates, including when both types of phage are plated separately on the same plate, so as to rule out any issues of having a particularly dry plate. Needless to say, this is a very confusing result, and will require further investigation.

4.4 Adsorption Rate on Solid Media

Here, we adapt the procedure for measuring phage adsorption rate in liquid culture (Sec. 4.2.1) to allow for the measurement of adsorption rate on solid media. At this point I would like to thank Chris Dickinson, who helped develop this protocol during his undergraduate work. The underlying idea of both procedures remains the same: phage are allowed to infect susceptible host, and after a given time t, chloroform is added to the system, killing the host bacteria and any phage already adsorbed to them, while any free phage are left unharmed, and can be subsequently plated with susceptible host to determine their number. The difference comes in the first steps of this process, where the liquid culture is now replaced by a lawn of bacteria grown overnight on the surface of an agar plate. Phage are added to the surface of this lawn, and after a time t the lawn is then removed and re-suspended in an LB/chloroform mixture (this is identical to Fig. 4.1 with the exception of part (a)).

4.4.1 Methods

Sample Preparation

6 well plates, with each well containing 5 ml of 20 g/l agar (VWR Chemicals) and 25 g/l LB (Invitrogen) - NaCl concentration 10 g/l - were prepared, and left to dry for several hours. Once dry, 30 μl droplets of overnight cultures of either *E. coli* BW25113 or eWM44 (grown from single colonies at 37 °C in 25 g/l LB) were pipetted into each well, and left to dry. For full details of the strains see Appendix A, but as a

brief reminder: BW25113 is the phage-susceptible wild-type, whereas eWM44 prevents phage adsorption and is used as a control. These plates were then incubated overnight at 37 °C to allow the bacteria to grow into a thick layer.

The following day 2 μl droplets of stock T7 phage (approx 10^7 pfu/μl) were pipetted onto the surface of the bacteria lawns. After a time period t from the phage being added, the bacteria layer was removed from the surface of the agar using a sterile inoculation loop, and deposited into an Eppendorf tube containing 970 μl of LB, and 30 μl of chloroform. This mixture was then briefly vortexed, stored in ice, and then serially diluted in LB, and 20 μl of each serial dilution was plated with 100 μl of overnight BW25113 and 5 ml of 7 g/l agar. These plates were then left overnight to allow plaques to form.

Determining Density of Bacteria Lawns

To determine the rate of phage adsorption, we must have an estimate for the density of bacteria in the lawns used in the experiment (i.e. after they have grown to a thick layer overnight). As the lawns used in the experiment had to be destroyed during the process of taking measurements, separate lawns were grown under the same conditions (Sec. 4.4.1), and these were used for the determination of the bacterial density.

To determine the density, two things had to be measured: the volume of the lawns and the number of bacteria they contained. The volume of the lawn was measured first by imaging the lawns to determine the cross-sectional area, and then by separately imaging the 'base' and the 'top' of the lawn, to determine their respective z locations, and consequently the height of the lawn. The base was imaged by focusing on the very edge of the lawn where the bacteria meet the surrounding agar. To image the top of the lawn, a 1 μl droplet of 1 μm fluorescent beads (3.6×10^{10} beads per ml) was added to the lawn surface, so as to generate a feature that could be used to find the correct focus height. Images were taken using a Zeiss Axio Zoom.V16 stereo microscope equipped with either a Zeiss PlanApo Z 0.5x/0.125 FWD 114 mm objective or a Zeiss PlanNeoFluar Z 2.3x/0.57 FWD 10.6 mm objective, and analysed using Fiji (v1.52h), an open source distribution of ImageJ focused on scientific image analysis [156, 157]. The volume determined here is an approximation, as it assumes that the lawn has a constant height.

To measure the number of bacteria cells in the lawns, the lawns were re-suspended in 1 ml of LB, and then serially diluted in LB. Then, 100 μl droplets of the 10^7, 10^8 and 10^9 dilutions were pipetted onto the surface of a 20 g/l agar plate and left to dry

(an approximate estimate of the density made by measuring the OD_{600} of the culture indicated that these dilutions would contain $\sim 10^2$, 10^1 and 10^0 cells respectively). These plates were then incubated overnight at 37 °C, and the following day colonies were counted to determine the number of cells present in the suspension (see Fig. 4.7).

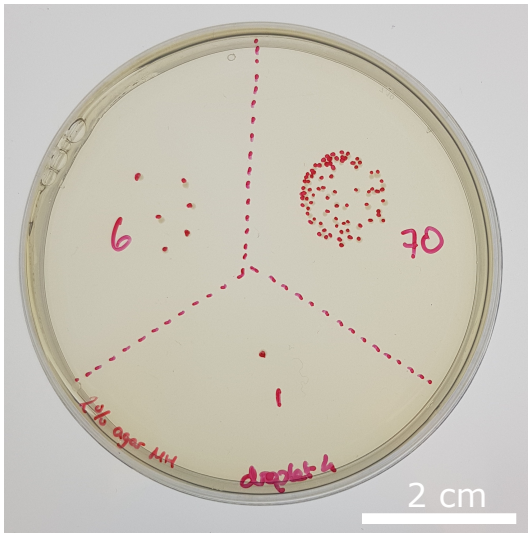
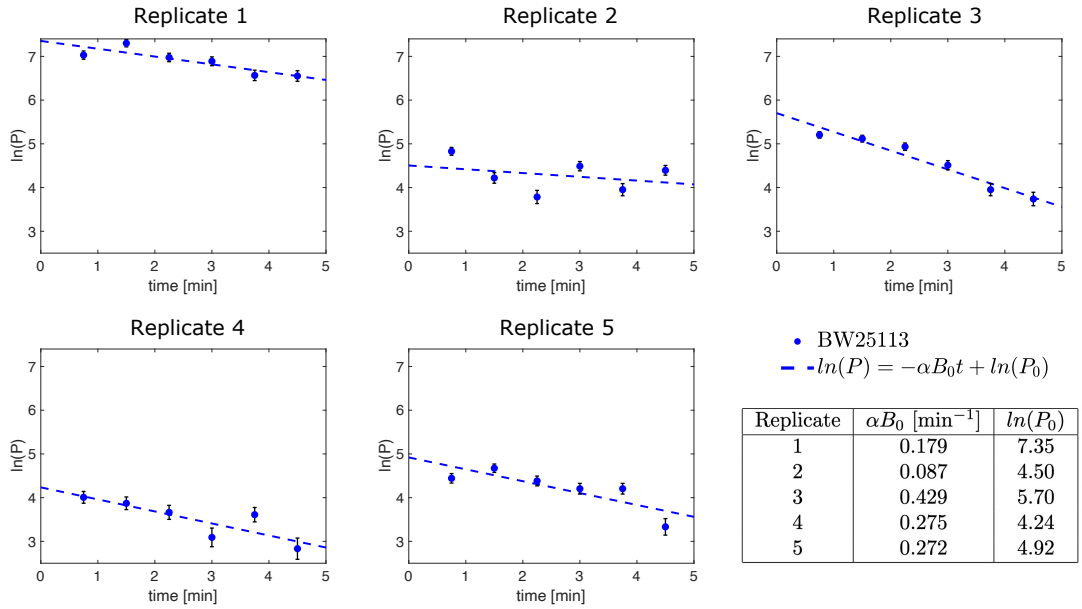


Fig. 4.7 Example of the colony spotting assay used to determine the bacteria concentration.

4.4.2 Results

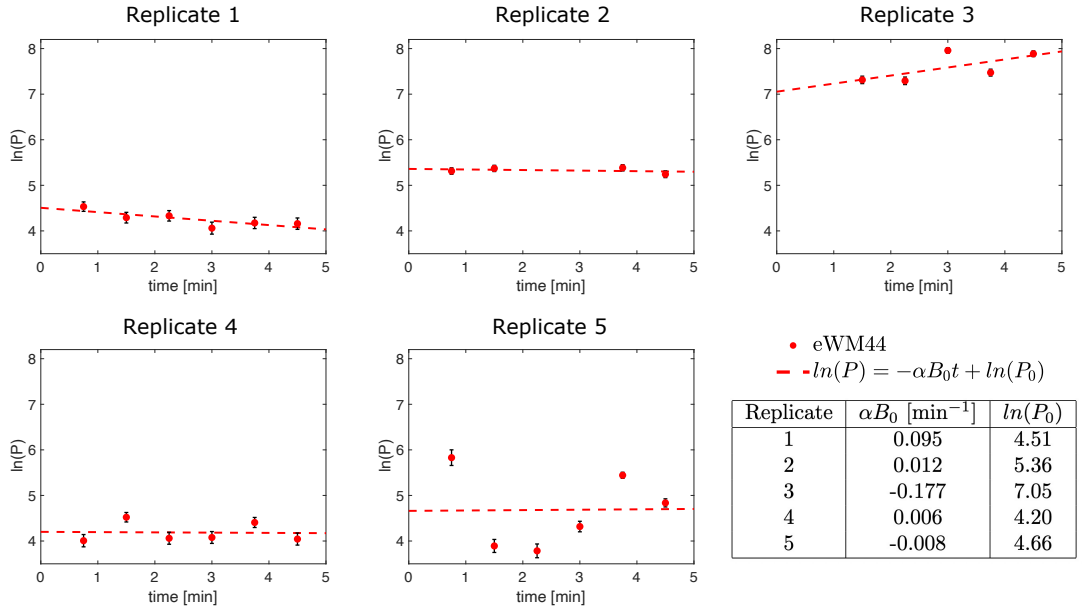
As expected, it can be seen in Fig. 4.8 that the number of free phage recovered from the surface of the adsorbing bacteria lawns (BW25113) consistently decreases as a function of time across replicates. This does not occur in the lawns of control bacteria (eWM44) where adsorption should be prohibited (Fig. 4.9), indicating that the fall in free phage recovered is primarily due to adsorption to bacteria.

It is worth taking a moment here to comment on the different sources of variability/error in these measurements. Working backwards, there is of course the Poisson sampling error on each of the individual data points that occurs when a small sample is plated from the liquid suspension of phage. This error is shown as the y error bar in Fig. 4.8 and Fig. 4.9. Within a single experiment (i.e. within a single plot in either Fig. 4.8 or Fig. 4.9) there is also variability in the number of phage which are successfully 'scooped-up' from the plate, due to this being an imperfect process. The scale of this error can be somewhat appreciated by observing the spread of points *within* a single plot in Fig. 4.9, although this spread will also be affected by the Poisson sampling



Replicate	αB_0 [min^{-1}]	$ln(P_0)$
1	0.179	7.35
2	0.087	4.50
3	0.429	5.70
4	0.275	4.24
5	0.272	4.92

Fig. 4.8 Number of free phage recovered as a function of time from the surface of adsorbing bacteria (BW25113) lawns. Error bars come from the Poisson sampling error, i.e. $\Delta P = \sqrt{P}$.



Replicate	αB_0 [min^{-1}]	$ln(P_0)$
1	0.095	4.51
2	0.012	5.36
3	-0.177	7.05
4	0.006	4.20
5	-0.008	4.66

Fig. 4.9 Number of free phage recovered as a function of time from the surface of non-adsorbing bacteria (eWM44) lawns. Error bars come from the Poisson sampling error, i.e. $\Delta P = \sqrt{P}$.

error. Finally, there is also variability between different replicates, i.e. between the multiple plots in either Fig. 4.8 or Fig. 4.9. This can be appreciated by noting that

the y-intercept of each of the plots is different, even though the same number (within sampling error) of phage were added in all cases. I believe that this error originates in differences between the bacteria used in different experiments. In each case, the strain of bacteria (BW25113 or eWM44) is of course the same, but cultures used to grow the lawns were established from different colonies in each case, which may each contain different minor mutations. Based on my anecdotal experience, this manifests in lawns which are slightly thicker or thinner, and slightly more or less rigid on different days. In my experience, more phage are recovered from lawns which are which are thicker or more rigid.

A summary of the 5 replicates is shown in Fig. 4.10. It can be seen that in the non-adsorbing control, the fitted gradient is on average $\alpha B_0 = -0.02 \pm 0.04$ min^{-1}, which is consistent with no adsorption. In the case where we expect adsorption to occur however, the average gradient is $\alpha B_0 = 0.25 \pm 0.06$ min^{-1}. To convert this into an adsorption rate alone we then require the density of the bacteria lawns. This was measured as described previously, and the average of 6 measurements was taken, yielding a density of $B_0 = 0.92 \pm 0.08$ μm^{-3} $= (0.92 \pm 0.08) \times 10^{12}$ ml^{-1} for BW25113 lawns. In the case of adsorbing bacteria, this leads to an adsorption rate of $\alpha = 0.27 \pm 0.07$ μm^3/min $= (0.27 \pm 0.07) \times 10^{-12}$ ml/min.

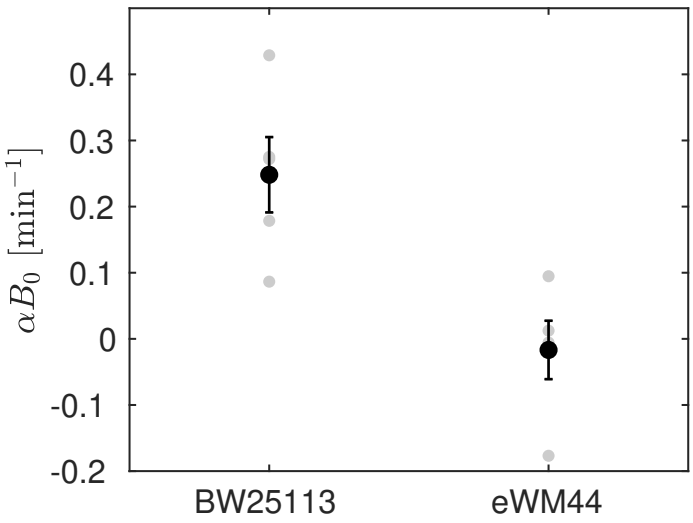


Fig. 4.10 Average adsorption rates determined from 5 replicates.

It is interesting to note here that the product $\alpha B_0 = 0.25 \pm 0.06$ min^{-1} that we determine on solid media is comparable to those typically found in liquid cultures ($\alpha B_0 \approx 0.1$ min^{-1} in Ref. [62]), but with significantly different values for α and B_0

separately. Ref. [62] reports $\alpha \approx 10^{-9}$ ml/min and $B_0 \approx 10^8$ ml^{-1}, meaning that our adsorption rate is several orders of magnitude lower than in liquid, with the density of bacteria being the corresponding orders of magnitude more dense. This is in good qualitative agreement with Eriksen *et al.* [170], who showed using computational models that adsorption rates were substantially lower in microcolonies than in well-mixed liquid cultures. Their explanation is that the spatial structure in microcolonies significantly reduces the hosts' potential for exposure to phage, both because the bacteria at the periphery of the microcolony shield the bacteria in the centre, and because the close packing of cells means that not all of the surface is easily accessible to the phage for adsorption.

4.5 Lysis Time on Solid Media

Much like adsorption rate, lysis time is traditionally measured in liquid media, since it requires periodic and precise sampling of the phage population (Sec. 4.2.2). This approach is not easily adapted to measuring lysis time on agar plates because, as we have shown in the previous section, adsorption is significantly slower on plates than in liquid, meaning that it is difficult to 'synchronise' infections at the start of the protocol. In this section, we therefore take a different approach.

The reaction-diffusion models of plaque growth presented in Chapter 2 produce a front which expands by the width of the infected region $\Delta\overline{x}_I$ during one lysis time interval. As a result, the dimensionless infection region width $\Delta\overline{x}_I$ is equivalent to the dimensionless speed $\overline{c}$, as predicted in all the models and confirmed by the numerics (Fig. 4.11). In physical units, this means that the plaque front travels the width of the infected region every lysis time.

By utilising phage engineered to result in fluorescent infected cells (see [163] and Chapter 3), we image a growing plaque of such phage over time in both fluorescent and brightfield channels to determine the distributions of infected and uninfected bacteria, and estimate the width of infected cells Δx_I. By simultaneously measuring the speed of the expansion c, the lysis time in solid media and its variation during the course of an expansion could be estimated from $\tau = \Delta x_I / c$.

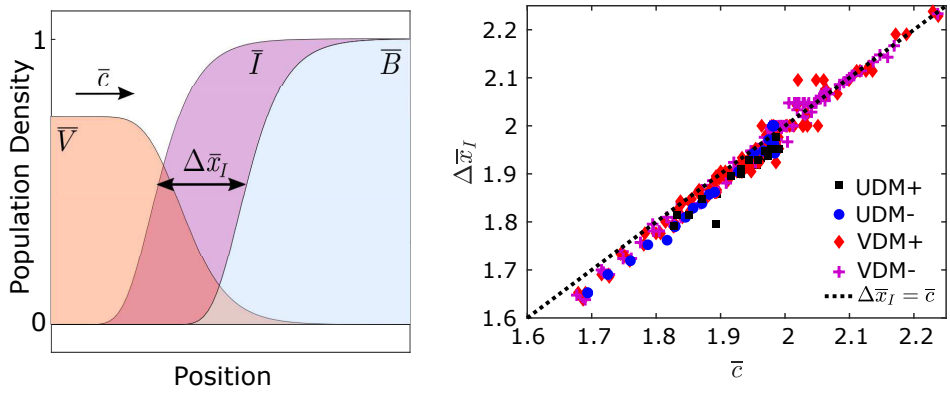


Fig. 4.11 Dimensionless width $\Delta\overline{x}_I$ shown to be equal to $\overline{c}$ for all models across the range of f and K_{max} investigated in Fig. 2.7, corresponding to an expansion that travels the width of the infected region every lysis time. We attribute the small discrepancy observed at lower values to the limited convergence of the front profile to its steady-state because of trade-offs between precision and computational cost. Figure made from material included in Chapter 2 and Ref. [171].

4.5.1 Methods

Sample Preparation

90 mm petri dishes containing 25 ml of 20 g/l agar (VWR Chemicals), with LB (Invitrogen) - NaCl concentration 10 g/l - were prepared and left to dry for several hours. Plates were then refrigerated if they were to be used at a later date.

Overnight liquid cultures of *E. coli* BW25113 were grown from single colonies at 37 °C in LB. Then, 20 μl droplets of bacteria were pipetted onto the surface of the agar and left to dry. These plates were then incubated for approximately 1 hour. 2 μl droplets of $T7_{\text{mEYFP}}$ were then pipetted onto the surface of the bacteria.

Data Acquisition

The plates were imaged using a Zeiss Axio Zoom.V16 stereo microscope equipped with a Zeiss PlanNeoFluar Z 2.3x/0.57 FWD 10.6 mm objective. Images of the sample were taken every 10 minutes for a period of 8 hours. During the imaging period, the sample was kept with its lid on at 37 °C using an ibidi Multi-Well Plate Heating System.

Image Analysis

An example of the type of images collected during the expansion of $T7_{mEYFP}$ are shown in Fig. 4.12. The bright area in the brightfield image shows the plaque in the bacterial lawn. In the fluorescence image, a bright band can be seen at the edge of the plaque, which corresponds to the region of infected cells. All of the images were analysed using Fiji (v1.52h), an open source distribution of ImageJ focused on scientific image analysis [156, 157].

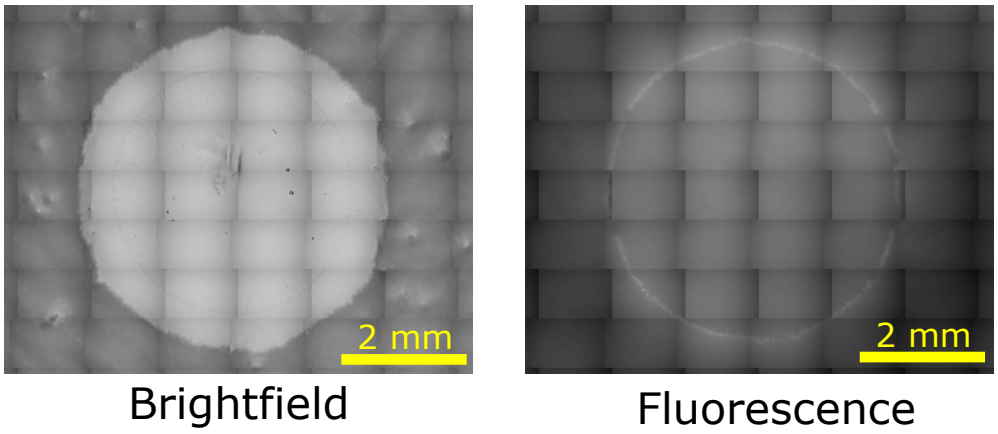


Fig. 4.12 A pair of example images. The bright area in the brightfield image shows the plaque in the bacterial lawn. In the fluorescence image, a bright band can be seen at the edge of the plaque. This corresponds to the region of infected cells.

The first step in the analysis is to identify the centre of the plaque. This is done by automatically thresholding the brightfield image, such that ideally the pixels corresponding to the plaque are white and everything else is completely black. The edge of the plaque is then identified using the wand tool, which traces around the boundary between white and black pixels. The centre of mass of this boundary is then measured, and taken to be an approximation of the centre of the plaque. This process is essentially the same as the initial steps of the protocol to identify phage sectors discussed in Chapter 3 - see Fig. 3.4 for illustration.

Next, the shortest distance r between the plaque centre and the edge of the image is found. A line is then drawn a distance r outward from the plaque centre, at an angle θ defined clockwise from the positive x direction.[3] The intensity profile in the fluorescent image is then found along this line, both as a function of position along the line and as a function of frame in the timelapse. This process is then repeated at 5° intervals around the plaque, producing a total of 72 profiles.

[3]How θ is defined here does not actually matter, so long as it is done so consistently.

Data Analysis

The profiles produced during the image analysis are then analysed using MATLAB. It can be seen in Fig. 4.13b that the profiles produced contain many unwanted features, such as sharp jumps in intensity at the boundaries between tiles, as well as generally larger intensities inside the plaque compared to outside the plaque.

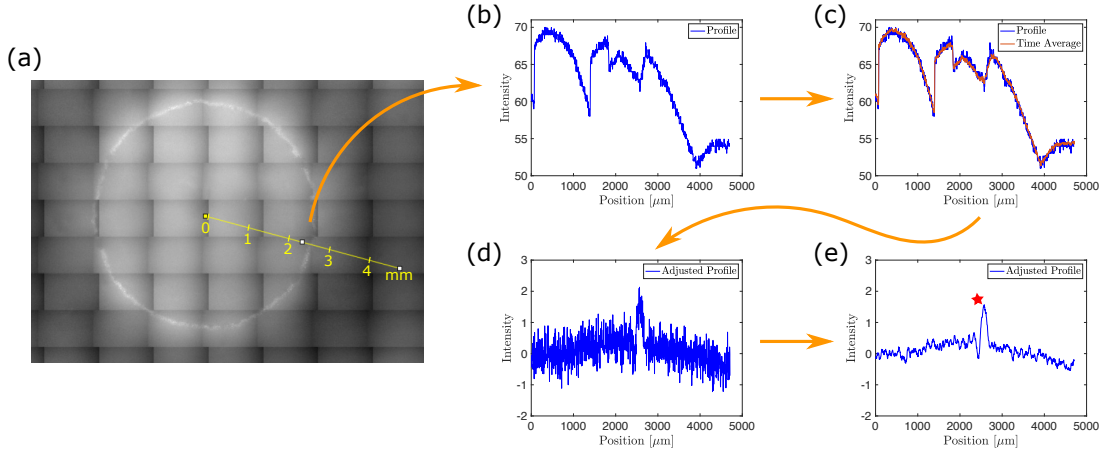


Fig. 4.13 An illustration of the way the data is analysed for an example image. (b): Initially, the fluorescent intensity profile along a line extending radially outward from the centre of the plaque is found. (c): The time average of this profile is then determined using the current frame, and three frames in the past/future. (d): This time averaged profile is then subtracted from the profile in the current frame to remove unwanted features such as the edges between tiles in the image. (e): A moving average is taken of the adjusted profile to reduce the level of noise in the signal, with the averaging window corresponding to 1% of the length of the profile. A clear peak corresponding to the region of infected cells is highlighted with a red star.

To remove such features, the time averaged background of the profile is calculated using the current frame, and three frames in the past/future. The choice of three was made by trial and error, as either too many or too few frames results in the peak simply being removed. This time average background is shown in Fig. 4.13c in comparison to the profile in a single frame. This background is then subtracted from the intensity profile resulting in an intensity profile which is approximately 0 in regions where the image does not change significantly from frame to frame (Fig. 4.13d). Finally, given this profile is itself very noisy, a moving average of the profile is taken, with the averaging window corresponding to a size of 1% of all of the data ($\pm$ $\sim$20 mm). After this final step is complete, a clear peak in intensity can be seen in the profile (Fig. 4.13e), corresponding to the peak in intensity observed in the infected region.

This peak then has to be identified, and its properties (i.e. position and width) need to be measured. This is done in two separate ways, although as we will see later both produce similar results.

The first involves using the MATLAB signal processing function *findpeaks* which finds local maxima in the input signal vector (i.e. fluorescent intensity as a function of position in our case). What is considered a "peak" is controlled by a set of optional inputs including maximum number of peaks, minimum peak height, minimum height difference between neighbouring peaks, minimum peak separation, minimum peak width, maximum peak width and minimum peak prominence, amongst others. These properties were discussed in more detail in Chapter 3, but for our purposes at the moment, the function is simply run without any of the optional inputs, meaning that it identifies *all* of the peaks in the signal. The peak corresponding to the infected region is then assumed to be the most prominent peak (the peak which stands out relative to its surrounding landscape the most - see Fig. 3.5) of those identified, and so the position and width of that peak are recorded, with the other peaks being discarded.

The second method essentially consists of fitting a Gaussian curve to the peak in the profile. I find this method to be less preferable because it requires more 'user input' than *findpeaks*, but I include it here because it provides a useful point of comparison.[4] The type of user input that I mean is that for MATLAB to fit a Gaussian curve to the *correct* peak in the data, one has to provide a reasonable starting point for where the peak can be found, otherwise the script will fit a curve somewhere in the noise of the signal. To find this starting point, one has to look at the initial frame and identify roughly where the peak is. To avoid doing this in each frame the starting point for the code moves along by the average speed of the front. This is an imperfect process, and sometimes the code still fits a curve to the wrong part of the signal. With a bit of trial and error however, it does produce similar results to the first method.

In both cases, the width of the peaks are measured along with their position as a function of time. From the position as a function of time the plaque speed can also be determined by simple linear regression.

[4]The reality is also that while I have presented it here as the second method, it's actually the way I originally thought to do it, before I discovered the *findpeaks* package.

4.5.2 Results

Lysis Time on Plate

The lysis time τ as a function of time during the course of the expansion are shown in Fig. 4.14 and Fig. 4.15. These figures show the radial average around the edge of the plaque. Fig. 4.14 shows the radial average of 56% of the profiles which were observed to show a clearly visible[5] fluorescent band in the original images, while Fig. 4.15 shows the radial average of all of the directions.

It can be seen that both the Gaussian method and the *findpeaks* method produce qualitatively very similar results. I believe the slight offset arises for the following reason. The *findpeaks* function determines the width at a reference line, which is by default positioned a distance below the peak equal to half of the peak's prominence. For comparison to the Gaussian approach, this height is then taken to be equal to the full width at half maximum (FWHM), and converted into a standard deviation σ that one would determine in an equivalent Gaussian curve. Fitting a Gaussian curve directly however usually results in the inclusion of some of the surrounding 'noise' of the peak, as the peak itself is never a perfect Gaussian curve. This results in a peak which is slightly wider than that determined from a direct measurement of the FWHM.

It is also worth reiterating my earlier preference for the *findpeaks* results. While both methods produce similar results, particularly on the 'good' data (the 56% of profiles with a clearly visible band, shown in Fig. 4.14), the *findpeaks* approach performs much better on the 'bad' data (the remaining 44%), despite requiring no user input. Even though both average out across many profiles to be similar, it's clear from the single profiles that the results are much more sensible using *findpeaks*. The Gaussian method will often fit to the 'wrong' peak, which in some cases is easy to identify, for instance when the height of the peak is negative, or the plot of peak position against time is not a straight line as expected since the plaque boundary moves at roughly constant speed.

Several aspects of the results we see are interesting. Firstly, it is worth noting that the earliest time measurements of lysis time are in line with those that we would expect from liquid measurements (i.e. $\sim$ 20 mins, see Fig. 4.16). As the expansion progresses, however, we see the lysis time initially increase, then return to its original value, before increasing again until the end of the expansion, eventually reaching a lysis time that is 2-3 times larger than the initial value.

[5] At least to me

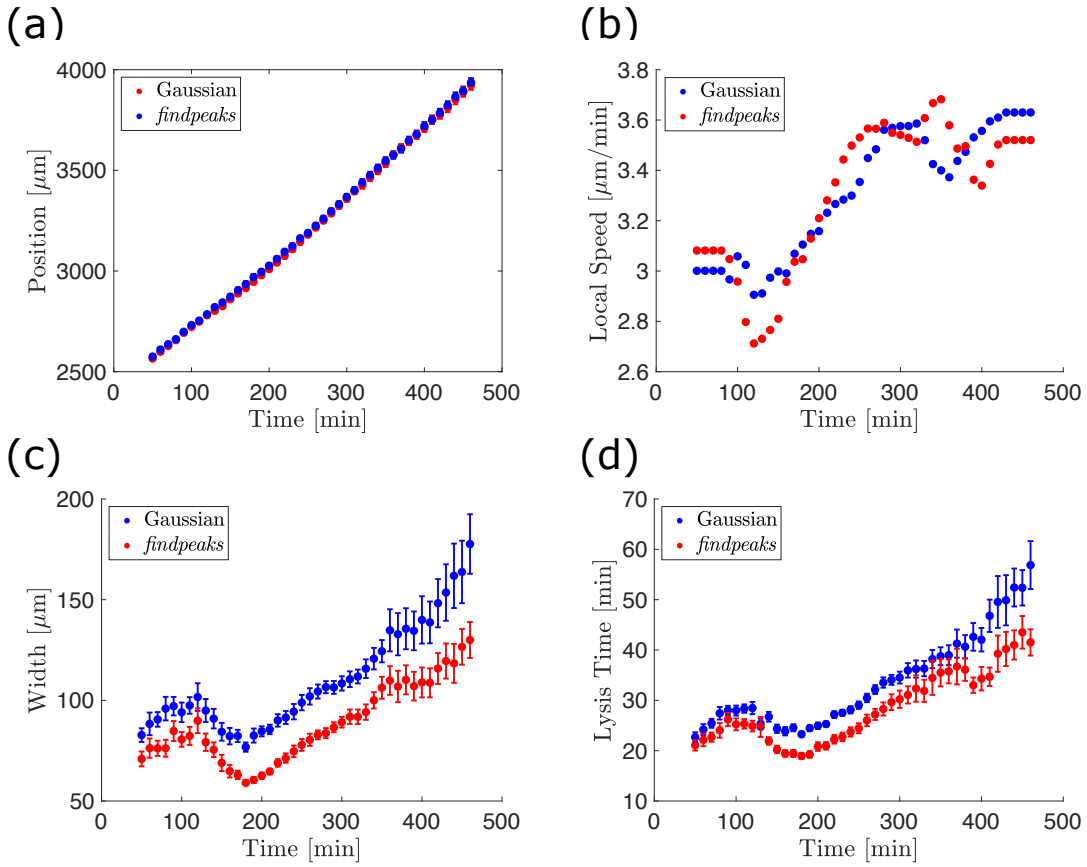


Fig. 4.14 The radially averaged lysis time at the boundary of the plaque over the course of the expansion. The average is taken over approximately 56% of the angles in which a strong band was clearly visible. (a) Shows the position of the plaque boundary as a function of time. (b) Shows the speed of the front as a function of time, calculated at each point from the position in a 50 min window (i.e. from 5 frames of the timelapse). (c) Shows the width of the infected region. (d) Shows the lysis time. A comparison between the two different methods are shown, with both showing qualitatively similar results. Error bars are given by the standard error on the mean across the different directions.

Impact of Host State

We wish to determine the degree to which the observed increase in lysis time during the course of the expansion Fig. 4.15 is due to a change in the metabolic state of the host. To this end, we grew a lawn of bacteria on an agar plate under the same conditions used for the main experiment. Then, this bacteria was re-suspended in fresh LB, and a lysis curve in liquid was performed, as described in Chapter 2. These results are shown in comparison to "regular" liquid lysis curve measurements in Fig. 4.16.

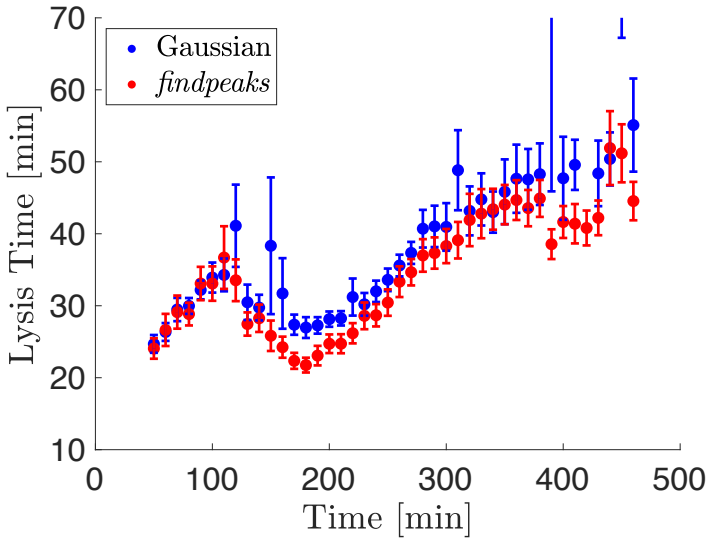


Fig. 4.15 The radially averaged lysis time over all directions around the plaque. A comparison between the two different methods are shown, with both showing qualitatively similar results. Error bars are given by the standard error on the mean across the different directions. Some data points for the Gaussian method at large times are outwith the y axis limit.

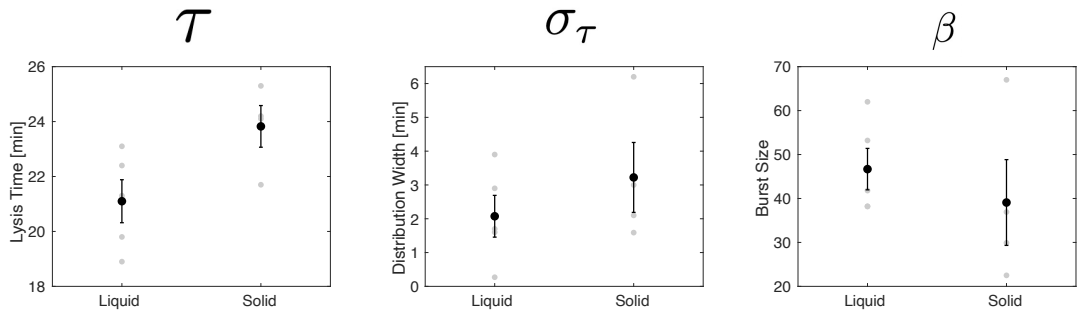
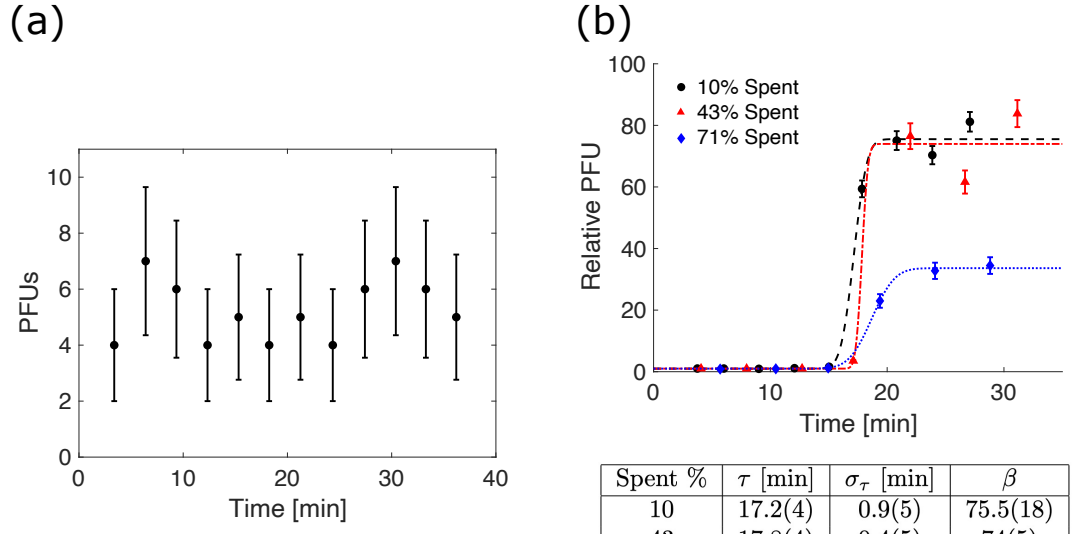


Fig. 4.16 A comparison of the single burst curve parameters found from the average of 5 measurements. "Liquid" measurements refer to measurements taken with exponentially growing bacteria, using the standard lysis curve assay described in Chapter 2. "Solid" measurements refer to the same protocol, but using bacteria that had been grown for 6 hours on an agar plate before being resuspended in fresh LB.

It can be seen that the mean lysis time τ is larger on average in the bacteria recovered from the agar plate. The increase in lysis time is significant ($P = 0.041$), whereas the changes to σ_τ ($P = 0.395$) and β ($P = 0.528$) are not. The increase in τ however is substantially less than the increase in lysis time that we observe during the course of the expansion on the plate Fig. 4.15.

This is obviously not a perfect control, as the plate bacteria are re-suspended in fresh LB, whereas on the plate some amount of the LB will have been used up. If absolutely no nutrients are provided in the liquid, however, for instance if the liquid measurements are carried out in water, then we found that no lysis/burst occurs (Fig. 4.17a). It's not clear how one would best assess the proportion of the nutrients that were used up in the plate over a given time period, as it is also likely that the bacteria will use nutrients globally and locally at different rates i.e., the nutrients initially directly beneath the lawn will likely be depleted more rapidly than those originating furthest from the bacteria. It's also true that in the plate after 8 hours there must still be *some* nutrients as the plaque is still growing at this time.

In an attempt to address this issue, we also performed lysis curves where the bacteria from the plate were re-suspended in partially 'spent' LB. To create this spent LB, liquid cultures of LB were inoculated with bacteria and allowed to grow until a target optical density OD_{600} was reached. By comparison to the OD_{600} of an overnight culture, we could then estimate the proportion of the available nutrients that had already been used at that point. This culture was then centrifuged to pellet the bacteria, and the supernatant (the spent LB) removed.



Spent %	τ [min]	σ_τ [min]	β
10	17.2(4)	0.9(5)	75.5(18)
43	17.8(4)	0.4(5)	74(5)
71	18.7(3)	1.5(5)	33.6(5)

Fig. 4.17 (a) A lysis curve performed in water, without the provision of any nutrients. No change in the number of PFUs can be seen. (b) Lysis curves performed in bacteria that had been re-suspended in 'spent' LB, i.e. LB that had already been used to grow bacteria to a target OD_{600}. The percentage spent indicates the target OD_{600} with respect to the density of an overnight culture.

The results of the lysis curves performed in the spent media can be seen in Fig. 4.17b. While the 71% spent media results in a significantly reduced burst size, little difference is made to the lysis time. This observation could potentially be explained by a limited portion of infected cells successfully lysing, due to the limited availability of nutrients. If, for instance, half of the cells which were infected lysed at the usual time, while the other half never lysed, this would produce a lysis curve where the fitted lysis time was the same, but the fitted burst size was halved. Regardless, these observations further suggest that the cause of the substantial increase in lysis time observed on plate is factors beyond the availability of nutrients.

4.6 Macrofluidic Channels in Agar Plates

One issue that has appeared multiple times throughout this Chapter is that of changing and uncertain availability of nutrients during experiments on plate. It would therefore be beneficial if the supply of nutrients to the plate could be carefully controlled, without disturbing the bacteria and phage that are growing on the surface. To achieve this, the idea is to create channels under the surface of the agar through which media can flow. The rate at which the media is flowed through can then be controlled, and maintained at a constant rate, in principle indefinitely.

4.6.1 Methods

Custom glass plates were designed as shown in Fig. 4.18. These plates have 8 × 3.2 mm diameter holes evenly spaced along each side of the plate. To prepare the plates for experiments, first everything is autoclaved to ensure it is sterile. Then, Tygon S3 E-Lab tubing (outer diameter 3.2 mm) is placed through the holes on opposite sides of the plate, such that the tubing forms a straightly line between opposite walls of the plate. It should be noted here that while the plates had the ability to accommodate 8 channels, only 4 were used in an attempt to reduce the risk of damaging the plate while it was being constructed (this will be discussed further later). Any holes that are not in use are sealed using either autoclave tape or blu tac. Next, 45 ml of 2% agar with *no* LB is poured into the plate. The lack of LB in the agar is to ensure that any growth of bacteria and phage on the surface is due to nutrients supplied exclusively through the channels. After the agar has dried, the tubing is removed, leaving an agar

plate with channels running under the surface that can be accessed via the side of the plate.

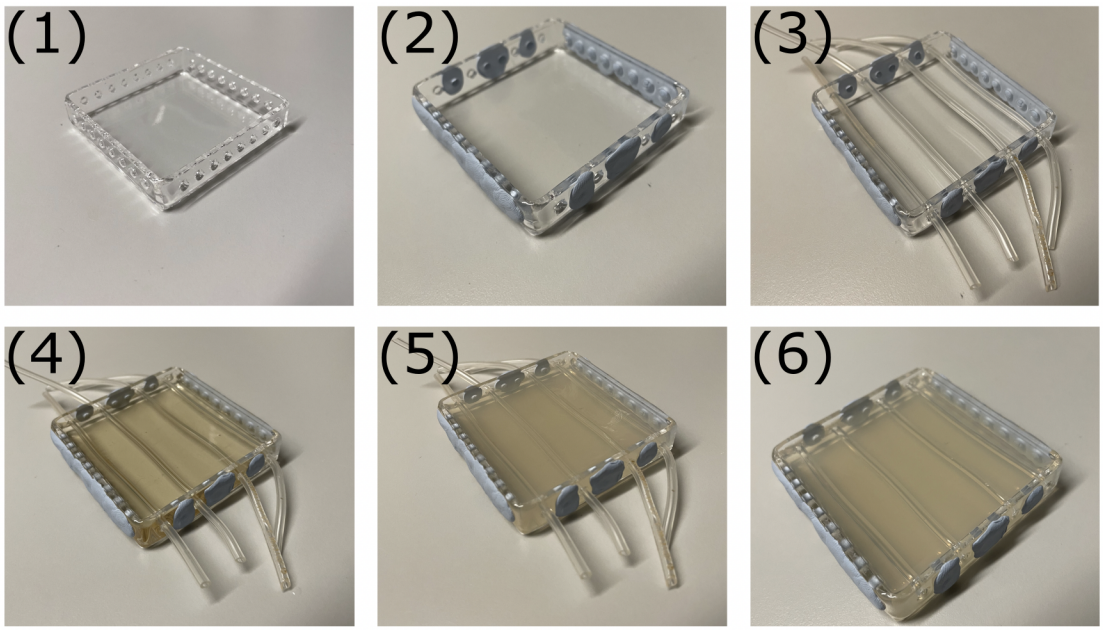


Fig. 4.18 An illustration of the the preparation procedure of agar plates with subsurface macrofluidic channels. For scale, the plates are a 7 cm square, and the holes have a diameter of 3.2 mm. Initially, any holes that will be unused are sealed using blu tac or autoclave tape. Then, Tygon S3 E-Lab tubing is used to connect holes on opposing sides of the plate. 45 ml of 2% agar without LB is then poured into the plate, and left to harden, at which point the tubing is removed.

The next step in preparing the plates is to reconnect tubing to each end of all channels, such that media can be flowed through. This step is very 'fiddly.' The tubing has to be inserted into the channels in such a way that the tubing forms a seal with the surrounding agar, so that pressure can be maintained within the channels and no liquid flows out. This is difficult to achieve without damaging the surrounding agar in some way, resulting in no seal (see Fig. 4.24, for instance). Similarly, once the tubing has been successfully inserted, one must be careful not to move it around too much for the same reason. This is the reason that 4 channels were used instead of 8, so as to increase the space between neighbouring channels, and reduce the risk of bumping previously inserted channels. In my experience, this step is easier if the ends of the tubing are cut in a way such that they are 'blunt,' rather than 'pointed' (Fig. 4.19). Pointed tubing feels easier to insert through the hole in the side of the plate, but that is because it does not need to be inserted straight, which increases the risk of damaging

the agar. While blunt tubing feels harder to insert, there is a lower risk of damaging the agar.

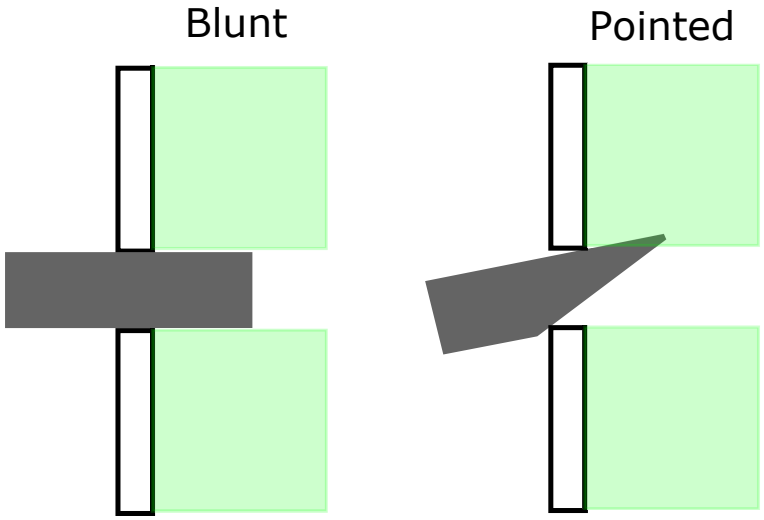


Fig. 4.19 A cartoon to illustrate the benefit of inserting 'blunt' rather than 'pointed' tubing. The tubing is shown in grey, the glass wall of the plate is shown as the white rectangles, with the agar appearing in green. It can be seen that while it is easier to insert the pointed tubing, it is more liable to rupture the surrounding agar.

Once inserted, the end of the tubes on the outlet side of the plate are placed into a bottle containing LB. The tubes on the inlet side of the plate are then merged into one tube using Y-splitters, and this tube is connected to a syringe. Media is then drawn into the syringe, through all of the tubing and the channels in the plate. The ends of the tubing in the outlet bottle must remain submerged, so that the tubing remains sealed end-to-end. The reason for drawing the media through in this fashion is that, from experience, if the media starts in the syringe, and the tubing is initially filled with air, then media does not flow down all of the channels unless a lot of pressure is applied, which risks damaging the plate. This is the same reason that the 4 outlet channels are not merged into one using the Y-splitters, as that would result in the same issue when drawing up the liquid.

If there are no leaks anywhere, the media can then be pumped slowly back through the channels to supply nutrients to the plate. For these initial measurements, the media was pumped through at a total rate of 1 ml/hr. As usual, overnight liquid cultures of *E. coli* BW25113 were grown from single colonies at 37 °C in LB. To create bacterial lawns, 20 μl droplets of overnight culture were pipetted onto the agar surface and left to dry. If the growth of plaques was also being observed, 1 μl droplets of wild-type T7 were pipetted onto the surface of the lawns.

The plates were then imaged using a Nikon D5300 camera equipped with a Nikon AF-P DX 18-55 mm f/3.5-5.6G VR lens. The camera is held on a stand pointed downward towards the sample, which is situated above a light table. The camera is connected to an intervalometer to allow for the acquisition of a time-lapse. During the imaging period, the sample was kept at 37 °C using an ibidi Multi-Well Plate Heating System. To allow the tubing to access to plate, there is a small (∼4 mm) air gap between the base and the lid of the heating system. A thermometer was used to check the temperature inside the chamber, and it was found that the air gap causes a reduction of 0.5 °C.

4.6.2 Results

Bacteria Grow Using Nutrients from Channels

It can be seen from the images in Fig. 4.20 that bacteria are able to grow using only the nutrients provided through the sub-surface channels.

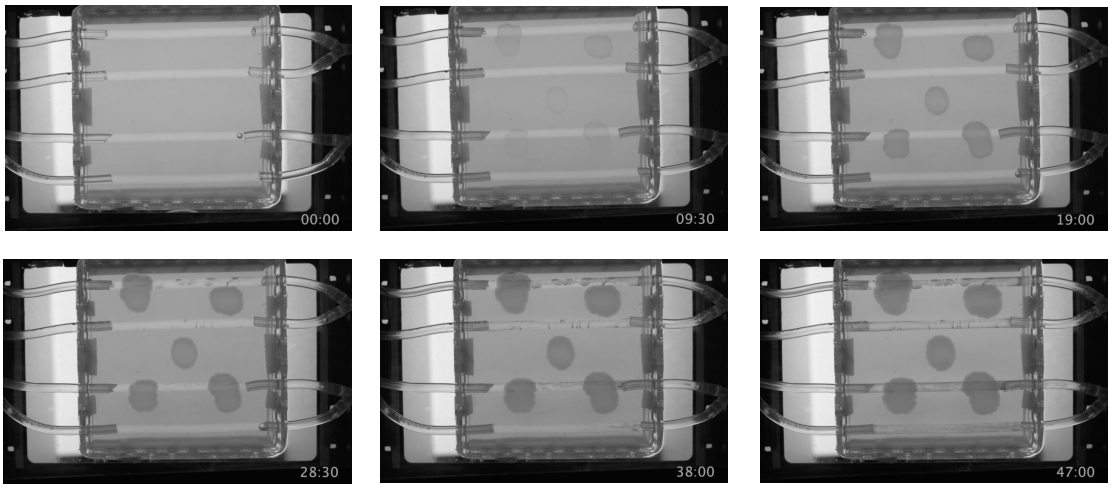


Fig. 4.20 Bacteria are able to grow on plates where nutrients are provided exclusive through sub-surface channels. The plate is made of 2% agar but contains no nutrients. LB is flowed into the plate at a total rate of 1 ml/hr. The time indicated in the bottom right of the image is of the format [hh:mm]. For scale, the plate is a 7 cm square.

We quantify growth in two ways: (i) pixel density and (ii) radius. For the first approach, the pixel intensity of the droplets was measured as a function of time. The pixel intensity of the plate background, i.e. regions with no bacteria, was also measured. The background profile is then subtracted, and then multiplied by -1. This multiplication occurs because in our set-up, the growth of cells decreases the

transmitted light, and therefore the intensity. This means that as the lawns grow, the intensity of the droplet minus the background becomes an increasingly negative value. Since we want the number to be a proxy for density, it makes more intuitive sense to make it positive. To account for the uneven light intensity across the image, the growth curves were then shift up or down such that they match at time $t = 0$. Averaged across the 5 droplets this results in the growth curve that can be seen in Fig. 4.21a.

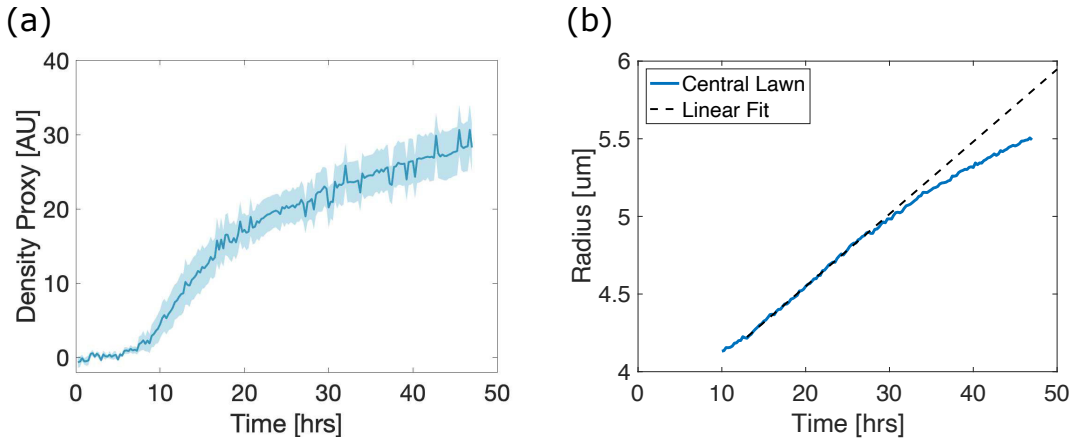


Fig. 4.21 (a) A proxy for the mean density of bacteria in the lawns as a function of time. The density proxy is calculated from the pixel intensity in the image. The shaded region indicates the standard deviation across the 5 droplets. (b) The equivalent radius of the central lawn as a function of time, measured from the area of the lawn assuming a circular geometry. A straight line fit of the period between 12 and 25 hours is shown for comparison.

From this it can be seen that the bacteria are still growing at 47 hours, albeit at a more limited rate. This is perhaps unsurprising, as over time the supply of oxygen to the bacteria at the base of the colony will become limited by all of the cells above them. Additionally, we might expect that as the colony becomes thicker and thicker, this will have a diminishing effect on the fraction of light that is transmitted.

The size of the central lawn as a function of time was also measured by thresholding the image such that the lawn and the background were completely black/white. This allowed easy measurement of the area of the lawn in pixels, which was converted into an equivalent radius assuming a circular geometry (Fig. 4.21b). This similarly confirms that the lawns are still growing after 47 hours, albeit also at a reduced rate. The reason that only the central lawn was used was that it does not overlap with any channels, which made the image analysis more straightforward.

Channels Prevent the Plate from Drying

As well as providing the bacteria with nutrients required for growth, another benefit of the channels is that they keep the agar hydrated and prevent it from drying out. This is evident by observing how rapidly plates which contain no channels, but which are otherwise kept in the same conditions, dry out (Fig. 4.22). The only difference between Fig. 4.22 and Fig. 4.20 is that the plates in Fig. 4.22 have no channels, and instead have LB supplied directly in the agar (i.e. they are 'normal' plates).

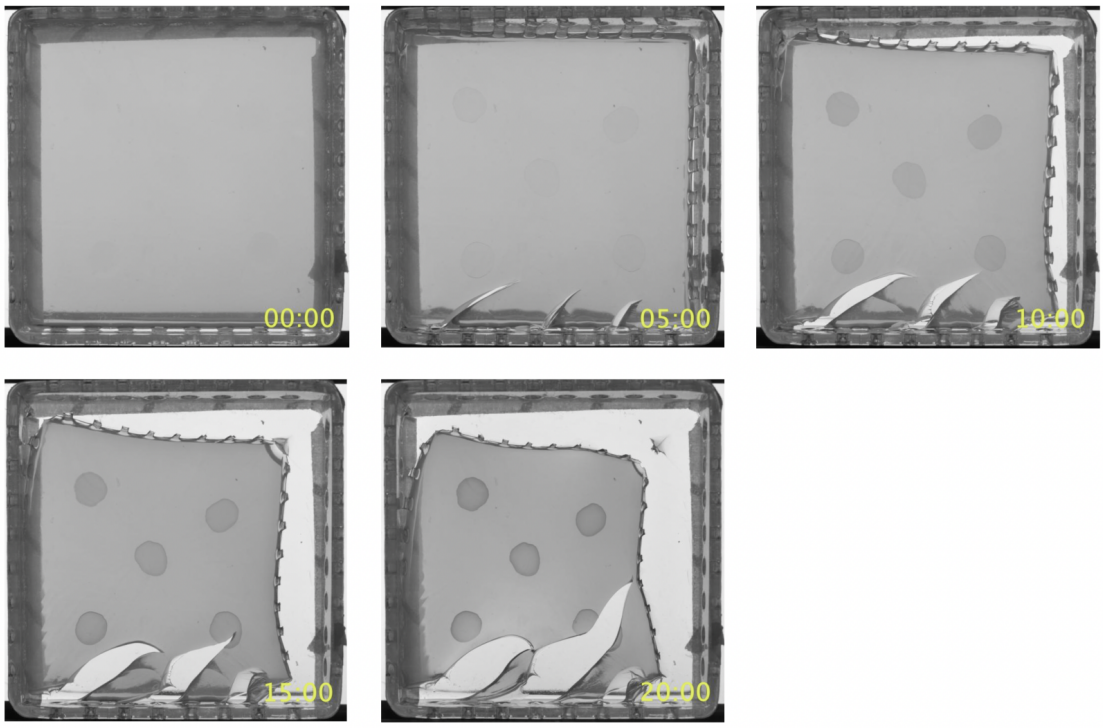


Fig. 4.22 Bacteria grown on plates with no channels, but with LB in the agar as is standard practice. The plates are kept in the same conditions as in Fig. 4.20. The time indicated in the bottom right of the image is of the format [hh:mm]. For scale, the plate is a 7 cm square.

Comparison to Bacterial Growth in 'Normal' Conditions

The conditions of this setup are themselves, however, admittedly not optimised to keep the plates from drying out. There are two main factors that contribute to this. Firstly, as noted previously, there is a small (∼4 mm) gap between the base and the lid of the heating system, which needed to be there to allow access to the tubing. Secondly, no

lids were placed over the plates inside the chamber. Both of these factors combined greatly increase the rate of evaporation from the plate.

So as to have a more meaningful control for the growth rate of the bacteria in normal conditions, we therefore compare the growth in the channelled plates to lawns grown on plates that are covered by a lid, and where the chamber itself has no unnecessary air gaps, as this prevents the plates from drying out so fast. Interestingly, we find that in these conditions, the control lawns appear to grow significantly faster in terms of our density proxy (Fig. 4.23a), but slightly slower in terms of the increase in radius of the lawn (Fig. 4.23b,c). These rates are likely dependent on the rate at which nutrients are supplied to the plate, and so perhaps this indicates that greater rates are required to better mimic the conditions of the control plate.

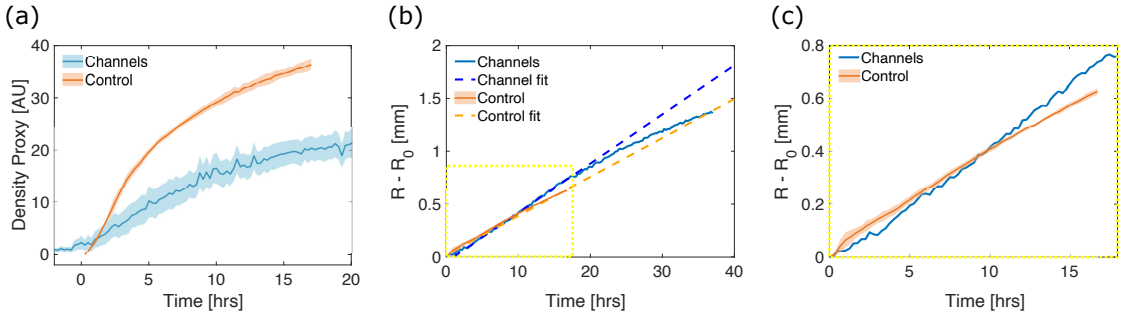


Fig. 4.23 Growth of bacterial lawns on plates with channels compared to on a control plate under normal conditions (i.e. with efforts made to minimise evaporation). (a) A proxy for the mean density of bacteria in the lawns as a function of time. The density proxy is calculated from the pixel intensity in the image. The shaded region indicates the standard deviation across multiple droplets. (b) The equivalent radius of the lawns as a function of time, measured from the area of the lawn assuming a circular geometry. Straight line fits are shown for comparison. (c) A closeup of the region highlighted by the yellow dotted rectangle in (b). Channels data is the same data as shown in Fig. 4.21, but shifted in time to aid comparison.

Plaques Also Form on Channel Plates

It can be seen in Fig. 4.24 that plaques are also able to form on bacteria lawns, where the nutrients are provided exclusively through the sub-surface channels.

In this plate, one of the channels was damaged when connecting the tubing. To create a proper seal, the tubing that would have connected to the outlet side of the plate was then connected directly to the Y-splitter on the inlet side, i.e. the broken channel was simply bypassed. This results in the plate drying out and shrinking during

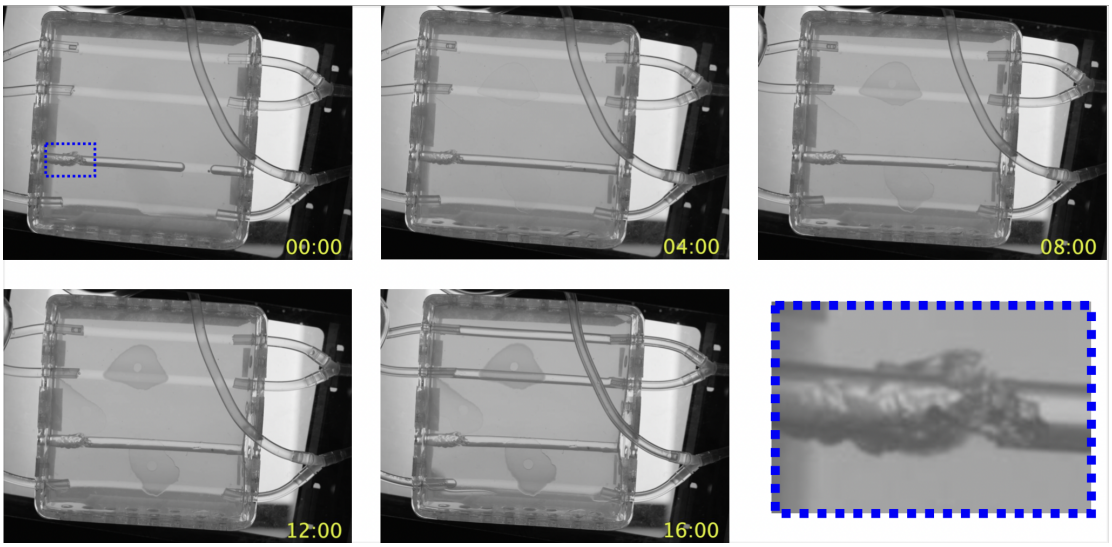


Fig. 4.24 Plaques are able to grow on plates where nutrients are provided exclusive through sub-surface channels. The plaques are visible as small, circular clearings in the centre of the bacterial lawns. The plate is made of 2% agar but contains no nutrients. LB is flowed into the plate at a total rate of 1 ml/hr. The time indicated in the bottom right of the image is of the format [hh:mm]. For scale, the plate is a 7 cm square. One of channels in the plate was damaged when connecting the tubing, meaning that it could not be used. This channel was bypassed, with the media being pumped directly to the reservoir at the end. The dotted blue rectangle highlights the damage to the channel.

the course of the experiment. Around the 14 hr mark, this shrinking brakes the seal of all of the other channels, and liquid leaks out of all of the channels (difference between hour 12 and hour 16 in Fig. 4.24).

Before this point however, it can be seen that small circular plaques had formed in the centre of the bacterial lawns, indicating that phage are able to grow in such a nutrient environment. Although not ideal, this 'broken' plate is also useful in that it indicates that at least 4 channels are necessary to prevent the plate from drying out, while 3 are insufficient.

4.7 Discussion

In this Chapter, I have presented several experimental protocols intended to address various technical challenges and shortcomings of existing techniques. At present, to measure phage phenotypic parameters one has to rely on measurements taken exclusively in liquid cultures, which are not necessarily representative of the phage

behaviour in bacterial lawns. This discrepancy is due to multiple effects, such as interactions that may occur between closely packed neighbouring cells in lawns, as well as the changing conditions that arise during growth on an agar plate. While the techniques that I have presented represent an important step forward, there is clearly still work to do in refining the methods in each case.

The protocol developed to measure adsorption rate in lawns appears to work well, as we are able to verify that no adsorption occurs in the non-adsorbing control. The protocol itself has the benefit of being fairly simple, and doesn't require any additional materials or equipment compared to existing protocols to measure adsorption rate in liquid. The main downside of this protocol is that there are many sources of quite significant error, meaning that a significant number of replicates are required to obtain a reliable average. This condition however is generally true of adsorption measurements in liquid too.

The protocol to measure lysis time in lawns also appears to work, as it does produce a lysis time which seems reasonable. One significant drawback of this protocol is that it requires a phage which is able to induce fluorescence in infected cells, meaning that the measurement cannot be made using *any* phage without genetic modification. There also remains some questions, both about the protocol and its results. For instance, it is not entirely clear how well the width that is measured in the fluorescent images actually maps to the infected cells. One of the reasons for this is purely practical - the widths measured through this procedure have a relatively large corresponding error. I expect, therefore, that while this technique should be able to give an estimate for the lysis time that is better than order of magnitude, it will not produce highly precise results.

Additionally, as we shall see later in Chapter 6, images taken in a microfluidic mother machine (Fig. 6.2) seem to suggest that cells only become visibly fluorescent $\sim$6 mins after infection, consistent with the timing of the expression of late genes in T7 [172]. While this could potentially alter the precise lysis time measured, the trend of increasing lysis time during the expansion should remain. This raises the question: why? Our controls in liquid would seem to suggest that this effect is not caused by a reduction in nutrients or the change in metabolic state. Further work will be required to better answer this question.

Finally, it was demonstrated that channels under the surface of the agar could be used to supply nutrients at a constant and controllable rate. Not only does this have the benefit of reducing uncertainty when it comes to levels of nutrients, but it also

prevents the plates from drying out, meaning that experiments can be run for much longer periods of time, potentially indefinitely. The main difficulty with this approach is purely practical - it's difficult to set up these plates without damaging something that prevents their use. Many of these issues could be fixed with simple modifications to the equipment, such as increased spacing between adjacent holes, or 3D printed tools that could act as connectors between the agar and the tubing, reducing the risk of damage. Beyond this, further work needs to be done to fully characterise the impact of all of the physical parameters of the system, e.g. the number and density of channels under the surface of the plate, the size of these channels, the density and thickness of the agar, the flow rate and concentration of nutrients, etc. In addition, while not explored here, this kind of set-up could be easily adapted to different purposes, such as supplying antibiotics, creating nutrient or antibiotic gradients, or looking at the impact of stopping/starting the supply of these products through time.

'The more we share, the more we have.'

Leonard Nimoy

'You can win the rat race but you're still a rat.'

Banksy

Chapter 5

The Evolutionary Impact of Superinfection

This chapter is based on

M. Hunter and D. Fusco, *Superinfection exclusion: A viral strategy with short-term benefits and long-term drawbacks*, PLOS Comput. Biol. **18**(5): e1010125, (2022)

5.1 Introduction

Here, we explore how allowing or preventing superinfection impacts the evolutionary fate of neutral and non-neutral variants in a simulated well-mixed phage population with constant, but limited, availability of host. We choose to focus on superinfection exclusion mechanisms that allow secondary adsorption events, but prevent DNA insertion, so that in isolation the phage growth dynamics are the same in the two cases and a direct comparison between the (dis)advantages of the two strategies is more straightforward. We first quantify the effective population size of superinfecting (S) and superinfection-excluding (SX) populations to estimate how these strategies affect genetic drift. We then turn our attention to the effect of non-neutral mutations on (i) the phage growth rate in isolation and (ii) their ability to out-compete the wild-type. Having characterised both the neutral dynamics and the fitness of different variants, we put both aspects together to explore the balance between drift and selection in superinfecting and superinfection-excluding populations, showing that selection is consistently more efficient in superinfecting populations. Finally, we study the evolutionary fate of a mutation which changes whether an individual is capable of preventing superinfection or not. Overall, this work establishes a baseline expectation for how the simple occurrence

of superinfection impacts fundamental evolutionary outcomes and provides insights into the selective pressure experienced by viral populations with limited, but constant host density.

5.2 Computational modelling framework

We study the evolutionary fate of phage mutants using a stochastic agent-based model. We simulate a well-mixed population of phages V interacting with a population of host bacteria that is kept at a constant density, similarly to a turbidostat [173, 174]. Each phage i has a defining set of life history parameters, namely an adsorption rate α_i, a lysis time τ_i and a burst size β_i, and each bacterium can either be in an uninfected B or an infected I state. Throughout most of this Chapter, the subscripts indicating phage i will be omitted to improve readability. Additionally, we shall use symbols with 'hat' notation (e.g., $\hat{\tau}$) to represent dimensional variables (e.g., $\hat{\tau} = 15$ min), and symbols without the 'hat' notation to represent non-dimensional/simulation variables.

In each simulation time-step Δt, adsorption, phage replication within the host and lysis occur. The number of infecting phage V_I in each step is drawn from a Poisson distribution whose mean corresponds to the expected value $\alpha V(B + I)\Delta t$ in a well-mixed population. The infecting phage are removed from the pool of free phage, and V_I bacteria, whether infected or uninfected, are chosen uniformly and with replacement to be the infection target. In both superinfecting and superinfection-excluding scenarios, the final lysis time τ of the host is set by the first phage to infect it and it is treated as deterministic to limit the number of model parameters. This choice was made for the sake of simplicity, given the complex and varied nature of superinfection mechanisms [74–76]. A preliminary analysis of the effect of stochasticity in lysis time is presented in Sec. 5.3.4. In the case where multiple phage infect the same host in a single time-step, the 'first' phage is chosen uniformly among those infecting the host. Phage replication within the host post-adsorption depends on whether superinfection is allowed or prevented:

Absence of superinfection: τ steps after the first adsorption event, the bacteria will lyse, releasing new phage into the pool of free phage. The number of phage released Y is drawn from a Poisson distribution with mean β.

Presence of superinfection: Pseudo-populations tracking the growth of phage inside the host are used (see Fig. 5.1b). Because here we focus on the case of two superinfecting phage populations (a and b), this results in two pseudo-populations p_a and p_b. During

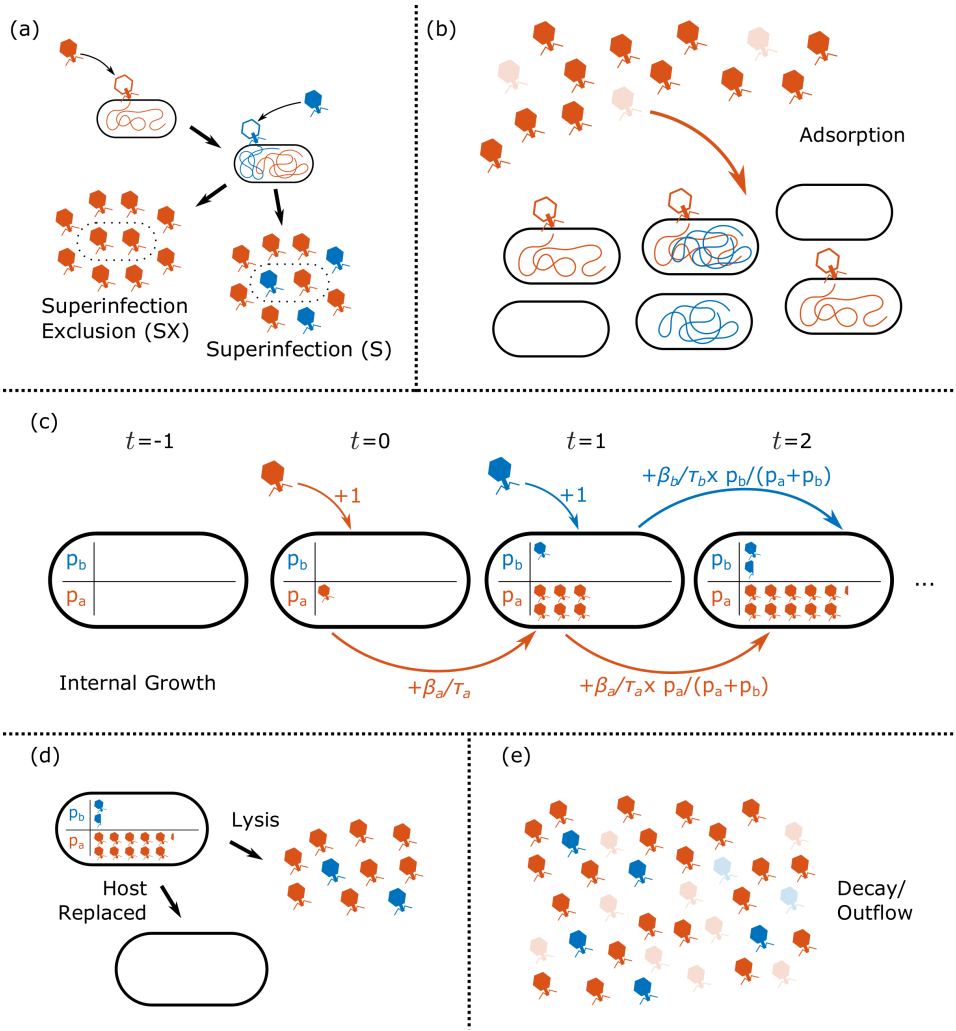


Fig. 5.1 (a): In superinfection-excluding scenarios, all of the progeny released as the cell lyses are copies of the initial infecting phage, whereas when superinfecting is permitted, the progeny are split between both types of phage. (b): In our model, adsorption of phage to host occurs each timestep Δt, with the number of infecting phage V_I being drawn from a Poisson distribution whose mean corresponds to the expected value $\alpha V(B+I)\Delta t$ in a well-mixed population. Note that phage are able to adsorb to both uninfected and infected hosts. (c): During superinfection, pseudo-populations p_a and p_b are used to represent the growth of two phage types (a and b) inside the host cells. These populations increase by 1 whenever a phage infects the host, and each population increases by some fraction of its rate β/τ determined by the relative size of the populations in the previous step. (d): Upon lysis, phage are released according to the relative size of the two pseudo-populations at that point (see text for details). Lysed cells are then immediately replaced with fresh hosts. (e): We also introduce a decay, or removal, of free phage at rate δ, which accounts for natural phage decay and the outflow of the turbidostat system.

the intermediate steps between the first adsorption event and lysis, in the case where there is only one type of phage inside the host, that population will grow at a constant rate β/τ, where β and τ are both specific to the type of phage (i.e. p_a grows at rate β_a/τ_a and p_b grows at rate β_b/τ_b)[1]. This is to reflect previous reports of a positive linear relationship between lysis time and burst size [51].

In the event where both types of phage are present within the host, to reflect the intracellular competition for the host's resources, each population increases by only a fraction of its potential. We assume that the host contains a fixed amount of resources that can be used in each step, and that the division of these resources is dictated by the sizes of the populations within the host at that time e.g., if there are 10 of one phage and 20 of another, the first will get 1/3 of the resources and the second will get 2/3 of the resources. This means that in a single time-step p_a increases by an amount $\beta_a/\tau_a \times p_a/(p_a + p_b)$ and p_b increases by an amount $\beta_b/\tau_b \times p_b/(p_a + p_b)$.

At the point of lysis, the total number of phage released Y is drawn from a Poisson distribution with mean $p_a + p_b - V_n$, where V_n represents the number that infected the host prior to lysis. This is to ensure that, in the event where a cell is only infected by 1 type of phage, its mean burst size remains β, regardless of how many phages had infected the cell until that point. The number of phage released of one type Y_a is then drawn from a binomial distribution with Y attempts and probability $p_a/(p_a + p_b)$ of success, with any remaining phage being the other type ($Y_b = Y - Y_a$).

Following lysis, the lysed bacteria are immediately replaced with a new, uninfected host, resulting in a bacterial population of constant size. We also introduce a decay, or removal, of free phage at rate δ, which accounts for natural phage decay and the outflow of the turbidostat system.

Our decision to model the phage-bacteria interactions in this way is motivated by a desire to have a relatively 'simple' model with which we can explore the baseline evolutionary impact of superinfection and superinfection-exclusion mechanisms. It simplifies various biological aspects by design. For instance, in a true turbidostat, bacteria are able to replicate, and some of the host cells will be removed through the outflow of the system along with the phage. Our system neglects these aspects, as we desire a truly *constant* host population size, rather than one in which the populations go through repeated cycles of growth and dilution, thereby introducing bottlenecks

[1]In Appendix 5.A we also present the results from a version of the model where lysis time *and* burst size are set by the initial phage.

that may obscure the impacts of superinfection. Some of the consequences of these decisions will be discussed further in Sec. 5.4.

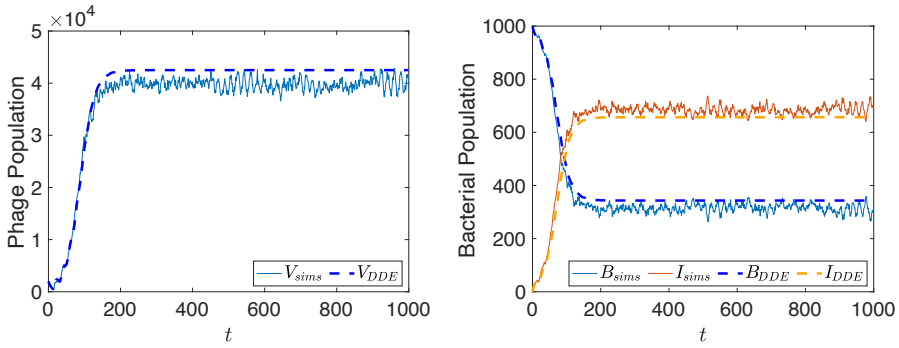


Fig. 5.2 An example realisation of the simulation. The resident phage population initially grows until it reaches a steady-state V_{ss}. The average behaviour of the model is mostly captured by a set of delay differential equations (DDEs), set out later in Eqs. 5.1. Slight discrepancies arise due to the fact that in the simulations, infection, decay and lysis must occur in discrete steps. Parameters used are $\alpha = 3 \times 10^{-6}$, $\beta = 100$, $\tau = 15$, $\delta = 0.1$ and $B_0 = 1000$.

Simulations were initialised with B_0 uninfected bacteria and $2B_0$ "resident" phage, and then run until the phage, uninfected bacteria and infected bacteria populations each reached steady-state values (V_{ss}, B_{ss} and I_{ss} respectively), determined by their running average (Fig. 5.2). This steady-state arises due to a balance between phage production and loss and it is independent of the initial number of phages (Fig. 5.3).

5.3 Results

5.3.1 Superinfection leads to a larger effective population size

First, we find that genetic diversity consistently declines faster in populations that prevent superinfection, indicating a smaller effective population size when compared to superinfecting populations (Fig. 5.4). This can be understood by considering that in the superinfecting scenario, each phage has more opportunity to successfully infect a host cell, since secondary infections can result in the production of some offspring when the cell lyses. Therefore, more phage are able to contribute to the next generation, thereby slowing the loss of diversity.

In addition, Fig. 5.5 shows that in both superinfecting and superinfection-excluding populations higher adsorption rate and burst size, and shorter lysis time result in larger

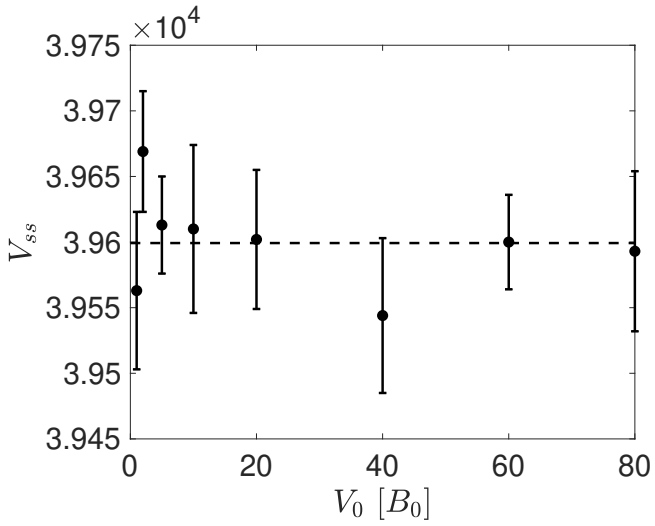


Fig. 5.3 The steady-state phage population V_{ss} reached does not depend on the initial number of phage V_0 in the simulations. In all, $\alpha = 3 \times 10^{-6}$, $\beta = 100$, $\tau = 15$, $\delta = 0.1$ and $B_0 = 1000$. Error bars show the standard error on the mean of 10 simulations.

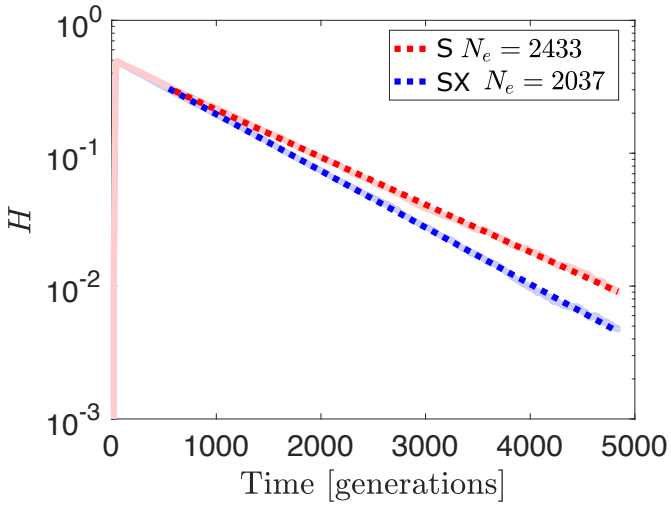


Fig. 5.4 Linear fit to log transformed heterozygosity data, with slope $\Lambda \equiv 2/N_e$ revealing that allowing superinfection (red, S) results in a larger effective population size compared to the case where superinfection is excluded (blue, SX). Parameters used were $\alpha = 3 \times 10^{-6}$, $\beta = 100$, $\tau = 15$, $\delta = 0.1$ and $B_0 = 1000$. Data obtained is the average of 1000 independent simulations

effective populations. This observation is, however, partially attributable to the change in total phage population $N_T = (V_{ss} + \beta I_{ss})$, where V_{ss} indicates the steady-state free phage population, I_{ss} indicates the steady-state number of infected bacteria, and so βI_{ss} represents the number of phage that inevitably will join the free phage population.

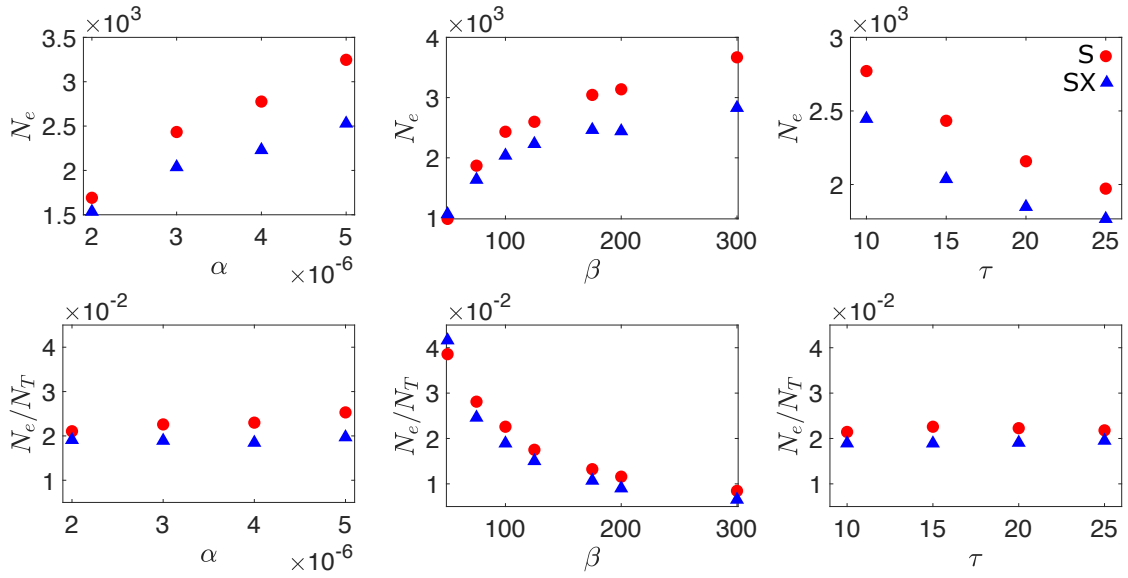


Fig. 5.5 The effective population size in both superinfecting (S) and superinfection excluding (SX) populations as a function of adsorption rate α, burst size β and lysis time τ. Populations also shown scaled by the size of the total phage population $N_T = (V_{ss} + \beta I_{ss})$. Parameters used were $\alpha = 3 \times 10^{-6}$, $\beta = 100$ and $\tau = 15$ unless otherwise stated. As throughout, $\delta = 0.1$ and $B_0 = 1000$. Error bars are plotted but are too small to see. The data is obtained from an average of at least 1000 independent simulations.

Indeed, adsorption rate and lysis time impact both the effective and actual population sizes in the same way (i.e. $N_e/N_T \approx$ const.). By contrast, larger burst sizes increase the effective population size less than the actual population size (Fig. 5.5), resulting in a decrease of N_e/N_T. This can be interpreted by noticing that while increasing burst size results in more phage, the number of phage that can actually contribute to the next generation (i.e. the effective population size) is limited by the number of bacteria that are available. Therefore, as burst size is increased, a larger fraction of phage become wasted.

5.3.2 Neutral mutants are consistently more likely to fix in superinfecting populations

To continue our characterisation of the neutral dynamics in both superinfecting and superinfection excluding populations, we turn to the fixation probabilities of neutral mutants (see Sec. 5.5.3), and determine how they are affected by the phage infection parameters.

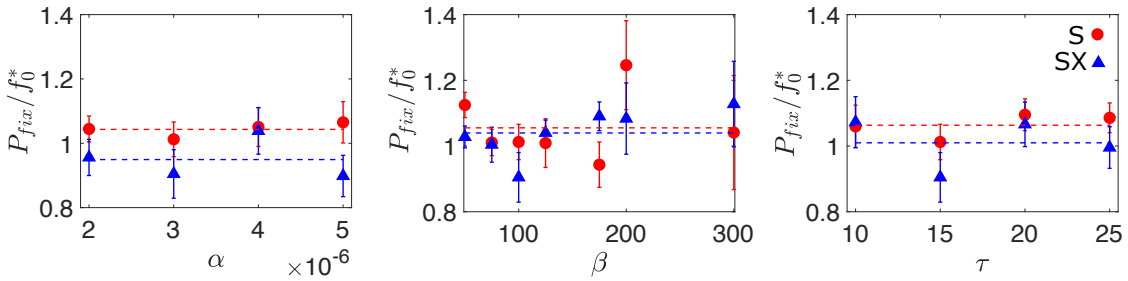


Fig. 5.6 Probability of mutant fixation P_{fix} in the superinfecting (S) and superinfection excluding (SX) scenarios, scaled by the initial frequency of the mutant $f_0^* = 1/(V_{ss} + \beta I_{ss})$, as a function of adsorption rate α, burst size β and lysis time τ. Dashed lines indicate the simple average of the data for both the superinfecting (blue) and non superinfecting (red) scenarios. These lines indicate that neutral mutants in superinfecting populations experience a small advantage over mutants in an equivalent non superinfecting population. Unscaled P_{fix} data can be seen in Supporting Information, Fig. S4. Unless otherwise stated, the parameters used were $\alpha = 3 \times 10^{-6}$, $\beta = 100$, $\tau = 15$, $\delta = 0.1$ and $B_0 = 1000$. The error in our estimate of the fixation probability ΔP_{fix} is given by $\Delta P_{fix} = \sqrt{n_{fix}}/n$, where n and n_{fix} represent the total number of simulations and the number of simulations where the mutant fixes respectively. The data is obtained from a minimum of 14 million independent simulations.

Because the total phage population size depends on the life history parameters, the initial mutant frequency corresponding to one mutant phage inoculated in the population also varies with life history parameters. To account for this effect, we rescale the fixation probability by the initial frequency of the mutant $f_0^* = 1/(V_{ss} + \beta I_{ss})$, which is the same in superinfecting and superinfection-excluding populations. Fig. 5.6 shows that $P_{fix}/f_0^* \approx 1$ as the parameters are varied, indicating that the total number of phages for a given set of parameters is the main controller of neutral dynamics.

Indeed, we find that the impact of the life history parameters on the probability of fixation is what one would intuitively expect (Fig. 5.7): larger adsorption rate and burst size, and shorter lysis time, increase the steady-state size of the phage population, and reduce P_{fix}. In the following Section (Sec. 5.3.3) we also describe the average behaviour of our simulations with a system of delay differential equations (DDEs). The DDE solution shows that the total phage population at steady-state N_T is the same as in the stochastic model (Fig. 5.7).

Fig. 5.6 also shows that, on average, neutral mutants in the superinfecting scenario are more likely to fix than mutants in an equivalent superinfection-excluding population (red and blue dashed lines in Fig. 5.6 respectively). This result agrees with that found by Wodarz *et al.* [131], who showed that in a superinfecting viral population, higher

multiplicities of infection slightly favoured rare neutral and disadvantageous mutants in the short term. The intuition behind this observation can be explained in the following way: at the moment that the mutant is introduced, all infected cells are infected by the resident phage. In the superinfecting scenario, the mutant population can therefore grow by infecting an uninfected cell, or by infecting an already infected cell, as this secondary infection will lead to some fraction of the burst size being allocated to the mutant type. While resident phage can replicate by infecting either types of host, the resident population cannot further grow by infecting previously infected cells. This is because all infected cells are already exclusively infected with resident phage, and superinfection of resident infected cells by more resident phage does not result in any more resident phage being produced. As a result, superinfection increases the mutant's chance of survival in the early stages in comparison to the superinfection-excluding counterpart, similarly to conditions of high vs. low MOI [131].

5.3.3 DDE Description of Model

The average behaviour of the model used in this Chapter can be described by a set of delay differential equations (DDEs):

$$\frac{dV}{dt} = -\alpha V(B+I) - \delta V + \beta\alpha V_{t-\tau}B_{t-\tau}, \tag{5.1a}$$

$$\frac{dB}{dt} = -\alpha VB + \alpha V_{t-\tau}B_{t-\tau}, \tag{5.1b}$$

$$\frac{dI}{dt} = \alpha VB - \alpha V_{t-\tau}B_{t-\tau}, \tag{5.1c}$$

where all of the symbols are defined the same as before (V, B and I indicate the concentrations of phage, uninfected bacteria and infected bacteria as a function of time respectively; α, β, τ and δ indicate the phage adsorption rate, burst size, lysis time and decay rate respectively). The subscript is used to indicate that those terms are calculated at time $t-\tau$. The positive term in Eq. 5.1b accounts for the instantaneous replacement of lysed cells in our turbidostat environment.

By numerically solving this DDE system, we can verify that for certain parameters, a steady-state solution is reached where $V = V_{ss}$, $B = B_{ss}$ and $I = I_{ss}$, in agreement with the average behaviour of the stochastic model used throughout this Chapter (Fig. 5.2 and Fig. 5.7). In Appendix 5.B we also present a mathematical analysis of

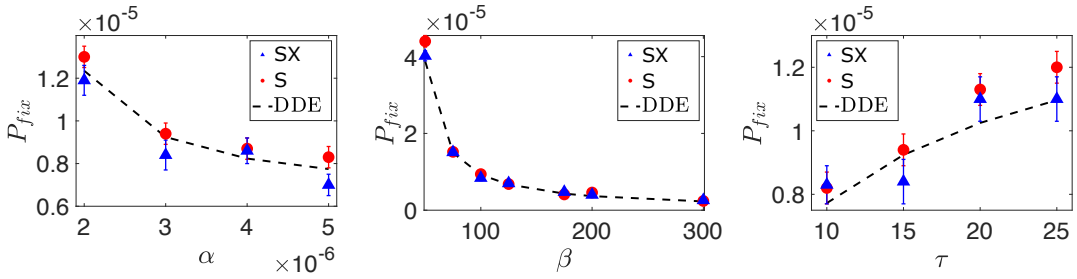


Fig. 5.7 Probability of mutant fixation P_{fix} in the superinfecting and non superinfecting scenarios as a function of adsorption rate α, burst size β and lysis time τ. This is the same as the data displayed in Fig. 5.6, prior to scaling by the initial frequency $f_0^* = 1/(V_{ss} + \beta I_{ss})$. Error bars are plotted, although in some instances may be too small to see. This data is compared with the solution of the system of DDEs, with the black dashed line represents the frequency f_0^* calculated from the steady-states reached.

this DDE system, and show that our numerically determined steady-state populations are consistent with the analytical expressions.

5.3.4 Stochastic lysis time

Here we will take a brief pause to explore our decision to implement stochasticity in both adsorption and burst size, but not in lysis time. The reason we make this choice is that to introduce stochasticity in lysis time in a realistic way would require additional parameters. For instance, Campos *et al.* compare descriptions of lysis using piece-wise, Gaussian and logistic-like functions [169]. Depending on the mathematical framing of these functions, each would require at least one parameter that controls the width or rise rate of the distribution. Given that our model already contains several parameters, we preferred to choose a simpler description, where lysis time is deterministic.

Nevertheless, here we implement a version of the model previously described where we introduce stochasticity in lysis time. In this version of the model, at the point of infection, a lysis time L is drawn from a Gaussian distribution with mean τ and standard deviation $\tau/10$. The choice of standard deviation is based on lysis time data collected for different variants of coliphage T7 (Fig. 5.8). In this case the variants are an isolate of wild-type T7 originally obtained as an aliquot from the Richardson Lab (Harvard Medical School, Boston, MA), and T7 mutant D111 (wild-type T7 background with deletions from base 532 to 1662) which was kindly donated by Dr. Vivek Mutalik from the Berkeley Lab (Berkeley, CA). The full procedure can be found

in Chapter 4 but briefly, the technique used here is adapted from the single-step growth curve protocol used originally by Ellis and Delbrück [166], and involves infecting an exponentially growing culture of bacteria with phage and then periodically sampling, with the samples being plated in soft agar with susceptible host to attain plaques. Counting the number of plaques yields the plaque forming units (PFUs) as a function of time, and from this the lysis time and burst size can be inferred.

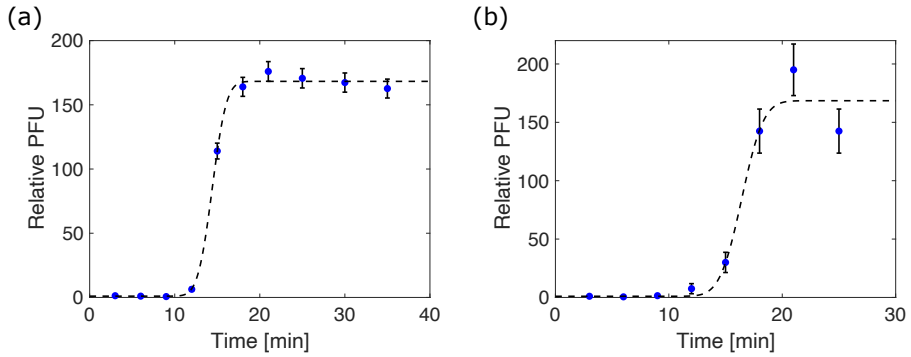


Fig. 5.8 One-step growth curves of two coliphage T7 variants. (a) corresponds to an isolate of wild-type T7, while (b) corresponds to T7 mutant D111 which is the wild-type T7 background with deletions from base 532 to 1662. Fit to the data is the Gaussian based function (Eq. 4.4), with mean lysis time $\hat{\tau}$, standard deviation $\hat{\sigma}$ and burst size $\hat{\beta}$. These fits yield $\hat{\tau}_a = 14.40(11)$ min, $\hat{\sigma_a} = 1.32(19)$, $\hat{\beta}_a = 168(2)$, $\hat{\tau}_b = 16.4(5)$ min, $\hat{\sigma_b} = 1.5(5)$ and $\hat{\beta}_b = 168(12)$. These values are typical for wild-type T7 in similar conditions [175], and so it can be seen that $\hat{\sigma} \sim \hat{\tau}/10$.

Using this version of the model, we re-run a subset of our simulations, namely the simulations used to determine the probability of fixation of neutral mutants. As can be seen in Fig. 5.9 and Fig. 5.10, the introduction of stochasticity in lysis time does not significantly alter the behaviour of the model. It is still the case that the probability of fixation of a neutral mutant is controlled by its initial frequency in the population (Fig. 5.9), and that the average behaviour of the model remains unchanged, with the simulations remaining consistent with an DDE description of the model (Fig. 5.10). We do note that the difference between superinfecting and superinfection excluding scenarios in Fig. 5.9 is less clear than in the corresponding figure without lysis stochasticity (Fig. 5.6). We speculate that this is caused by the increased level of stochasticity introduced here, but a more detailed analysis would be required to test this systematically.

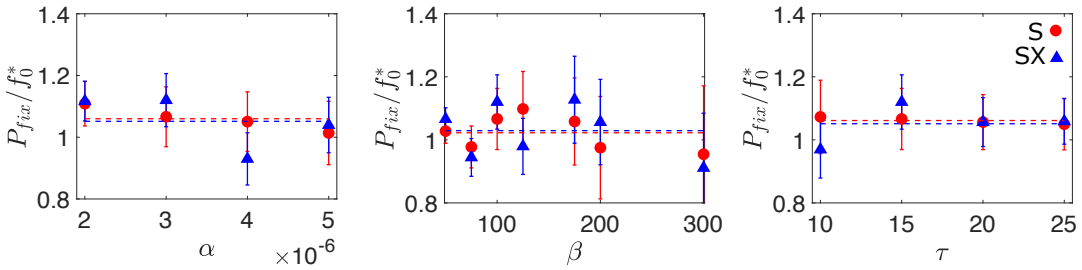


Fig. 5.9 Probability of mutant fixation P_{fix} in the superinfecting (S) and superinfection excluding (SX) scenarios, scaled by the initial frequency of the mutant $f_0^* = 1/(V_{ss} + \beta I_{ss})$, as a function of adsorption rate α, burst size β and lysis time τ. Dashed lines indicate the simple average of the data for both the superinfecting (blue) and superinfection-excluding (red) scenarios. Unless otherwise stated, the parameters used were $\alpha = 3 \times 10^{-6}$, $\beta = 100$, $\tau = 15$, $\delta = 0.1$ and $B_0 = 1000$. The data is obtained from a minimum of 10 million independent simulations.

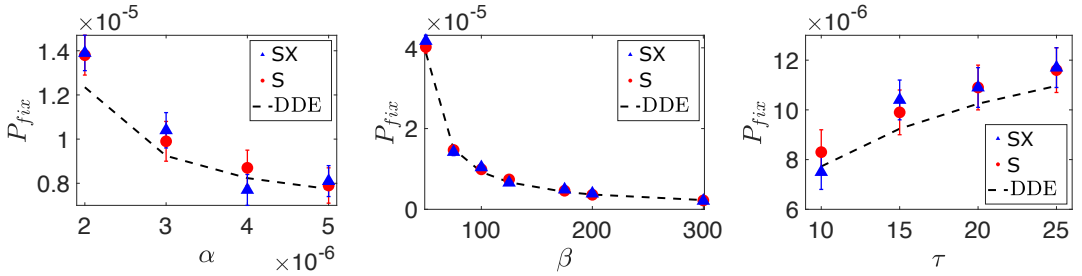


Fig. 5.10 Probability of mutant fixation P_{fix} in the superinfection (S) and superinfection-exclusion (SX) scenarios as a function of adsorption rate α, burst size β and lysis time τ. Error bars are plotted, although in some instances may be too small to see. This data is compared with the solution of a system of DDEs used to describe the average behaviour of the model, where the black dashed line represents the frequency f_0^* calculated from the steady-state values. Unless otherwise stated, the parameters used were $\alpha = 3 \times 10^{-6}$, $\beta = 100$, $\tau = 15$, $\delta = 0.1$ and $B_0 = 1000$. The data is obtained from a minimum of 10 million independent simulations.

5.3.5 Higher growth rate does not translate into competitive advantage

Now returning to the model with no stochasticity in lysis time, to investigate the evolutionary fate of non-neutral mutations, we first characterise how phage growth rate and competitive fitness is affected by changes to the phage life history parameters, i.e., adsorption rate α, burst size β and the lysis time τ, relative to the values used in our neutral simulations (see Sec. 5.5.2).

Fig. 5.11 shows that increasing burst size β or adsorption rate α results in a larger selective advantage s both in isolation and in direct competition. However, while

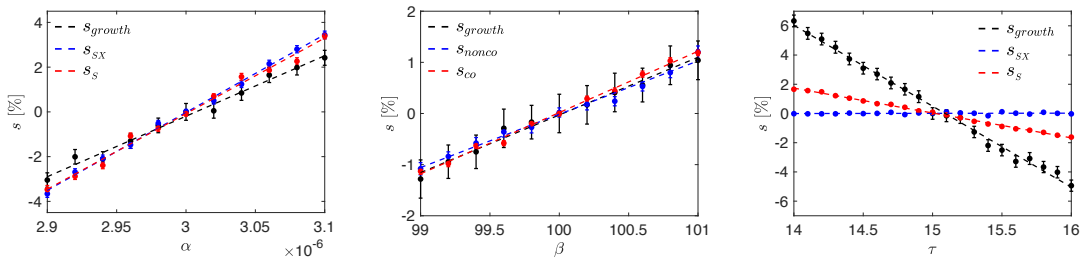


Fig. 5.11 The selective advantage s relative to a resident phage that results from a change to adsorption rate α, burst size β and lysis time τ. This is measured both in terms of the effect on the isolated growth rate of the mutant (s_{growth}, Eq. 5.17), and in terms of the change in frequency in a population initiated with 50% mutant and 50% resident (s_{SX} and s_S, Eq. 5.18). Resident parameters used were $\alpha = 3 \times 10^{-6}$, $\beta = 100$ and $\tau = 15$. As before $\delta = 0.1$ and $B_0 = 1000$. s_{growth} determined from 500 simulations, and s_{comp} determined from 200 simulations. Error bars are given by the standard error on the mean of the simulations.

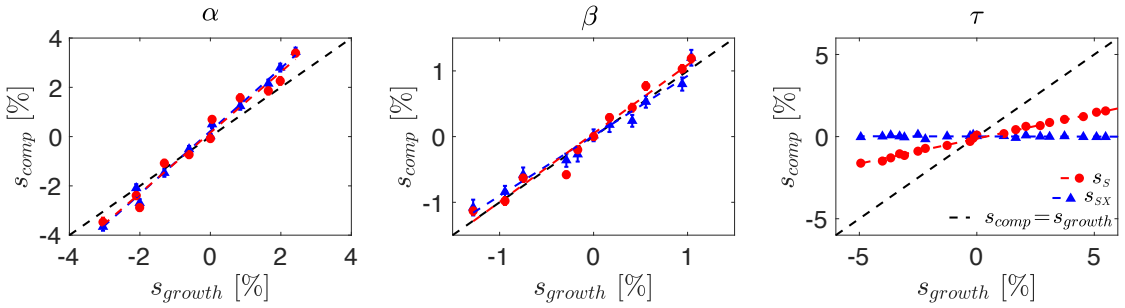


Fig. 5.12 The selective advantage in a competitive setting s_{comp} as a function of the change in growth rate s_{growth}, when changing adsorption rate α, burst size β and lysis time τ. Straight line fits are shown as dashed lines, with gradient σ such that $s_{comp} = \sigma s_{growth}$. From the above data we find $\sigma_{S\alpha} = 1.23(6)$, $\sigma_{SX\alpha} = 1.28(4)$, $\sigma_{S\beta} = 1.04(6)$, $\sigma_{SX\beta} = 0.91(5)$, $\sigma_{S\tau} = 0.306(6)$ and $\sigma_{SX\tau} \approx 0$. Resident parameters used were $\alpha = 3 \times 10^{-6}$, $\beta = 100$ and $\tau = 15$. As before $\delta = 0.1$ and $B_0 = 1000$. s_{growth} determined from 500 simulations, and s_{comp} determined from 200 simulations. Error bars are given by the standard error on the mean of the simulations. Error bars on x axis have been omitted for clarity, but are shown in Fig. 5.11.

variations in burst size affect similarly the growth rate of the viral population in isolation and its (dis)advantage in a competitive setting ($s_{growth} \approx s_{comp}$, Fig. 5.12), variations in adsorption rate lead to a stronger competitive (dis)advantage than what would be determined by the growth rate ($|s_{growth}| < |s_{comp}|$). The intuition behind this result is that increasing adsorption rate becomes particularly advantageous in a competitive environment, as being the *first* virus to infect a host allows the virus

to have largely (superinfection scenario) or completely (non superinfection scenario) exclusive access to the host resources.

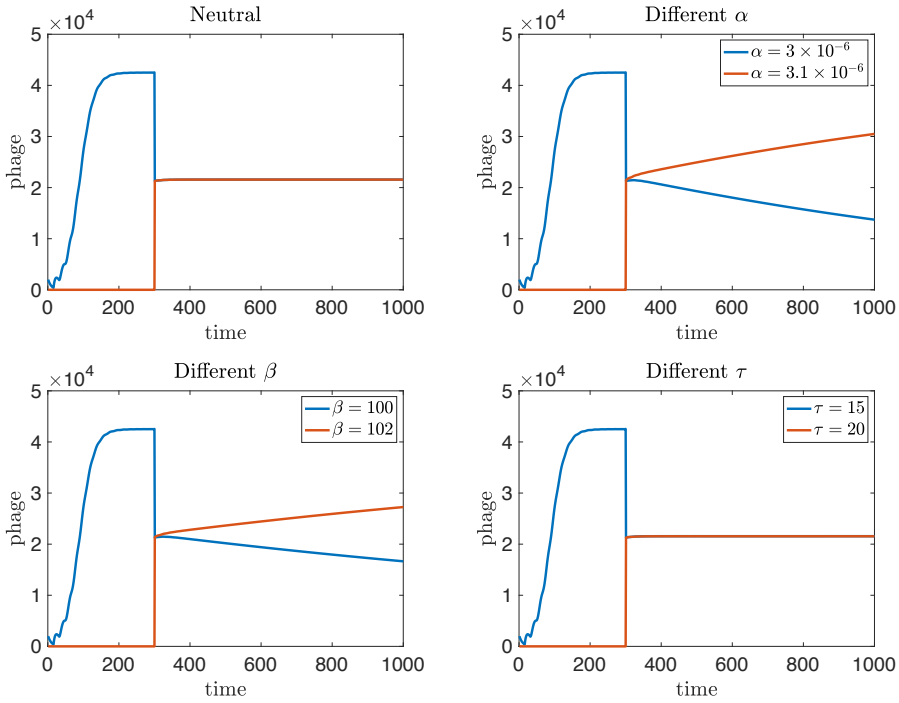


Fig. 5.13 The relative change in frequency of two populations in the DDE model (indicating the average behaviour in the stochastic model). It can be seen that once at steady-state, changing lysis time τ has no effect. Parameters used were $\alpha = 3 \times 10^{-6}$, $\beta = 100$ and $\tau = 15$ unless otherwise stated. As throughout, $\delta = 0.1$ and $B_0 = 1000$.

The impact of altering lysis time τ is surprising. 5.11 shows that increasing τ results in a reduced growth rate, as intuition suggests. Yet, in the superinfection-excluding scenario no discernible impact on s_{comp} is observed (Fig. 5.12). This result is supported by our DDE model (Fig. 5.13), which shows that once the system is at steady-state, alterations to lysis time offer no advantage to one phage over the other (Fig. 5.13). We believe that this is a special feature of a turbidostat setting, as lysed hosts are immediately replaced by uninfected cells, providing the same number of viable hosts independently of the time needed by the phage to lyse them. By contrast, in the superinfecting case, we are able to observe a selective (dis)advantage in direct competition, although at a significantly reduced level compared to the change in growth rate. We believe that this effect arises because, while the extracellular competition is limited by the turbidostat setup, in the superinfecting scenario there is the opportunity for some intracellular competition to occur, as mutants will grow at different rates inside

the host, resulting in different numbers of phage (both in total and proportionally) being released upon lysis.

5.3.6 Superinfection results in more efficient selection

Having characterised how changes to the phage infection parameters alter first genetic drift and second fitness, we now put both ingredients together and investigate the dynamics of non-neutral mutants. To this end, we simulate a resident phage population to steady-state, introduce a single non-neutral mutant and then run the simulation until extinction or fixation occurs.

In agreement with our observations regarding the difference between growth rate and competitive fitness, we find that the value of s_{growth} is not sufficient to determine the fixation probability of the corresponding mutant (Fig. 5.14): a mutation associated with a higher adsorption rate α increases the mutant's chance to fix more than a mutation which alters the burst size β and leads to the same growth rate. We also find that beneficial mutations are consistently more likely to fix (and deleterious mutations more likely to go extinct) in superinfecting populations (red) than superinfection-excluding populations (blue). This suggests that superinfection improves selection efficiency, by more readily fixing beneficial mutations and purging deleterious ones.

To provide a theoretical framework to our findings, we compare the simulation data to the fixation probabilities one would expect in a corresponding Moran model. For small selective advantage s_{comp}, the probability of fixation is given by

$$P_{fix} = \frac{1 - e^{-N_e s_{comp} f_0}}{1 - e^{-N_e s_{comp}}}, \tag{5.2}$$

where f_0 is the initial frequency of the mutant in the population with effective population size N_e [10, 176] - see Sec. 5.5.4 for derivation. Our earlier results on neutral dynamics and fitness provide independent measurements of the parameters in Eq. 5.2 for different values of α, β and τ: $f_0 = f_0^*$ from our initial condition (i.e., $1/N_T$, where N_T is the steady-state phage population size when the mutant is introduced); N_e is measured from the decay of heterozygosity (Fig. 5.5); and $s_{comp} = \sigma s_{growth}$ is derived from our measurements of the relationship between competitive and growth rate advantage (Fig. 5.12). These theoretical predictions are plotted without additional fitting parameters as lines in Fig. 5.14.

Fig. 5.14 shows that theoretical prediction from the appropriately parameterised Moran model matches the simulation data remarkably well, despite the complex internal

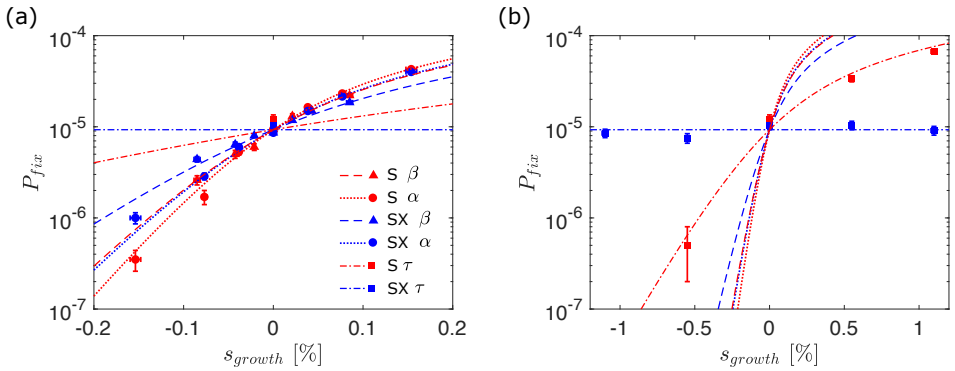


Fig. 5.14 Probability of mutant fixation P_{fix} as a function of selective growth advantage s_{growth}. Points indicate simulation results, while lines indicate theoretically predicted values in a Moran model with equivalent parameters (Eq. 5.2). Data points for the α and β mutants have been omitted from the right hand panel for clarity. The error in our estimate of the fixation probability ΔP_{fix} is given by $\Delta P_{fix} = \sqrt{n_{fix}}/n$, where n and n_{fix} represent the total number of simulations and the number of simulations where the mutant fixes respectively. Error bars in the x-axis represent the errors on the growth rate fitness s_{growth} that each burst size corresponds to. These are calculated by fitting a linear relation to growth rate measurements such that $s_{growth} = m(\beta_{mut} - \beta_{res})$. The fractional error on the s_{growth} is then equal to the fractional error on the fitted gradient m. The data is obtained from a minimum of 5 million independent simulations.

infection dynamic. If we introduce a free scaling parameter ϕ, such that $s_{comp} = \phi s_{growth}$, and optimally fit Eq. 5.2 to the data, we can compare the resulting values for ϕ with our previous estimates of σ to measure of the quality of agreement between simulations and theory. It should be noted that while we have described the scaling in terms of s_{comp}, it is mathematically identical to scaling N_e. Through this optimal fitting we find that $\phi_{S\alpha} = 1.22$, $\phi_{SX\alpha} = 1.34$, $\phi_{S\beta} = 1.03$, $\phi_{SX\beta} = 0.94$, $\phi_{S\tau} = 0.27$ and $\phi_{SX\tau} = 0.01$ with subscripts indicating scenario and parameter combinations. This indicates that the data maps well to an equivalently parameterised Moran model, with an average difference of $\sim 5\%$.

We note, however, that the simulation data consistently fails to intersect at the same point when $s_{growth} = 0$ in the superinfecting scenario. This is because of the effect outlined in Fig. 5.6, where rare mutants initially experience a slight advantage in the superinfecting scenario because they are able to increase in number by infecting both uninfected and infected cells.

5.3.7 $\beta_{res} = 70$ Measurements

To test the validity of our findings across parameter space, we also perform all of the above analysis with different resident parameters. Here we repeat a subset of the measurements carried out previously for different resident phage parameters, in this instance $\beta_{res} = 70$, with all other parameters remaining the same as in the rest of the Chapter. First, the effective population size is measured in both superinfecting and non superinfecting populations (Fig. 5.15), demonstrating that N_e is larger in superinfecting populations.

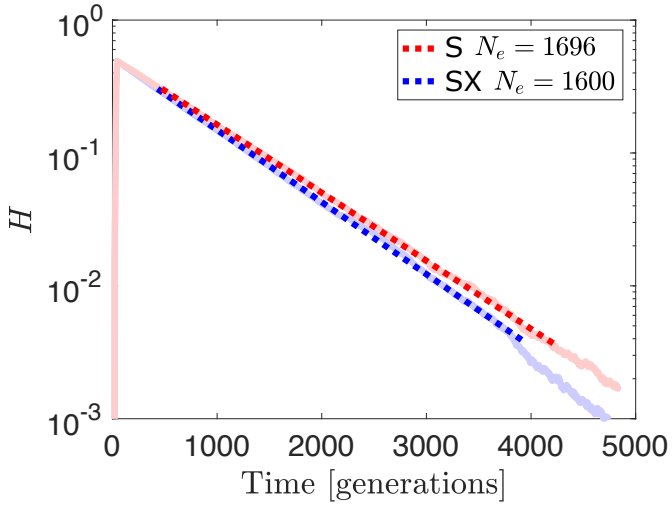


Fig. 5.15 Linear fit to log transformed heterozygosity data, with slope $\Lambda \equiv 2/N_e$ revealing that allowing superinfection (red) results in a larger effective population size compared to the case where superinfection is prevented (blue). Parameters used were $\alpha = 3 \times 10^{-6}$, $\beta = 70$, $\tau = 15$, $\delta = 0.1$ and $B_0 = 1000$.

We then move on to characterise the fitness of non-neutral mutants, in this instance only varying burst size β (Fig. 5.16). Again, we find a positive linear relationship between burst size and fitness, both in terms of the effect on growth rate in isolation and in a competitive setting. Interestingly here we find that alterations to burst size make slightly less difference in a competitive setting, as compared to the effect on growth rate. This could potentially be because, at lower burst sizes, any small change in β has a large impact on the growth rate, but has a smaller impact in a population already at steady-state.

Finally, we put both aspects together and measure the probability of fixation of non-neutral mutants in both superinfecting and superinfection-excluding populations (Fig. 5.17). As before, we find fairly good, although slightly worse, agreement between

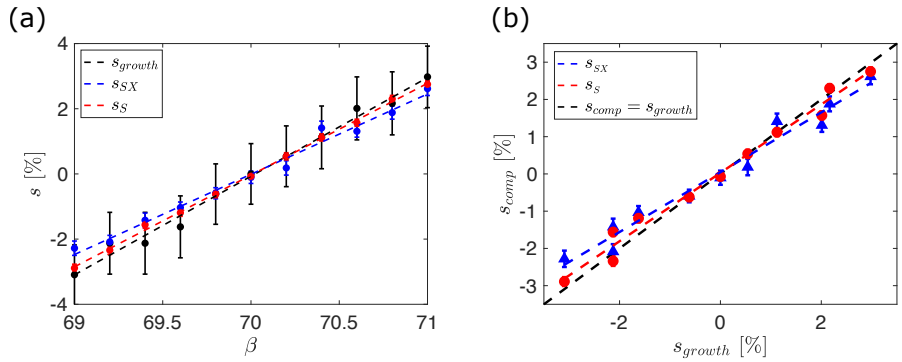


Fig. 5.16 (a): The selective advantage s relative to a resident phage that results from a change to burst size β. This is measured both in terms of the effect on the isolated growth rate of the mutant (s_{growth}, Eq. 5.17), and in terms of the change in frequency in a population initiated with 50% mutant and 50% resident (s_{SX} and s_S, Eq. 5.18). (b): The fitness in a competitive setting s_{comp} is then shown as a function of the fitness in an isolated setting s_{growth}. Straight line fits are shown as dashed lines, with gradient σ such that $s_{comp} = \sigma s_{growth}$. From the above data we find $\sigma_S = 0.92(4)$ and $\sigma_{SX} = 0.80(5)$. Resident parameters used were $\alpha = 3 \times 10^{-6}$, $\beta = 70$ and $\tau = 15$. As before $\delta = 0.1$ and $B_0 = 1000$. s_{growth} determined from 500 simulations, and s_{comp} determined from 200 simulations. Error bars are given by the standard error on the mean of the simulations.

our simulation results and the prediction from a Moran model with our independently measured parameters (Fig. 5.15 and Fig. 5.16). In terms of the additional fitting parameter introduced earlier, we find here that $\phi_S = 0.80$ and $\phi_{SX} = 0.73$. It's possible that this discrepancy is caused by imprecision in the measurements of fitness as a function of burst size. Indeed, over the whole range of β we would not expect a perfect linear relationship between burst size and fitness, with the benefits of increased burst size being larger for small β, and so at these lower values of β we find that the linear fit is less of a good approximation.

5.3.8 Superinfection exclusion slows down adaptability in the long run, but is a winning strategy in the short term

Our findings imply that, even in the absence of intra-cellular processes such as recombination, superinfection results in more efficient selection, so that beneficial mutations are relatively more likely to fix, and deleterious ones are more likely to be purged, leading to a fitter overall population in the long run. From the point of view of viral adaptation, allowing superinfection ultimately seems like the better long-term

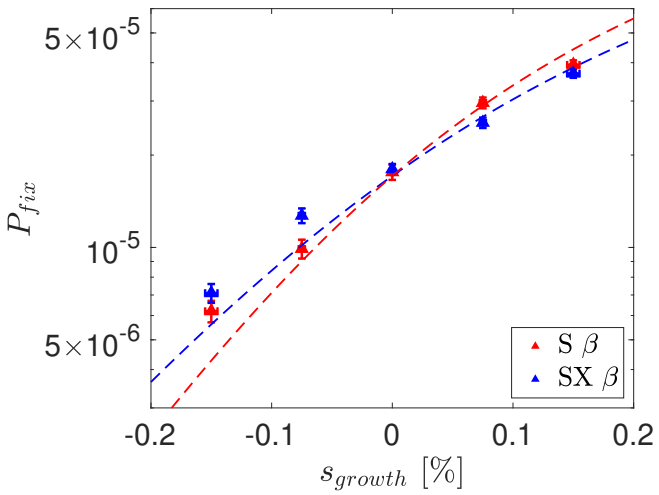


Fig. 5.17 Probability of mutant fixation P_{fix} as a function of selective growth advantage s_{growth}. Points indicate simulation results, while lines indicate theoretically predicted values in a Moran model with equivalent parameters (Eq. 5.2). The error in our estimate of the fixation probability ΔP_{fix} is given by $\Delta P_{fix} = \sqrt{n_{fix}}/n$, where n and n_{fix} represent the total number of simulations and the number of simulations where the mutant fixes respectively. Error bars in the x-axis represent the errors on the growth rate fitness s_{growth} that each burst size corresponds to. These are calculated by fitting a linear relation to growth rate measurements such that $s_{growth} = m(\beta_{mut} - \beta_{res})$. The fractional error on the s_{growth} is then equal to the fractional error on the fitted gradient m. The data is obtained from a minimum of 20 million independent simulations.

strategy. It is therefore puzzling why several natural phage populations have developed sophisticated mechanisms to prevent superinfection, particularly given that employing these mechanisms is expected to come with a biological cost, such as reduced burst size [128, 177] or increased lysis time [178].

To address this question, we consider the fate of mutations that either (i) remove the mutant's ability to prevent superinfection in a resident superinfection-excluding population or (ii) provide the mutant the ability to prevent superinfection in a resident superinfecting population. Fig. 5.18 shows that if the mutant is neutral ($\beta_{mut} = \beta_{res} = 100$), then the superinfection-excluding mutant is two orders of magnitude more likely to fix than the expectation based on its initial frequency f_0^*, and that, by contrast, the superinfecting mutant is at least two orders of magnitude more likely to go extinct. It should be noted that we actually find no instances of mutant fixation in this case, but our detection power is limited by the number of simulation runs. Here, we run at least 20 million simulations, and we can thus infer that $P_{fix} \ll 10^{-7}$. This indicates that

mutants which are able to prevent superinfection experience a very strong selective advantage over their superinfecting counterparts, and vice-versa.

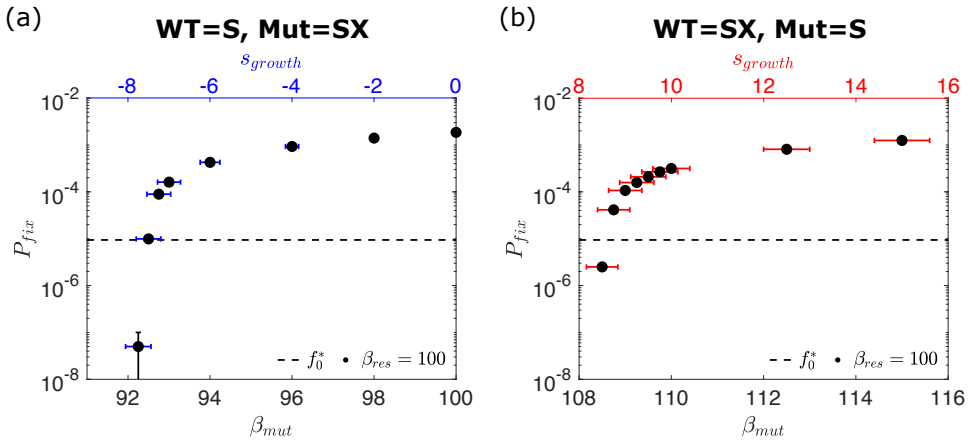


Fig. 5.18 (a) The probability P_{fix} of a mutant which prevents superinfection fixing in a population that allows it, as a function of mutant burst size β_{mut}. (b) The probability P_{fix} of a mutant which allows superinfection fixing in a population that prevents it, as a function of mutant burst size β_{mut}. It can be seen that the superinfecting mutant requires a significantly increased burst size to fix, and conversely the superinfection-excluding mutant can fix, even if its burst size is greatly reduced. The error in our estimate of the fixation probability ΔP_{fix} is given by $\Delta P_{fix} = \sqrt{n_{fix}}/n$, where n and n_{fix} represent the total number of simulations and the number of simulations where the mutant fixes respectively. Error bars in the x-axis represent the errors on the growth rate fitness s_{growth} that each burst size corresponds to. These are calculated by fitting a linear relation to growth rate measurements such that $s_{growth} = m(\beta_{mut} - \beta_{res})$. The fractional error on the s_{growth} is then equal to the fractional error on the fitted gradient m. The fixation data is obtained from a minimum of 20 million independent simulations.

To account for the possibility that superinfection exclusion comes at a cost in phage growth, as preventing superinfection likely requires the production of extra proteins, the resources for which could otherwise have gone to the production of more phage, we consider the case where superinfection exclusion is associated with a reduction in burst size [177]. Remarkably, we find that even when preventing superinfection carries a burden of 7% reduction in burst size ($s_{growth} < -7\%$), the non superinfecting mutant still fixes more often than a neutral superinfecting mutant (Fig. 5.18). Conversely, a minimum of 8% increase in burst size ($s_{growth} > 8\%$) is necessary to give a superinfecting mutant any chance of fixing in a superinfection excluding population. This indicates that while allowing superinfection increases selection efficiency at the population level,

preventing it is ultimately a winning strategy in the short term, partially explaining why superinfection exclusion is so common in nature [59, 78].

5.4 Discussion

In this work, we have considered the impact of either allowing or preventing superinfection on the evolution of viral populations. Using a stochastic agent-based model of viral infection, we have shown that allowing superinfection reduces the strength of genetic drift, leading to an increase in effective population size. Weaker fluctuations result in a higher efficiency of selection in viral populations, with beneficial mutations fixing more frequently, and deleterious ones more readily being purged from the population. Despite the long term, population-wide benefit of allowing superinfection, we find that if a mutant arises which is capable of preventing superinfection, it will fix remarkably easily, even if its growth rate is heavily compromised. Conversely, if the whole population is capable of preventing superinfection, mutants which allow it will have almost no chance of ever succeeding.

The evolutionary impact of superinfection (and more generally multiple infections) has most often focused on the role of intracellular interactions and competition [91–94, 96–99], such as genetic recombination and reassortment [100–103], and viral complementation [103–107]. A prevalent finding (amongst others) is that recombination and reassortment can improve the efficiency of selection in viral populations which do not exclude superinfection. Remarkably, our work demonstrates that the basic occurrence of superinfection alone, absent of any recombination or reassortment, is capable of increasing the selection efficiency. In this context, our results provide a useful baseline for comparison when trying to assess the significance of each of these more complex effects, which may or may not be present in different situations.

An unexpected finding of this work is that in the turbidostat system we consider, while increased adsorption rate and burst size both increase the fitness of the phage population in all respects, in the superinfecting scenario lysis time plays a significantly reduced role in the competitive (dis)advantage experienced once the system has reached a steady-state, and in the superinfection-excluding scenario it plays no role whatsoever. While it has been demonstrated previously that changes to fecundity and generation time can have different impacts on mutation fixation probability, even when they have the same impact on long-term growth rate [179], our result is somewhat in contrast with previous studies showing that well-mixed liquid cultures with an abundance of

hosts generally select for higher adsorption rates and lower lysis times [49, 51, 84, 180]. The key difference between such liquid cultures and the turbidostat system we model here is that in the former host cells are not maintained at a constant density, but the phage population continues to grow until no bacteria remain.

This finding illustrates how the presence or absence of a co-existing steady-state between phage and bacteria completely alters the selective pressure on the phage with important implications for studies into the co-evolution of phage and bacteria populations using continuous culturing set-ups [181–183]. In particular, our results suggest that in an evolutionary experiment in a turbidostat, the virus should evolve towards very large burst size even if this feature comes at the cost of longer lysis times, especially if superinfection exclusion occurs [49]. Reciprocally, detecting a selective pressure on lysis time could be used to identify potential phages that allow superinfection, as, in this case, a shorter lysis time is slightly advantageous all else being equal.

As noted previously, however, our model does not fully capture all of the aspects of a true turbidostat. In a true turbidostat, unless there is a filter over the outflow, bacteria can also be diluted out of the system. I expect that this could have significant consequences for the impact of lysis time, since phage with a long lysis time would be more likely to be diluted out while still inside an infected host, thereby reinstating the advantage of short lysis times. On the other hand, in a true turbidostat, dilutions typically occur as periodic and discrete events, rather than at a continuous rate. I expect that the frequency of such events, and so the degree to which there is a continuous steady-state, would have a significant impact on the overall dynamics.

Following this work, it is natural to wonder how the (dis)advantages and impact of either strategy depends on the selective pressure experienced in different environments. The relationship between viral fitness and the phage life-history parameters (adsorption rate, lysis time and burst size) has been shown to be very context-dependent in both well-mixed and spatially structured settings. For instance, as noted previously, well-mixed settings generally favour higher adsorption rates [184], but in spatially structured settings phage with lower adsorption rates are more successful [89, 90]. Additionally, it has been shown previously that eco-evolutionary feedbacks at the edge of expanding viral populations can result in travelling waves with vastly different effective population sizes [171]. Given that competition for resources (i.e. viable hosts) in spatially structured environments is local rather than global, phage are more likely to be in competition with other genetically identical phage released by nearby cells.

It is therefore possible that superinfection exclusion proves less useful in this context than in well-mixed environments where competition is global and phage are more likely to encounter other genetically different viruses. All of this points at the role of superinfection strategies and other social viral behaviour on the eco-evolutionary dynamics of spatially expanding viral populations as an exciting avenue for future research.

5.5 Methods

5.5.1 Measuring effective population size of the phage population

Consistent with previous work [10], we expect that the genetic diversity of the phage population, quantified by the heterozygosity H, will decay exponentially at long times due to genetic drift, so that $H(t) \propto e^{-2t/N_e}$, with the decay rate in units of inverse generations being expressed in terms of an effective population size $2/N_e$ (Moran model [10]).

We track the viral heterozygosity H as a function of time, which in a biallelic viral population is given by

$$H = 2\langle f(1-f)\rangle, \tag{5.3}$$

where f and $(1-f)$ represent the frequencies of two neutral viral alleles in the population, and $\langle \dots \rangle$ indicates the average over independent simulations. $H(t)$ can be understood to be the time-dependent probability that two individuals chosen from the population are genetically distinct.

To determine the generation time T, we borrow from demographic analysis, and first calculate the *net reproduction rate* R_0, which represents the number of offspring an individual would be expected to produce if it passed through its lifetime conforming to the age-specific fertility and mortality rates of the population at a given time (i.e. taking into account the fact that some individuals die before reproducing) [185].[2] R_0 can be calculated as

$$R_0 = \sum l_t m_t, \tag{5.4}$$

[2]In demography, the net reproduction rate R_0 applies specifically to the offspring born to *females*, rather than all individuals. Given that all phage are capable of producing offspring, however, we adapt the approach here to include all phage. I also note that the demographic concept of the net reproduction rate R_0, used here, should not be confused with the epidemiological concept of the basic reproduction number, also given the symbol R_0. Finally, I note that despite being named the net reproduction rate, I do not believe it is a rate.

where l_t represents the proportion of individuals (in our case, phage) surviving to age t, and m_t represents the average number of offspring produced at age t.

There are two mechanisms in our simulations by which phages can 'die' when superinfection exclusion applies: either by decaying with rate δ, or by adsorbing to a host with rate αB_0. In a sufficiently small timestep Δt, these rates correspond to a proportion $\delta\Delta t$ and $\alpha B_0 \Delta t$ of the total phage, respectively. Equivalently, these can be considered to be the probability that any single phage will die in the same period. As a result, the probability of a phage surviving to age t is $l_t = (1 - \delta\Delta t - \alpha B_0 \Delta t)^{t/\Delta t}$.

The average number of offspring m_t produced at age t is 0 if $t < \tau$, because we assume that no phage is released before the lysis time. For $t > \tau$, m_t is given by the probability of successfully infecting a viable host in a timestep Δt, τ time earlier ($\alpha B_{ss} \Delta t$), multiplied by the yield of new phage ($\beta - 1$).

In the limit where $\Delta t \to 0$, this will result in a net reproductive rate of the form

$$R_0 = \lim_{\Delta t \to 0} \sum_{t=0}^{\infty} m_t l_t = \lim_{\Delta t \to 0} \sum_{t=\tau}^{\infty} \Delta t \alpha B_{ss} (\beta - 1)(1 - \Delta t(\delta + \alpha B_0))^{(t-\tau)/\Delta t}, \tag{5.5}$$

$$= \int_{t=\tau}^{\infty} \phi e^{\theta\tau} e^{-\theta t} \mathrm{d}t, \tag{5.6}$$

$$= \phi e^{\theta\tau} \left[-\frac{1}{\theta} e^{-\theta t} \right]_{\tau}^{\infty}, \tag{5.7}$$

$$= \phi e^{\theta\tau} \left(-\frac{1}{\theta e^{\infty}} + \frac{1}{\theta e^{\theta\tau}} \right), \tag{5.8}$$

$$= \frac{\alpha B_{ss}(\beta - 1)}{\delta + \alpha B_0}, \tag{5.9}$$

where the integral starts at τ because no offspring are produced prior to that point.

Then the generation time T, defined as the average interval between the birth of an individual and the birth of its offspring, is

$$T = \lim_{\Delta t \to 0} \frac{\sum t l_t m_t}{R_0} = \frac{1}{R_0} \int_{t=\tau}^{\infty} t \alpha B_{ss} (\beta - 1) e^{-(\delta + \alpha B_0)(t-\tau)} \mathrm{d}t, \tag{5.10}$$

$$= \frac{1}{R_0} \int_{t=\tau}^{\infty} t \phi e^{\theta\tau} e^{-\theta t} \mathrm{d}t, \tag{5.11}$$

$$= \frac{1}{R_0} \left(\phi e^{\theta\tau} \left[t . \frac{1}{\theta} e^{-\theta t} \right]_{\tau}^{\infty} + \frac{\phi e^{\theta\tau}}{\theta} \int_{t=\tau}^{\infty} e^{-\theta t} \mathrm{d}t \right), \tag{5.12}$$

$$= \frac{1}{R_0} \left(\frac{\phi e^{\theta\tau}}{\theta} \tau e^{-\theta\tau} + \frac{R_0}{\theta} \right), \tag{5.13}$$

$$= \frac{1}{R_0}\left(\tau R_0 + \frac{R_0}{\theta}\right), \tag{5.14}$$

$$= \tau + \frac{1}{\theta}, \tag{5.15}$$

$$= \tau + \frac{1}{\delta + \alpha B_0}. \tag{5.16}$$

Here, we will use resident phage parameters $\alpha = 3 \times 10^{-6}$, $\tau = 15$, $\delta = 0.1$ and a total bacterial population of $B_0 = 1000$, which leads to a generation time of $T = 24.8$. Throughout this work, we use the same generation time for both superinfecting and superinfection-excluding populations (more details shortly).

For comparison, coliphage T7 in liquid culture typically has parameters of $\hat{\tau} \approx 10 - 20$ min and $\hat{\alpha} \approx 10^{-9}$ ml/min, while $\hat{B_0} \approx 10^6 - 10^8$ ml^{-1}, thereby yielding an $\hat{\alpha}\hat{B_0} \approx 10^{-3} - 10^{-1}$ min^{-1} [49, 62]. These values are comparable to our own if we equate 1 timestep = 1 min, and so $\tau = 15 \equiv 15$ min and $\alpha B_0 = 3 \times 10^{-3} \equiv 3 \times 10^{-3}$ min^{-1}, such that the relative timescales in our simulation remain consistent. The reason behind choosing a larger adsorption rate and smaller bacteria population is purely practical, as the alternative would lead to unreasonably long computational times. Given these values, our choice of decay rate δ is made such that steady-state population sizes are reached.

This generation time is also supported by stochastic simulations of the phage adsorption and death processes. We simulate a single phage, which in each time-step Δt has a probability of successfully adsorbing to an uninfected host ($\alpha B_{ss}\Delta t$), and a probability of *dying* ($(\delta + \alpha I_{ss})\Delta t$). In the event that the phage successfully adsorbs to a host before it dies, the number of steps t_{steps} taken for this to occur is noted, and the time $T = \tau + t_{steps}\Delta t$ is recorded (representing the time between the 'birth' of the original phage and the 'birth' of its offspring). A schematic representation of this process is shown in Fig. 5.19. This process was repeated 10 million times, with the time T being recorded in all of the instances where the phage successfully reproduced. This yields an average generation time of $\overline{T} = 24.78(3)$ in agreement with the analytical calculation.

Additionally, throughout this work we use the above generation time for both superinfection-excluding and superinfecting populations. However, the superinfection scenario differs from that laid out above in that adsorption to infected cells does not result in death, and relatedly, the time between successful host infection and offspring production may be less than τ. To evaluate the error we introduce with

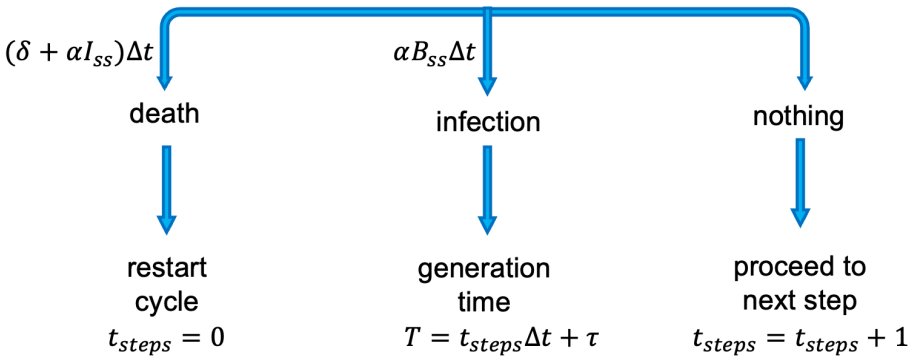


Fig. 5.19 Schematic diagram illustrating the processes used to verify the generation time T in superinfection-excluding populations.

our approximation of generation time, we modify the simple stochastic simulations above to take into account these differences. In this scenario, in a single time-step the phage has a probability $\delta\Delta t$ of dying, and a probability $\alpha B_0\Delta t$ of infecting a host (either infected or uninfected). In the case where infection occurs, the phage has a probability B_{ss}/B_0 of infecting an uninfected host, which as before, results in a generation time $T = \tau + t_{steps}\Delta t$. In the remainder of infection cases, phage will infect already infected hosts. Because of the nature of the process of within-host replication, secondary infections that occur too late after the initial infection generate almost no offspring of the superinfecting phage. We account for this observation by assuming that only secondary infections occurring within the first n steps post initial infection will successfully produce offspring[3]. Given that we are considering populations at steady-state, we assume that infected cells are equally likely to be found any number of steps post-infection ($< \tau$), and so infection of a cell in the first n steps post initial infection simply occurs a fraction n/τ of the times that secondary infection occurs. In this case, the generation time is given by $T = t_{steps}\Delta t + \tau - \Delta t\frac{\sum_1^n k}{n}$, where the final two terms represent the average number of steps between secondary infection and lysis. This final term can be simplified by noting that $\frac{\sum_1^n k}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$. A schematic representation of this process is shown in Fig. 5.20.

Using an example value of $n = 3$, this process was again repeated 10 million times, yielding an average generation time of $\overline{T} = 24.11(3)$. It can be seen that the difference in generation time is marginal. If we instead take a value of $n = \tau - 1$ we find an average generation time of $\overline{T} = 19.7(2)$. While this neglects the fact that late superinfections

[3]This is, in itself, also a simplification of the process, as the precise number of offspring produced depends not only on *when* superinfection occurs, but also on how many phage of the other type have infected, and how many of each type infect in the future.

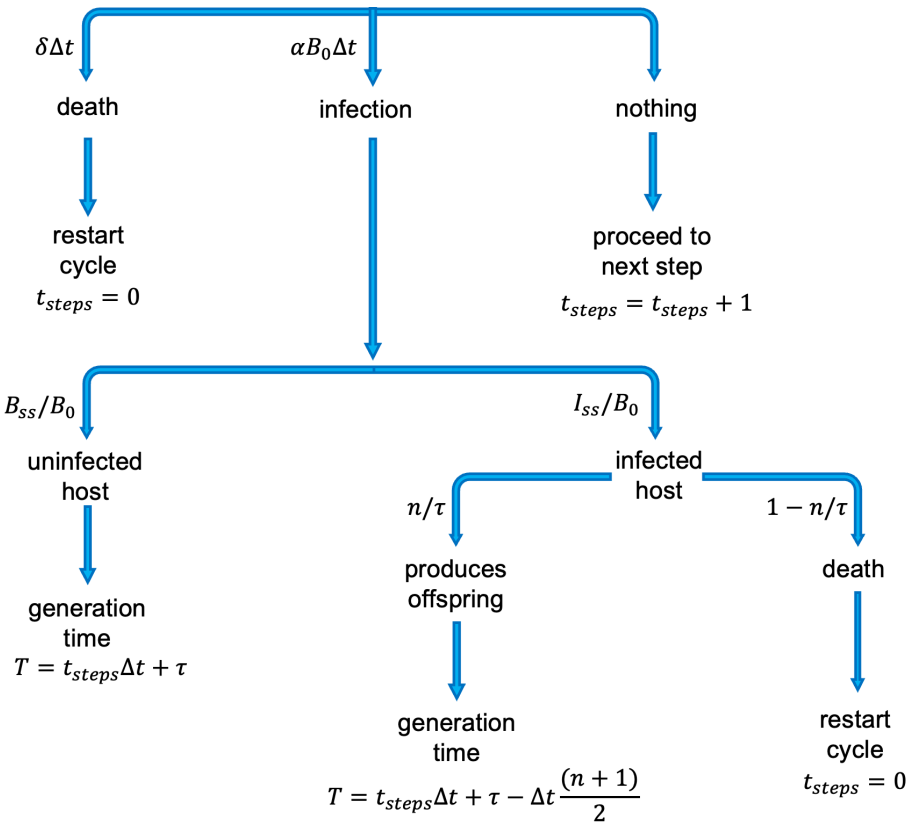


Fig. 5.20 Schematic diagram illustrating the processes used to measure the generation time T in superinfecting populations.

are unlikely to yield any offspring, and so represents a significant underestimation of the generation time, it is worth noting that were we to fully account for the shorter generation time in superinfecting populations, it would result in an even larger effective population size, further emphasising our main findings.

5.5.2 Measuring mutant fitness and growth rate

We start by defining a selective advantage s_{growth} in terms of the exponential growth rate r_{mut} of the mutant phage population relative to that of the resident phage r_{res} [9]:

$$s_{growth} = \frac{r_{mut}}{r_{res}} - 1. \tag{5.17}$$

The exponential growth rate is determined by simulating the growth of the corresponding phage population in isolation, and performing a linear fit to the log-transformed phage number as a function of time, which is then averaged over 500 independent simu-

lations. It should be noted that as there is only one type of phage in these simulations, the growth rate of both superinfecting and superinfection-excluding populations is the same.

We also characterised the fitness of mutants in a competitive setting, by simulating a resident population until steady-state, and then replacing 50% of the population with the mutant. In this direct competition scenario, we determine the selective (dis)advantage s_{comp} of the mutant phage by tracking the relative growth of mutant and resident populations, so that

$$\frac{V_{mut}}{V_{res}} = \frac{V_{mut}(t=0)e^{r_{res}(1+s_{comp})t}}{V_{res}(t=0)e^{r_{res}t}} = e^{r_{res}s_{comp}t}, \tag{5.18}$$

as $V_{mut}(t=0) = V_{res}(t=0)$. s_{comp} is determined from the average of 200 simulations. Importantly, in contrast to s_{growth}, this competitive selective advantage (s_{comp}) can in principle differ between superinfecting (s_S) and superinfection-excluding (s_{SX}) phage populations. In the absence of any interactions between the two competing phage populations, s_{growth} and s_{comp} are typically expected to be the same.

5.5.3 Measuring mutant probability of fixation

To measure fixation probabilities of individual mutations, we allow our simulations to reach steady-state, we then introduce a single mutant phage into the free phage population, and run the simulation until either fixation or extinction occurs. This process is repeated at least 14 million times for each set of parameters. The probability of mutant fixation P_{fix} is determined from the fraction of simulations where the mutant fixed, n_{fix}, over the total number of simulations run, n (i.e. $P_{fix} = n_{fix}/n$). Assuming a binomial distribution, the error in our estimate of the number of fixation events Δn_{fix} is given by $\Delta n_{fix} = \sqrt{nP_{fix}(1-P_{fix})}$. Consequently, our error in the estimate of fixation probability ΔP_{fix} is given by $\Delta P_{fix} = \sqrt{P_{fix}(1-P_{fix})/n}$. It can be easily verified that in the case where $n_{fix} \ll n$, as we have here, the error approaches $\Delta P_{fix} = \sqrt{n_{fix}}/n$ as would be found in a Poisson distribution.

5.5.4 Probability of Fixation in Moran Model

The Moran model describes a population of fixed size N in which two alleles A and B compete for dominance of the population [10]. In each step, a random individual is chosen for reproduction, and a random individual is chosen for death, thus ensuring a

constant population size. At any given time-step, let i denote the number of individuals of type A in the population. Since the number of A individuals can at most change by 1 in a single time-step, only transitions between state i and $i-1$, and i and $i+1$ are possible. In the neutral case, this leads to transition probabilities of the form:

$$P_{i,i-1} = \underbrace{\frac{N-i}{N}}_{\text{reproduce}} \cdot \underbrace{\frac{i}{N}}_{\text{die}}, \tag{5.19a}$$

$$P_{i,i} = 1 - P_{i,i-1} - P_{i,i+1}, \tag{5.19b}$$

$$P_{i,i+1} = \underbrace{\frac{i}{N}}_{\text{reproduce}} \cdot \underbrace{\frac{N-i}{N}}_{\text{die}}, \tag{5.19c}$$

where $P_{i,j}$ represents the probability of transitioning from state i to state j.

We can then consider the non-neutral case, where type A has a fitness advantage s over type B. To align with the bulk of this Chapter, we will assume that this fitness increases the chance of being chosen to reproduce, without having an effect on the individual's chance of dying. In this case the above transition probabilities become:

$$P_{i,i-1} = \underbrace{\frac{N-i}{(1+s)i + (N-i)}}_{\text{reproduce}} \cdot \underbrace{\frac{i}{N}}_{\text{die}}, \tag{5.20a}$$

$$P_{i,i} = 1 - P_{i,i-1} - P_{i,i+1}, \tag{5.20b}$$

$$P_{i,i+1} = \underbrace{\frac{(1+s)i}{(1+s)i + (N-i)}}_{\text{reproduce}} \cdot \underbrace{\frac{N-i}{N}}_{\text{die}}. \tag{5.20c}$$

The probability of fixation x_i when starting in state i can then be defined by recurrence:

$$x_i = \begin{cases} 0 & \text{when } i = 0 \\ P_{i,i-1}x_{i-1} + (1 - P_{i,i-1} - P_{i,i+1})x_i + P_{i,i+1}x_{i+1} & \text{when } 1 \leq i \leq N-1 \\ 1 & \text{when } i = N \end{cases} \tag{5.21}$$

This can then be rearranged as follows:

$$x_i = P_{i,i-1}x_{i-1} + (1 - P_{i,i-1} - P_{i,i+1})x_i + P_{i,i+1}x_{i+1}, \tag{5.22}$$

$$P_{i,i-1}(x_i - x_{i-1}) = P_{i,i+1}(x_{i+1} - x_i), \tag{5.23}$$

$$\gamma_i y_i = y_{i+1}, \tag{5.24}$$

where we have defined a new variables $y_i = x_i - x_{i-1}$ and $\gamma_i = P_{i,i-1}/P_{i,i+1}$.

Now we must consider two properties of our variable y_i. The first is that

$$\sum_{i=1}^{m} y_i = \sum_{i=1}^{m}(x_i - x_{i-1}), \tag{5.25}$$

$$= (x_1 - x_0) + (x_2 - x_1) + \cdots + (x_{m-1} - x_{m-2}) + (x_m - x_{m-1}), \tag{5.26}$$

$$= x_m. \tag{5.27}$$

The second property is that

$$y_k = x_1 . \prod_{l=1}^{k-1} \gamma_l. \tag{5.28}$$

This can be verified by considering

$$\prod_{l=1}^{k-1} \gamma_l = \prod_{l=1}^{k-1} \frac{y_{i+1}}{y_i}, \tag{5.29}$$

$$= \frac{y_2}{y_1} . \frac{y_3}{y_2} . \cdots . \frac{y_{k-1}}{y_{k-2}} . \frac{y_k}{y_{k-1}}, \tag{5.30}$$

$$= \frac{y_k}{y_1}. \tag{5.31}$$

Therefore

$$x_1 . \prod_{l=1}^{k-1} \gamma_l = x_1 . \frac{y_k}{y_1}, \tag{5.32}$$

$$= x_1 . \frac{x_k - x_{k-1}}{x_1 - x_0}, \tag{5.33}$$

$$= x_k - x_{k-1}, \tag{5.34}$$

$$= y_k. \tag{5.35}$$

Now using Eq. 5.27 and Eq. 5.28 we can write that

$$\sum_{i=1}^{m} y_i = x_m = \sum_{i=1}^{m} x_1 \prod_{l=1}^{i-1} \gamma_l, \tag{5.36}$$

$$= x_1 + x_1 \sum_{j=1}^{m-1} \prod_{k=1}^{j} \gamma_k, \tag{5.37}$$

where we have used the fact that the first term of the sum ($i = 1$) equals x_1, as $\prod_{l=1}^{0} \gamma_l = 1$. If we now combine this with the fact that $x_N = 1$ we find that

$$x_N = x_1 \left(1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} \gamma_k \right) = 1, \tag{5.38}$$

$$x_1 = \left(1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} \gamma_k \right)^{-1}, \tag{5.39}$$

and so

$$x_i = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^{j} \gamma_k}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} \gamma_k}. \tag{5.40}$$

It can be easily verified that in our system

$$\gamma_k = \frac{P_{i,i-1}}{P_{i,i+1}} = \frac{1}{1+s}, \tag{5.41}$$

and so if we combine this with Eq. 5.40 we find that

$$x_i = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^{j} \gamma_k}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} \gamma_k}, \tag{5.42}$$

$$= \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^{j} (\frac{1}{1+s})}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} (\frac{1}{1+s})}, \tag{5.43}$$

$$= \frac{1 + \sum_{j=1}^{i-1} (\frac{1}{1+s})^j}{1 + \sum_{j=1}^{N-1} (\frac{1}{1+s})^j}, \tag{5.44}$$

$$= \frac{1 + \frac{1-(\frac{1}{1+s})^i}{1-(\frac{1}{1+s})} - 1}{1 + \frac{1-(\frac{1}{1+s})^N}{1-(\frac{1}{1+s})} - 1}, \tag{5.45}$$

$$= \frac{1 - (1+s)^{-i}}{1 - (1+s)^{-N}}. \tag{5.46}$$

where we have used the properties of geometric series to perform the sum. This derivation is similar to that presented in many textbooks, for instance [186].

Finally, if s is small, $e^s \approx (1+s)$, and so we can restate this expression as

$$x = \frac{1 - e^{-sNf_0}}{1 - e^{-sN}}, \tag{5.47}$$

where we have rewritten in terms of the initial frequency $i = Nf_0$. This is the expression used in this Chapter for the probability of fixation in the Moran model.

Appendix 5.A Alternative model

Here we present results from an alternative version of the model to that outlined in the main body of this Chapter. The key difference in this version of the model relates to how the pseudo-populations of phage grow within the host prior to lysis. In this version of the model, in both superinfecting and superinfection-excluding scenarios, the lysis time τ *and* the final burst size β of the host are set by the first phage to infect it. In the superinfecting scenario, this means that during the intermediate steps between the first adsorption event and lysis, the total number of phage inside the host increases at a constant rate β/τ, determined by the initial infecting phage. As before, these β/τ phages are allocated proportionally to each phage type inside the host at that time, in line with previous reports of a positive linear relationship between lysis time and burst size [51], and to reflect the intracellular competition for the host's resources. This version of the model therefore limits the ability for selection to act on mutants with different burst sizes and lysis times, as these differences will not impact their growth rate inside the cell post-superinfection, as this rate is set by the initial infecting phage.

It should also be noted that this version of the model does not impact neutral dynamics, or any dynamics in the superinfection-excluding scenario. It only impacts non-neutral, superinfecting dynamics. We therefore present the relevant results below.

In the superinfecting scenario, alterations to burst size now make marginally less difference in a competitive setting compared to the main model (Fig. 5.21 and Fig. 5.22). This is perhaps not surprising, as we have removed the ability for selective (dis)advantages to manifest inside the host, while maintaining the (dis)advantage gained when the mutant is the first to infect a cell. This effect is more striking however when we consider alterations to lysis time. Now, also in the superinfecting scenario alterations to lysis time have almost no impact in a competitive setting, once the population has reached steady-state (Fig. 5.21 and Fig. 5.22).

These differences between the models are also reflected in the fixation of single non-neutral mutants (Fig. 5.23). Here, we find that selection of superinfecting mutants is slightly less efficient when alterations are made to burst size, and significantly less efficient when alterations are made to lysis time, when compared to the main model. Crucially however, even in this alternative model where the power of selection is limited, selection is still always more efficient in superinfecting populations than in superinfection-excluding populations. We also find that the results of our alternative model can still be mapped well to an equivalently parameterised Moran model.

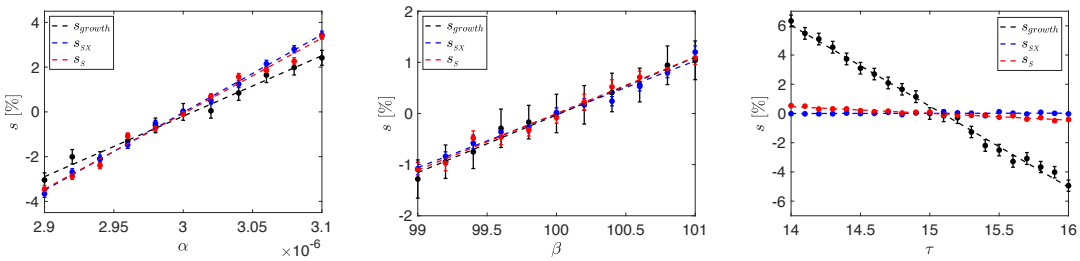


Fig. 5.21 The selective advantage s relative to a resident phage that results from a change to adsorption rate α, burst size β and lysis time τ. This is measured both in terms of the effect on the isolated growth rate of the mutant (s_{growth}, Eq. 7), and in terms of the change in frequency in a population initiated with 50% mutant and 50% resident (s_{SX} and s_S, Eq. 9). Resident parameters used were $\alpha = 3 \times 10^{-6}$, $\beta = 100$ and $\tau = 15$. As before $\delta = 0.1$ and $B_0 = 1000$.

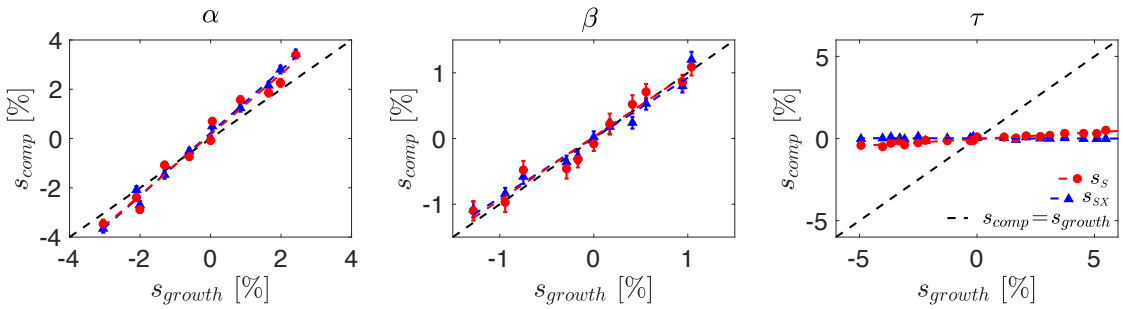


Fig. 5.22 The selective advantage in a competitive setting s_{comp} as a function of the change in growth rate s_{growth}, when changing adsorption rate α, burst size β and lysis time τ. Straight line fits are shown as dashed lines, with gradient σ such that $s_{comp} = \sigma s_{growth}$. Resident parameters used were $\alpha = 3 \times 10^{-6}$, $\beta = 100$ and $\tau = 15$. As before $\delta = 0.1$ and $B_0 = 1000$. s_{growth} determined from 500 simulations, and s_{comp} determined from 200 simulations. Error bars on x axis have been omitted for clarity.

Finally, when we consider mutations which change a mutant's ability to prevent superinfection, our results remain qualitatively the same (Fig. 5.18), although there are slight quantitative differences, with superinfection-excluding mutants able to sustain a slightly more reduced burst size without becoming deleterious (-8% vs -7% in the main model).

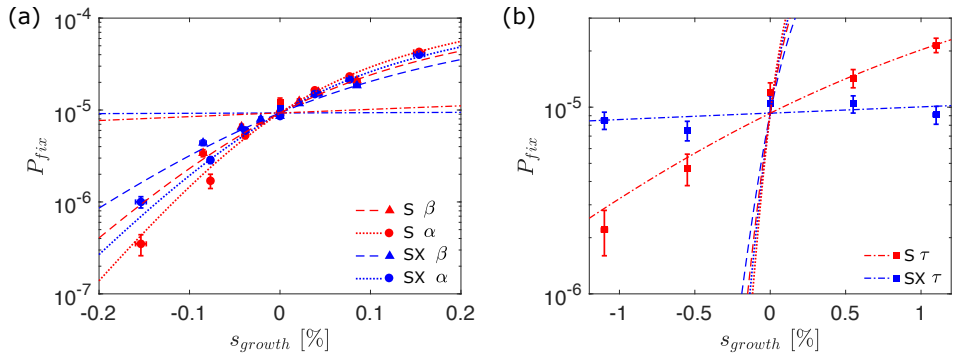


Fig. 5.23 Probability of mutant fixation P_{fix} as a function of selective growth advantage s_{growth}. Points indicate simulation results, while lines indicate theoretically predicted values in a Moran model with equivalent parameters (Eq. 5.2). This is with the exception of the superinfection-excluding lysis time (SX τ) line, which shows an optimised fit to the Moran model, with $s_{moran} = \sigma s_{growth}$, as we were unable to accurately measure the relationship between s_{growth} and s_{comp} in this case (see Fig. 5.12). We find $\sigma_{SX\tau} = 0.008$. Data points for the α and β mutants have been omitted from the right hand panel for clarity. The data is obtained from a minimum of 5 million independent simulations.

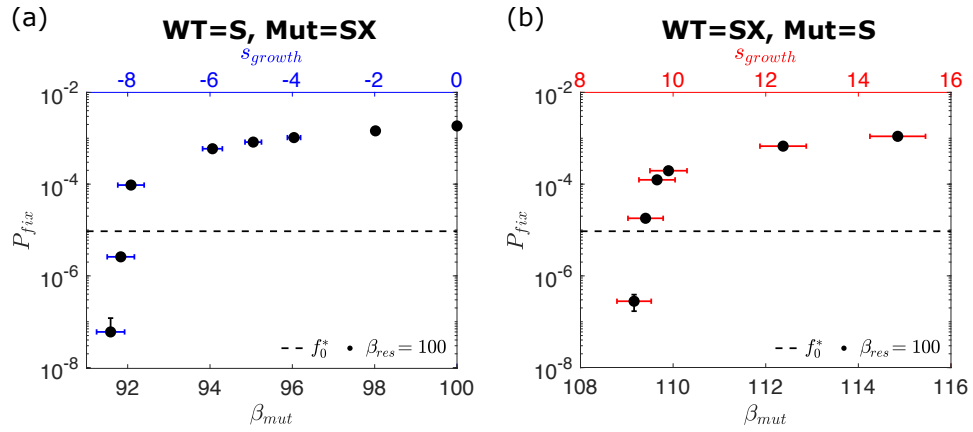


Fig. 5.24 (a) The probability P_{fix} of a mutant which prevents superinfection fixing in a population that allows it, as a function of mutant burst size β_{mut}. (b) The probability P_{fix} of a mutant which allows superinfection fixing in a population that prevents it, as a function of mutant burst size β_{mut}. It can be seen that the superinfecting mutant requires a significantly increased burst size to fix, and conversely the superinfection-excluding mutant can fix, even if its burst size is greatly reduced. Error bars in the x-axis represent the errors on the growth rate fitness s_{growth} that each burst size corresponds to. These are calculated by fitting a linear relation to growth rate measurements such that $s_{growth} = m(\beta_{mut} - \beta_{res})$. The fractional error on the s_{growth} is then equal to the fractional error on the fitted gradient m. The fixation data is obtained from a minimum of 20 million independent simulations.

Appendix 5.B Mathematical Analysis of DDE Model

While not directly relevant for the main focus of this Chapter, I present here a brief mathematical analysis of our delay differential equation (DDE) model, which is reproduced below:

$$\frac{dV}{dt} = -\alpha V(B+I) - \delta V + \beta\alpha V_{t-\tau}B_{t-\tau}, \tag{5.48a}$$

$$\frac{dB}{dt} = -\alpha VB + \alpha V_{t-\tau}B_{t-\tau}, \tag{5.48b}$$

$$\frac{dI}{dt} = \alpha VB - \alpha V_{t-\tau}B_{t-\tau}, \tag{5.48c}$$

All of the symbols are defined the same as before (V, B and I indicate the concentrations of phage, uninfected bacteria and infected bacteria as a function of time respectively; α, β, τ and δ indicate the phage adsorption rate, burst size, lysis time and decay rate respectively). The subscript is used to indicate that those terms are calculated at time $t-\tau$. The positive term in Eq. 5.48b accounts for the instantaneous replacement of lysed cells in our 'turbidostat' environment.

We start by noting that once the system has reached steady-state and remained there for a time $t > \tau$, the rate of change of the populations will equal zero (e.g., $\frac{dV}{dt} = 0$) and the populations will equal their steady-state values, both at time t and time $t-\tau$ (e.g., $V = V_{t-\tau} = V_{ss}$). If we focus on Eq. 5.48a, at steady-state this yields

$$\frac{dV}{dt} = 0 = -\alpha V_{ss}(B_{ss}+I_{ss}) - \delta V_{ss} + \beta\alpha V_{ss}B_{ss}, \tag{5.49}$$

$$0 = -\alpha(B_{ss}+I_{ss}) - \delta + \beta\alpha B_{ss}, \tag{5.50}$$

$$0 = -\alpha B_0 - \delta + \beta\alpha B_{ss}, \tag{5.51}$$

$$\beta\alpha B_{ss} = \alpha B_0 + \delta, \tag{5.52}$$

$$B_{ss} = \frac{\alpha B_0 + \delta}{\beta\alpha}, \tag{5.53}$$

where we have used the fact that $B_0 = B_{ss} + I_{ss}$. Using this same relation it can easily be shown that

$$I_{ss} = B_0 - B_{ss}, \tag{5.54}$$

$$= B_0 - \frac{\alpha B_0 + \delta}{\beta\alpha}, \tag{5.55}$$

$$= \frac{\beta\alpha B_0}{\beta\alpha} - \frac{\alpha B_0 + \delta}{\beta\alpha}, \tag{5.56}$$

$$= \frac{\alpha B_0(\beta - 1) - \delta}{\beta\alpha}. \tag{5.57}$$

For the resident phage parameters $\alpha = 3 \times 10^{-6}$, $\beta = 100$, $\tau = 15$, $\delta = 0.1$, with a total bacterial population of $B_0 = 1000$, this yields values of $B_{ss} = 343$ and $I_{ss} = 657$, in agreement with our numerical calculations (Fig. 5.25b). It is interesting to note that B_{ss} and I_{ss} are independent of the lysis time τ.

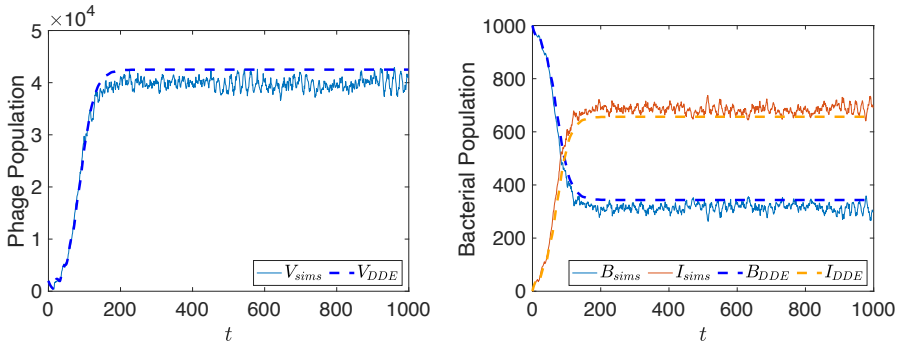


Fig. 5.25 An example realisation of the simulation. The resident phage population initially grows until it reaches a steady-state V_{ss}. The average behaviour of the model is mostly captured by a set of delay differential equations (DDEs), set out later in Eqs. 5.1. Slight discrepancies arise due to the fact that in the simulations, infection, decay and lysis must occur in discrete steps. Parameters used are $\alpha = 3 \times 10^{-6}$, $\beta = 100$, $\tau = 15$, $\delta = 0.1$ and $B_0 = 1000$. This figure is the same as Fig. 5.2.

To solve for the viral population, consider the rate at which cells *become* infected αVB. At steady-state this rate is constant at $\alpha V_{ss} B_{ss}$, and so the number of cells infected in a time window Δt is simply equal to $\alpha V_{ss} B_{ss} \Delta t$. We also know that the number of cells which are infected at time t is the sum of all the cells that became infected between the period of $t - \tau$ to t. Once at steady-state, this means that the number of infected cells is simply given by $I_{ss} = \alpha V_{ss} B_{ss} \tau$. We can then use this relationship, along with our expressions for B_{ss} and I_{ss} to show that

$$V_{ss} = \frac{I_{ss}}{\tau\alpha B_{ss}}, \tag{5.58}$$

$$= \frac{\frac{\alpha B_0(\beta-1)-\delta}{\beta\alpha}}{\tau\alpha\frac{\alpha B_0+\delta}{\beta\alpha}}, \tag{5.59}$$

$$= \frac{\alpha B_0(\beta - 1) - \delta}{\tau\alpha(\alpha B_0 + \delta)}. \tag{5.60}$$

Using the same resident parameters as before this gives $V_{ss} = 42,502$ in agreement with our numerical calculations (Fig. 5.25a).

'The closer you look at something, the more complex it seems to be.'

Vint Cerf

Chapter 6

Single Cell Imaging of Phage Infection

All of the analysis presented in this Chapter is my own. The single cell imaging data analysed was collected by Charlie Wedd, in the group of Dr. Somenath Bakshi.

6.1 Introduction

In Chapter 4 we attempted to address shortcomings in current experimental phage techniques, whereby phenotypic parameters are exclusively measured in liquid cultures rather than in lawns on the surface of a plate. An additional issue with current assays, and indeed the new assays presented in Chapter 4, however, is that they are based on population averages, which obscure many of the more stochastic and heterogeneous aspects of the biological system.

For instance, the degree to which the phage life-history parameters vary from cell-to-cell, and phage-to-phage, remains unclear. A concrete example of this came in Chapter 4, when our bulk lysis curve measurements performed in partially spent media resulted in a fitted burst size half of what is found under normal conditions. As was pointed out at the time, however, we are unable using our existing techniques to say whether this reflects a scenario where each of the infected cells produces half the normal number of phage, or whether it reflects a scenario where only half of the infected cells lyse and release the normal number of phage.[1]

Not only this, but if the environment (for instance the lawn of bacteria) is heterogeneous, one cannot distinguish the effects of specific local conditions. Furthermore, using existing techniques it would be difficult to observe and quantify the trade-offs

[1] Or indeed, something in between these two scenarios.

between different interactions, as these will often be averaged out in population based measurements.

These are all important issues, because in addition to sampling stochasticity (i.e. genetic drift), the evolutionary dynamics of populations are also affected by phenotypic stochasticity. For instance, in bacteria stochasticity in gene expression and other biochemical reactions can lead to cell-to-cell variability in traits such as cell size and division time, which in turn impact the growth rate of the population [187–190]. Classical theory would suggest that if two strains of bacteria with different population growth rates were to compete in a Moran type process, the strain with the higher growth rate should come to dominate [10, 186]. More recent work has showed however that a more detailed understanding of the distribution of phenotypes in the population is required to predict fitness [191, 192].

Concerning phage, it has been shown previously using bulk measurements that the host's metabolic state and growth rate impact burst size and lysis time [129]. Given this, it seems reasonable to assume that within a bacterial population, variability from cell-to-cell or in the environment will impact the phage infection cycle. Such phenotypic heterogeneity could be particularly important for the dynamics at the front of an expanding population where, as has been discussed previously, the number of individuals is relatively very small. For instance, if a cell close to the front of the expansion were to lyse even slightly earlier or later than average, this could conceivably have a significant impact on the likelihood of the phage it releases reaching fixation.

The issue goes beyond this, however. The concept of biological stochasticity in phage infection was partially explored in Chapter 5, but as discussed at the time, it is not clear at present how best to accurately model the variability in these processes, even in a well-mixed setting. It is entirely possible, if not likely, that even in well-mixed settings there will be significant consequences of phenotypic heterogeneity to the phage population.

Being able to observe and measure parameters such as the lysis time at a single cell level will therefore provide critical information that can be used in stochastic models of phage infection. Here, we utilise a high-throughput, continuous, liquid culturing microfluidic device known as a 'mother machine' [193, 194]. The mother machine consists of a series of narrow channels, in which bacteria grow end-to-end in chains, allowing for the identification and characterisation of single cells. By introducing phage $T7_{mEYFP}$ (which induces expression of fluorescent proteins in infected cells) to this setup, we are therefore able to observe infection in single cells.

6.2 Methods

These experiments were performed by Charlie Wedd in the group of Dr. Somenath Bakshi, who kindly shared the protocols used in the Bakshi lab for the manufacture of microfluidic devices.

6.2.1 Feature Design

The first step in making the microfluidic 'mother machine' devices is to produce a master, which is a template that is repeatedly used to fabricate polydimethylsiloxane (PDMS) replicas. In this case, these masters come in the form of silicon wafers which are produced using photolithography offsite.

A schematic of the design of the mother machine can be seen in Fig. 6.1. In the mother machine there are many 'trenches,' which consist of 75 μm long, $\sim$1.4 μm square channels in which bacteria grow end-to-end in chains. The cells in these channels are also surrounded by $\sim$0.5 μm deep side channels, which facilitate efficient access to media (and phage) over long times [194]. The channels are closed on one end, and on the other they empty into a feeding channel that supplies fresh medium and washes away excess cells as they are pushed out by growth.

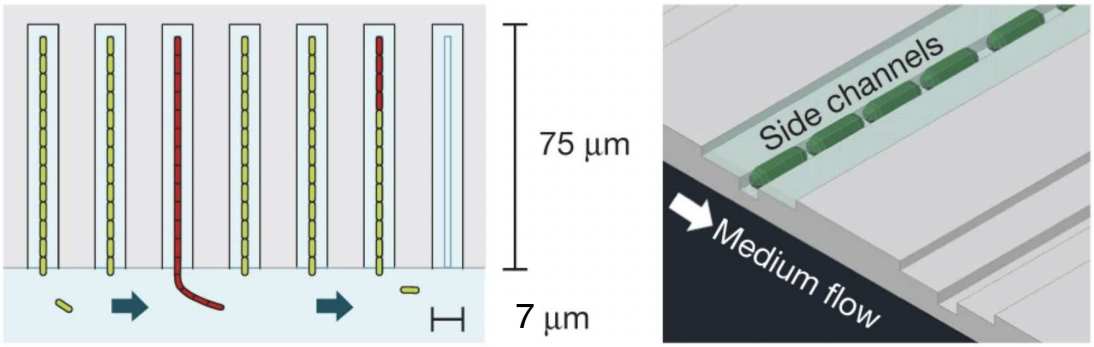


Fig. 6.1 Top and isometric schematics of microfluidic trenches in which bacteria are held. The trenches are characterised by a central $\sim$1.4 μm wide square channel, in which the cells sit, as well as $\sim$0.5 μm deep side channels that surround the cells, creating a 'bath' of media and phage, that enables efficient access over long times. Channels connect to a larger channel through which media and phage is continually replaced, and excess cells are washed away. Figure and caption adapted from Ref. [194].

6.2.2 PDMS Replica Fabrication

The protocol for producing the PDMS replica is as follows. First, 30 g of PDMS base elastomer is mixed in a disposable plastic cup with 6 g curing agent elastomer. The reagents are then mixed thoroughly, either using a milk frother at $\approx$ 1000 rpm for two minutes, or alternatively using a spatula. This mixture is then degassed for $\approx$ 30 mins in a desiccator to remove the largest air bubbles.

While this occurs, a curing bowl is prepared. This is achieved by spreading aluminium foil over a glass dish and working into a "bowl shape." The master silicon wafer is then placed in the centre of the bowl, and pressed down using a pipette tip such that it lies flat.

The PDMS mixture is then spread rapidly over the wafer while maintaining constant contact between the PDMS and the bowl to avoid the production of large bubbles during pouring. The curing bowl with the wafer and the PDMS is then degassed again in the desiccator to get rid of any remaining air bubbles. Once degassed, the PDMS and wafer are baked at 65 °C for at least an hour.

Once baked, the PDMS is peeled carefully off of the silicon wafer, and holes are punched from the side of the PDMS that contains the features of the microfluidic device. These holes will act as inlet and outlet channels to allow media/bacteria/phage to flow into the microfluidic device.

6.2.3 Bonding and Chip Treatment

Initially, the PDMS replica is cleaned by sonicating in isopropanol for 30 mins to remove any debris produced during the hole punching. After this, the replica is dried with an air-gun, then rebaked for an hour at 65 °C, and then rinsed and sonicated in Milli-Q water for 30 mins.

The next step is to clean the glass cover slip that the replica will be bonded to. This procedure is similar to the PDMS replica, and involves first sonicating for 20 mins in 1 M KOH then rinsing multiple times with Milli-Q water. The slip is then sonicated in Milli-Q water for 20 mins, dried with an air-gun and then baked for 1 hour at 65 °C.

The PDMS replica then needs to be bonded to the glass cover slip. The PDMS replica and cover slip are placed in a plasma cleaner with the surfaces to be bonded facing up. They are then left to react with the plasma on high power for approximately 60 s, before being removed from the cleaner. The PDMS replica is then placed carefully on the cover slip, ensuring no air gets trapped between the two surfaces. The PDMS

is then pressed down firmly to allow covalent bonds to form between the PDMS and the cover slip. The bonded chip is then placed on a hot plate at 95 °C for 5 mins, before being based again for 3 hours at 65 °C. This final step is because the plasma treatment makes the PDMS very hydrophilic, which makes loading the cells into the small trenches in the device difficult. The time in the oven allows the PDMS return to it's normal hydrophobic state, which makes loading the cells easier.

6.2.4 Running an Experiment

Overnight cultures of *E. coli* strain SB7 (MG1655 constitutively expressing mCherry, see Appendix A and Ref. [195]) are grown in EZRDM[2] + 1% by volume pluronic solution at 37 °C. On the morning of the experiment, the overnight culture is diluted 1:100 into EZRDM + pluronic solution and allowed to grow for 2-3 hours so that the cells come out of stationary phase.

The lanes of the chip are checked to ensure that the inlets and outlets are bonded accurately and well, and to look for debris or deformations in the lanes and trenches. The lane with the fewest defects is then chosen, and needles are inserted into its inlet and outlet. EZRDM + pluronic is then flowed through the lane to 'passivate' it, i.e. to make it easier for fluid to easily enter the lane.

Next, ∼1 ml of the diluted culture is spun down in a centrifuge at 1000g for 3 mins to genetly sediment the cells. The supernatant is then poured away, and the cells resuspended in the residual volume to create a very dense culture. A small amount of this dense culture is taken in a gel loading tip, and pushed into the lane in the mother machine chip. The chip is then left for $\sim$ 20 mins to allow the cells to diffuse into the trenches.

After this, the needles are reattached and media is flowed at 20 μl/min through the lane. The chip is placed on the microscope stage, and checked for leaks. The mother machine lane is then aligned with the stage's x coordinate, such that when moving the stage horizontally the lane remains in view. The trenches are then imaged using a Nikon Eclipse Ti2 microscope equipped with a Plan Apo λ 40× Ph2 DM objective (1.5× post objective magnification used, yielding 60× total). Phase contrast and fluorescence images are then taken at each location every 3 min intervals. In total, 32 fields of view (FOVs) were sampled, with each FOV containing 15 individual

[2]EZRDM stands for EZ Rich Defined Medium, and is a preparation of Neidhardt Supplemented MOPS Defined Medium [196]. The 'EZ' is used to indicate that modifications have been made to streamline the protocol given in Ref [196].

trenches. The bacteria are then left to grow in the mother machine for ∼1.5 hours to establish a baseline growth behaviour, before switching the media to that containing phage.

6.3 Results

At present, this analysis will be limited to highlighting various, qualitatively different, behaviours that are observed in the mother machine when the bacteria are exposed to phage. At this point I will also attempt to clarify the terminology that will be used. Throughout this Chapter, we will be using the expression and observation of mEYFP as our 'infection' marker, as it occurs only in cells which have successfully been infected. However, the expression of mEYFP does not begin at the moment of adsorption or DNA injection, it begins as the phage genome is transcribed. As we shall see shortly, this means there is a delay between the injection of the phage genome and the moment in which cells become observably fluorescent. Therefore, when I say in this Chapter that we observe 'infection,' I do not mean that we observe the initial moment in which infection occurs, but that we observe cells in an infected state. Similarly, when cells begin to fluoresce, we can say that 'infection' has occurred, but not in that precise moment, rather at some point (likely several minutes) in the past. Relatedly, whenever I say that we observe 'lysis,' I do not mean that we see the precise moment in which lysis occurs. If a lysis event were to occur in one of the trenches during the experiment, because of the frequency of image capture (once every 3 mins) this would simply appear as an infected cell at time t, which had disappeared by time t_{+1}.

6.3.1 Infection and Lysis

The first qualitative behaviour we observe is exactly what one might typically expect: infection and subsequent lysis of the host bacteria. Fig. 6.2 shows a single trench in the mother machine as a function of time. Initially, no cells in the trench are infected (at least as far as we can tell), as indicated by the lack of mEYFP expression from any of the cells. One of the cells then begins to fluoresce, indicating that infection has occurred. The intensity of this fluorescence then increases for approximately 12 mins, before suddenly disappearing when the cell lyses. Approximately 6 mins after this lysis event, *all* of the cells in the trench near the lysed cell then go through the same process, ultimately leading to an emptying of the trench.

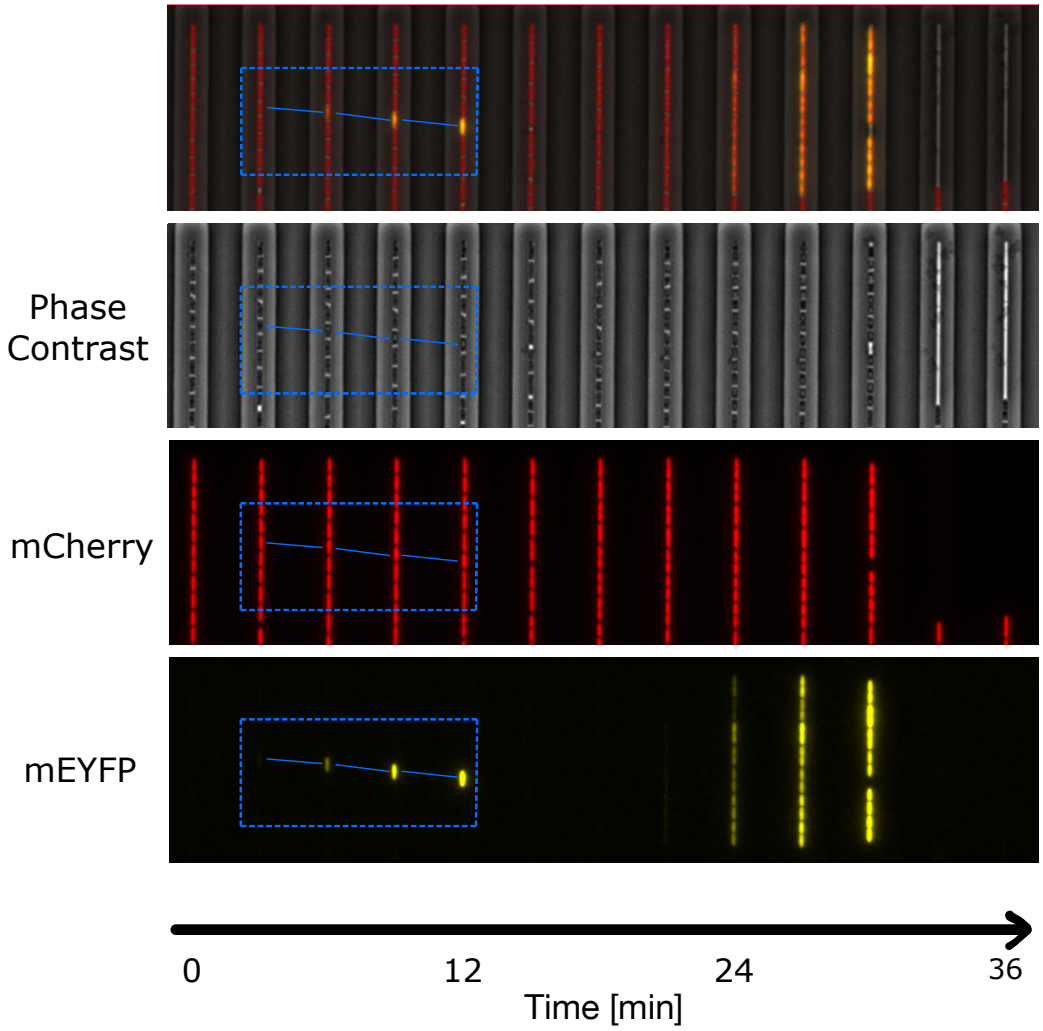


Fig. 6.2 An example kymograph showing infection and lysis in the mother machine. The blue dashed rectangle highlights the initial cell which is infected and subsequently lyses. Blue lines are drawn to aid the viewer in tracing the cell from frame to frame. For scale, the trenches are 7 μm wide in total.

This pattern of 'mass infection' following lysis allows us both to easily identify lysis events, and to estimate the timing of lysis inside the mother machine. We can place the time of the initial lysis event somewhere in the 3 min window after the last frame in which it is visible. Then, if we assume that infections of all of the subsequent cells occurs rapidly, i.e. within the same 3 min window (which seems like a reasonable assumption given the simultaneous appearance of fluorescence and subsequent lysis), then we can say that it takes approximately 6 mins for some fluorescence to become

visible, and a further approximately 12 mins for the cells to lyse. This is broadly in line with our expectations of a ∼6 min long eclipse period, and a total lysis time of ∼ 18 mins.

However, while the type of situation shown in Fig. 6.2 was what we were expecting to see in the mother machine, in actual fact it represents quite a rare case. The data collected so far consists of 434 initially bacteria-filled trenches imaged over a 3 hour period. During this period, at least one infection event occurs in 387 (89%) of these trenches, as indicated by at least one yellow fluorescent cell. Of these 387 trenches however, only 16 (4%) exhibit the kind of lysis followed by mass infection pattern shown in Fig. 6.2. It's also worth pointing at that compared to the total number of infections this percentage would be even smaller, since many of the trenches with at least one infection in fact contain multiple infections. This then naturally raises the question: what happens to all the other infected cells?

6.3.2 Outflow

One of the most common fates of infected cells is that they flow out of the end of the trench and are lost. This occurs because the growth and division of cells further up the trench results in the cells further down the trench being pushed out. The only cell to remain where it is in the trench is the cell at the very top, the so called 'mother' cell (hence the name mother machine). An example of this can be seen in Fig. 6.3. Because of this effect, unless the cell which is infected is the mother cell, the infected cell can be pushed out of the trench before it lyses. This is particularly true of cells which are infected when they are already near the bottom of the trench.

In Fig. 6.3 however, a cell is shown as starting in the upper half of the trench, and you will notice takes >36 mins[3] to be forced out of the trench, which is double the amount of time we would expect for lysis to occur in normal conditions (see Fig. 6.2).

6.3.3 Infection Without Lysis

This brings us onto the next phenotype which we find is common in the mother machine: cells which are clearly infected, but never lyse. An example of this can be seen in Fig. 6.4. In this example, infection occurs in the mother cell shortly after the start of

[3]It takes longer than this 36 mins as infection occurred several frames prior to the start of the figure, but these are not shown to better illustrate the outflow effect.

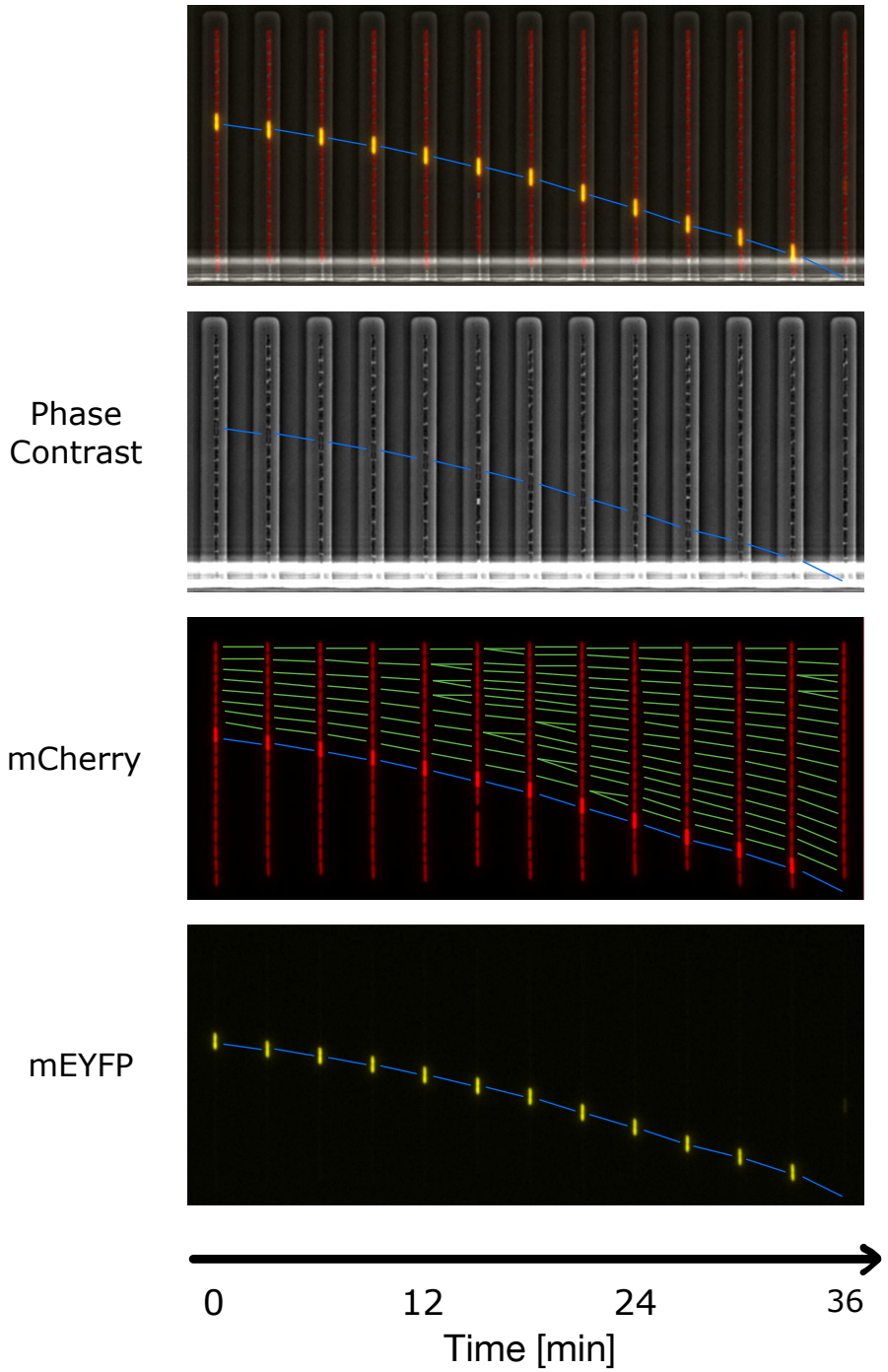


Fig. 6.3 An example kymograph showing infection of a cell which then flows out the end of the trench prior to lysis. Blue lines are drawn to aid the viewer in tracing the infected cell from frame to frame, with green lines tracing the lineages of cells further up the trench. For scale, the trenches are 7 μm wide in total.

the 3 hour imaging period, but lysis never occurs, and as such the cell stays visible in the trench for the entirety of the experiment.

Crucially, the mEYFP signal remains strong throughout this period. This point is important because we also observe a second phenotype where the cell never lyses, but does stop expressing mEYFP (Fig. 6.5). In this example some mEYFP signal does remain, but at a much fainter level than at the height of infection.

Both of these examples also emphasise that while many infected cells are pushed out of the end of the trench before they lyse, it seems likely that many of them never would. This is evidenced by the fact that many of the infected cells, like the example shown in Fig. 6.3, are in an infected state in the trench for a period much longer than the expected lysis time.

6.3.4 Other Observations

Another interesting observation is that infected cells are often identifiable from the phase contrast image alone. It can be seen in Fig. 6.6 that in phase contrast images, uninfected cells appear relatively uniform in intensity, whereas infected cells are characterised by a brighter patch in the centre of the cell. It's possible that this occurs due to DNA condensation and increased protein concentration within infected cells. This could be very significant, as if this occurred during infection with *any* phage, rather than only phage $T7_{mEYFP}$, then there would be no need to genetically modify the phage to observe infection.

A further observation is that in the mother machine, infections very often come in pairs, i.e. two neighbouring cells appear as infected simultaneously (Fig. 6.7). Given the frequency of infections in the trenches, it seems unlikely that this would happen by chance (excluding the 'mass infection' events discussed previously, although these also do not occur by chance). In cases where such paired infections occur, shortly before fluorescence becomes visible the two cells divide from the same mother cell. This suggests that the mother cell was infected with phage shortly before division, resulting in two infected offspring. I think this division must come early in the infection however, i.e. before the cell begins to express mEYFP, as I observe no instances where an already yellow fluorescent cell divides. This observation could potentially be used to better understand the finer timing of the progress of phage infection within the cell. For instance, for both daughter cells to be infected there presumably must have been at least two copies of the phage genome within the parent cell at the point of division, thereby giving us an indication of the timing of genome replication.

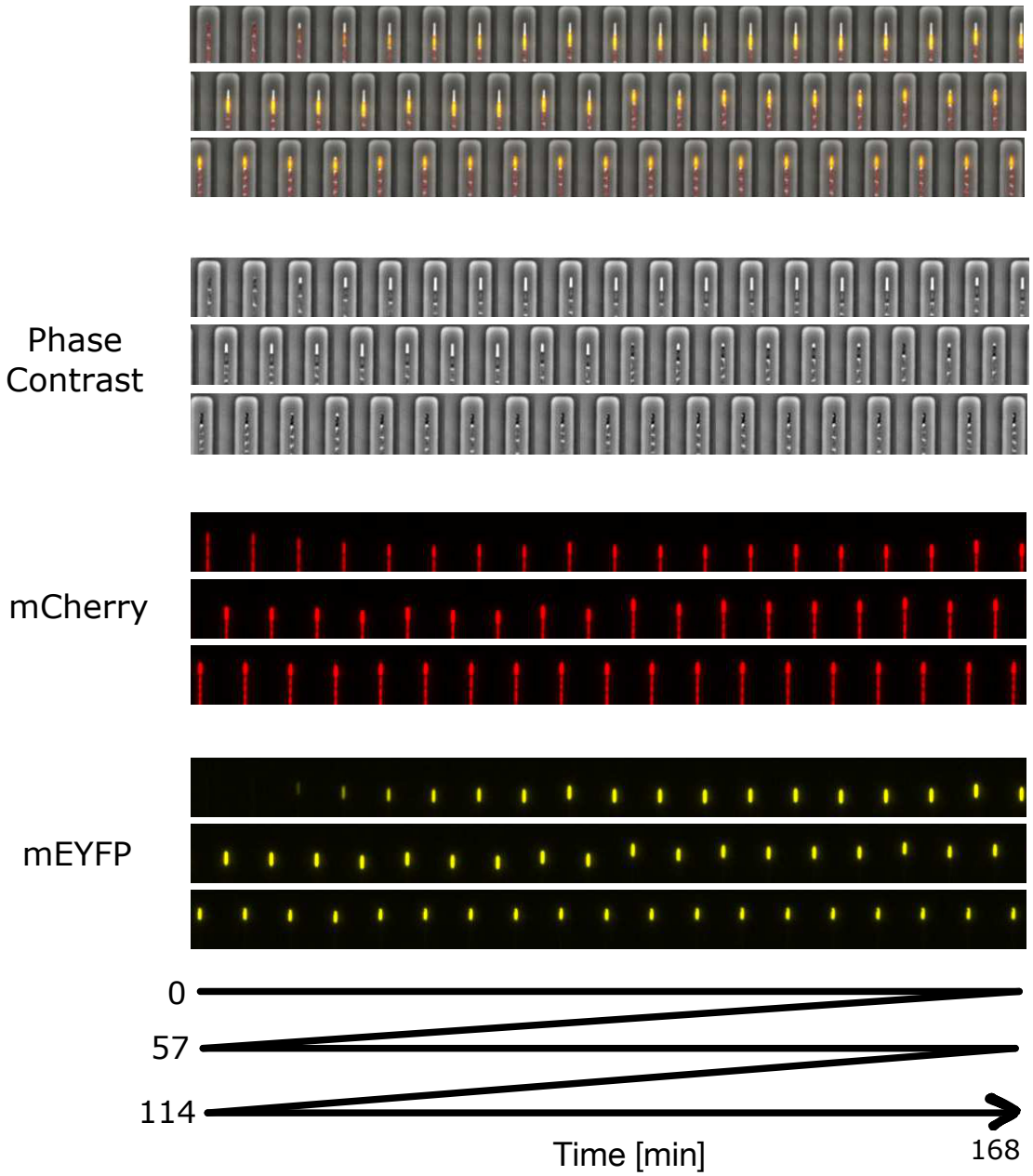


Fig. 6.4 An example kymograph showing infection of a mother cell which then remains infected without lysing for approximately 3 hours. The intensity of the mEYFP signal remians strong throughout this time. For scale, the trenches are 7 μm wide in total.

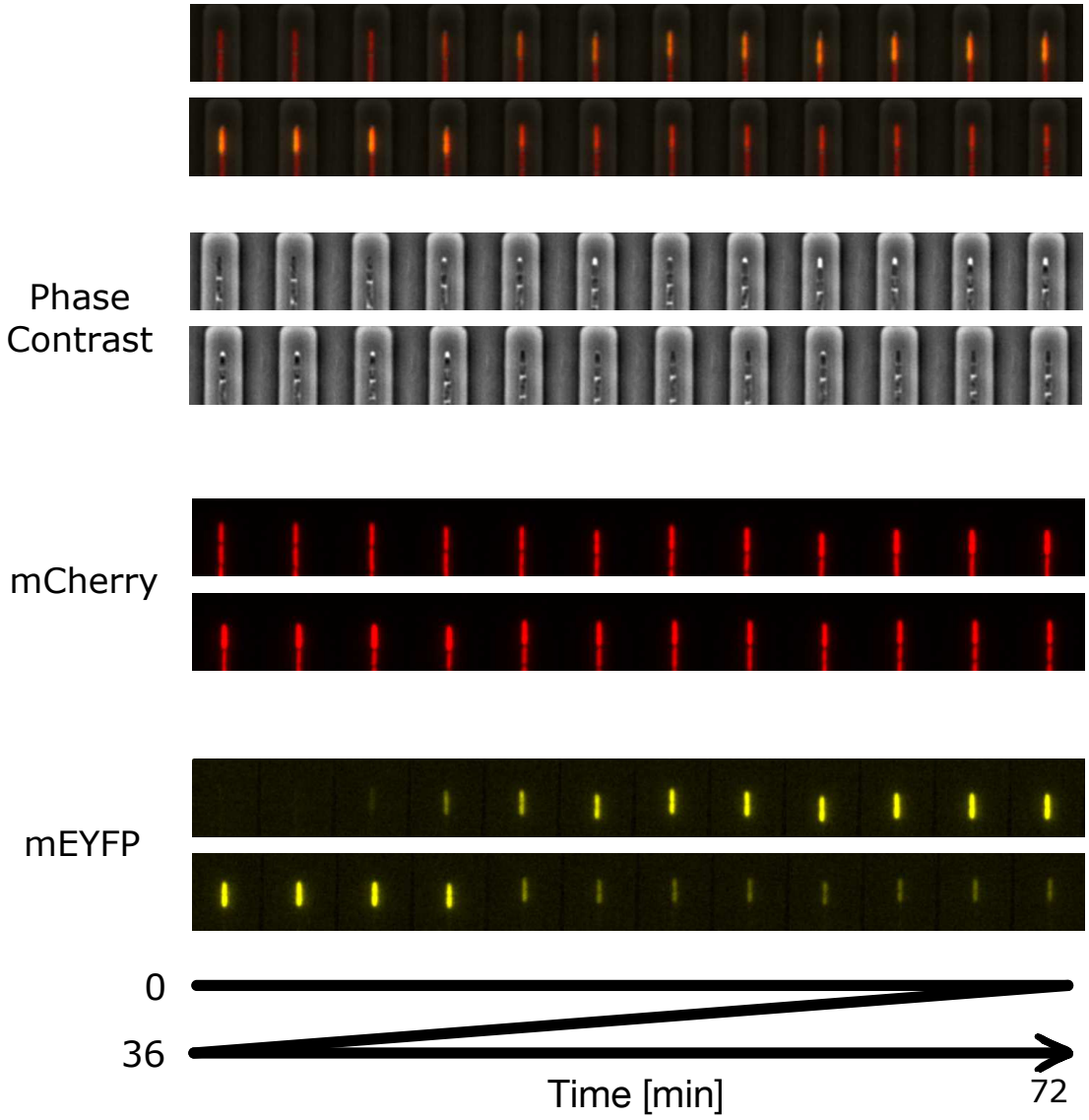


Fig. 6.5 An example kymograph showing infection of a mother cell which then never lyses, but where the intensity of the mEYFP signal reduces after approximately 30 mins. For scale, the trenches are 7 μm wide in total.

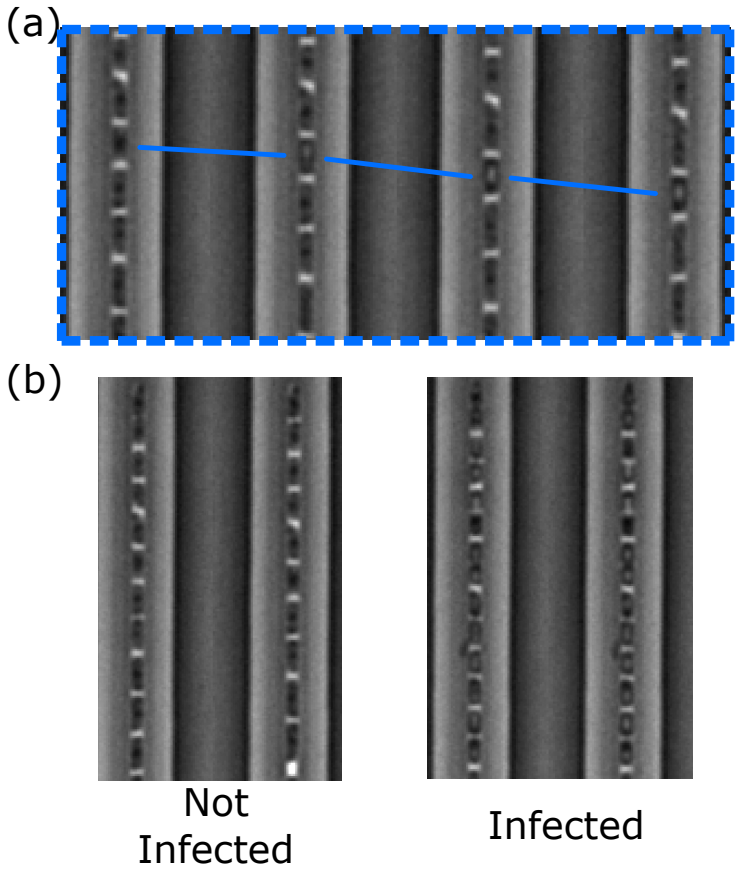


Fig. 6.6 Infected cells appear different in the phase contrast images. (a) shows the blue rectangle highlighted in Fig. 6.2, while (b) shows further examples of infected and uninfected cells. In infected cells, the centre of the cell appears brighter in phase contrast images, whereas uninfected cells are uniform in intensity. For scale, the trenches are 7 μm wide in total.

6.4 Discussion

In this Chapter it was demonstrated that we are able to observe phage infection at a single cell level in the mother machine. In addition to cell lysis, as was anticipated, we observe many instances where cells never lyse, and either remain fluorescent or partially recover. The observation of these phenotypes raises many interesting questions. Perhaps the most obvious question is why don't the cells lyse? My instinct is that the answer to this is, at least partly, related to the media used in the mother machine. We have seen in Chapter 4 that the availability of nutrients can have a significant impact on lysis (Fig. 4.17). I would suggest that an obvious next step would be to perform

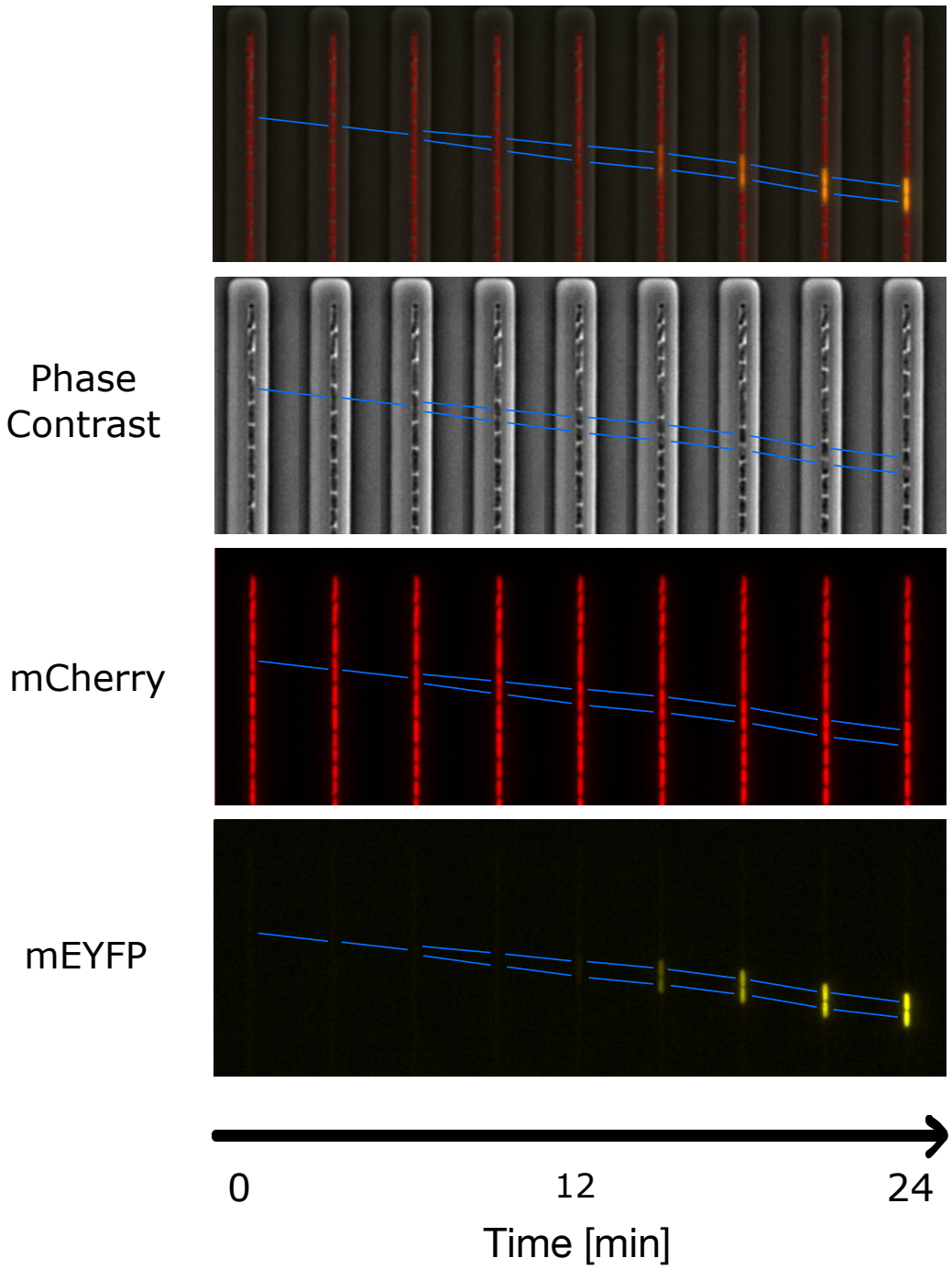


Fig. 6.7 An example kymograph showing infection of neighbouring cells in the mother machine. Shortly prior to the cells beginning to express the infection marker, the pair of cells had come from the same mother cell, indicating that the infection may have occurred prior to cell division. For scale, the trenches are 7 μm wide in total.

experiments in different media, and at different media concentrations, to ascertain what impact this has on the frequency of lysis.

Another interesting question is, given how rarely infected cells seem to lyse in the current set-up, why is it that *all* of the secondarily infected cells in Fig. 6.2 *do* lyse? Is it because those instances represent more 'effective' phage mutants in the population, or because high multiplicity infections increase the likelihood of lysis? If infection multiplicity is a factor, this could potentially be explored by flowing a *much* higher concentration of phage through the feeding lane, and seeing if there is a corresponding increase in the relative proportion of infections which lead to lysis. Another possibility is that enzymes and other proteins released when the initial cell lyses remain in the trench environment, and aid with the lysis of the later cells. This could perhaps be explored by incorporating phage lysate into the media used in the mother machine, to see if it increases the frequency of lysis.

One caveat to the seeming rarity of lysis is that, in this work, lysis events have been identified retrospectively based on the simultaneous appearance of many infected cells in a trench. It is possible, however, that there are instances where infected cells do lyse, but for some reason do not lead to many subsequent infections. In such instances, I expect that it would be difficult to distinguish between cells which had lysed, and cells which had fully recovered, particularly given the current 3 min gap between frames. To address this, I would propose the development or use of a lineage tracking algorithm, which could match cells between frames based on their size, position, growth rate, etc. Lysis could then be identified more quantitatively if a cell cannot be matched to a high degree of certainty with any cell in the subsequent frame.

There are also questions which may be more challenging to answering. In the cells which continue to fluoresce strongly, are phage continually being produced inside the host? In the cells which don't continue to fluoresce strongly, are phage no longer being produced? Does this indicate some kind of 'recovery' of the cell? On this point, it is worth noting that I have not observed an instance where a 'recovered' cell began growing/dividing again. Perhaps instead the reduction in fluorescence indicates that the cell membrane has become partially compromised, such that the cell remains intact, but alters the cell environment in such a way that the mEYFP can be degraded. Interestingly, in these cells the mCherry signal appears to be unaffected (Fig. 6.5). These observations could be explained by the cell environment becoming more acidic. The acid sensitivity of fluorescent proteins is measured using a pKa value, which corresponds to the pH at which the intensity of the fluorescent protein drops to 50% of

its maximum value. mEYFP has a pKa=6.9 [197], whereas mCherry has a pKa<4.5 [198], meaning that mCherry maintains its intensity better in acidic environments.

In addition to these questions, as this work continues there will also be a more rigorous and quantitative analysis of the frequency of events, and the properties of the cells during infection such as length of the cell and length of infection. Beyond this, it would be useful to observe wild-type phage, along with other mutants, in the mother machine to determine whether lysis is observed and if cells appear differently in phase contrast shortly before lysis. If this was a general feature then it would be incredibly useful as it would mean, in principle, that phage infection could be observed and identified in the mother machine without the need to induce fluorescence. Finally, other phage mutants could be created to observe different steps in the phage infection cycle, such as adsorption and injection of viral DNA. For instance, a phage could be engineered to itself be fluorescent by attaching fluorescent proteins to its capsid. It's possible that as this phage diffused through the mother machine, its movement would result in no clear signal, but at the moment of adsorption the phage becomes fixed in place, resulting in a visible signal.

Still, thou art blest, compar'd wi' me!
The present only toucheth thee:
But Och! I backward cast my e'e,
On prospects drear!
An' forward, tho' I canna see,
I guess an' fear!

from To A Mouse, by Robert Burns

Chapter 7

Conclusion

7.1 Summary

We have seen in this thesis a wide variety of interrelated findings regarding the impact of virus-host interactions on the evolutionary dynamics of phages in general, and T7 more specifically. Throughout much of this thesis, there has been a particular focus on interactions that occur during growth in bacterial lawns, i.e. during spatial range expansions. It was known previously, that when individuals in single-species expansions grew or dispersed better cooperatively, this could result in 'pushed' expansions which are much more resilient to stochastic fluctuations, making the population better equipped to adapt to changes in the environment. One of the initial questions posed in this thesis was whether or not a similar pushed dynamic could arise in the expansion of phage, not through cooperativity (since the phage do not interact directly with one another), but through the unavoidable physical interactions the phage has with the host bacteria.

In Chapter 2, we have seen that indeed it can. By combining experimental measurements of phage diffusion in a bacterial lawn with a reaction-diffusion model of plaque growth, we saw that the bacteria hinder phage diffusion through steric interactions, and that this can result in pushed expansions at high bacterial density. Not only that, but we also discovered that pushed expansions could occur due to a second independent effect, which originates in the fact that phage are unable to disperse while trapped inside the host during incubation.

This not only highlights the fact that the genetic diversity and adaptability of *many* expanding viral populations may often be much higher than is currently assumed, but it also indicates that bacteriophage offer researchers a potential system in which to

pulled and pushed dynamics in the laboratory. One approach to such an endeavour is through the quantification of the properties of the large monoclonal sectors that form during spatial growth. In Chapter 3, an image analysis pipeline was developed to identify these sectors from fluorescent images of plaques where phage are able to induce fluorescence in infected cells.

The next question posed in the introduction was how best to interpret plaque growth experiments in light of our models and theories. In part, our ability to do this accurately remains limited because of the experimental tools at our disposal. As discussed in the introduction, the experimental study of bacteriophage relied almost exclusively on measurements carried out in a well-mixed liquid culture of exponentially growing bacteria. While such measurements are certainly useful, they are unlikely to capture the complex infection dynamics that occur in spatially structured environments, where bacteria with varying metabolic states are often densely packed together with an irregular supply of nutrients. In Chapter 4, I presented various novel experimental techniques designed to address these issues, and to facilitate better comparisons between models and experiments of plaque growth.

Next, we turned our attention to the evolutionary impact of superinfection and superinfection-exclusion mechanisms. Originally, our ideas and questions about this topic were also rooted in a spatial context. Indeed, such mechanisms were already touched upon in Chapter 2 in relation to the four model variants we looked at, and it was shown that the presence or absence of a mechanism that prevents adsorption to previously infected cells greatly alters the expansion dynamic. Our interest in superinfection-exclusion mechanisms goes beyond this, however. In a spatially structured environment, phage will mainly be in competition with clonal phage that were released by the same or neighbouring hosts (picture the monoclonal sectors discussed in Chapter 3). This raises the question of whether or not having a superinfection-exclusion mechanism is less useful here? At what point does the cost of maintaining a superinfection-exclusion mechanism become intolerable for the virus? What impact does allowing or excluding superinfection have on the balance between selection and genetic drift in expanding viral populations?

Before we are able to understand the impact of spatial structure on these dynamics, however, we first have to understand how they operate in an unstructured, well-mixed environment. To our surprise, even in the comparatively simple case of a well-mixed environment, the baseline impact on evolution of superinfection and superinfection-exclusion remained relatively poorly understood. We first therefore had to address this

gap, and in Chapter 5 we present the results of an extensively stochastic simulation study characterising the impact of superinfection on the forces of selection and genetic drift in a phage population grown in a turbidostat. While these results provide a useful baseline for works investigating the impact of various intracellular interactions in viral populations, and will provide a necessary point of comparison for future work exploring the impact of spatial structure, arguably one of the most pertinent findings concerns simply how different selective pressures can act in a turbidostat compared to a more traditional well-mixed culture. Our finding that lysis time is under no selective pressure in superinfection-excluding populations, and that it is only weakly selected for in superinfecting populations has significant consequences for researchers attempting to study the evolutionary dynamics of viral populations using continuous culturing setups.

Finally, another issue raised implicitly in Chapter 5 is the impact of biological stochasticity, e.g. cell-to-cell variations in the phage life-history parameters, how those parameters depend on the chemical environment, and how they depend on the cell metabolism. The level of such variability is currently not well understood, as both existing phenotypic assays and the newer assays presented in Chapter 4 rely on bulk measurements of population averages, rather than observations made at the level of single cells. In Chapter 6 therefore, we began to investigate the extent of phenotypic variability during phage infection by imaging the infection of single cells in a microfluidic mother machine. While this work is still in its early stages, several surprising behaviours and a significant amount of variation was observed, indicating that phenotypic heterogeneity could play a very significant role in the evolution of bacteriophage populations.

7.2 Outlook

7.2.1 Single Cell Imaging

There are many possible avenues future research in this area could take, as has been highlighted in each of the Chapters. Perhaps one of the most obvious and fruitful is a continuation of the work presented in Chapter 6, where phenotypic variability was observed at a single cell level in a microfluidic mother machine. The use of mother machine devices to study bacteria appears to be becoming increasingly prevalent, owing to their ability to gather *vast* amounts of data, and their use can clearly also be adapted

for the study of phage infection. At present, we possess a phage that, as its genome is transcribed within the host, induces the expression of fluorescent proteins. This is an incredibly useful marker, but it indicates only one step in the infection cycle. Further genetic modifications could also be made to the phage with a view to observing other steps. For instance, a phage could be engineered to itself be fluorescent by attaching fluorescent proteins to its capsid, and the signal could potentially be used to identify the moment of adsorption.

Mother machines also offer the potential to observe and quantify *many* different phage simultaneously. This could be achieved by labelling a population of bacteria cells with genetically encoded fluorescent barcodes that are identifiable in the mother machine. These labelled cells could then be loaded into the mother machine, each infected with a different phage. At the start of the experiment, we could therefore tell which phage were in which trenches (within infected cells). Uninfected bacteria could then be loaded into the trenches below the infected cells, such that when those initial cells lyse there are bacteria in the trench for the phage to infect. The timing of infection for many phage could therefore be monitored in a single experiment. One potential challenge that may be faced here is how to prevent the infected cells from lysing *before* they are successfully loaded into the trenches. I would propose that a solution to this issue potentially lies in suspending and loading the cells in water, or some other liquid lacking nutrients. As we saw in Fig. 4.17, this appears to prevent cells from lysing, although whether or not this is because it also prevents infection remains unclear. If cells were infected without lysing, however, this would allow us to safely load them into the mother machine, and then induce their lysis at a timing of our choosing by flowing in fresh media.

7.2.2 The Role of Phenotypic Heterogeneity

While using the mother machine to probe the nature of phage infection is undoubtedly highly promising, if I am honest, it is not where my heart lies. This is perhaps a more philosophical point, and it may seem strange given the contents of this thesis, but I have never felt like my PhD was about phage. I was interested in understanding how viruses evolved, particularly in space, but in a more general sense. The focus on phage, at least to my mind, was driven simply by the fact that it is a 'model' viral system (i.e. it is relatively easy to control and manipulate in the laboratory) which could be used to test predictions. It was not, in itself, the thing which was of interest.

Therefore, the questions that can be answered using mother machines, such as what is the precise timing of various steps in the T7 infection process, and what is the cell-to-cell variability in the length of each of those steps, are at least to me not by themselves interesting. What compels me is the broader question of how phenotypic variability alters the evolutionary dynamics of viral populations. What does it *mean* for a viral population? Does increased variability help, hinder, or indeed make no difference to one virus in competition with another? If we are to use phage T7 as a model system to help us explore the impact of phenotypic heterogeneity, then the levels of phenotypic heterogeneity in T7 must of course be categorised. I do not dispute this. If it were me, however, I would first attempt to determine what role I *expected* phenotypic heterogeneity to play in the evolutionary dynamics of the populations.

Initially, I would propose to investigate the role of phenotypic heterogeneity using similar stochastic simulations to those described in Chapter 5. In that Chapter, we briefly looked at the effect of stochasticity in lysis time by drawing the lysis time from a Gaussian distribution with some mean τ and standard deviation σ_τ, but we did not explore the implications of varying σ_τ. If two mutants have the same τ, but a different σ_τ, are their chances of success the same, or different, and under what conditions? If one virus has much greater levels of variability compared to its competitor, perhaps the first will gain the upper hand because *some*, albeit a minority, of phage in the population will lyse faster than their competitors. On the other hand, if the lysis time of those progeny are not correlated with their parent, then any advantage would be short lived.

This is obviously a relatively simple first step, but I think it would be a worthwhile one. If we found that, in this context, biological stochasticity has little to no impact on the evolutionary dynamics, I would not prioritise trying to quantify it using a mother machine.

7.2.3 More Complex Spatial Dynamics

Much in the same way that I have focused on phage T7 as a model viral system, throughout this thesis I have focused on plaques as a model system for viruses spreading through space. This is because plaques represent a relatively 'simple' case, e.g. they occur on a flat, homogeneous surface where the bacteria are unable to move. In nature, however, bacteria and phage commonly exist in environments with more complex dynamics and spatial structure. Moving forward, I would be inclined to explore the impact that these have on the evolutionary dynamics.

For instance, what happens if we relax the condition that the bacteria are unable to move? If the bacteria cells were motile, perhaps they could carry the phage significant distances before lysing. What impact does this have on genetic drift? Presumably some of the phage would be carried significantly beyond the front and act as 'founders,' but some will presumably also be carried parallel to or far behind the front, thereby increasing the degree of mixing. The movement of bacteria also allows for even more complex behaviours, as it has been shown previously that bacteria can exhibit phenotypic resistance to phage by aggregating [163, 199] or propelling themselves away from phage [200]. How do such mechanisms evolve, and how do the phage react and adapt to such changes? Additionally, what happens if we relax the condition that our phage are growing on a flat, homogeneous lawn of bacteria? In nature bacteria and phage commonly exist in more fragmented environments like the human gut or in soil. How does infection spread in these environments, and how are the evolutionary dynamics altered?

Similar to my proposed investigation of phenotypic heterogeneity, I would suggest that initially the most straightforward way to tackle some these questions would be to utilise a model similar to that discussed in Chapter 2. That model could be adapted to account for bacterial motility and an inhomogeneous initial distribution of bacteria. Following the same approach as in Chapter 2, the speed of the front could be determined and compared to the speed of the linearised system, and the genetic drift could be characterised using stochastic simulations.

This could then be followed up by experiments similar to those presented in Chapter 3, where phage used to induce fluorescence in infected cells were used to measure and quantify the formation and growth of monoclonal sectors during plaque growth. These experiments could be repeated using motile bacteria embedded in soft agar, so that the bacteria are able to move. Similarly, the bacteria could be inoculated heterogeneously on the plate. I would anticipate that these experiments would, however, pose some technical challenges. Since the motile bacteria would need to be embedded in a several mm thick layer of agar, I would expect that any fluorescence signal would appear to be quite blurry, since it would be originating from multiple z heights. On top of that, it's possible that the movement of bacteria would result in a much more diffuse front, where the boundaries between sectors are hard to identify. The extent of these issues will not be known, however, until the experiments are attempted.

7.2.4 The 'Big' Challenge

While there will clearly be technical obstacles to overcome, the more profound challenge I foresee in pursuing the lines of enquiry that I have outlined here, is how we ensure that our models actually tell us something useful about reality. As we try to account for more and more complex aspects of phage infection, decisions will need to be made about how to implement those aspects in our model. I do not think it is always obvious how to do that. As an example, our simple description in Chapter 5 does not account for an eclipse period (where infection has occurred, but no phage have been produced inside the cell). In the case where only one type of phage infects, this would be straightforward to implement, but what if there are two types of phage? In the event where a superinfecting phage infects a host prior to the end of the original phage's eclipse period, how should we model the internal dynamics? Should both phage begin production at the end of the first's eclipse period, or at the end of their own? Should the phage which infects first gain some advantage, and does this depend on when *exactly* during the eclipse period the second phage infects, or does competition only begin at the end of the eclipse period?

In part, some insight as to the answer to these questions can obviously be gained from experiments, although even through the use of mother machines I would anticipate that extracting the *intracellular* dynamics will be difficult. At present, this then leaves us to make decisions based on what seems 'sensible' and widely applicable. My biggest fear, however, is not that we will be unable to eventually answer these questions, or even that the decisions we make in the meantime will turn out to be wrong; my biggest fear is that there simply is no way of describing these processes that *is* widely applicable. As was noted in Chapter 5, superinfection-exclusion mechanisms, and more broadly the specifics of the infection process, are highly diverse. As we get into the finer and finer details, it may be that what is true for one virus is completely irrelevant to another. I suppose that in itself would be interesting, but as a physicist I *want* to be able to distil these complex behaviours down into a few simple ingredients. Perhaps, however, we will discover that this is simply not possible.

That being said, while my interest as a physicist working on this topic has been in the general, there are obviously many researchers in this area spanning biology, engineering, mathematics, theory, and experiment who each have their own approaches and questions, some more general and some more specific. Progress will be made by a combination of all of them, and I will watch with interest as it is.

References

1. Charlesworth, B. & Charlesworth, D. *Evolution: A Very Short Introduction* (Oxford University Press, 2017).
2. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nature Reviews Genetics* **8,** 610–618 (2007).
3. Higham, T. *et al.* The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* **512,** 306–309 (2014).
4. O'Neill, J. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations* tech. rep. (The Review on Antimicrobial Resistance, 2016).
5. Darwin, C. *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life* (John Murray, London, 1859).
6. Messer, P. W. in *Encyclopedia of Evolutionary Biology* 119–123 (Academic Press, 2016).
7. Wright, S. Evolution in Mendelian Populations. *Genetics* **16,** 97–159 (1931).
8. Wright, S. Inbreeding and Homozygosis. *Proceedings of the National Academy of Sciences of the United States of America* **19,** 411 (1933).
9. Fisher, R. A. & Bennett, J. H. *The Genetical Theory of Natural Selection: A Complete Variorum Edition* (OUP Oxford, 1999).
10. Moran, P. A. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society* **54,** 60–71 (1958).
11. Wang, J., Santiago, E. & Caballero, A. Prediction and estimation of effective population size. *Heredity* **117,** 193–206 (2016).
12. Fusco, D., Gralka, M., Kayser, J., Anderson, A. & Hallatschek, O. Excess of mutational jackpot events in expanding populations revealed by spatial Luria-Delbrück experiments. *Nature Communications* **7,** 1–9 (2016).
13. Birzu, G., Hallatschek, O. & Korolev, K. S. Genealogical structure changes as range expansions transition from pushed to pulled. *Proceedings of the National Academy of Sciences of the United States of America* **118,** e2026746118 (2021).
14. Colautti, R. I. & Barrett, S. C. H. Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science* **342,** 364–366 (2013).
15. Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B. & Thomas, C. D. Rapid range shifts of species associated with high levels of climate warming. *Science* **333,** 1024–1026 (2011).

16. Hewitt, G. M. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58,** 247–276 (1996).
17. Hewitt, G. The genetic legacy of the quaternary ice ages. *Nature* **405,** 907–913 (2000).
18. Templeton, A. R. Out of Africa again and again. *Nature* **416,** 45–51 (2002).
19. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. Demic expansions and human evolution. *Science* **259,** 639–646 (1993).
20. Rosenberg, N. A. *et al.* Genetic Structure of Human Populations. *Science* **298,** 2381–2385 (2002).
21. Ramachandran, S. *et al.* Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 15942–15947 (2005).
22. Hallatschek, O. & Nelson, D. R. Gene surfing in expanding populations. *Theoretical Population Biology* **73,** 158–170 (2008).
23. Provine, W. B. Ernst Mayr: Genetics and Speciation. *Genetics* **167,** 1041–1046 (2004).
24. Klopfstein, S., Currat, M. & Excoffier, L. The Fate of Mutations Surfing on the Wave of a Range Expansion. *Molecular Biology and Evolution* **23,** 482–490 (2006).
25. Birzu, G., Hallatschek, O. & Korolev, K. S. Fluctuations uncover a distinct class of traveling waves. *Proceedings of the National Academy of Sciences of the United States of America* **115,** E3645–E3654 (2018).
26. Fisher, R. A. The Wave of Advance of Advantageous Genes. *Annals of Eugenics* **7,** 355–369 (1937).
27. Kolmogorov, A., Petrovskii, I. & Piscunov, N. A study of the equation of diffusion with increase in the quantity of matter, and its application to a biological problem. *Moscow University Mathematics Bulletin* **1,** 1–25 (1937).
28. Birzu, G., Matin, S., Hallatschek, O. & Korolev, K. S. Genetic drift in range expansions is very sensitive to density dependence in dispersal and growth. *Ecology Letters* **22,** 1817–1827 (2019).
29. Van Saarloos, W. Front propagation into unstable states. *Physics Reports* **386,** 29–222 (2003).
30. Murray, J. D. *Mathematical Biology* (Springer, 2013).
31. Brunet, É. & Derrida, B. Effect of Microscopic Noise on Front Propagation. *Journal of Statistical Physics* **103,** 269–282 (2001).
32. Allee, W. C. Co-Operation Among Animals. *American Journal of Sociology* **37,** 386–398 (1931).
33. Courchamp, F., Clutton-Brock, T. & Grenfell, B. Inverse density dependence and the Allee effect. *Trends in Ecology & Evolution* **14,** 405–410 (1999).
34. Kramer, A. M., Dennis, B., Liebhold, A. M. & Drake, J. M. The evidence for Allee effects. *Population Ecology* **51,** 341–354 (2009).

35. Roques, L., Garnier, J., Hamel, F. & Klein, E. K. Allee effect promotes diversity in traveling waves of colonization. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 8828–8833 (2012).

36. Johnson, D. M., Liebhold, A. M., Tobin, P. C. & Bjørnstad, O. N. Allee effects and pulsed invasion by the gypsy moth. *Nature* **444,** 361–363 (2006).

37. Veit, R. R. & Lewis, M. A. Dispersal, population growth, and the allee effect: Dynamics of the house finch invasion of eastern North America. *The American Naturalist* **148,** 255–274 (1996).

38. Phillips, B. L., Brown, G. P., Webb, J. K. & Shine, R. Invasion and the evolution of speed in toads. *Nature* **439,** 803 (2006).

39. Shine, R., Brown, G. P. & Phillips, B. L. An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 5708–5711 (2011).

40. Gandhi, S. R., Yurtsev, E. A., Korolev, K. S. & Gore, J. Range expansions transition from pulled to pushed waves as growth becomes more cooperative in an experimental microbial population. *Proceedings of the National Academy of Sciences* **113,** 6922–6927 (2016).

41. Gandhi, S. R., Korolev, K. S. & Gore, J. Cooperation mitigates diversity loss in a spatially expanding microbial population. *Proceedings of the National Academy of Sciences of the United States of America* **116,** 23582–23587 (2019).

42. Abedon, S. T. *Bacteriophage Ecology Population Growth, Evolution, and Impact of Bacterial Viruses* (Cambridge University Press, Cambridge, 2008).

43. Twort, F. An Investigation on the Nature of Ultra-Microscopic Viruses. *The Lancet* **186,** 1241–1243 (1915).

44. d'Herelle, F. An invisible antagonist microbe of dysentery bacillus. *Comptes Rendus Hebdomadaires des Seances de L'Acadamie des Sciences* **165,** 373–375 (1917).

45. Abedon, S. T., Kuhl, S. J., Blasdel, B. G. & Kutter, E. M. Phage treatment of human infections. *Bacteriophage* **1,** 66–85 (2011).

46. Bruynoghe, R. & Maisin, J. Therapeutic testing using the staphylococcus bacteriophage. *Comptes Rendus des Seances de la Societe de Biologie et de ses Filiales* **85,** 1120–1121 (1921).

47. Morozova, V. V., Vlassov, V. V. & Tikunova, N. V. Applications of Bacteriophages in the Treatment of Localized Infections in Humans. *Frontiers in Microbiology* **9,** 1696 (2018).

48. Moye, Z. *et al.* Bacteriophage Applications for Food Production and Processing. *Viruses* **10,** 205 (2018).

49. Heineman, R. H. & Bull, J. J. Testing optimality with experimental evolution: lysis time in a bacteriophage. *Evolution* **61,** 1695–1709 (2007).

50. Kutter, E. & Sulakvelidze, A. *Bacteriophages: Biology and Applications* (CRC Press, 2004).

51. Wang, I.-N. Lysis Timing and Bacteriophage Fitness. *Genetics* **172,** 17–26 (2006).

52. Hay, I. D. & Lithgow, T. Filamentous phages: masters of a microbial sharing economy. *EMBO Reports* **20,** e47427 (2019).

53. Dunn, J. J., Studier, F. W. & Gottesman, M. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *Journal of Molecular Biology* **166,** 477–535 (1983).

54. Serwer, P., Wright, E. T., Hakala, K. W. & Weintraub, S. T. Evidence for bacteriophage T7 tail extension during DNA injection. *BMC Research Notes* **1,** 36 (2008).

55. Martin, S. J. *The biochemistry of viruses* (Cambridge University Press, Cambridge, 1978).

56. Burch, C. L. & Chao, L. Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature* **406,** 625–628 (2000).

57. Koch, A. L. The growth of viral plaques during the enlargement phase. *Journal of Theoretical Biology* **6,** 413–431 (1964).

58. Yin, J. & McCaskill, J. S. Replication of Viruses in a Growing Plaque - A Reaction-Diffusion Model. *Biophysical Journal* **61,** 1540–1549 (1992).

59. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* **8,** 317–327 (2010).

60. Braun, V., Killmann, H. & Herrmann, C. Inactivation of FhuA at the cell surface of Escherichia coli K-12 by a phage T5 lipoprotein at the periplasmic face of the outer membrane. *Journal of Bacteriology* **176,** 4710–4717 (1994).

61. Pedruzzi, I., Rosenbusch, J. P. & Locher, K. P. Inactivation in vitro of the Escherichia coli outer membrane protein FhuA by a phage T5-encoded lipoprotein. *FEMS Microbiology Letters* **168,** 119–125 (1998).

62. Fort, J. & Méndez, V. Time-delayed spread of viruses in growing plaques. *Physical Review Letters* **89,** 178101 (2002).

63. Stollar, D. & Levine, L. Two-dimensional immunodiffusion. *Methods in Enzymology* **6,** 848–854 (1963).

64. Ackermann, H. W. Classification of tailed enterobacteria phages. *Pathologie-biologie* **24,** 359 (1976).

65. Fricke, H. A mathematical treatment of the electric conductivity and capacity of disperse systems I. The electric conductivity of a suspension of homogeneous spheroids. *Physical Review* **24,** 575–587 (1924).

66. Alvarez, L. J., Thomen, P., Makushok, T. & Chatenay, D. Propagation of fluorescent viruses in growing plaques. *Biotechnology and Bioengineering* **96,** 615–621 (2007).

67. Abedon, S. T. & Culler, R. R. Bacteriophage evolution given spatial constraint. *Journal of Theoretical Biology* **248,** 111–119 (2007).

68. You, L. & Yin, J. Amplification and Spread of Viruses in a Growing Plaque. *Journal of Theoretical Biology* **200,** 365–373 (1999).

69. Fort, J. & Méndez, V. Time-delayed theory of the neolithic transition in europe. *Physical Review Letters* **82,** 867–870 (1999).

70. Jones, D. A., Smith, H. L., Thieme, H. R. & Röst, G. On spread of phage infection of bacteria in a petri dish. *SIAM Journal on Applied Mathematics* **72,** 670–688 (2012).

71. De Rioja, V., Fort, J. & Isern, N. Front propagation speeds of T7 virus mutants. *Journal of Theoretical Biology* **385,** 112–118 (2015).

72. Ortega-Cejas, V., Fort, J., Méndez, V. & Campos, D. Approximate solution to the speed of spreading viruses. *Physical Review E* **69,** 031909 (2004).

73. Amor, D. R. & Fort, J. Virus infection speeds: Theory versus experiment. *Physical Review E* **82,** 061905 (2010).

74. Adams, M. H. *Bacteriophages* (Interscience Publishers, 1959).

75. Roux, S., Hallam, S. J., Woyke, T. & Sullivan, M. B. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *eLife* **4,** e08490 (2015).

76. Díaz-Muñoz, S. L. Viral coinfection is shaped by host ecology and virus-virus interactions across diverse microbial taxa and environments. *Virus Evolution* **3,** vex011 (2017).

77. Turner, P. E. & Duffy, S. in *Bacteriophage Ecology* 195–216 (Cambridge University Press, Cambridge, 2008).

78. Van Houte, S., Buckling, A. & Westra, E. R. Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiology and molecular biology reviews* **80,** 745–763 (2016).

79. Lu, M. J., Stierhof, Y. D. & Henning, U. Location and unusual membrane topology of the immunity protein of the Escherichia coli phage T4. *Journal of Virology* **67,** 4905–4913 (1993).

80. Lu, M. J. & Henning, U. Superinfection exclusion by T-even-type coliphages. *Trends in Microbiology* **2,** 137–139 (1994).

81. McAllister, W. T. & Barrett, C. L. Superinfection exclusion by bacteriophage T7. *Journal of Virology* **24,** 709 (1977).

82. Hirsch-Kauffmann, M., Pfennig-Yeh, M. l., Ponta, H. & Herrlich, P. A virus-specified mechanism for the prevention of multiple infection—T7- and T3-mutual and superinfection exclusion. *Molecular and General Genetics* **149,** 243–249 (1976).

83. Abedon, S. T., Hyman, P. & Thomas, C. Experimental Examination of Bacteriophage Latent-Period Evolution as a Response to Bacterial Availability. *Applied and Environmental Microbiology* **69,** 7499–7506 (2003).

84. Wang, I. N., Dykhuizen, D. E. & Slobodkin, L. B. The evolution of phage lysis timing. *Evolutionary Ecology* **10,** 545–558 (1996).

85. Bull, J. J., Millstein, J., Orcutt, J. & Wichman, H. A. Evolutionary Feedback Mediated through Population Density, Illustrated with Viruses in Chemostats. *The American Naturalist* **167,** 39–51 (2006).

86. Bull, J. J. Optimality models of phage life history and parallels in disease evolution. *Journal of theoretical biology* **241,** 928–938 (2006).

87. Gallet, R., Kannoly, S. & Wang, I. N. Effects of bacteriophage traits on plaque formation. *BMC Microbiology* **11,** 181 (2011).

88. Yin, J. Evolution of bacteriophage T7 in a growing plaque. *Journal of Bacteriology* **175,** 1272–1277 (1993).

89. Roychoudhury, P., Shrestha, N., Wiss, V. R. & Krone, S. M. Fitness benefits of low infectivity in a spatially structured population of bacteriophages. *Proceedings of The Royal Society B: Biological Sciences* **281,** 20132563 (2014).

90. Kerr, B., Neuhauser, C., Bohannan, B. J. M. & Dean, A. M. Local migration promotes competitive restraint in a host-pathogen 'tragedy of the commons'. *Nature* **442,** 75–78 (2006).

91. Frank, S. A. All of life is social. *Current Biology* **17,** R648–R650 (2007).

92. Weitz, J. S., Mileyko, Y., Joh, R. I. & Voit, E. O. Collective decision making in bacterial viruses. *Biophysical Journal* **95,** 2673–2680 (2008).

93. Refardt, D. Within-host competition determines reproductive success of temperate bacteriophages. *ISME Journal* **5,** 1451–1460 (2011).

94. Ojosnegros, S., Perales, C., Mas, A. & Domingo, E. Quasispecies as a matter of fact: Viruses and beyond. *Virus Research* **162,** 203–215 (2011).

95. Díaz-Muñoz, S. L., Sanjuán, R. & West, S. Sociovirology: Conflict, Cooperation, and Communication among Viruses. *Cell Host and Microbe* **22,** 437–441 (2017).

96. Koelle, K., Farrell, A. P., Brooke, C. B. & Ke, R. Within-host infectious disease models accommodating cellular coinfection, with an application to influenza. *Virus Evolution* **5,** 18 (2019).

97. Iranzo, J., Faure, G., Wolf, Y. I. & Koonin, E. V. Game-Theoretical Modeling of Interviral Conflicts Mediated by Mini-CRISPR Arrays. *Frontiers in Microbiology* **11,** 381 (2020).

98. Vafadar, S., Shahdoust, M., Kalirad, A., Zakeri, P. & Sadeghi, M. Competitive exclusion during co-infection as a strategy to prevent the spread of a virus: A computational perspective. *PLoS ONE* **16,** e0247200 (2021).

99. Sanjuán, R. & Domingo-Calap, P. in *Encyclopedia of Virology* 53–61 (Elsevier, 2021).

100. Bretscher, M. T., Althaus, C. L., Müller, V. & Bonhoeffer, S. Recombination in HIV and the evolution of drug resistance: for better or for worse? *BioEssays* **26,** 180–188 (2004).

101. Vijay, N. N., Vasantika, Ajmani, R., Perelson, A. S. & Dixit, N. M. Recombination increases human immunodeficiency virus fitness, but not necessarily diversity. *Journal of General Virology* **89,** 1467–1477 (2008).

102. Weller, S. K. & Sawitzke, J. A. Recombination Promoted by DNA Viruses: Phage λ to Herpes Simplex Virus. *Annual Review of Microbiology* **68,** 237–258 (2014).

103. Gao, H. & Feldman, M. W. Complementation and epistasis in viral coinfection dynamics. *Genetics* **182,** 251–263 (2009).

104. Froissart, R. *et al.* Co-infection Weakens Selection Against Epistatic Mutations in RNA Viruses. *Genetics* **168,** 9–19 (2004).

105. García-Arriaza, J., Manrubia, S. C., Toja, M., Domingo, E. & Escarmís, C. Evolutionary Transition toward Defective RNAs That Are Infectious by Complementation. *Journal of Virology* **78,** 11678–11685 (2004).

106. García-Arriaza, J., Ojosnegros, S., Dávila, M., Domingo, E. & Escarmís, C. Dynamics of Mutation and Recombination in a Replicating Population of Complementing, Defective Viral Genomes. *Journal of Molecular Biology* **360,** 558–572 (2006).

107. Gelderblom, H. C. *et al.* Viral complementation allows HIV-1 replication without integration. *Retrovirology* **5,** 60 (2008).

108. Turner, P. E. & Chao, L. Sex and the Evolution of Intrahost Competition in RNA Virus ϕ6. *Genetics* **150,** 523–532 (1998).

109. Turner, P. E. & Chao, L. Prisoner's dilemma in an RNA virus. *Nature* **398,** 441–443 (1999).

110. Turner, P. E. & Chao, L. Escape from prisoner's dilemma in RNA phage ϕ6. *The American Naturalist* **161,** 497–505 (2003).

111. Dennehy, J. J., Duffy, S., O'Keefe, K. J., Edwards, S. V. & Turner, P. E. Frequent Coinfection Reduces RNA Virus Population Genetic Diversity. *Journal of Heredity* **104,** 704–712 (2013).

112. Donahue, D. A., Bastarache, S. M., Sloan, R. D. & Wainberg, M. A. Latent HIV-1 Can Be Reactivated by Cellular Superinfection in a Tat-Dependent Manner, Which Can Lead to the Emergence of Multidrug-Resistant Recombinant Viruses. *Journal of Virology* **87,** 9620–9632 (2013).

113. Chao, L. Evolution of sex in RNA viruses. *Trends in Ecology and Evolution* **7,** 147–151 (1992).

114. Asatryan, A., Wodarz, D. & Komarova, N. L. New virus dynamics in the presence of multiple infection. *Journal of Theoretical Biology* **377,** 98–109 (2015).

115. Dixit, N. M. & Perelson, A. S. Multiplicity of Human Immunodeficiency Virus Infections in Lymphoid Tissue. *Journal of Virology* **78,** 8942–8945 (2004).

116. Dixit, N. M. & Perelson, A. S. HIV dynamics with multiple infections of target cells. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 8198–8203 (2005).

117. Althaus, C. L. & Bonhoeffer, S. Stochastic Interplay between Mutation and Recombination during the Acquisition of Drug Resistance Mutations in Human Immunodeficiency Virus Type 1. *Journal of Virology* **79,** 13572–13578 (2005).

118. Fraser, C. HIV recombination: What is the impact on antiretroviral therapy? *Journal of the Royal Society Interface* **2,** 489–503 (2005).

119. Wodarz, D. & Levy, D. N. Effect of different modes of viral spread on the dynamics of multiply infected cells in human immunodeficiency virus infection. *Journal of the Royal Society Interface* **8,** 289–300 (2011).

120. Cummings, K. W., Levy, D. N. & Wodarz, D. Increased burst size in multiply infected cells can alter basic virus dynamics. *Biology Direct* **7,** 16 (2012).

121. May, R. M. & Nowak, M. A. Superinfection, metapopulation dynamics, and the evolution of diversity. *Journal of Theoretical Biology* **170,** 95–114 (1994).

122. Nowak, M. A. & May, R. M. Superinfection and the evolution of parasite virulence. *Proceedings of the Royal Society B: Biological Sciences* **255,** 81–89 (1994).

123. Van Baalen, M. & Sabelis, M. W. The dynamics of multiple infection and the evolution of virulence. *The American Naturalist* **146,** 881–910 (1995).

124. Alizon, S. & Van Baalen, M. Multiple infections, immune dynamics, and the evolution of virulence. *The American Naturalist* **172,** 150–168 (2008).

125. Alizon, S., de Roode, J. C. & Michalakis, Y. Multiple infections and the evolution of virulence. *Ecology Letters* **16,** 556–567 (2013).

126. Leeks, A., Segredo-Otero, E. A., Sanjuán, R. & West, S. A. Beneficial coinfection can promote within-host viral diversity. *Virus Evolution* **4,** vey028 (2018).

127. Gourley, S. A. & Kuang, Y. A delay reaction-diffusion model of the spread of bacteriophage infection. *SIAM Journal on Applied Mathematics* **65,** 550–566 (2005).

128. Karam, J. D., Drake, J. W., Kreuzer, K. N., Hall, D. H. & Mosig, G. *Molecular Biology of Bacteriophage T4* (American Society for Microbiology, 1994).

129. Choua, M. & Bonachela, J. A. Ecological and evolutionary consequences of viral plasticity. *The American Naturalist* **193,** 346–358 (2019).

130. Phan, D. & Wodarz, D. Modeling multiple infection of cells by viruses: Challenges and insights. *Mathematical Biosciences* **264,** 21–28 (2015).

131. Wodarz, D., Levy, D. N. & Komarova, N. L. Multiple infection of cells changes the dynamics of basic viral evolutionary processes. *Evolution Letters* **3,** 104–115 (2019).

132. Matthysen, E. Density-dependent dispersal in birds and mammals. *Ecography* **28,** 403–416 (2005).

133. Möbius, W., Murray, A. W. & Nelson, D. R. How Obstacles Perturb Population Fronts and Alter Their Genetic Structure. *PLoS Computational Biology* **11,** e1004615 (2015).

134. Qimron, U., Marintcheva, B., Tabor, S. & Richardson, C. C. Genomewide screens for Escherichia coli genes affecting growth of T7 bacteriophage. *Proceedings of the National Academy of Sciences* **103,** 19039–19044 (2006).

135. Crank, J. & Crank, E. P. J. *The Mathematics of Diffusion* (Clarendon Press, 1979).

136. Narayanan, J., Xiong, J. Y. & Liu, X. Y. Determination of agarose gel pore size: Absorbance measurements vis a vis other techniques. *Journal of Physics: Conference Series* **28,** 83–86 (2006).

137. Amor, D. R. & Fort, J. Cohabitation reaction–diffusion model for virus focal infections. *Physica A: Statistical Mechanics and its Applications* **416,** 611–619 (2014).

138. Park, H. J. & Gokhale, C. S. Ecological feedback on diffusion dynamics. *Royal Society Open Science* **6,** 181273 (2019).

139. Cates, M. E., Marenduzzo, D., Pagonabarraga, I. & Tailleur, J. Arrested phase separation in reproducing bacteria creates a generic route to pattern formation. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 11715–11720 (2010).

140. Perry, L. L. *et al.* Sequence analysis of Escherichia coli O157:H7 bacteriophage Φv10 and identification of a phage-encoded immunity protein that modifies the O157 antigen. *FEMS Microbiology Letters* **292,** 182–186 (2009).

141. Newton, G. J. *et al.* Three-component-mediated serotype conversion in Pseudomonas aeruginosa by bacteriophage D3. *Molecular Microbiology* **39,** 1237–1247 (2004).

142. Bondy-Denomy, J. *et al.* Prophages mediate defense against phage infection through diverse mechanisms. *ISME Journal* **10,** 2854–2866 (2016).

143. Walsh, D. & Naghavi, M. H. Exploitation of Cytoskeletal Networks during Early Viral Infection. *Trends in Microbiology* **27,** 39–50 (2019).

144. Doceul, V., Hollinshead, M., Van Der Linden, L. & Smith, G. L. Repulsion of superinfecting virions: A mechanism for rapid virus spread. *Science* **327,** 873–876 (2010).

145. Gal-On, A. & Shiboleth, Y. M. in *Natural Resistance Mechanisms of Plants to Viruses* 261–288 (Springer Netherlands, 2006).

146. Folimonova, S. Y. Superinfection Exclusion Is an Active Virus-Controlled Function That Requires a Specific Viral Protein. *Journal of Virology* **86,** 5554–5561 (2012).

147. Bergua, M. *et al.* A Viral Protein Mediates Superinfection Exclusion at the Whole-Organism Level but Is Not Required for Exclusion at the Cellular Level. *Journal of Virology* **88,** 11327–11338 (2014).

148. Panja, D. Effects of fluctuations on propagating fronts. *Physics Reports* **393,** 87–174 (2004).

149. Matsuyama, T., Komatsu, K., Nakahara, A. & Matsushita, M. Experimental Investigation on the Validity of Population Dynamics Approach to Bacterial Colony Formation. *Journal of the Physical Society of Japan* **63,** 1205–1211 (1994).

150. Giometto, A., Rinaldo, A., Carrara, F. & Altermatt, F. Emerging predictable features of replicated biological invasion fronts. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 297–301 (2014).

151. Peischl, S., Kirkpatrick, M. & Excoffier, L. Expansion load and the evolutionary dynamics of a species range. *The American Naturalist* **185,** 81–93 (2015).

152. Bosshard, L. *et al.* Accumulation of Deleterious Mutations During Bacterial Range Expansions. *Genetics* **207,** 669–684 (2017).

153. Kaczmarczyk, A., Vorholt, J. A. & Francez-Charlot, A. Cumate-inducible gene expression system for sphingomonads and other Alphaproteobacteria. *Applied and Environmental Microbiology* **79,** 6795–6802 (2013).

154. Kapoor, G., Saigal, S. & Elongavan, A. Action and resistance mechanisms of antibiotics: A guide for clinicians. *Journal of Anaesthesiology, Clinical Pharmacology* **33,** 300 (2017).

155. Volkmer, B. & Heinemann, M. Condition-Dependent Cell Volume and Concentration of Escherichia coli to Facilitate Data Conversion for Systems Biology Modeling. *PLoS ONE* **6,** e23126 (2011).

156. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9,** 676–682 (2012).

157. Rueden, C. T. *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18,** 529 (2017).

158. Klein, G. Mean first-passage times of Brownian motion and related problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **211,** 431–443 (1952).

159. Ebert, U. & Van Saarloos, W. Front propagation into unstable states: Universal algebraic convergence towards uniformly translating pulled fronts. *Physica D: Nonlinear Phenomena* **146,** 1–99 (2000).

160. Hallatschek, O., Hersen, P., Ramanathan, S. & Nelson, D. R. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 19926–19930 (2007).

161. Korolev, K. S., Avlund, M., Hallatschek, O. & Nelson, D. R. Genetic demixing and evolution in linear stepping stone models. *Reviews of Modern Physics* **82,** 1691–1718 (2010).

162. Korolev, K. S., Xavier, J. B., Nelson, D. R. & Foster, K. R. A quantitative test of population genetics using spatiogenetic patterns in bacterial colonies. *The American Naturalist* **178,** 538–552 (2011).

163. Vidakovic, L., Singh, P. K., Hartmann, R., Nadell, C. D. & Drescher, K. Dynamic biofilm architecture confers individual and collective mechanisms of viral protection. *Nature Microbiology* **3,** 26–31 (2018).

164. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9,** 62–66 (1979).

165. Storms, Z. J. & Sauvageau, D. Modeling tailed bacteriophage adsorption: Insight into mechanisms. *Virology* **485,** 355–362 (2015).

166. Ellis, E. L. & Delbrück, M. The Growth of Bacteriophage. *The Journal of general physiology* **22,** 365–384 (1939).

167. You, L., Suthers, P. F. & Yin, J. Effects of Escherichia coli physiology on growth of phage T7 in vivo and in silico. *Journal of Bacteriology* **184,** 1888–1894 (2002).

168. Rabinovitch, A., Hadas, H., Einav, M., Melamed, Z. & Zaritsky, A. Model for bacteriophage T4 development in Escherichia coli. *Journal of Bacteriology* **181,** 1677–1683 (1999).

169. Campos, D., Méndez, V. & Fedotov, S. The effects of distributed life cycles on the dynamics of viral infections. *Journal of Theoretical Biology* **254,** 430–438 (2008).

170. Eriksen, R. S., Mitarai, N. & Sneppen, K. On Phage Adsorption to Bacterial Chains. *Biophysical Journal* **119,** 1896–1904 (2020).

171. Hunter, M., Krishnan, N., Liu, T., Möbius, W. & Fusco, D. Virus-Host Interactions Shape Viral Dispersal Giving Rise to Distinct Classes of Traveling Waves in Spatial Expansions. *Phys. Rev. X* **11,** 21066 (2021).

172. McAllister, W. T. & Wu, H. L. Regulation of transcription of the late genes of bacteriophage T7. *Proceedings of the National Academy of Sciences of the United States of America* **75,** 804–808 (1978).

173. Bryson, V. & Szybalski, W. Microbial selection. *Science* **116,** 45–51 (1952).

174. Gresham, D. & Dunham, M. J. The enduring utility of continuous culturing in experimental evolution. *Genomics* **104,** 399–405 (2014).

175. Nguyen, H. M. & Kang, C. Lysis Delay and Burst Shrinkage of Coliphage T7 by Deletion of Terminator Tϕ Reversed by Deletion of Early Genes. *Journal of Virology* **88,** 2107–2115 (2014).

176. Dinh, K. N., Corey, S. J. & Kimmel, M. Application of the Moran Model in Estimating Selection Coefficient of Mutated CSF3R Clones in the Evolution of Severe Congenital Neutropenia to Myeloid Neoplasia. *Frontiers in Physiology* **11,** 806 (2020).

177. Delbrück, M. Interference between bacterial viruses; the mutual exclusion effect and the depressor effect. *Journal of bacteriology* **50,** 151–170 (1945).

178. Abedon, S. T. Lysis of lysis-inhibited bacteriophage T4-infected cells. *Journal of Bacteriology* **174,** 8073–8080 (1992).

179. Wahl, L. M. & DeHaan, C. S. Fixation probability favors increased fecundity over reduced generation time. *Genetics* **168,** 1009–1018 (2004).

180. Shao, Y. & Wang, I. N. Bacteriophage adsorption rate and optimal lysis time. *Genetics* **180,** 471–482 (2008).

181. Spanakis, E. & Horne, M. T. Co-adaptation of Escherichia coli and coliphage γvir in continuous culture. *Journal of General Microbiology* **133,** 353–360 (1987).

182. Mizoguchi, K. *et al.* Coevolution of bacteriophage PP01 and Escherichia coli O157:H7 in continuous culture. *Applied and Environmental Microbiology* **69,** 170–176 (2003).

183. Koskella, B. & Brockhurst, M. A. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews* **38,** 916–931 (2014).

184. Bull, J. J., Heineman, R. H. & Wilke, C. O. The phenotype-fitness map in experimental evolution of phages. *PLoS ONE* **6,** 27796 (2011).

185. Preston, S., Heuveline, P. & Guillot, M. *Demography: Measuring and Modeling Population Processes* (Wiley, 2000).

186. Nowak, M. A. *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, 2006).

187. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297,** 1183–1186 (2002).

188. Lin, J. & Amir, A. The Effects of Stochasticity at the Single-Cell Level and Cell Size Control on the Population Growth. *Cell Systems* **5,** 358–367 (2017).

189. Ho, P.-Y., Lin, J. & Amir, A. Modeling Cell Size Regulation: From Single-Cell-Level Statistics to Molecular Mechanisms and Population-Level Effects. *Annual Review of Biophysics* **47,** 251–271 (2018).

190. Barber, F., Min, J., Murray, A. W. & Amir, A. Modeling the impact of single-cell stochasticity and size control on the population growth rate in asymmetrically dividing cells. *PLoS Computational Biology* **17,** e1009080 (2021).

191. Hashimoto, M. *et al.* Noise-driven growth rate gain in clonal cellular populations. *Proceedings of the National Academy of Sciences of the United States of America* **113,** 3151–3156 (2016).

192. Levien, E., Kondev, J. & Amir, A. The interplay of phenotypic variability and fitness in finite microbial populations. *Journal of the Royal Society Interface* **17,** 0190827 (2020).

193. Wang, P. *et al.* Robust Growth of Escherichia coli. *Current Biology* **20,** 1099–1103 (2010).

194. Norman, T. M., Lord, N. D., Paulsson, J. & Losick, R. Memory and modularity in cell-fate decision making. *Nature* **503,** 481–486 (2013).

195. Bakshi, S. *et al.* Tracking bacterial lineages in complex and dynamic environments with applications for growth control and persistence. *Nature Microbiology* **6,** 783–791 (2021).

196. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture medium for enterobacteria. *Journal of Bacteriology* **119,** 736–747 (1974).

197. Stoddard, A. & Rolland, V. I see the light! Fluorescent proteins suitable for cell wall/apoplast targeting in Nicotiana benthamiana leaves. *Plant Direct* **3,** e00112 (2019).

198. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology* **22,** 1567–1572 (2004).

199. Eriksen, R. S., Svenningsen, S. L., Sneppen, K. & Mitarai, N. A growing microcolony can survive and support persistent propagation of virulent phages. *Proceedings of the National Academy of Sciences of the United States of America* **115,** 337–342 (2017).

200. Taylor, T. B. & Buckling, A. Bacterial motility confers fitness advantage in the presence of phages. *Journal of Evolutionary Biology* **26,** 2154–2160 (2013).

Appendix A

Bacteria Strains

Strain	**Genotype/Characteristics**	**Source**
BW25113	*E. coli* BW25113 WT (CGSC# 7636)	-
eWM43	BW25113, with plasmid (see [133]) conferring resistance to ampicillin, and expressing venus YFP	Ref. [133]
eMTH43	eWM43, with second plasmid pAK591 (Addgene# 48107) conferring resistance to chloramphenicol	Chapter 2
eWM44	*E. coli* ΔwaaC, is resistant to phage infection, with plasmid (see [133]) conferring resistance to ampicillin and expressing mCherry	Ref. [133]
SB8	Constructed by a P1 transduction from the Keio collection strain CGSC#:9565, ΔmotA743::kan into *E. coli* MG1655, and flippase was used to remove the kanamycin resistance	Ref. [195]
SB7	Constructed by P1 transducing *glmS*::P_{RNAI}-$mCherry_{1-11}$-mKate-T1 terminator-FRT Kan FRT::pstS into SB8. Expresses mCherry.	Ref. [195]