An Examination on Phages as a Naturally Composable System

by

Bryan Wu

October 2025

A thesis submitted to the
Graduate School
of the
Institute of Science and Technology Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Committee in charge:

Caroline Muller, Chair
Calin Guet
Gašper Tkačik
Stephen Abedon

The thesis of Bryan Wu, titled *An Examination on Phages as a Naturally Composable System*, is approved by:

Supervisor: Calin Guet, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Gašper Tkačik, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Stephen Abedon, Ohio State University, Ohio, USA

Signature: _____

Defense Chair: Caroline Muller, ISTA, Klosterneuburg, Austria

Signature: _____

Signed page is on file

ISTA Thesis, ISSN: 2663-337X

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I accept full responsibility for the content and factual accuracy of this work, including the data and their analysis and presentation, and the text and citation of other work.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: _____

Signed page is on file

Abstract

Systems design has classically relied on composable systems, in which individual subsystems have defined inputs, outputs, and interactions with each other; however, attempts at designing complex systems in synthetic biology has often run in to issues of crosstalk and interference, given that these systems must function within the context of the host. In nature, mobile genetic elements are systems that have evolved to travel between hosts, and thus appear to be a good candidate with which to evaluate composability. Selecting temperate phages as a model system, I used mathematical modelling to identify sources of information that temperate phages should respond to. I found that essential proteins of temperate phages can interfere with potential hosts, indicating limitations to composability. I also designed a lysogeny reporter construct and characterize its behavior across various laboratory and environmental strains, finding differences in phage lambda lysogens, and potential interference from prophages that already exist within the environmental strains. Although the information gathered is not conclusive, it suggests that composability is not a key property of temperate phages, implying that biological systems may not be composable, and that other system design principles should be considered when designing synthetic systems.

Acknowledgments

I have been fortunate enough to have worked with a large number of non-mutually exclusive mentors, collaborators and friends over the course of my PhD, without whom this thesis would not exist.

I would like to thank Remy Chait, Kirti Jain and Bor Kavčič for their guidance and help when I was first starting my PhD. Many thanks to Moritz Lang, Pavel Payne, Maros Pleška, Mato Lagator, Rok Grah, and Claudia Igler for discussions on the direction of the project. I greatly appreciate Nathalie Gruber, Antoine Loussouarn and Roderich Römhild for putting up with my many relevant and irrelevant questions.

And finally thanks to Mike and Maura Hennessey-Wesen, for keeping me sane so far from home.

About the Author

Bryan Wu completed his BSc in Biochemistry from McMaster University in Hamilton, Ontario, Canada. His MSc in Microbiology at McMaster was on quorum sensing systems in *Streptococcus*, and continued in his research interests on decision-making in bacteria both *in silico* and *in vivo* during his PhD. He presented his research results at the Viruses of Microbes 2024 conference in Cairns, Australia.

List of Collaborators and Publications

Antoine Loussouarn performed experiments with non-lambda phages, and aided in the experiments determining lysogeny rates. The defenseless MG1655 strain was developed by Nathalie Gruber, and the ΔIS strains were developed by Dr. Roderich Römhild and Athina Diakogianni. Collaborators in the publications are noted there.

Wu, B. & Guet, CC. 2025. Responsive Lysogeny under Nonproductive Phage Binding. Under Revision for Evolution (Oxford Publishing).

Igler, C., **Wu, B.**, Fourcade, C., Kavčič, B., Waldminghaus, T., Pauler, F.M., Santhanam, B., Tkačik, G., Guet, CC. 2025. Non-cognate molecular interactions impact bacterial growth. Under Revision. Previous version available on BioRxiv at https://www.biorxiv.org/content/10.1101/2021.10.20.465141v1.

Table of Contents

List of Tables

List of Terms

Composability	The ability to combine individual components into larger systems.
Bacteriophage	A virus of bacteria.
Phage	See bacteriophage.
Lysis	Bursting of the cell, often caused by phage reproduction.
Lysogeny	The process by which the phage genome integrates itself into the host
Induction	The process by which the prophage begins producing phage particles.
Lysogen	A host cell that has been colonized by a prophage.
Prophage	A phage genome integrated into a host genome.
Cryptic prophage	A prophage that can no longer induce.
Temperate phage	A phage capable of undergoing lysogeny.
Lysogenic phage	See temperate phage.
Virulent phage	A phage that can only undergo lysis.
Lytic phage	See virulent phage.

1 Introduction

Cells are machines. Complex machines, more complex than a single person can fully understand, but machines nevertheless. A cell fundamentally functions as a series of physical steps, each entailing the next, carrying out a program put in motion by the initial state of its internal components and its environment. As biologists we study this machine, identifying pieces and figuring out how they fit together, how components respond to the environment, how one step leads to the next, all in order to build an understanding of how the machine functions.

And as humans, we believe that we can control this machine.

1.1 Composable Systems in Synthetic Biology

In synthetic biology, we try to bend nature to our will. We believe that we have sufficient knowledge to have cells perform the instructions we transfer into them, to program them with the messages we write in DNA.

And, on a first order, it works. We can delete genes, and put new ones back in. We can create simple genetic circuits, which respond to chemical cues we add to the medium. We can also create simple logic gates, to perform simple processing on the inputs before expressing an output.

But any farther and it falters. Two circuits, that work independently, fail when interfaced with each other. A circuit that functions perfectly in one cell fails when placed in a different cell. Complex genetic circuits have been designed, but require multiple rounds of evolutionary iterations before they function anywhere close to as predicted (Castle *et al.*, 2024).

This is all an issue of composability: the ability to combine components into larger systems. Individual components must be able to receive their inputs and produce their proper outputs, independent of the functioning of the other components in the system. And currently in synthetic biology this fails. We are limited in the complexity and scope of our designs, as our tools do not work when we put them together.

Some of these issues are foreseeable, while other issues have only been identified in retrospect. If two components use the same transcription factors, then it is obvious that there will be crosstalk between them. If you overexpress too many proteins, eventually the cell will be unable to keep up. But if multiple components use the same degradation machinery, how many of those components can you include before the degradation machinery is overwhelmed? How many transcription factors can you express before off-target binding becomes an issue? Designing complex biological circuits is a blind reach into the dark, navigating a labyrinth only by touch.

Biology is not a wire guiding an output directly to an input; biology is a bag of marbles, inputs and outputs and internal states all colliding with one another, interfacing and interfering. More complex natural biological systems have solved this problem by creating membrane-bound organelles to separate one system from another, or by separating systems into completely different, differentiated cells. But for synthetic biology, and especially for the engineering of systems within single cells, the problems remain: are there limits to the

complexity of genetic circuits? Do composable systems even exist, or are we deluding ourselves into thinking that we can create systems unaffected by the noise of the context that surrounds them?

1.2 Mobile Genetic Elements

Nature has not been idle. Nature, too, has evolved systems that insert themselves into other cells, and try to carry out their program to the best of their ability. These are the mobile genetic elements, often referred to as selfish genetic elements, as many of them often carry out their program at a cost to their host.

But in many cases, the cost to the host cannot be too high. As these mobile genetic elements live within the host, death to the host also means death to the genetic elements that inhabit it. And thus, mobile genetic elements often exist in an uneasy truce with their host: selfish genetic elements carrying out their programming, while minimizing any interference with critical host functions (Haudiquet *et al.*, 2022).

What is the programming that the genetic elements carry out? As with all replicators, they exist to make copies of themselves. Insertion sequences and transposons are short sequences of DNA that cut, copy and paste themselves into other pieces of DNA, and sometimes that DNA is released into the environment and taken up by another cell, spreading the genetic element to a new host. Plasmids can encode their own systems to spread; as circular pieces of DNA, they can encode more complex systems such conjugation systems, with which they can inject a copy of themselves to bacteria in physical contact (Frost *et al.*, 2005).

And they want their host to be reproductively successful. Any daughter cells of the host should contain copies of any genetic elements that are present (and some selfish genetic elements work hard to ensure this), and thus success for the host translates into success for the genetic element.

Mobile genetic elements carry out their programs, while avoiding undue stress to their hosts. They are transferred from host to host, and function the best they can wherever they find themselves. As such, they appear to be a perfect system to examine for natural composability.

1.3 Phages

Phages are another class of mobile genetic element with an interesting twist; their fitness is less strictly tied to that of their host. For all the previously mentioned mobile genetic elements, their fitness is strictly tied to their host; if they cannot find another host to infect, then they remain within their host, and if their host dies, then they also die.

Phages, on the other hand, can be virulent. Virulent phages decouple their fitness from that of their host. Once a host is infected by a virulent phage, the phage hijacks the host's cellular machinery to make more copies of itself, before killing and lysing the host to spread its copies into the environment in search of new hosts. There is no sustained transmission of virulent phages vertically to any daughter cells[1], and there is no reason to keep the host alive.

[1] Note that this is not strictly true; in the case of pseudolysogeny a lytic phage is capable of delaying lysis for an extended period of time, being passed on to daughter cells; however, I consider this a form a lysogeny, and is similar to strategies deployed by temperate phages, as described below.

This strategy does not work in every environment. In environments with few hosts, or environments that are harsh to the free phage particles, a sizable fraction of the phage's offspring will be unable to find a host and reproduce. Virulent phages are therefore always at the mercy of their environment, doing well during the good times and badly during the hard times.

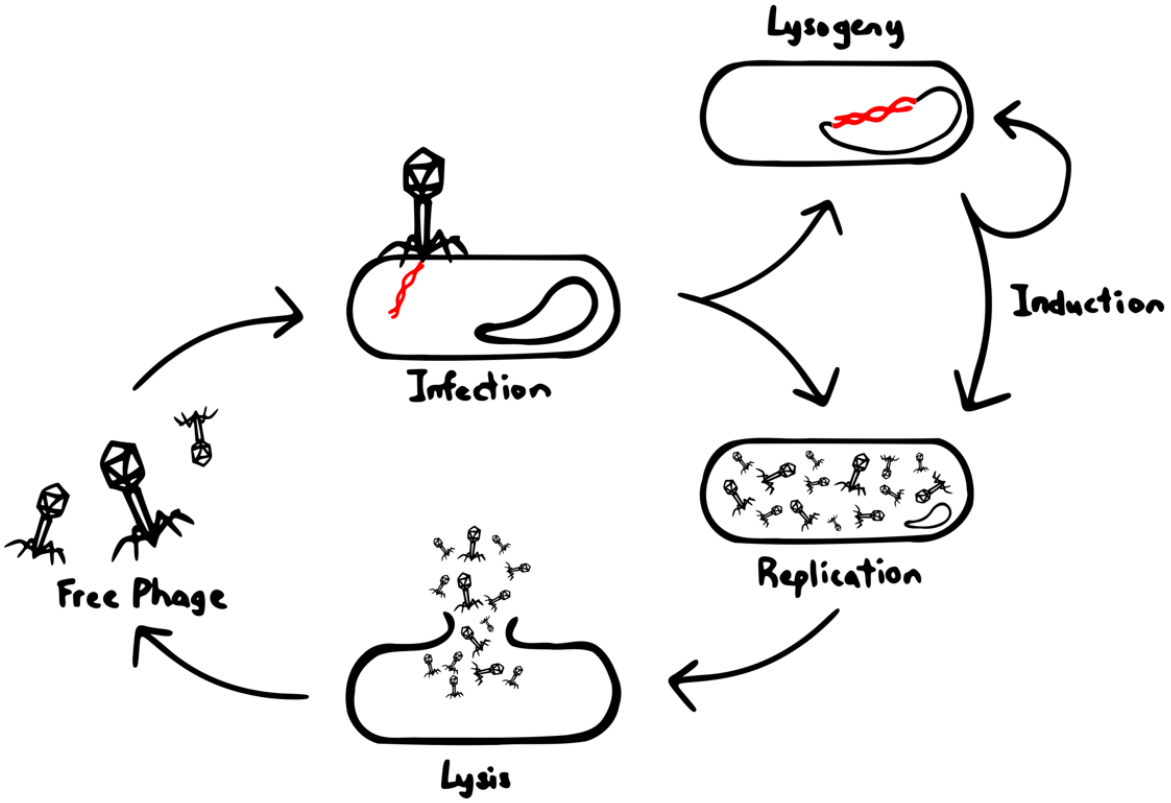


Figure 1: Lytic and lysogenic cycles of the temperate bacteriophage.

Free phage infect susceptible cells. In the lytic cycle, the phage genome (red) is replicated and produces a number of phage particles. Eventually the phages lyse the host cell, releasing the phage back into the environment. In the lysogenic cycle, the phage genome integrates into the genome of the host as a prophage, replicating when the host cell divides. The prophage can be induced by environmental conditions to exit the lysogenic cycle, excising its DNA from the host genome and leading to the replication and release of phages. Adapted from Oppenheim et al. (2005).

Temperate phages, on the other hand, can choose. Upon infecting the host, they can choose to kill and lyse their host and produce as many copies of themselves as possible, or if the environments seems unsuitable, they can go quiet, integrating themselves into the host genome or maintaining themselves as phage-plasmids, transforming the host into a lysogen and tying their own fitness to that of their host. Eventually, environmental conditions can change to conditions that support virulence, and the lysogen can induce, producing free phage particles to release into the environment once more.

Upon infection, temperate phages can choose between lysis and lysogeny, and that makes all the difference.

1.4 The Lysis-Lysogeny Decision

Upon infection of a host, a temperate phage must decide: shall I undergo lysis, or shall I undergo lysogeny? This is a decision with only two outcomes, as the two outcomes are mutually exclusive, and the outcomes are easily distinguishable: either the phage undergoes lysis and the host cell dies, or the phage undergoes lysogeny and the host survives.

Which outcome is better? It depends on the environment. In environments with few hosts, or environments that would rapidly kill any free phage particles, it may be more prudent to lysogenize the host instead of releasing more phages into such a harsh environment. Contrariwise, in environments teeming with sensitive hosts, it makes little sense to undergo lysogeny when there are so many sensitive hosts to infect easily within reach.

But in order for the decision to be a choice and not just acting at random, the phage must gather information on which it bases its decision. It has been shown that phages can gather information from the environment through a variety of mechanisms, responding to signals in the environment as well as the internal state of the cell (Shao *et al.*, 2019), and use this information to determine whether to undergo lysis or lysogeny[2].

1.5 The Lysis-Lysogeny Decision in Lambdoid Phages

The lysis-lysogeny decision is most well studied in lambdoid phages, due to the historical usage of phage lambda as a model organism. The key players of the lysis-lysogeny decision are located in a region called the lambda switch. The structure of regulation of the lambda switch is well conserved among lambdoid phages, despite the variability in operator sites and protein identity.

After injection of the phage genome into the host cell, phage lambda begins transcription at the P_L and P_R promoters, initially transcribing the *N* and *cro* genes. The Cro protein suppresses the expression of CI, and the N protein functions as an antiterminator, which allows for the transcription of *cIII*, the CII transcriptional activator and the Q antiterminator. The Q antiterminator extends the transcription of P_R to begin the expression of lytic genes, committing the phage to lytic replication; however, CII suppresses the expression of *cro* and *Q*. CII has a short half-life, but is protected from degradation by CIII. If CII is able to reach a sufficiently high concentration, CII suppresses the expression of the lytic genes, allowing the phage to enter the lysogenic pathway.

In the lysogenic pathway, CII activates the P_{RE} promoter, transcribing RNA antisense to *cro*. P_{RE} also drives the expression of CI. CI represses both P_L and P_R, shutting down expression of the lambda switch and allowing P_{RM} to maintain CI levels. At sufficiently high levels of CI, CI can repress the expression of P_{RM}, maintaining an intermediate level of CI in the lysogen.

[2] This is true of some virulent phages as well; some virulent phages, such as T-even phages, can delay their lysis if they detect that environmental conditions are unfavorable, termed lytic inhibition, lysis delay, or lysis timing, often leading to delayed lysis but larger burst size (Bode, 1967). The core decision-making appears to be the same: make a decision dependent upon environmental conditions.

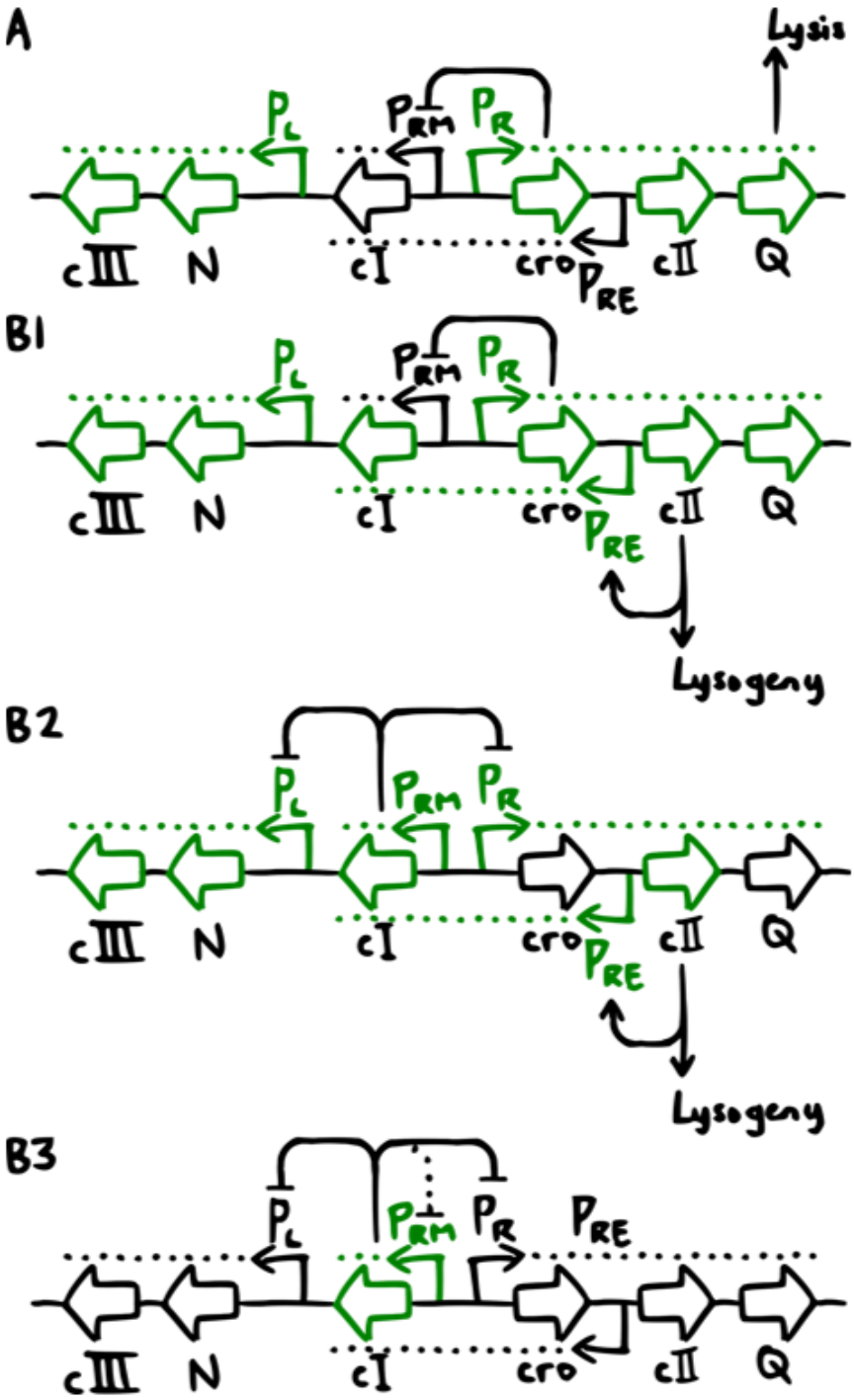


Figure 2. The lambda switch.

During late-early expression (A), the P_L and P_R promoters are active, leading to the transcription of multiple genes. The expressed Cro protein represses the P_{RM} promoter. If the CII protein does not reach sufficiently high levels, the Q antiterminator protein turns on genes which result in phage replication and cell lysis. If CII proteins reach sufficiently high concentrations (B), the CII protein activates the P_{RE} promoter (B1), as well as other promoters for genes required for lysogeny. CII also activates the transcription of an antisense promoter to Q, repressing its expression (not shown). P_{RE} transcribes the cI *gene and RNA antisense to* cro *(B2), resulting in the silencing of cro translation, which allows P_{RM} to be transcribed. The CI protein represses transcription from the P_L and P_R promoters, preventing the transcription of* cro *and* cII *(B3). At sufficiently high concentrations, CI represses P_{RM}, regulating its own expression. The N and CIII proteins are also part of the switch, as the N protein is required for the transcription of* cII *and* cIII*, and CIII protects CII from degradation. Adapted from Oppenheim et al. (2005).*

So how does this dance of repressors, activators, and antiterminators function to integrate information? Much of the initial decision of the switch depends upon the concentration of CII; sufficiently high concentrations of CII represses the lytic pathway and allows the lysogenic pathway to proceed. Thus, anything that influences the concentration of CII can influence the lysis-lysogeny decision. Smaller cells have been shown to have an increased chance of lysogeny (St-Pierre & Endy, 2008), as does lower temperatures (Obuchowski *et al.*, 1997) and low nutrient environments, where fewer host proteases protect CII from degradation (Ptashne, 2004).

CII is also a key player in mediating the effect of the multiplicity of infection on the lysis-lysogeny decision. In high-phage low-host environment, it is likely that multiple phage bind concurrently to the same host cell, before the decision is made. This leads to an increase in the number of phage genome copies inside the cell, increasing the CII concentration and driving the phage towards the lysogenic pathway (Weitz *et al.*, 2008).

An important distinction to note is that the CI repressor is primarily involved in the maintenance of lysogeny, rather than the initiation of lysogeny. During lysogeny, if CI levels drop below a critical threshold, the P_R and P_L promoters become unrepressed, leading to induction and a return to the lytic pathway. CI is cleaved during the SOS response by RecA*; however, this does not necessarily mean that the SOS response shifts the lysis-lysogeny decision towards lysis, rather, the lysis-lysogeny decision could have made the decision to under lysogeny, but was immediately induced, being unable to maintain the lysogenic state.

1.6 Using Temperate Phages as a Platform to Examine Biological Composability

If a naturally-occurring composable system exists, it is almost surely a mobile genetic element. They transfer between cell to cell, host to host, and must carry out their program within the context in which they find themselves. And if they are to process information to make a decision, then they require the means to gather information and the discrimination to differentiate between useful signals and useless noise, and the means to process that information without being disrupted by the functions of the other systems within the cell[3].

Temperate phages, specifically lambdoid phages and phage lambda, are well-characterized examples of mobile genetic elements, and are amendable to exploration in their properties of composability. In this thesis, I have examined the composability of phage lambda. In Chapter 2 I use mathematical models to examine what sources of information in the environment the phage should respond to. In Chapter 3 I examine how the phage could inadvertently interfere with host systems, and find evidence that phage receptors interfere with their own hosts less than they interfere with non-hosts. In Chapter 4 I experimentally examine how the system behaves in various host backgrounds, and explore the interference by pre-existing prophages in environmental strains. With these experiments, I hope to shed insight on the composability of temperate phages, and the limits thereof.

[3] Of course evolution is not that simple; mechanisms evolve specifically to counter mobile genetic elements, as their selfish behavior is often detrimental to that of the host. These defense systems, and anti-defense systems, and phage-encoded anti-phage defense systems, add additional complexity to notions of composability; if something does interfere with the system, is it because the system is fragile and sensitive to noise, or is it because it evolved to specifically interfere with the system? Further discussion can be found in section 4.6.

2 Responding to Signals

2.1 Preamble

In order to make a decision, a system must process information, and in order to gather information, a system must be able to distinguish between the signal and noise. But systems do not have all access to all information; when a phage genome is injected into host cell, it no longer has direct access to the external environment. The phage genome is limited to receiving information that exists within the host, and thus mechanistic explanations of the lysis-lysogeny decision tell us what signals the phage responds to, but not necessarily what information the phage should respond to in order to make its decision.

Mechanistic explanations have provided some insight into the information that the phages care to sense, but also some confusion. The number of concurrent infections is potentially the most studied responsive factor of the lysis-lysogeny decision, as there is a sharp increase in the probability of lysogeny when more than one phage binds to the same cell; however, it is still debated what information this concurrent binding provides. It appears to provide information on the ratio of phages to susceptible hosts, as concurrent infections can only occur if there are a high number of phages and a low number of cells, but some have argued that concurrent binding additionally reports on the decay rate of phages in the environment, or that the only important information is the number of sensitive hosts, and concurrent binding is the only mechanistic way of obtaining that information.

Novel mechanisms of phage-phage communication have been discovered which influence lysis-lysogeny decision, which raise further questions about what information would help in making such a decision. The *Bacillus subtilis* phage phi3T produces a quorum-sensing peptide called “arbitrium” during both lysis and lysogeny, and exposure to the peptide shifts the decision towards lysogeny, while also suppressing prophage induction (Bruce *et al*., 2021). The peptide is also taken up and degraded by both lysogens and sensitive cells, resulting in a complex response that counts lysogens, lytic events, total number of cells, and the degradation rate of the environment.

In order to determine the most important sources of information that a phage should respond to, I designed a model in which the phage has complete access to information. With this model, I examined which sources of information the phage should respond to in order to maximize fitness, independent of known mechanisms. By determining what the phage should respond to, instead of what it does respond to, we gain knowledge that can be used to drive exploration towards further mechanisms that impact the lysis-lysogeny decision.

The following paper has been submitted for publication in *Evolution*, and is undergoing revisions.

2.2 Responsive Lysogeny under Nonproductive Phage Binding

Bryan Wu and Călin C. Guet

ABSTRACT

Upon infecting a cell, temperate phages must make a decision between lysis and lysogeny. While research has explored how phages sense environmental information to make this choice, most studies have focused on modelling known mechanisms that impact the decision. These mechanisms tell us what environmental information the phage responds to, but not the sources of information that it should respond to, as the signals sensed by the phage may serve as proxies for other sources of information. Here, using a mechanism-agnostic population dynamics model, we find that irreversible phage binding to immune lysogens protects sensitive host cells from infection, and additionally results in lysogens being an environmental factor that the phage should sense while making its decision, alongside the number of sensitive cells. Using this model, we derive a lysogeny probability for phages that responds to both sensitive cell and lysogen densities optimized towards invading phage-occupied systems, and show that it is more capable of invading and resisting invasion than phage with fixed lysogeny probabilities across different environmental conditions.

INTRODUCTION

Bacteriophages are viruses of bacteria that generally kill their hosts in order to produce more of themselves through the process of lysis. However, temperate phages have an additional strategy to replicate their genome alongside that of their host as a prophage, thus tying the phage's reproductive fitness to that of their host. These prophage-carrying cells, known as lysogens, can reactivate at a later point in time, ultimately releasing more phage particles through the lysis and death of the lysogen (Ptashne, 2004).

The existence of these two different modes of replication, the lytic and lysogenic cycles, suggests that conditions exist where each mode of replication is advantageous over the other. It has been suggested that the lytic mode is favored at high susceptible host densities and low resources, whereas the lysogenic mode is favored at low host densities and high resources (Li *et al*., 2020, Wahl *et al*., 2018, Steward & Levin, 1984). Mathematical models of phage-bacteria populations have sought to examine these conditions, as well as conditions that favor intermediate strategies, where phages choose lysogeny or lysis with some probability (Stewart & Levin, 1984, Maslov & Sneppen, 2015, Wahl *et al.*, 2018, Sinha *et al.*, 2018).

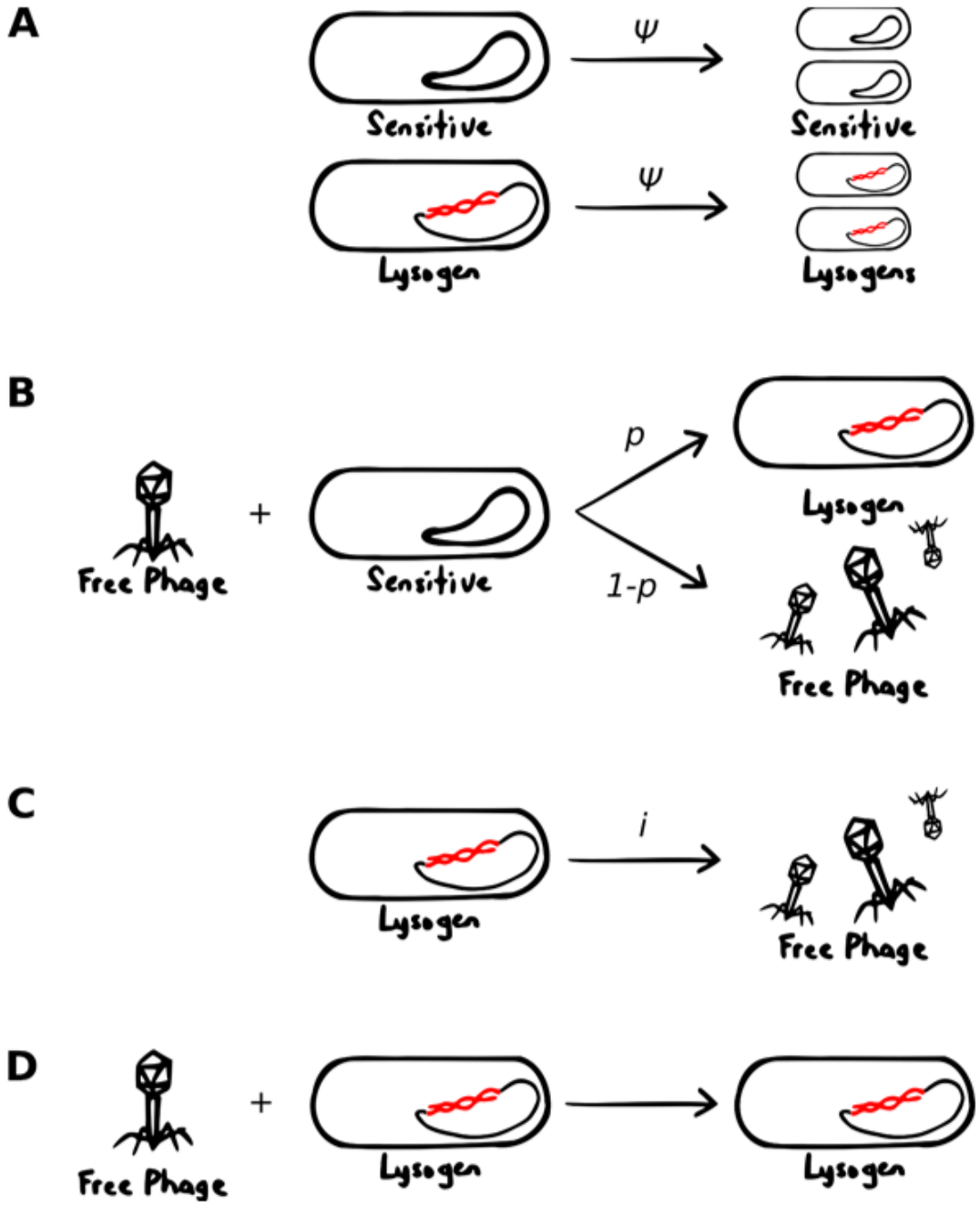
Temperate phages, however, do not necessarily use a fixed probability of lysogeny. Among the different species of phages, the process of undergoing lysis or lysogeny is a responsive choice, decided upon during the infection of a host, which is capable of integrating various sources of information, the most salient of which is the number of co-infecting phages (Zeng *et al.*, 2010). However, various phages use a variety of other mechanisms to gather information from environmental factors, such as those tied to the state of the cell (St-Pierre & Endy, 2008) and the concentration of phage quorum-sensing molecules (Erez *et al.*, 2017, Bruce *et al.*, 2021). Responsive mechanisms to all these different sources of information demonstrate that the lysis-lysogeny decision can be a dynamic process, changing with environmental conditions as the system changes.

Phages also have a variety of mechanisms to confer superinfection inhibition, where upon infecting a cell, the phage prevents the host from being infected by identical or similar phages. Superinfection inhibition can be broadly grouped into two categories: superinfection exclusion, where superinfection is interrupted prior to the injection of the superinfecting phage DNA, and superinfection immunity, which inhibits the replication of the post-translocated DNA (Biggs *et al*., 2021). Although the host is protected from superinfection by either superinfection exclusion or immunity, the superinfecting phage experiences drastically different outcomes. Phage adsorption is a three-step process, starting with initial contact, which leads to reversible binding of tail fibres and finally ending with irreversible binding, which is generally tied to the injection of phage DNA into the host (Silva *et al.*, 2016). With superinfection exclusion, the invading phage is either unable to inject the DNA or unable to bind at all, and thereby generally has the opportunity to unbind from the protected host and search for another susceptible target; however, with superinfection immunity, the phage has already irreversibly injected its DNA into the immune host and has no opportunity to back out; it is already dead. This process is known as sorptive scavenging, originally termed to describe the binding of phages to sediment particles and removal from the seawater (Hewson & Fuhrman, 2003) has later been expanded to describe the sequestration of phages by resistant cells (Simmons *et al.* 2020). Sorptive scavenging increases the decay rate of phages in environments with lysogens that can irreversibly bind to the phage, and has implications on the lysis-lysogeny decision and the dynamics of a phage-bacteria system (Berryhill & Levin, 2025).

A previous model examining the conditions for lysis and lysogeny made the assumption that the phage cannot bind to lysogens, and that lysogen resistance was conferred by downregulating the receptor (Wahl *et al.*, 2018). However, this model did not examine how the choice of inhibition mechanism affects the behavior of the lysis-lysogeny decision. As superinfection immunity would create an environment where phages continue to bind and inject their DNA into immune lysogens and are removed from the population, this would likely alter the optimal decision choice as the presence of lysogens decreases the survival of free phages, and should shift the decision towards lysogeny. Some models do include the nonproductive binding of phages to lysogens, but do not focus on the impact that sorptive scavenging has, instead extending the model to focus on additional mechanisms that impact the lysis-lysogeny decision (Bruce *et al.*, 2021).

We began by examining the impact of nonproductive binding of phages to lysogens on the lysis-lysogeny decision in order to determine a responsive function that can be used by the phage to determine whether to undergo lysis or lysogeny. We then calculate responsive rates of lysogeny optimizing for invasion when rare in environments with and without binding to lysogens, and show that phages that do not respond to lysogens perform poorly in environments with binding to lysogens. Finally, we show that the responsive lysogeny rate is capable of invading systems under more conditions than a fixed lysogeny probability, suggesting the number of lysogens in the environment is an important source of information in determining the outcome of the lysis-lysogeny decision in environments with sorptive scavenging by lysogens.

METHODS



Box 1. Major transitions in the mathematical model. A) In the model, both sensitive and lysogenic cells self-replicate with rate ψ, which is dependent on the total number of cells and the carrying capacity of the system. **B)** When a phage binds with a sensitive cell (with a rate of b), the phage is consumed and the cell turns into a lysogen with a probability p, otherwise the cell lyses and becomes β phages. **C)** Lysogens have a probability i of inducing, lysing and producing β phages. **D)** Phages also bind with lysogens at rate b, which destroys the phage. Not pictured: all species decay with a rate ω, and new sensitive cells are added with a rate c.

Epidemiological model

Our model is a simple no-delay mass-action kinetic model with implicit resources in a chemostat-like setup (Box 1). We track uninfected, sensitive host cells *S*, lysogenic cells *L* and free phage particles *V*. All cells grow at the same rate ψ, which is dependent on the total number of living cells and the carrying capacity *K* of the system, with a maximum rate of τ (Box 1A). All species decay at the same rate ω, and there is a constant inflow of sensitive cells at a rate of *c*, as in a two-stage chemostat. Phage particles bind to both sensitive cells and lysogens at rate *b* and are removed from the system. If a phage particle binds to a sensitive cell, the cell either transforms into a lysogen with a probability *p*, or immediately undergoes lysis, producing β new phage particles with a probability of *1-p* (Box 1B). Additionally, lysogens induce at a rate *i*, dying and producing β new phage particles (Box 1C). We assume that lysogens cannot be reinfected, and are perfectly immune to all other phages in the model (Box 1D). This model can be extended for up to *N* types of phages, but in this study are limited to 1 or 2 phages.

This model was chosen to emphasize intermediate lysis-lysogeny rates, as it has been shown that pure lysogeny is preferred in any system without an inflow of sensitive cells (Wahl *et al.*, 2018). The inflow of sensitive cells may occur in different environments such as rivers, sewage systems, or any such environment with a unidirectional flow, in addition to environments with a resistant population of bacteria that loses resistance at high rates, such as under phase variation (Shkoporov *et al.*, 2021).

This results in the following model:

$$\frac{\delta S}{\delta t} = \psi S - \omega S - bS\sum_{x=1}^{N} V_x + c \tag{1}$$

$$\frac{\delta L_x}{\delta t} = \psi L_x - \omega L_x + p_x bSV_x - iL_x \tag{2}$$

$$\frac{\delta V_x}{\delta t} = (1-p_x)\beta bSV_x - \omega V_x - b(S+\sum_{x=1}^{N} L_x)V_x + \beta iL_x \tag{3}$$

where the bacterial growth rate ψ is equal to:

$$\psi = \tau(1 - (S + \sum_{n=1}^{N} L_n)K^{-1})$$

(4)

For ease of modeling, we assume a no-delay model, where lysis and lysogeny are both instantaneous upon binding to a sensitive cell. Adsorption is modelled as a function combining an adsorption rate constant and the concentration of phages and bacteria in the system, and follow standard mass-action kinetics. These assumptions are in line with similar models (Wahl *et al.* 2018, Bruce *et al.* 2021).

For our calculations and simulations, the following parameters were used unless otherwise stated:

Parameter	*Description*	*Value*
τ	Maximum bacterial growth rate	1
K	System carrying capacity	10000
ω	System flow rate	0.2
β	Burst size	100
b	Binding rate	0.01
i	Induction rate	0.01
c	Inflow of sensitive cells	150

During invasion analysis, 0.001 of invading phage and lysogens were added to the system.

Analytical calculations and simulations were done in Mathematica 11.1 (Wolfram Research, Inc., 2017).

RESULTS

Nonproductive binding protects sensitive cells

In previous studies, it has been experimentally shown that bacteria which are immune to phage but can still bind those phages can have a protective effect on sensitive cells found in the same community, with results akin to herd immunity (Brown *et al.*, 2022, Payne *et al.*, 2018), but limited their models to epidemic infections. We therefore examined whether sorptive scavenging produces similar protective effects in a chemostatic model, and found similar results. We find that nonproductive phage binding to lysogens greatly increases the number of sensitive cells at equilibrium which is driven by a marked reduction in free phage concentrations, whereas the number of lysogens remained relatively unaffected (Figure 1).

This suggests that lysogens serving as sinks for phage particles lowers the free phage levels and protects sensitive cells in the environment, even under continuous flow conditions.

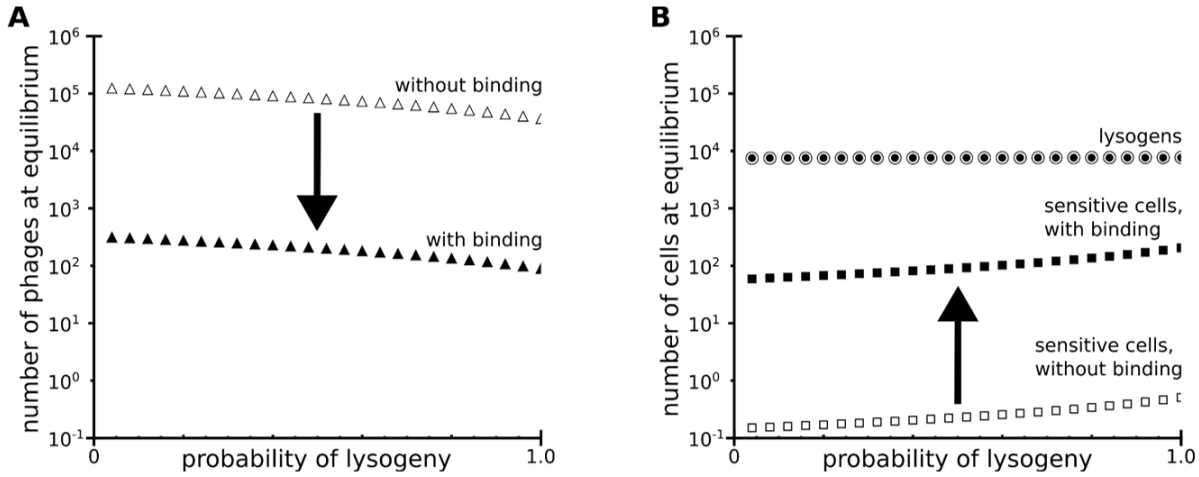


Figure 1. Nonproductive binding of phages to lysogens increases number of sensitive cells at equilibrium, while decreasing number of free phage. A) The number of free phages at equilibrium decreases when phage binding to lysogens is allowed (black triangles), compared to an equilibrium where phage binding to lysogens is prohibited (white triangles). **B)** The number of sensitive cells at equilibrium increases when phage binding to lysogens is allowed (black squares) compared to without phage binding to lysogens (white squares), though the number of lysogens are relatively unaffected (circles, overlapping).

Numerical simulations of fixed responses for invasion when rare

As nonproductive binding of phages to lysogens has an effect on the equilibrium of the system (Figure 1), we examined how this affects how well phages can invade equilibria of systems already occupied by a phage, identical to the invading phage other than a different fixed lysogeny probability. To this end we simulated pairwise combinations of phages with fixed lysogeny probabilities over extensive periods of time, where a small amount of the invading phage and corresponding lysogen is added to a system at equilibrium already occupied by a resident phage (Figure 2A). The result are similar to a pairwise invasibility plot (Brännström *et al.*, 2013), but additionally plots the degree of coexistence.

We find that an optimal intermediate probability of lysogeny exists, which is capable of both resisting invasion and invading any non-optimal lysogeny probability. A resident phage with a lysogeny probability less than the optimum can be invaded by any phage with a greater lysogeny probability, even *p=1* (Figure 2A, red regions on the left), whereas resident phages with a lysogeny probability greater than the optimal value can be invaded by any phage with a lower lysogeny probability, even *p=0* (Figure 2A, red regions on the right). Additionally, two coexistence regions exist, where both phage strains are mutually invasible and coexist to varying degrees.

Our results are qualitatively similar to that of the pairwise invasibility plot described in Wahl et al. (2018), which modeled epidemic infections with an inflow of sensitive cells but did not include sorptive scavenging to lysogens. This suggests that nonproductive binding to lysogens does not greatly impact the dynamics of phage invasion when rare, as both resident and

invading phages are equally impacted by binding to lysogens. A closer comparison between environments with and without nonproductive binding to lysogens unexpectedly shows that the optimal fixed probability of lysogeny is slightly greater in environments without binding to lysogens (Figure 2B). This may be due the number of sensitive cells at equilibrium; in environments where phages bind to lysogens, the number of sensitive cells at equilibrium increases (Figure 1), and an equilibrium with a higher number of sensitive cells may be more easily invaded by a phage that tends more towards lysis.

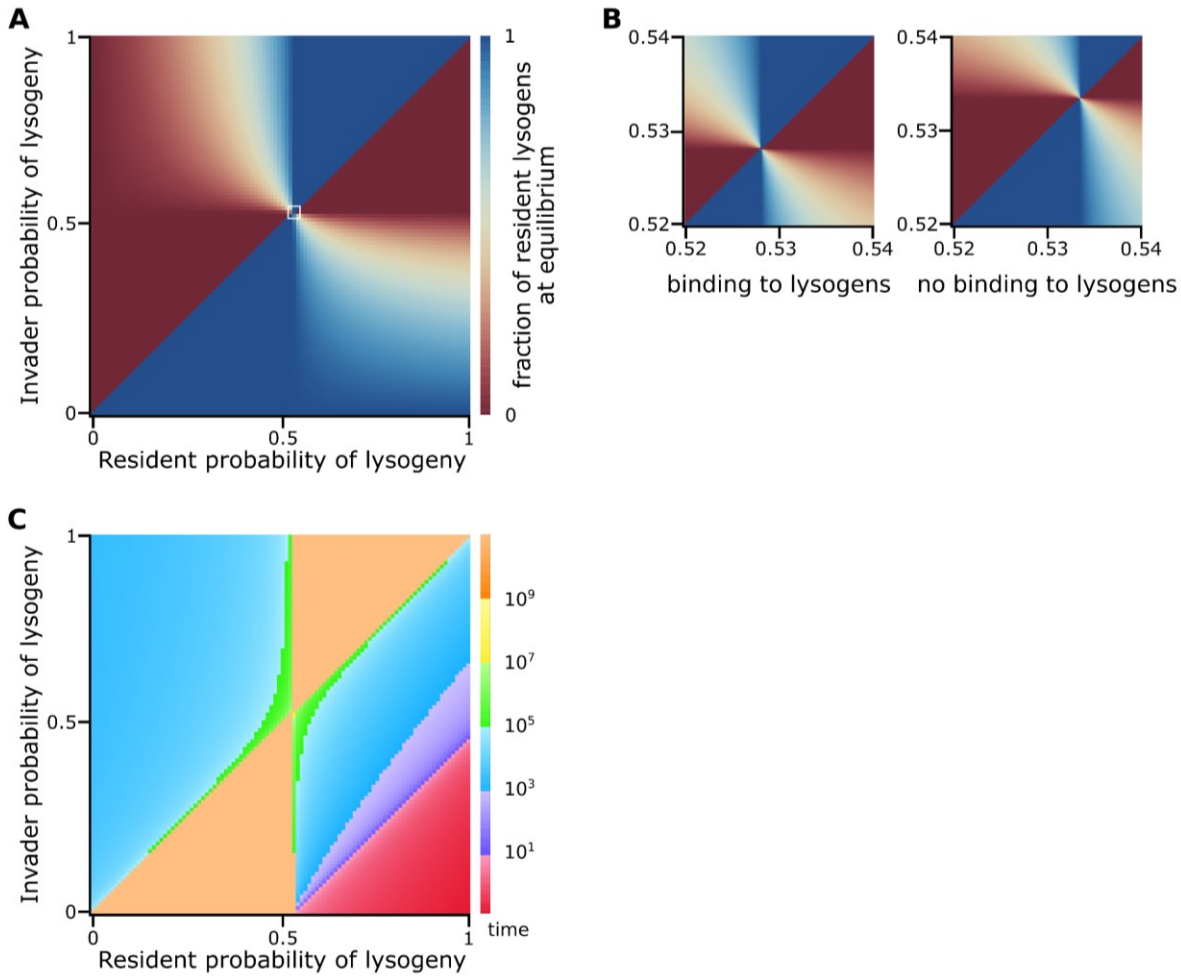


Figure 2. Invasion and invasibility in phages with fixed probability of lysogeny. A) A pairwise invasibility plot of phages with fixed probability of lysogeny in an environment where phages bind nonproductively to lysogens. The equilibrium of the system with a single species of resident phage was solved, then a small number of invading phage and lysogens were added and the system was simulated to $t=10^9$. The percent of lysogens that contain the non-invading (resident) phage genome were then plotted. Red indicates a successful invasion. The white square indicates the section represented in panel B. **B)** A zoomed-in section of panel A (left), compared to the matching area of a pairwise invasibility plot where phages do not bind to lysogens (right). **C)** Amount of time required for invading a phage to increase by a thousand-fold, in an environment with phage binding to lysogens, simulated to a maximum time of $t=10^{12}$.

With the numerical simulations, we were concerned with the amount of time it takes the invader to reach equilibrium, as more time at low numbers leaves the invader vulnerable to extinction via stochastic fluctuations. We tracked the time taken for the phage population to increase one thousand-fold from inoculation (Figure 2C), and find that generally a lower probability of lysogeny invades faster when capable of invasion. Additionally, the closer the resident phage's lysogeny probability is to the optimal lysogeny probability, the more time it takes to invade.

Although our simulations borrow the "invasion when rare" assumption from adaptive dynamics and yield similar results to the pairwise invasibility plot (Wahl *et al.* 2018), our interest does not solely lie with long-term evolutionary outcomes. We are also interested in the space of all possible equilibria, and the subset of which it is possible to invade.

Invasibility of equilibrium states by lytic phages and lysogens

As all single-phage occupied equilibria have some number of phages, lysogens and sensitive cells, we wished to determine the set of these equilibria susceptible to invasion by phages and lysogens. Since the resident phages do not directly interact with the invading phages or lysogens at the moment of invasion (as the initial growth of invading phages and lysogens does not directly depend on resident phages, see Appendix 2.2-2.3), invasion is dependent solely on the number of sensitive cells and resident lysogens at equilibrium. We then analytically calculated the regions in which lytic phages and lysogens are capable of invading when rare (Appendix 2), for system both with and without nonproductive phage binding (Figure 3).

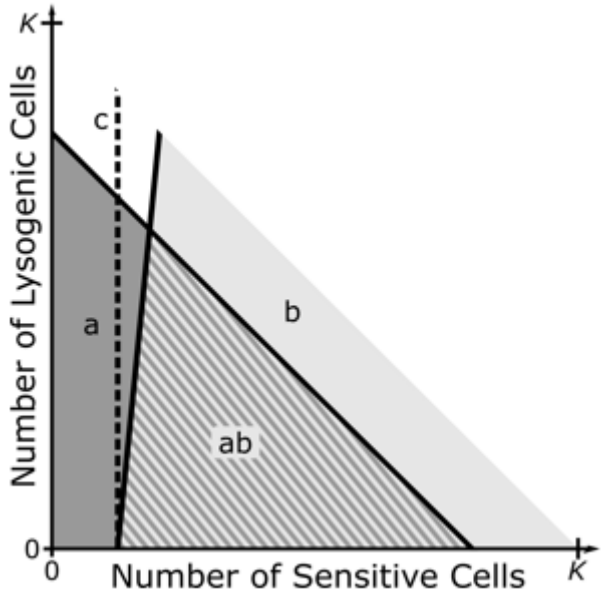


Figure 3. The invasibility of a phage-occupied system. With nonproductive phage binding, the system can be invaded by invading lysogens in the dark gray regions (a), and by invading lytic phages in the light gray regions (b), bounded by equation (7). The system cannot be invaded in the white area (c). The dotted line represents the location of the line demarcating b without nonproductive binding, representing equation (5). The graph is truncated to regions where the cell population is less than the carrying capacity *K*. Note that cell populations never

reach the theoretical carrying capacity *K* due to the constant decay of the chemostat system, and region (a) is bounded by the maximum achievable population, $K(1-\frac{i}{\tau}-\frac{\omega}{\tau})$.

Without nonproductive binding of phage particles to lysogens and fixed environmental parameters the ability for a lytic phage to invade is solely dependent on the number of sensitive hosts at equilibrium (Appendix 2.1.2), where the lytic phage can invade under the following circumstances:

$$S > \frac{\omega}{b(\beta-1)} \tag{5}$$

This can be understood as the invading lytic phages do not directly interact with the resident phages or lysogens, and only directly interact with the number of sensitive cells. Thus, the ability to invade is solely dependent on the number of sensitive cells and fixed environmental variables. The number of sensitive cells required for invasion by lytic phage increases as the decay rate increases, as the phages decay more quickly, and decreases as the binding rate or burst size increase, as under those conditions a phage burst is more likely to produce a phage which binds to a sensitive cell.

This corresponds with conclusion of Wahl *et al*. (2018), where lysis is favored if host cells can be maintained at a sufficiently high level. When phages can bind nonproductively to lysogens, however, the ability for a lytic phage to invade additionally depends on the number of lysogens at the moment of invasion (Appendix 2.1.1), resulting in a varying threshold of sensitive cells (Figure 3) with the following equation:

$$L < (\beta-1)S - \frac{\omega}{b} \tag{6}$$

Or, alternatively rearranged

$$S > \frac{L}{\beta-1} + \frac{\omega}{b(\beta-1)} \tag{7}$$

The second term in equation 7 is identical to that of lytic invasion when lysogens do not bind to phages. The difference lies in the additional first term, which is the ratio of lysogens to the burst size. Thus, the number of sensitive cells required for lytic invasion increases as the number of lysogens increase due to sorptive scavenging, but the impact of the number of lysogens decreases as the burst size increases. Binding rate b does not play a role in this term as the binding rate to sensitive cells and lysogens are identical, and thus cancel out.

In our model, lysogens have an identical growth rate to that of sensitive cells, and thus lysogens can invade under most conditions where the total cell number does not exceed the carrying capacity, less the dilution and induction rate (Figure 3, Appendix 2.3).

Using invasibility as a responsive function for lysogeny

It is known that the probability of lysogeny of phages is not fixed, but instead depends on a number of external factors (Herskowitz & Hagen, 1980; St-Pierre & Endy, 2008, Zeng *et al.*, 2010; Erez *et al.*, 2017). Instead of implementing any particular mechanism, we wished to ask the following: if a phage genome, once injected into a susceptible host, had perfect information about the external environment, what information should it respond to? and how well does it perform? The previous equations show that, in environments with sorptive scavenging to lysogens, the number of lysogens partly determines if lytic phages can invade. This is in addition to the variables required in environments where phages do not bind to lysogens: number of sensitive cells, burst size, binding rate, and decay rate.

The molecular mechanisms by which phages can sense this information are out of the scope of this paper, we only note that the mathematics show that this information is necessary to accurately determine whether lytic invasion of an equilibrium is possible, which we suggest is beneficial for the phage to determine; however, we do provide examples in the discussion that suggest such mechanisms are physically possible.

Using the equations above, we decided to implement a responsive lysogeny function where the phage undergoes lysis under conditions where lytic phages can invade, and otherwise undergo lysogeny. In line with previous examinations of responsive functions (Sinha *et al.*, 2017), we fix the probability of lysogeny to 1 in the regions where lytic phage cannot invade, and to 0 where lytic phage can invade, with a smooth transition fitted to the logistic function, such that the invasion-optimal responsive lysogeny function for environments with no binding (nb) to lysogens is

$$p_{nb} = \frac{1}{1 + e^{-\left(\frac{\omega}{b(\beta-1)} - S\right)}} \tag{8}$$

and the invasion-optimal (io) responsive lysogeny function for environments with binding to lysogens is

$$p_{io} = \frac{1}{1 + e^{-\left(L - \left((\beta-1)S - \frac{\omega}{b}\right)\right)}} \tag{9}$$

Simulations show that the phage with p_{nb} fares poorly in environments where phages bind to lysogens. It is only capable of invading systems where the resident phages are lytic or have a probability of lysogeny greater than the optimal fixed probability, and even then only coexists as free phages with no lysogens (Figure 4A, left). It is also susceptible to invasion by phages with any nonzero fixed probability of lysogeny (Figure 4A, right). This can be understood as a mismatch between the responsive function and the environment: the phage systematically undercalculates the decay rate of the environment as it is unable to respond to the number of lysogens, and consistently chooses lysis in environments where lytic replication is not

supported. p_{nb} performs better in a matched environment where phages do not bind to lysogens (Appendix, Figure A1).

On the other hand, the phage with p_{io} fares well in environments with binding to lysogens, and can invade all non-optimal fixed lysogeny probability phage-occupied systems at equilibrium (Figure 4B, left). Additionally, the invasion-optimized lysogeny function fully resists invasion by phages with a fixed probability of lysogeny less than the optimal fixed probability, but can still be invaded by phages with a fixed probability of lysogeny above the optimal fixed probability. (Figure 4B, right).

Although phages with an optimal fixed lysogeny probability can outperform those with a responsive lysogeny probability, this is strongly environment-dependent as the optimal fixed lysogeny probability changes alongside changes to the environmental parameters (Figure 4C). The responsive lysogeny probability phages, in contrast, change their behavior in response to changes in the environmental parameters, and thereby can outperform the fixed lysogeny probability phage under a wider range of conditions (Figure 4D).

We note that phages with p_{io} can still be invaded by phages with fixed lysogeny probability greater than the optimal fixed lysogeny probability. This is a similar behavior to a phage approaching the optimal fixed lysogeny probability (Figure 2A), where approaching from below allows invasion by phages with a greater lysogeny probability but total resistance to lower lysogeny probability phages, whereas approaching from above gives the opposite.

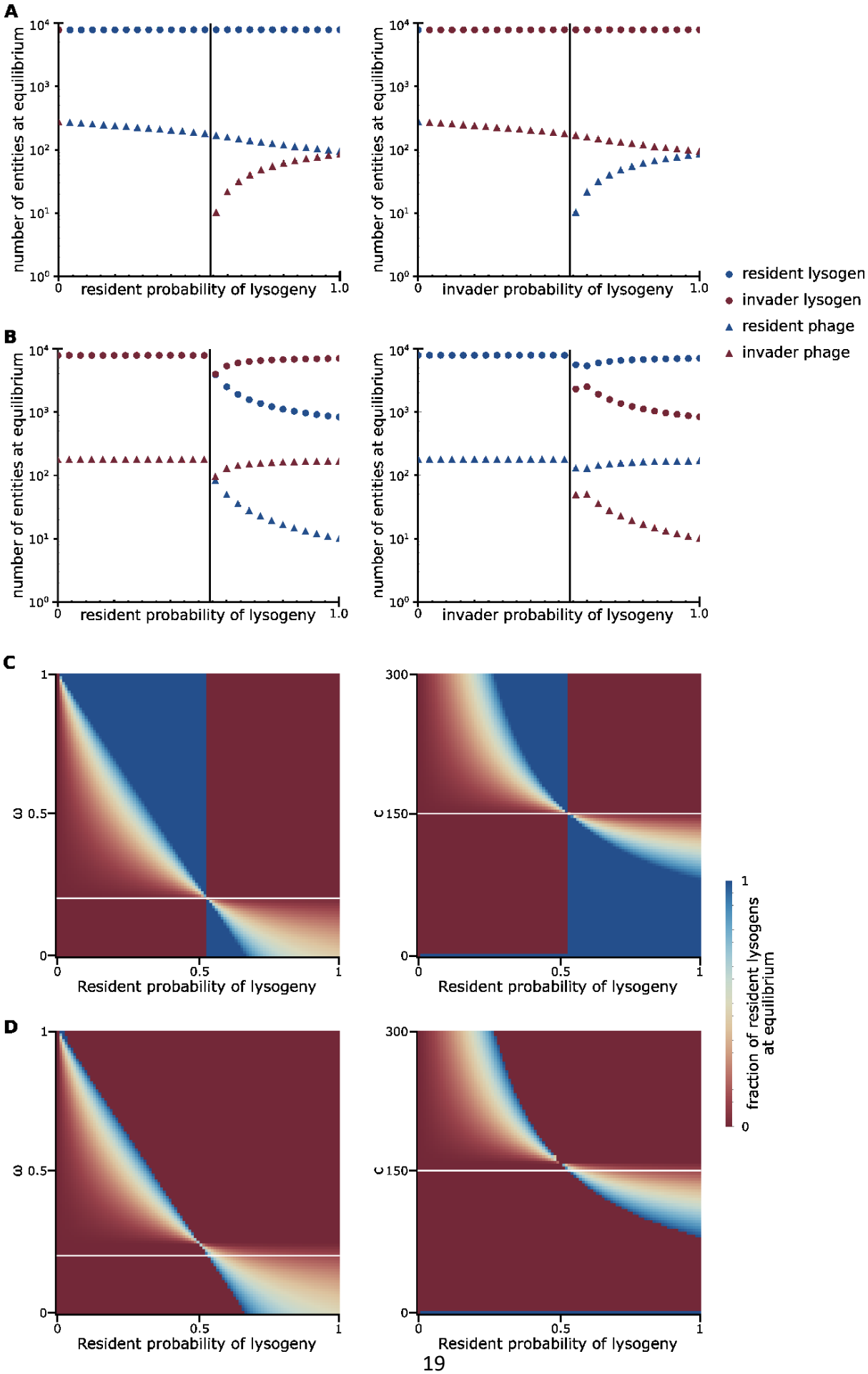

A
B
C
D
number of entities at equilibrium
resident probability of lysogeny
invader probability of lysogeny
resident lysogen
invader lysogen
resident phage
invader phage
Resident probability of lysogeny
fraction of resident lysogens at equilibrium

Figure 4. Invasibility plots of fixed and responsive lysogeny rates. A) Phages and lysogens with responsive lysogeny rate that does not respond to lysogens (p_{nb}) invade systems of phages with a fixed lysogeny probability at analytical equilibrium (left), or are simulated to equilibrium and invaded by phages and lysogens with a fixed lysogeny probability (right). **B)** Phages and lysogens with responsive lysogeny rate that does respond to lysogens (p_{io}) invade a system of phages with a fixed lysogeny probability at analytical equilibrium (left), or are simulated to equilibrium and invaded by phages and lysogens with a fixed lysogeny probability (right). Black lines in A and B correspond to the optimal fixed lysogeny probability, $p\approx0.527$. **C)** Phages and lysogens with a fixed probability of lysogeny $p=0.527$ invades a system of phages with fixed lysogeny probability at analytical equilibrium at different flow rates (ω, left) or influx of sensitive cells (c, right). **D)** Phages and lysogens with responsive lysogeny rate p_{nb} invade a system of fixed lysogeny phages at analytical equilibrium at different flow rates (ω, left) or influx of sensitive cells (c, right). All systems were simulated to $t=10^{12}$ post-invasion. White lines in C and D correspond with the parameter set ($\omega=0.2$, $c=150$), which are the same parameters used in A and B.

DISCUSSION

The lysis-lysogeny decision in phages, in particular phage lambda, has been studied in great detail since the advent of molecular biology, unveiling the molecular machinery and genetic regulatory networks involved in making the fate choice between the lytic and lysogenic cycles (Zeng *et al.*, 2010). At the same time, several models have been proposed to justify the existence of lysogeny, often invoking changes in environmental conditions which shift from favoring one life history strategy to the other (Cheong *et al.*, 2022; Maslov & Sneppen, 2015, Stewart & Levin, 1984).

In order to make any decision, however, a system requires information. It has been shown that a number of factors impact the lysis-lysogeny decision (Herskowitz & Hagen, 1980; St-Pierre & Endy, 2008, Zeng *et al.*, 2010; Erez *et al.*, 2017), and suggests that the mechanism has evolved to integrate information from a broad swath of sources, but it can be difficult to determine which factors provide useful information and which factors are merely noise. Information received by the mechanism is additionally constrained by the physical limitations of gathering the information from inside the cell, resulting in integrating indirect information about the state of the external environment. Though much work has been done to examine the molecular mechanisms of how all these factors interplay in the lysis-lysogeny decision, nevertheless the question remains: what information does the system actually care about?

It is easy to suggest maximizing the number of successful offspring in the next generation as the goal of natural selection, but this requires the ability to retrieve information from the future, for which no biologically plausible mechanism has yet been proposed. The next best thing is to maximize the predicted number of successful offspring in the next generation, which is more easily done as the prediction does not necessarily have to match reality (but the more accurate the prediction the better).

For a phage, successful offspring can be defined as progeny phages that find a new host, or, in the case of lysogeny, lysogen daughter cells. While the number of lysogen daughter cells is dependent on a fairly limited number of factors – such as growth rate, dilution rate and decay rate – finding a new host is dependent on a number of factors in the environment, including

the binding and decay rate of phages, the flow rate of the system, and the number of available hosts.

In our work we find that the introduction of sorptive scavenging of free phages by lysogens increases the number of sensitive cells and decreases the number of phages at equilibrium. This leads to a decrease in the optimal fixed lysogeny probability, as equilibria with more sensitive cells can be more readily exploited by phages that tend more towards lysis.

We show that the addition of sorptive scavenging by lysogens results in the invasibility of an equilibrium becoming additionally dependent upon the density of lysogens. We then found that, under these conditions, phages with a responsive lysogeny probability that responds to the number of lysogens is more capable of invading and resisting invasion, compared to phages that do not respond to the number of lysogens. The responsive strategy cannot invade the optimal fixed strategy; however, the optimal fixed strategy is only optimal for a single set of environmental parameters, whereas the responsive functions can respond to changes in the environment, and thus performs well across different environments.

These results, however, have been obtained from a mechanism-agnostic approach. How does it compare to known biological mechanisms? The most well-studied mechanism, where the concurrent binding of multiple phages to a single cell, is often simplified as merely the ratio of phages to sensitive cells. Whereas both p_{io} and the phage:cell ratio respond to the number of sensitive cells, p_{io} does not respond to the number of phages, and the phage:cell ratio does not respond to the number of lysogens. It is possible that the number of phages serves as a proxy for the number of lysogens (as increased sorptive scavenging by lysogens would reduce phage numbers), as well as providing information about the decay rate of the system (as high decay rates would reduce phage numbers). However, environments with high phage decay rates would reduce the phage:cell ratio and bias the decision towards lysis, which seems counterproductive, as environments which are hostile to phage particles should bias towards lysogeny.

Concurrent binding of phages can be considered a costly signal, as it is generally associated with irreversible binding, and concurrently-bound phage leads to fewer free phages in the environment compared to superinfection exclusion. Sorptive scavenging of phages by lysogens also has a cost, as it makes it more difficult for related phages to find a sensitive host. Additionally, superinfection immunity is not perfect, and environments with high phage concentrations may select for lysogens that have either lost the receptor, or with mechanisms to downregulate the receptor upon infection (Berryhill e*t al.*, 2023).

Another mechanism that has been shown to modify the lysis-lysogeny decision are quorum sensing systems, which serve as a cheaper alternative for measuring host availability and the flow rate of the system, but may have difficulty discriminating between non-hosts, lysogens and sensitive cells (Silpe & Bassler, 2020; Igler & Abedon, 2019) unless the phage can modulate the host quorum sensing system, or supplement with its own (Erez *et al.*, 2017). For example, in the arbitrium system in Bacillus phage phi3T, the arbitirum signal peptide is produced both by actively lysing phages and lysogens, and the signal peptide is degraded by both lysogens and uninfected hosts, thus providing information about the relative number of lytic events and lysogens compared to the number of hosts. At high lysogen densities the lysogen is capable of producing sufficient arbitrium to suppress prophage reactivation, even

in the absence of lytic events. It is thought that an inflow of sensitive hosts would rapidly decay the extracellular signal peptide, suggesting that the signal peptide provides information on the fraction of lysogens in the population (Bruce *et al*., 2021), which may serve a potential mechanism by which the phage can sense the number of lyosgens. Additionally, some phages appear to make their lysis-lysogeny decision independently of the multiplicity of infection, which suggests that they either have an alternative means of gathering information, or that the information in the environment is not useful for making the decision (Zhang *et al.*, 2021).

We attempted to evaluate the behavior of lysogeny responding to the phage:cell ratio and arbitrium concentrations, but both mechanisms rely on the existence of a threshold to switch from lysis to lysogeny. Similar to fixed lysogeny probabilities, although we could find an optimal threshold for any given set of environmental parameters, the optimal threshold changes when environmental parameters change (Appendix 4, Figures A3, A4), and thus compare poorly to our proposed responsive lysogeny functions, which adapt to changing environmental parameters.

Integrating information from the environment to make decisions in phages is not limited to the lysis-lysogeny decision. Lysis timing is another decision that can vary based off environmental conditions (Moussa *et al.,* 2012, Bednarek *et al*., 2022) and can be examined under a similar framework. Our system models a continuous mass-action kinetics chemostatic environment, with a constant inflow of sensitive cells. This was chosen to model steady-state dynamics and invasion at equilibrium, as it has been previously shown that without the inflow of sensitive cells, pure lysogeny is preferred (Wahl *et al*., 2018). Environments with a constant inflow of sensitive cells can exist in rivers, sewage systems, or any such environment with a unidirectional flow, in addition to environments with a resistant population which loses resistance at high rates, such as under phase variation (Shkoporov *et al.*, 2021). But given the variety of natural environments we must ask: what is the natural environment of a phage? The natural environment would drive the phage to evolve differing responses depending on whether phages spend most of their time in chemostatic conditions, epidemic conditions, or fluctuating ones.

Mechanisms of responsive lysogeny have been shown to provide a benefit under epidemic conditions; for instance, arbitrium systems have been shown to be beneficial under serial passages (Doekes *et al.*, 2021) or a fluctuating inflow of sensitive cells, leading to repeated epidemics (Bruce *et al.*, 2021). Under epidemic conditions, initial sensitive cell populations are high, and the number of sensitive cells changes massively over the course of the epidemic, resulting in virulence being favored in early epidemic conditions whereas lysogeny is favored during late epidemic conditions (Berngruber *et al.*, 2013), and phages can benefit from sensing and responding to the state of the infection. Under chemostatic invasion conditions, however, the initial number of sensitive cells has already been driven low by the resident phage, and changes to the number of sensitive cells are much more subtle compared to the changes in epidemic conditions (Figure A4). Our results show a responsive lysogeny decision that is capable of invading these phage-occupied systems under non-epidemic conditions. It is possible that known biological mechanisms have evolved under epidemic and fluctuating conditions, and thus provide benefit under those circumstances and not under chemostatic conditions.

However, all of these models are that of well-mixed environments, opposed to structurally-organized environments such as biofilms and microcolonies. These environments may be more biologically relevant, but are more complex to analyze.

Our model is not without caveats, as we used a simple model for demonstration. In reality, phage infections take time to progress and do not occur instantaneously. During that time, it is possible for superinfections to occur, which would both sorptively scavenge more phages and can provide information to the phage to make its lysogeny decision. Binding of phages to these intermediate infected cells likely increase the decay of free phages in the environment, which would thereby increase the number of sensitive cells required for lytic invasion, though the magnitude of this effect likely depends on how long the intermediate infected state persists.

We additionally assume that lysogens have perfect immunity, that the cells do not evolve resistance, that phages, lysogens and cells all have identical decay rates, and that lysogens and cells have identical growth rates. We also use concepts borrowed from adaptive dynamics to evaluate our phages, which we recognize may be of limited applicability in phages as phages have a high mutation rate and can greatly vary in phenotype with small mutations.

In conclusion, lysogens that sorptively scavenge phages from the environment protect sensitive cells from infection. This results in a decreased optimal fixed lysogeny probability as the number of sensitive cells increase, and additionally increases the number of sensitive cells required at equilibrium for lytic phages to invade. We further show that, under these conditions, phages that with a probability of lysogeny that responds to the number of lysogens outperform those that cannot, suggesting that it may be beneficial for a phage to sense the number of lysogens in its environment.

BIBLIOGRAPHY

Bednarek A, Cena A, Izak W, Bigos J, Łobocka M. Functional Dissection of P1 Bacteriophage Holin-like Proteins Reveals the Biological Sense of P1 Lytic System Complexity. Int J Mol Sci. 2022 Apr 11;23(8):4231. doi: 10.3390/ijms23084231. PMID: 35457047; PMCID: PMC9025707.

Berngruber TW, Weissing FJ, Gandon S. Inhibition of superinfection and the evolution of viral latency. J Virol. 2010 Oct;84(19):10200-8. doi: 10.1128/JVI.00865-10. Epub 2010 Jul 21. PMID: 20660193; PMCID: PMC2937782.

Berryhill BA, Garcia R, McCall IC, Manuel JA, Chaudhry W, Petit MA, Levin BR. The book of Lambda does not tell us that naturally occurring lysogens of Escherichia coli are likely to be resistant as well as immune. Proc Natl Acad Sci U S A. 2023 Mar 14;120(11):e2212121120. doi: 10.1073/pnas.2212121120. Epub 2023 Mar 7. Erratum in: Proc Natl Acad Sci U S A. 2023 Apr 18;120(16):e2303941120. doi: 10.1073/pnas.2303941120. PMID: 36881631; PMCID: PMC10089163.

Berryhill BA, Levin BR. Semantics Count in the Description of the Interactions Between Bacteria and Bacteriophage. Phage. 2025 Mar 17. doi: 10.1089/phage.2024.0063.

Biggs KRH, Bailes CL, Scott L, Wichman HA, Schwartz EJ. Ecological Approach to Understanding Superinfection Inhibition in Bacteriophage. Viruses. 2021 Jul 17;13(7):1389. doi: 10.3390/v13071389. PMID: 34372595; PMCID: PMC8310164.

Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL. Prophages mediate defense against phage infection through diverse mechanisms. ISME J. 2016 Dec;10(12):2854-2866. doi: 10.1038/ismej.2016.79. Epub 2016 Jun 3. PMID: 27258950; PMCID: PMC5148200.

Brännström Å, Johansson J, von Festenberg N. The Hitchhiker's Guide to Adaptive Dynamics. Games. 2013. 4, no. 3: 304-328. https://doi.org/10.3390/g4030304

Brown S, Mitarai N, Sneppen K. Protection of bacteriophage-sensitive Escherichia coli by lysogens. Proc Natl Acad Sci U S A. 2022 Apr 5;119(14):e2106005119. doi: 10.1073/pnas.2106005119. Epub 2022 Mar 28. PMID: 35344423; PMCID: PMC9168506.

Bruce JB, Lion S, Buckling A, Westra ER, Gandon S. Regulation of prophage induction and lysogenization by phage communication systems. Curr Biol. 2021 Nov 22;31(22):5046-5051.e7. doi: 10.1016/j.cub.2021.08.073. Epub 2021 Sep 24. PMID: 34562385; PMCID: PMC8612742.

Cheong KH, Wen T, Benler S, Koh JM, Koonin EV. Alternating lysis and lysogeny is a winning strategy in bacteriophages due to Parrondo's paradox. Proc Natl Acad Sci U S A. 2022 Mar 29;119(13):e2115145119. doi: 10.1073/pnas.2115145119. Epub 2022 Mar 22. PMID: 35316140; PMCID: PMC9060511.

Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, Melamed S, Leavitt A, Savidor A, Albeck S, Amitai G, Sorek R. Communication between viruses guides lysis-lysogeny

decisions. Nature. 2017 Jan 26;541(7638):488-493. doi: 10.1038/nature21049. Epub 2017 Jan 18. PMID: 28099413; PMCID: PMC5378303.

Herskowitz I, Hagen D. The lysis-lysogeny decision of phage lambda: Explicit programming and responsiveness. Annu. Rev. of Genetics. 1980. 14: 399–445.

Hewson I, Fuhrman JA. Viriobenthos production and virioplankton sorptive scavenging by suspended sediment particles in coastal and pelagic waters. Microb Ecol. 2003 Oct;46(3):337-47. doi: 10.1007/s00248-002-1041-0. Epub 2003 Sep 17. PMID: 14502409.

Igler C, Abedon ST. Commentary: A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. Front Microbiol. 2019 Jun 3;10:1171. doi: 10.3389/fmicb.2019.01171. PMID: 31214137; PMCID: PMC6557168.

Leavitt JC, Woodbury BM, Gilcrease EB, Bridges CM, Teschke CM, Casjens SR. 2024. Bacteriophage P22 SieA-mediated superinfection exclusion. mBio15:e02169-23.https://doi.org/10.1128/mbio.02169-23

Maslov S, Sneppen K. Well-temperate phage: optimal bet-hedging against local environmental collapses. Sci Rep. 2015. 5, 10523. https://doi.org/10.1038/srep10523

Moussa SH, Kuznetsov V, Tran TA, Sacchettini JC, Young R. Protein determinants of phage T4 lysis inhibition. Protein Sci. 2012 Apr;21(4):571-82. doi: 10.1002/pro.2042. Epub 2012 Mar 2. PMID: 22389108; PMCID: PMC3375757.

Payne P, Geyrhofer L, Barton NH, Bollback JP, CRISPR-based herd immunity can limit phage epidemics in bacterial populations. eLife. 2018. https://doi.org/10.7554/eLife.32035.

Ptashne MA. Genetic Switch: Phage Lambda Revisited. CSHL Press. 2004.

Shkoporov AN, Khokhlova EV, Stephens N, Hueston C, Seymour S, Hryckowian AJ, Scholz D, Ross RP, Hill C. Long-term persistence of crAss-like phage crAss001 is associated with phase variation in Bacteroides intestinalis. BMC Biol. 2021 Aug 18;19(1):163. doi: 10.1186/s12915-021-01084-3. PMID: 34407825; PMCID: PMC8375218.

Silpe JE, Bassler BL. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. Cell. 2019 Jan 10;176(1-2):268-280.e13. doi: 10.1016/j.cell.2018.10.059. Epub 2018 Dec 13. PMID: 30554875; PMCID: PMC6329655.

Silva JB, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. FEMS Microbiol Lett. 2016 Feb;363(4):fnw002. doi: 10.1093/femsle/fnw002. Epub 2016 Jan 10. PMID: 26755501.

Sinha V, Goyal A, Svenningsen SL, Semsey S, Krishna S. In silico Evolution of Lysis-Lysogeny Strategies Reproduces Observed Lysogeny Propensities in Temperate Bacteriophages. Front Microbiol. 2017 Jul 26;8:1386. doi: 10.3389/fmicb.2017.01386. PMID: 28798729; PMCID: PMC5526970.

Stewart FM, Levin BR. The population biology of bacterial viruses: why be temperate. Theor Popul Biol. 1984. 26(1):93–117.

St-Pierre F, Endy D. Determination of cell fate selection during phage lambda infection. Proc Natl Acad Sci U S A. 2008 Dec 30;105(52):20705-10. doi: 10.1073/pnas.0808831105. Epub 2008 Dec 19. PMID: 19098103; PMCID: PMC2605630.

Wahl, LM, Betti, MI, Dick, DW, Pattenden T, Puccini AJ. Evolutionary stability of the lysis-lysogeny decision: Why be virulent?. Evolution. 2019, 73: 92-98. https://doi.org/10.1111/evo.13648

Wolfram Research, Inc., Mathematica, Version 11.1, Champaign, IL (2017).

Wolfram Research, Inc., Mathematica, Version 13.2, Champaign, IL (2022).

Zeng L, Skinner SO, Zong C, Sippy J, Feiss M, Golding I. Decision making at a subcellular level determines the outcome of bacteriophage infection. Cell. 2010 May 14;141(4):682-91. doi: 10.1016/j.cell.2010.03.034. PMID: 20478257; PMCID: PMC2873970.

Zhang K, Pankratz K, Duong H, Theodore M, Guan J, Jiang A, Lin Y, Zeng L. Interactions between Viral Regulatory Proteins Ensure an MOI-Independent Probability of Lysogeny during Infection by Bacteriophage P1. mBio. 2021. 12:10.1128/mbio.01013-21. https://doi.org/10.1128/mbio.01013-21

Appendix

1. System Model

1.1. System

$$\frac{\partial S}{\partial t} = \psi S - \omega S - bS\sum_{n=1}^{N} V_n + c$$
$$\frac{\partial L_x}{\partial t} = \psi L_x - \omega L_x + p_x bSV_x - iL_x$$
$$\frac{\partial V_x}{\partial t} = (1-p_x)\beta bSV_x - \omega V_x - b(S + \sum_{n=1}^{N} L_n)V_x + \beta iL_x \quad (1)$$

1.2 Growth Rate

$$\psi = \tau(1 - (S + \sum_{n=1}^{N} L_n)K^{-1} \quad (2)$$

S	The number of sensitive cells.
L_x	The number of phage *x* lysogen cells.
P_x	The number of phage *x* virions.
ψ	The growth rate of cells.
ω	The decay rate of the system.
b	The binding rate of the phage.
p_x	The probability of lysogeny for phage *x*.
i	The induction rate of the lysogen.
c	Inflow of sensitive cells.
β	The burst size of the phage.
τ	The maximum growth rate.

2. Derivation of Invasibility Regions

2.1 Lytic Phage Invasibility Regions

2.1.1. Lytic Phage with Nonproductive Binding

In order for a lytic phage (V_2) to invade a system with *S,* V_1, L_1, at equilibrium,

$$0 < (1 - p_2)\beta bSV_2 - \omega V_2 - b\left(S + \sum_{n=1}^{N} L_n\right)V_2 + \beta iL_2 \tag{3}$$

Since a lytic phage has $p_2 = 0$, $L_2 = 0$

$$0 < \beta bSV_2 - \omega V_2 - b(S + L_1)V_2$$
$$0 < V_2(\beta bS - \omega - bS - bL_1) \tag{4}$$

At the moment of invasion, $V_2 = \epsilon > 0$

$$0 < \beta bS - \omega - bS - bL_1$$
$$0 < (\beta b - b)S - \omega - bL_1$$
$$0 < (\beta - 1)bS - \omega - bL_1$$
$$bL_1 < (\beta - 1)bS - \omega$$
$$L_1 < (\beta - 1)S - \frac{\omega}{b} \tag{5}$$

2.1.2. Lytic Phage with No Nonproductive Binding

In the specific case where phages do not bind to lysogens, we remove the $-b\sum_{n=1}^{N} L_n V_2$ term. In order for the lytic phage to invade a system with *S,* V_1, L_1, at equilibrium,

$$0 < (1 - p_2)\beta bSV_2 - \omega V_2 - bSV_2 + \beta iL_2 \tag{6}$$

Since a lytic phage has $p_2 = 0$, $L_2 = 0$

$$0 < \beta bSV_2 - \omega V_2 - bSV_2$$
$$0 < V_2(\beta bS - \omega - bS) \tag{7}$$

At the moment of invasion, $V_2 = \epsilon > 0$.

$$0 < \beta bS - \omega - bS$$
$$\omega < Sb(\beta - 1)$$
$$S < \frac{\omega}{b(\beta - 1)} \tag{8}$$

2.2. Lysogenic Phage Invasibility Regions

For a lysogenic phage of $p_2 \neq 0$,

$$0 < (1 - p_2)\beta bSV_2 - \omega V_2 - b\left(S + \sum_{n=1}^{N} L_n\right)V_2 + \beta iL_2 \tag{9}$$

For lysogens $L_2 = \epsilon > 0$,

$$\begin{aligned} 0 &< (1 - p_2)\beta bSV_2 - \omega V_2 - b(S + L_1)V_2 \\ 0 &< V_2((1 - p_2)\beta bS - \omega - bS - bL_1 \end{aligned} \tag{10}$$

At the moment of invasion, $V_2 = \epsilon > 0$,

$$\begin{aligned} 0 &< (1 - p_2)\beta bS - \omega - bS - bL_1 \\ bL_1 &< \big((1 - p_2)\beta b - b\big)S - \omega \\ L_1 &< \big((1 - p_2)\beta - 1\big)S - \frac{\omega}{b} \end{aligned} \tag{11}$$

2.3. Lysogen Invasibility Region

In order for a lysogen (L_2) to invade a system with V_1, L_1, at equilibrium,

$$0 < \psi L_2 - \omega L_2 + p_2 bSV_2 - iL_2 \tag{12}$$

For $V_2 = \epsilon > 0$,

$$\begin{aligned} 0 &< \psi L_2 - \omega L_2 - iL_2 \\ 0 &< L_2(\psi - \omega - i) \end{aligned} \tag{13}$$

and since $L_2 = \epsilon > 0$,

$$\begin{aligned} 0 &< \psi - \omega - i \\ 0 &< \tau\left(1 - S + \sum_{n=1}^{N} L_n\right)K^{-1} - \omega - i \\ 0 &< \tau\left(1 - \frac{S + L_1}{K}\right) - \omega - i \\ 0 &< \tau - \frac{\tau S}{K} - \frac{\tau L_1}{K} - \omega - i \\ \frac{\tau L_1}{K} &< \tau - \frac{\tau S}{K} - \omega - i \\ L_1 &< K - S - \frac{K\omega}{\tau} - \frac{Ki}{\tau} \end{aligned} \tag{14}$$

3. Derivation of responsive probability of lysogeny functions

3.1. Invasion-Optimal Responsive Lysogeny with and without binding to lysogens

For the invasion-optimal responsive lysogeny functions, we will use the equation that defines the region that can be invaded by lytic phages, and use that to define the region to undergo lysis, as that maximizes invasibility. These equations are

$$S > \frac{\omega}{b(\beta - 1)} \tag{15}$$

for environments without binding to lysogens and

$$L_1 < (\beta - 1)S - \frac{\omega}{b} \tag{16}$$

with binding to lysogens.

We use this equation to define the boundary between choice between lysis and lysogeny. The probability of lysogeny is a sigmoid function centered around the equation, with a steepness of 1. This results in

$$p_{nb} = \frac{1}{1 + e^{-(\frac{\omega}{b(\beta-1)} - S)}} \tag{17}$$

for environments with no binding to lysogens, and

$$p_{io} = \frac{1}{1 + e^{-(L_1 - (\beta-1)S - \frac{\omega}{b})}} \tag{18}$$

for environments with binding to lysogens.

4 Mechanistically-informed responsive lysogeny functions

4.1 Multiplicity of Infection

We use the simplification of the phage:bacteria ratio as a stand-in for multiplicity of infection (MOI). Above a certain ratio of phages to sensitive cells, the phage shifts from lysis to lysogeny. Hence, lysis occurs when

$$threshold > \frac{\sum_{x=1}^{N} V_x}{S} \tag{19}$$

and, when fitted to a logistic function for ease of simulation,

$$p_{MOI} = \frac{1}{1 + e^{-(\frac{\sum_{x=1}^{N} V_x}{S} - threshold)}} \tag{20}$$

Given the existence of the free *threshold* parameter, we can find an optimal value for *threshold* in a given set of environmental parameters (Figure A2A); however, changing the environmental parameters changes the optimal *threshold* and any given optimal *threshold* does not necessarily perform well under different environmental parameters (Figure A2B).

In theory it may be possible to create a *threshold* that responds to environmental parameters, but that is out of scope of this paper, and a step too far from its mechanistic inspiration.

4.2 Arbitrium

The arbitrium system also functions by shifting the phage from lysis to lysogeny once the concentration of arbitrium (A) is above a certain threshold. Thus, lysis occurs when

$$threshold > A \tag{21}$$

and

$$p_{arbitrium} = \frac{1}{1 + e^{-(A - threshold)}} \tag{22}$$

Additionally, arbitrium must be explicitly tracked in the environment:

$$\frac{\partial A}{\partial t} = a_l \left(\sum_{x=1}^{N} L_x \right) + a_p \left(\sum_{x=1}^{N} (1 - p_x) b S V_x \right) - b_a \left(S + \sum_{x=1}^{N} L_x \right) A - \omega A \tag{23}$$

which introduces three additional parameters: the production rate of arbitrium by lysogens (a_l) and lytic events (a_p), and the rate at which arbitrium is consumed by cells (b_a).

We explicitly ignore the responsive induction that arbitrium is known to regulate as our focus is on the responsive lysis-lysogeny decision. Additionally, this assumes that the other competing species also produces arbitrium, even if it does not respond to it.

We can find an optimal *threshold* for a given set of environmental parameters (Figure A3A), but changing the environmental parameters also changes the optimal threshold (Figure A3B), and any given threshold does not perform well across all environments. It compares similarly

to that of a fixed lysogeny probability (Figure 4C), possibly due to only low fluctuations in arbitrium concentration.

Supplemental Figures

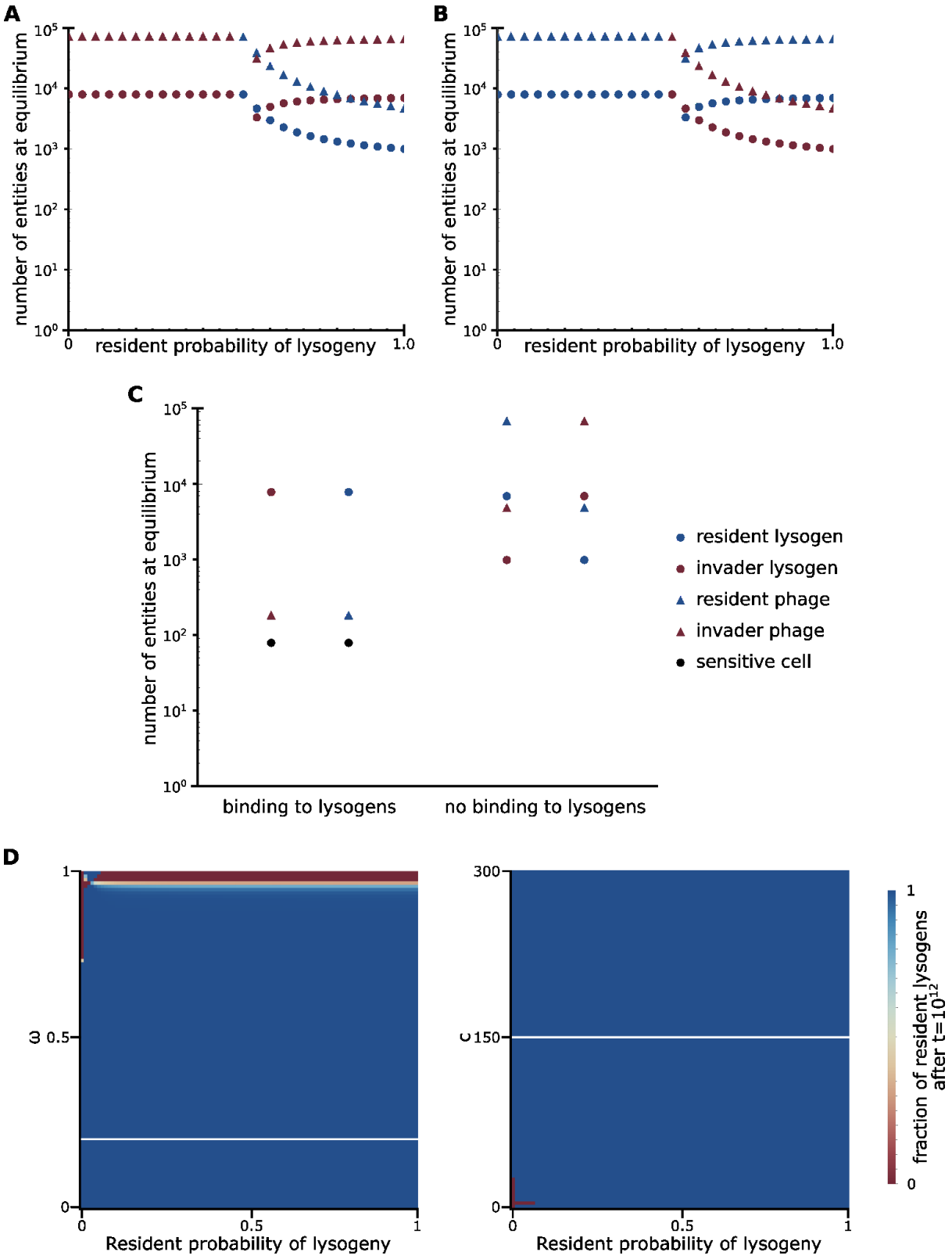


Figure A1. Phages with responsive lysogeny that do not respond to lysogens performs better in environments without binding to lysogens. A) In an environment with no phages binding to lysogens, phages and lysogens with the responsive lysogeny function p_{nb} invade systems with fixed p at analytical equilibrium, and then simulated to $t = 10^7$. **B)** In an environment

with no phages binding to lysogens, phages and lysogens with fixed rates of lysogeny invade systems with phage and lysogens with the responsive lysogeny function p_{nb} at simulated equilibrium, and then simulated to to $t = 10^7$. **C)** Phages with either p_{nb} or p_{io} invade each other in systems at simulated equilibrium with and without binding to lysogens. Left: phages with p_{io} invade a system occupied by p_{nb}. Right: phages with p_{nb} invade a system occupied by p_{io}. Systems were simulated to $t = 10^6$. p_{nb} is incapable of resisting invasion, and is only capable of invading in environments with no binding to lysogens. p_{io} is capable of invading p_{nb} under all conditions, and are able to resist invasion in environments with binding to lysogens. **D)** p_{nb} invading systems of phages and lysogens with a fixed lysogeny probabilities and binding to lysogens, under different flow rates (ω, left) or influx of sensitive cells (c, right), simulated to $t = 10^{12}$.

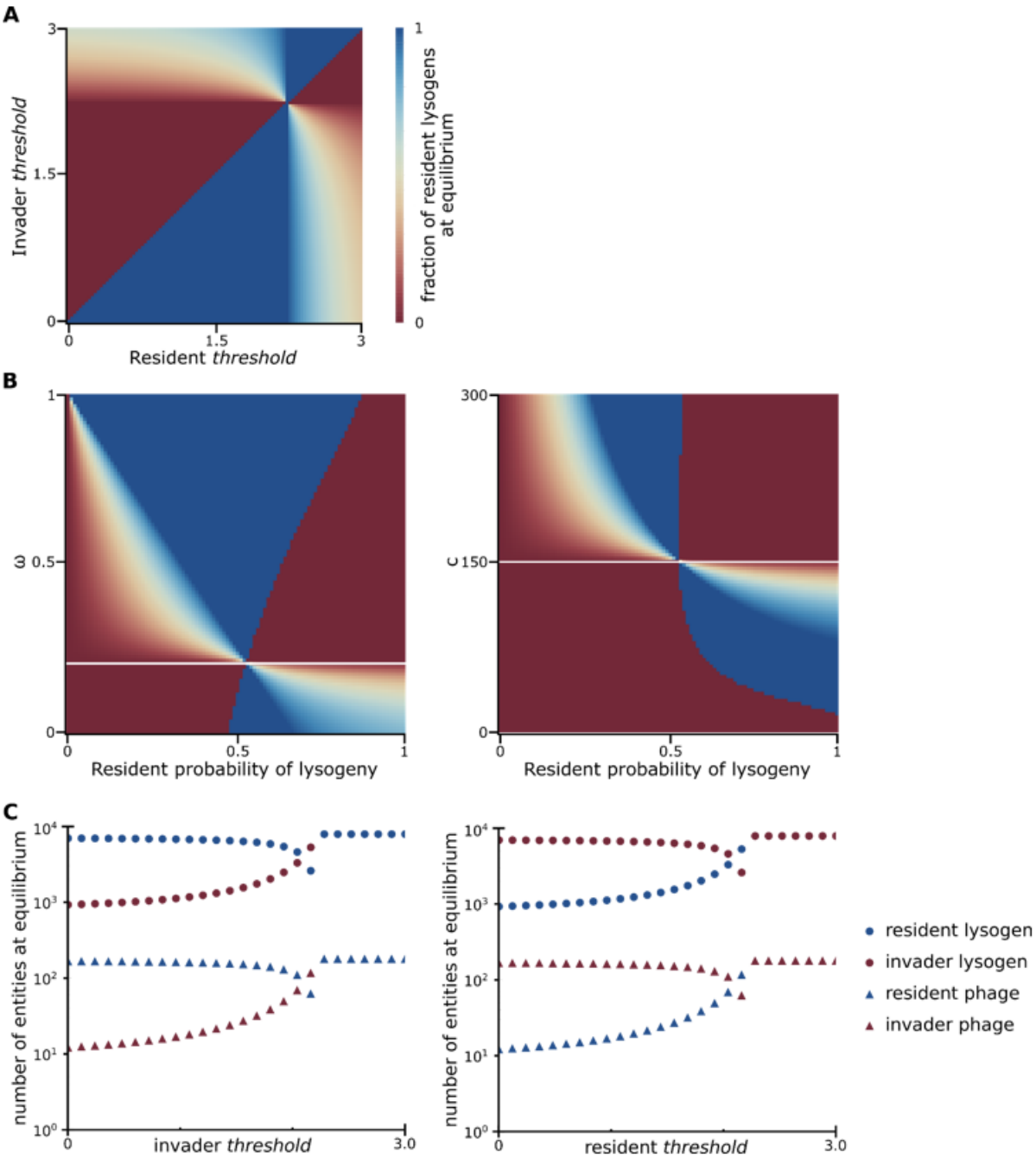


Figure A2. Phages with p_{MOI} are dependent on the threshold at which the response switches. **A)** A system occupied with p_{MOI} lysogens and phages are simulated to $t = 10^{12}$, and invaded by p_{MOI} lysogens and phages with a different threshold, and then simulated to $t = 10^{12}$. **B)** Systems occupied by phages and lysogens with a fixed lysogeny probability at equilibrium are invaded by phages and lysogens with p_{MOI}, with *threshold=2.25*, in environments with different flow rate (ω, left) or inflow of sensitive cells (c, right), and simulated to $t = 10^{12}$. The white bar represents the same parameters in panel A, ω*=0.2* and *c=150*. **C)** A system occupied by p_{io} phages and lysogens is invaded by p_{MOI} with different thresholds (left), or a system occupied by p_{MOI} at different thresholds is invaded by p_{io} phages and lysogens (right).

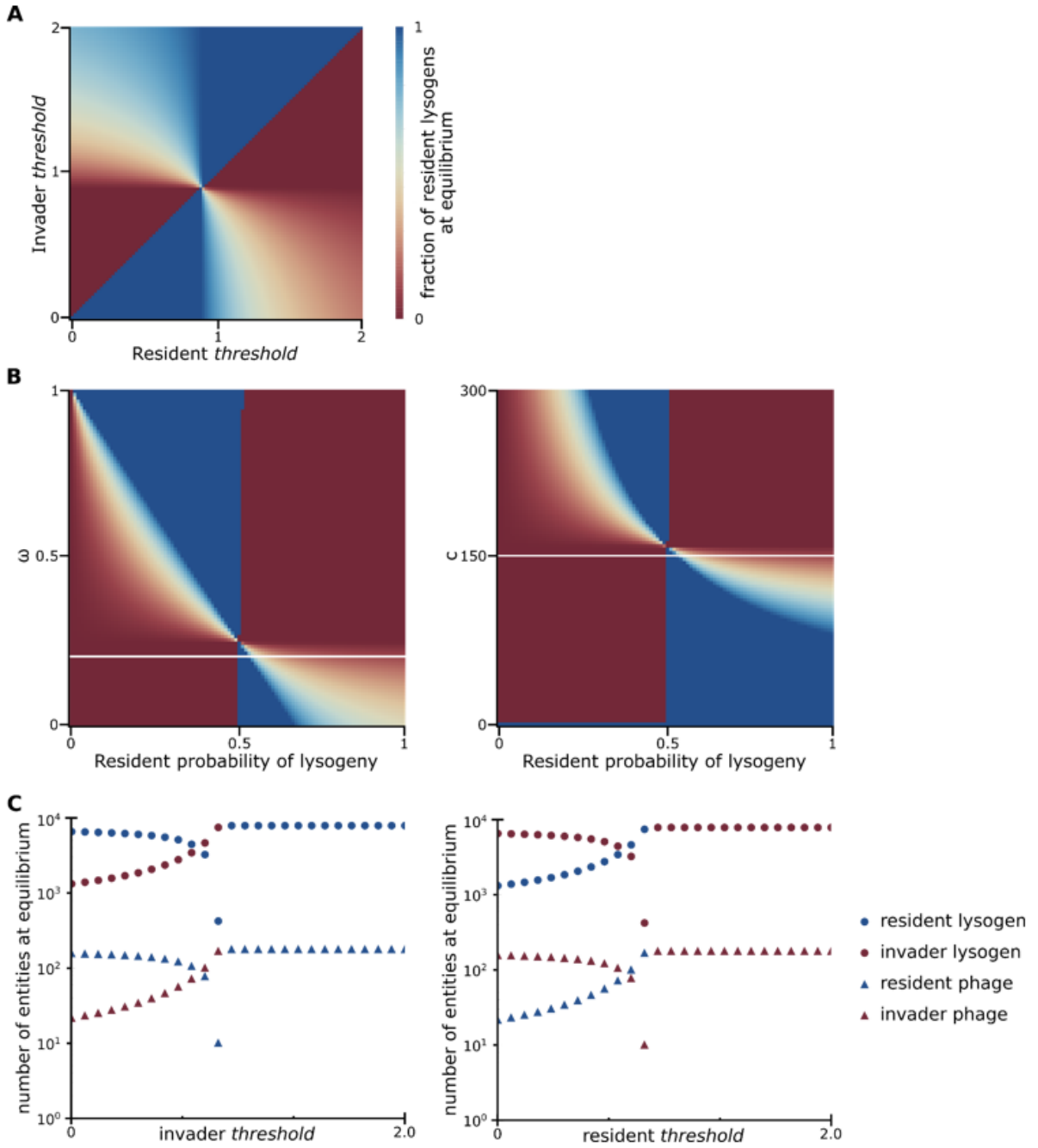


Figure A3. Phages with $p_{arbitrium}$ are dependent on the threshold at which the response switches. A) A system occupied with $p_{arbitrium}$ lysogens and phages are simulated to $t = 10^{12}$, and invaded by $p_{arbitrium}$ lysogens and phages with a different threshold, and then simulated to $t = 10^{12}$. **B)** Systems occupied by phages and lysogens with a fixed lysogeny probability at equilibrium are invaded by phages and lysogens with $p_{arbitrium}$ with *threshold=1*, in environments with different flow rate (ω, left) or inflow of sensitive cells (c, right), and simulated to $t = 10^{12}$. The white bar represents the same parameters in panel A, $\omega=0.2$ and $c=150$. **C)** A system occupied by p_{io} phages and lysogens is invaded by $p_{arbitrium}$ with different thresholds (left), or a system occupied by $p_{arbitrum}$ at different thresholds is invaded by p_{io} phages and lysogens (right). In all systems, $a_l = a_p = b_a = 1$.

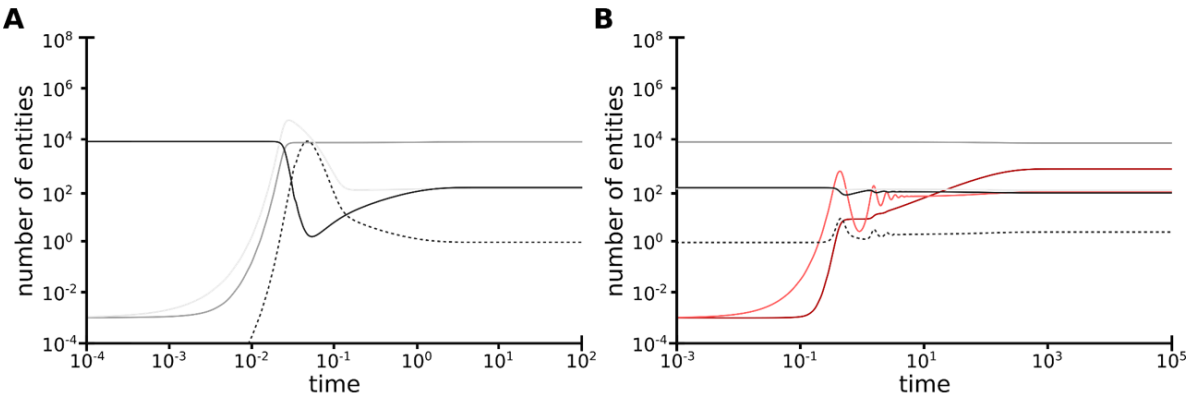


Figure A4. Population structure changes more when invading unoccupied systems compared to invading occupied systems. A) A system with sensitive cells (black) at equilibrium are invaded with phages (light gray) and lysogens (dark gray) with *p=0.9*. **B)** Phages (light red) and lysogens (dark red) with *p=0.1* are introduced to invade the system in panel A. The dotted line represents the ratio of phages to sensitive cells.

2.3 Postamble

A major focus of the paper was focused upon the binding of phages to lysogens, and how this impacts the optimal lysis-lysogeny decision. Given that lysogens will exist in the same environment as the phage, it makes sense that the phage should take into account that it may bind to a lysogen. Some phage lysogens downregulate their receptors, preventing the conspecific killing of phages, but also implies that there is a benefit in counting the number of lysogens in the environment, separate from counting the number of sensitive cells.

I would also like to note that this can apply to cells that are immune to the phage, which would also function as a sink by binding and removing phages from the environment. However, it is unlikely that phages have any ability to detect the immunity status of cells in the environment without a signal being sent from the immune cell.

Additionally, I would like to draw attention to the following term, which is present in the sensitive cell threshold equations for both environments with and without binding to lysogens:

$$\frac{\omega}{b(\beta - 1)}$$

If the number of sensitive cells is greater than this term then lytic replication is net positive, minus adjustments for binding to lysogens.

The key parameters are ω, the decay rate; b, the binding rate of phage to the cell; and β, the burst size of the phage. Of these three parameters, only the burst size is under direct regulatory control of the phage, as phages can tune their burst size to suit the environment through lysis timing (not included within the model), although there are tradeoffs between generation time and burst size.

The binding rate of the phage is dependent on environmental conditions, and some of this information is provided by multiplicity of infection, as the chance that multiple phages bind to the same cell is dependent on the binding rate, though confounded by the number of cells. Many phages are dependent the presence of divalent cations to bind to their target; perhaps sensing cation concentration of the environment can provide information on the binding rate.

In this model, the decay rate is equated to the flow rate of the system, as phages and cells are washed out. With this approach, parallels can be drawn to quorum sensing systems, whose quorum signals sense not only the quorum, but the rate at which secreted molecules flow away from the local environment (Mukherjee & Bassler, 2019), which provides information on whether metabolically expensive effectors that are secreted into the environment will remain in the local environment to provide a benefit. One could expect phages to benefit from gathering the same information, either by encoding their own quorum-sensing machinery, or tapping into existing machinery within their host.

On the other hand, effectors that are washed out of the local environment provide no benefit to the secreting host, whereas phages that are washed out of the local environment likely spread to a novel environment. Given that we model an open system, we do not track the fates of the phages and cells that are washed out of the system. Under chemostatic conditions

those that are washed out are destined to the autoclave, but in a natural environment the fate of those that leave the environment are strongly dependent upon the structure of the environment. For instance, as phages often coexist with their hosts in the gastrointestinal tract of humans, any bacteria or phages that are washed out generally enter a much harsher environment of the soil; however, it is also important to note that only by being washed out can the phage or bacteria find colonize a new human host, otherwise the bacteria and phage will die alongside their human host, which often occurs in chronic infections such as with cystic fibrosis (Winstanley *et al.*, 2016).

There are layers of complexity and different timescales of selection. This paper modelled a simple one, and found important sources of information to which the phage should respond. Simplifications of the model, such as instantaneous infection and identical growth rates of lysogens and uninfected cells, likely hide other sources of information that would also benefit the phage to detect and respond to. Does it hold true over all scales of selection? Unclear, but in this limited, local model, these parameters are key.

The paper elides the mechanism by which the phage can gain access to such information. It was meant to determine what sources of information the phage should respond to, and to suggest avenues of exploration on finding more mechanisms that influence the lysis-lysogeny decision. Physical mechanisms are likely unable to confer such information with perfect fidelity, and instead function as a proxy for information. These physical limitations of information transfer make it difficult for us to differentiate between signal and noise, as known mechanisms often combine multiple sources of information. Perhaps all the sources of information conveyed by the physical mechanism are important to decision-making, but some information must be more important than others, otherwise there is no signal, only noise.

3 Interfering with the Host

3.1 Preamble

A component of a system has a defined function that it must carry out, but just as importantly it must not interfere with the other components that make up the system. This holds true for prophages as well, as once they have decided to undergo lysogeny as opposed to lysis, the prophage ties its fitness to that of its host, and must minimally interfere with its host's processes. The prophage still carries out its own function, at minimum monitoring conditions that would signal induction, but cannot interfere with its host while doing so.

Which is a similar bind that synthetic biology finds itself in. Synthetic biology strives to design systems that function within the context of a cell without disrupting the functions of the cell. Standard issues, such as overexpression of costly proteins, often have negative effects on the cell, which results in a selective pressure to remove or disrupt the system that was added. The system and the host need to be able to coexist.

Lambdoid prophages rely on expressing a phage repressor protein, which repress lytic genes to maintain their lysogenic state, and loss of this repressor triggers the induction of the prophage. Lambdoid phage repressors are under autorepression in order to maintain a balanced level of phage repression that can induce in response to changes in the environment.

Prophages may also maintain a low level of phage repressor to prevent off-target effects. Repressors, like most transcription factors, bind with high affinity to DNA at specific sequences; however, they can also bind to similar sequences, with lower affinity dependent on how dissimilar the sequence is from the target sequence. Thus, reducing the amount of phage repressor present inside the cell also reduces the prevalence of off-target binding, which reduces the chance of inadvertently interfering with host cell processes.

Concurrently, there may also be selection for the host to minimize the amount of interference from the prophage. Hosts that have coevolved with the prophage have also likely reduced the amount of similar sequences present for binding, at least in critical regions which would be deleterious to interfere with; however, this depends upon both the cells coexisting with the lysogen over long periods, but also the later excision of the prophage without the death of the host cell. Thus, through these two processes, we would expect that phage repressors that coevolved with their host would be less deleterious than novel phage repressors to which the host has never been exposed.

In the following paper, we examine the effects of overexpression of transcription factors in cognate and non-cognate hosts, and find that low-affinity off-target DNA binding by transcription factors can lead to growth inhibition.

The following paper has been submitted to *Molecular Systems Biology* and is currently undergoing revisions. My contributions to the paper include data for Figures 1CD, Figure 2, Figure S1A and Figure S10, as well as revisions to the manuscript and experiments vital to revisions such tracking down the provenance of the strains used in the paper.

3.2 Non-Cognate Molecular Interactions Impact Bacterial Growth

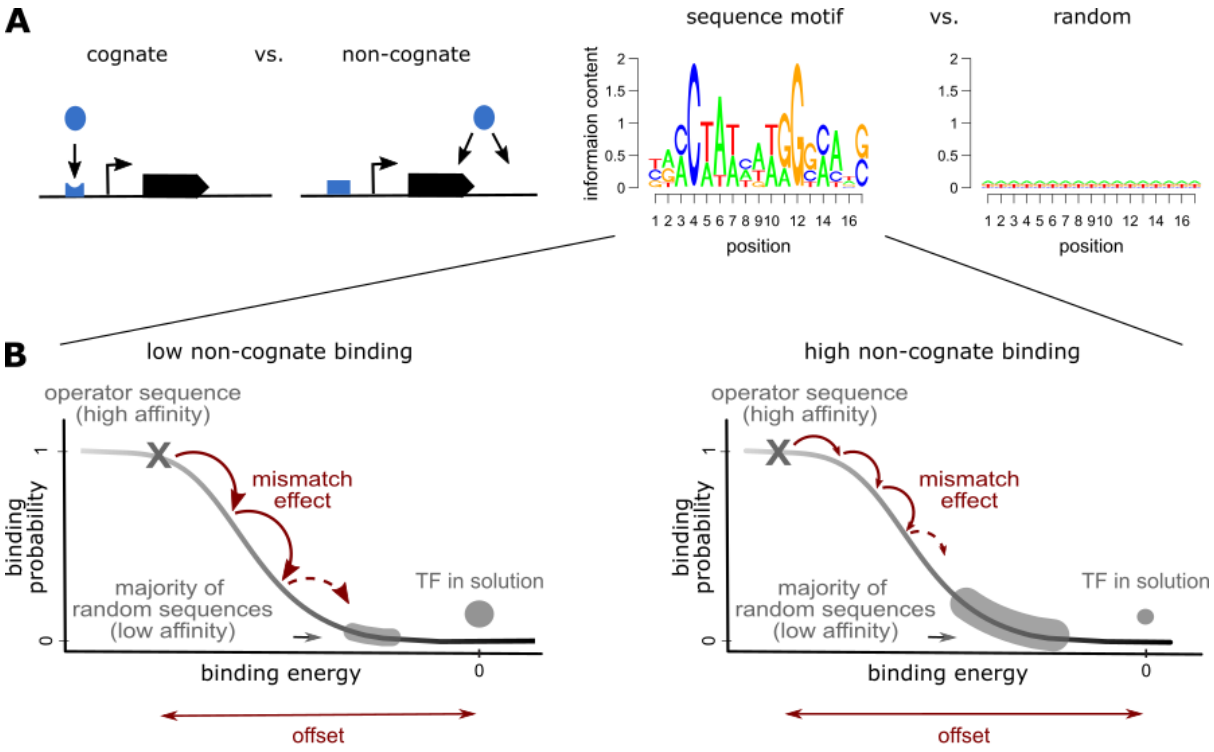
Claudia Igler, Bryan Wu, Claire Fourcade, Bor Kavčič, Torsten Waldminghaus, Florian M. Pauler, Balaji Santhanam, Gašper Tkačik, Călin C. Guet

ABSTRACT

Reliable operation of cellular programs depends upon cognate interactions between biomolecules. In gene regulatory networks, appropriate expression of genes is generally determined through high-affinity binding of transcription factors (TFs) to their cognate DNA sequences. However, the large genomic background likely contains many DNA sequences with varying similarity to TF target motifs, potentially allowing for ample low-affinity non-cognate TF binding. Whether and how non-cognate TF binding impacts cellular function and growth remains unclear. We show that increased expression of native and non-native transcriptional regulators in *Escherichia coli* and *Salmonella enterica* can significantly inhibit population growth across different conditions by inhibiting cell division. This effect depends upon (i) low-affinity TF binding to many DNA sequences, (ii) TF cooperativity, and (iii) TF concentration. Non-cognate TF binding to DNA can thus severely impact fitness, giving rise to fundamental biophysical and physiological constraints on gene regulatory design and evolution.

INTRODUCTION

Biology at all levels depends on the timely recognition and interaction between cognate biomolecules (Box 1A). The importance of molecular encounters between intended targets in the cell is highlighted by the intricate mechanisms that ensure cognate – and suppress non-cognate – interactions, a classic example being that of kinetic proofreading in the loading of amino acids onto tRNAs (Hopfield, 1974). In gene regulation, expression of genes at the right time and place is determined by transcription factors (TFs) binding to their respective cognate operator sites. This binding is usually with high affinity, although functional binding with low affinity to "weak" binding sites has also been recognized, especially in eukaryotes (Kribelbauer et al., 2019; Shahein et al., 2022). Additionally, non-cognate low-affinity interactions can occur at many other DNA sequences because TFs will typically have some propensity to bind even sequences far away from their consensus, as such sequences are likely to be encountered many times within a genome, given typical genome sizes of several million base pairs. These non-cognate interactions are not necessarily detrimental; for example, non-cognate DNA binding was found to speed up the TF search process by sliding along the DNA in a process called "facilitated diffusion" (Elf et al., 2007; Mirny et al., 2009), where conformational changes upon encounter of the target sequence result in cognate binding (Mirny et al., 2009). In prokaryotes, many transcriptional regulators appear to be bound to non-cognate DNA most of the time (Elf et al., 2007; Flyvbjerg et al., 2006; Stracy et al., 2021) and this state could play an important role in gene regulation (Bakk & Metzler, 2004b; Kao-Huang et al., 1977; von Hippel et al., 1974).



C

P22 C2	LacI	λ CI	
-19.4 kT	-22.2 kT	-21.4 kT	offset (operator binding)
>2kT/mismatch	<2kT/mismatch	<2kT/mismatch	average effect
-0.74 kT	0.29 kT	0.43 kT	basepair bias
			mismatch effect to operator
yes	no	yes	cooperativity (between TF molecules)

Box 1. The biophysics and molecular biology of TF crosstalk. (A) In order to elicit an appropriate function, TFs (blue ellipsoid) have to recognize their cognate DNA sites known as operators (blue trough) among a large background of non-cognate sites (blue rectangle), where binding is non-functional (i.e., off-target) and potentially interferes with cellular programs. Binding of TFs to cognate or non-cognate sites can occur by recognizing a TF-specific sequence motif on the DNA (with binding affinity ranging from high to low, depending on the similarity of the target site with the TF's consensus sequence (Gerland et al., 2002)) or through generic, inherently low-affinity interactions with any DNA sequence. **(B)** Schematics of binding probability curves for TFs with lower (e.g. P22 C2, left panel) or higher (e.g. λ CI, right panel) likelihood to bind non-cognate DNA targets are shown. Binding with high affinity usually occurs to a TF-specific DNA motif and is determined by three main factors: (i) the offset gives the binding energy (lower energy = stronger binding) to a single operator sequence (usually the strongest wild-type operator sequence, gray cross) relative to the unbound state (TF in solution; gray dot above the sigmoid). Random DNA sequences are usually bound with low affinity and are located at the lower end of the sigmoid (gray curved box) with energy values approaching that of the TF in solution. The number of TFs being bound to random sequences or free in solution are indicated by the size of the gray box and dot. (ii) Basepair

mismatches to the wild-type operator sequence incur an energy penalty and increase the binding energy (i.e. weaken the binding; shown by the red arrows for each mismatch) based on motif-dependent basepair binding energy mismatches and overall basepair bias (given here as average AT preference – average GC preference). Sequences with a higher number of mismatches to the operator sequence will be more abundant than ones with few mismatches, which is symbolized by the shading of the binding sigmoid. Hence, for a TF with large negative offset, such as λ CI, which is aided by (iii) high cooperativity, and an energy matrix characterized by small mismatch effects, many random sequences will not be far from or even on the rise of the sigmoid, making higher occupancy likely (see C). **(C)** Table compares these three factors (offset, basepair mismatches and cooperativity) for three well-characterized TFs, showing that all of them are fulfilled for λ CI, but only partially for P22 C2 and LacI (Barnes et al., 2019; Hilchey et al., 1997; Igler et al., 2018; Vilar & Saiz, 2005), making λ CI an obvious candidate for substantial non-cognate low-affinity binding.

Non-cognate TF-DNA interactions are often referred to as non-specific, as they are likely very weak and transient (Bakk & Metzler, 2004a; Elf et al., 2007). It is important to realize, however, that specificity is a biophysical property of a TF, describing – depending on the field – either the TF's relative ability to discriminate high and low binding affinity DNA sites by its residence time when bound, or the TF's binding mode and structural conformation upon interacting with the DNA (Jonge et al., 2022) (Box 1). Whether any such binding is biologically functional and has been selected for (cognate) or spurious (non-cognate) is a separate question. In this manuscript, we will focus on binding with high or low affinity at either cognate or non-cognate DNA sites. Low-affinity TF-DNA interactions at a non-cognate DNA site could be mediated by various molecular mechanisms: sequence-independent, purely electrostatic TF-DNA interactions, sequence-dependent interactions with low overlap with the consensus motif, or perhaps even short-lived stochastic conformational transition of the TF that can affect its sequence-dependent binding affinity. As TFs have to search for their cognate DNA targets among an extensive genomic background, they may encounter many non-cognate sites with sufficient target sequence similarity to allow for binding — either with high or low affinity — thus potentially trapping TF molecules (Gerland et al., 2002; Mirny et al., 2009). While non-cognate binding of TFs could be neutral, it could also incur substantial growth costs for the cell through (i) high-affinity binding to few DNA sites with high similarity to the target sequence; or (ii) low-affinity binding to many sites across the genome that accrues to a large overall effect. While in (i) the growth effect is more likely to stem from regulatory interference, the consequences of (ii) remain largely unknown. However, overexpression of one of the *E. coli* nucleoid-associated proteins (NAPs) involved in DNA organization, histone-like nucleoid-structuring protein (H-NS), can lead to growth arrest (Spurio et al., 1992), suggesting that structural changes in chromosome organization could play a role.

In eukaryotes, which have shorter operator sites, low-affinity binding of TFs seems to be an integral part of gene regulatory function and evolution (Burger et al., 2010; Crocker et al., 2016; Kribelbauer et al., 2019; Shahein et al., 2022). In a theoretical study that explored the impact of non-cognate binding we suggested that genomic low-affinity sites impose the existence of a global biophysical constraint, termed "crosstalk" (Friedlander et al., 2015). Some forms of TF cooperativity and combinatorial regulation can limit this problem (Friedlander et al., 2015), but additional mechanisms, such as out-of-equilibrium proofreading

mechanisms and chromatin-based regulation, may also be at work in eukaryotes (Cepeda-Humerez et al., 2015; Grah et al., 2020; Perkins et al., 2024). In prokaryotes, the importance of non-cognate binding to low-affinity sites has been acknowledged in regulatory interference (Sasson et al., 2012) but has been rarely investigated, especially with regard to its physiological consequences. Experiments addressing the fitness impact of low-affinity binding are lacking entirely as it is generally very difficult to disentangle non-cognate binding effects from physiological or cognate effects with native proteins, which rarely just have one target binding site (Shimada et al., 2018).

RESULTS

Experimental setup

We set out to test the consequences of non-cognate TF binding on bacterial growth by expressing TFs lacking known binding sites in native and non-native host cells. The main challenge in addressing this general question results from the fact that it is not trivial to disentangle physiological effects from non-physiological effects, as many TFs have several cognate binding sites across the bacterial genome and hence multiple or complex functions. Based on these considerations, we decided to use phage repressors, which are present for long periods at low numbers in bacterial hosts during the lysogenic cycle (Oppenheim et al., 2005). As temperate phages likely spend most of their existence as lysogens (Howard-Varona et al., 2017), we expect these repressors to be well adapted to the genomic background of their host. Consequently, these repressors should show few non-cognate binding effects in their hosts — although they might bind cognate binding sites to regulate host genes (Chen et al., 2005) — but non-cognate binding could be more prevalent in bacteria that do not serve as host to the particular phage. Therefore, we use transcriptional regulators from the same protein family from different lambdoid phages, with two of them native to *E. coli* (*λ cI, 434 cI*) and one to *S. enterica* (*P22 c2*), and explore their effects on growth in both bacterial hosts. Transcriptional regulation through activation or repression is well-studied for all three TFs (Fattah et al., 2000; Ptashne, 2004). The phage repressors bind to their cognate operator sites in the phage genome as a dimer, but they can also bind cooperatively to adjacent operators or form short- and long-distance loops (up to 10kbp for λ CI (Priest et al., 2014)) involving two to four dimers (Donner et al., 1997) (Fig. 1A). This type of TF cooperativity was previously found to facilitate non-cognate binding (Pray et al., 1998). Notably, for λ CI significant non-cognate binding has been predicted theoretically (Bakk & Metzler, 2004a; Pray et al., 1998) and substantiated experimentally (Igler et al., 2018), whereas for P22 C2 binding was rapidly lost at just a few mismatches from target motif (Igler et al., 2018), making it less likely to bind non-cognate sites (Box 1B).

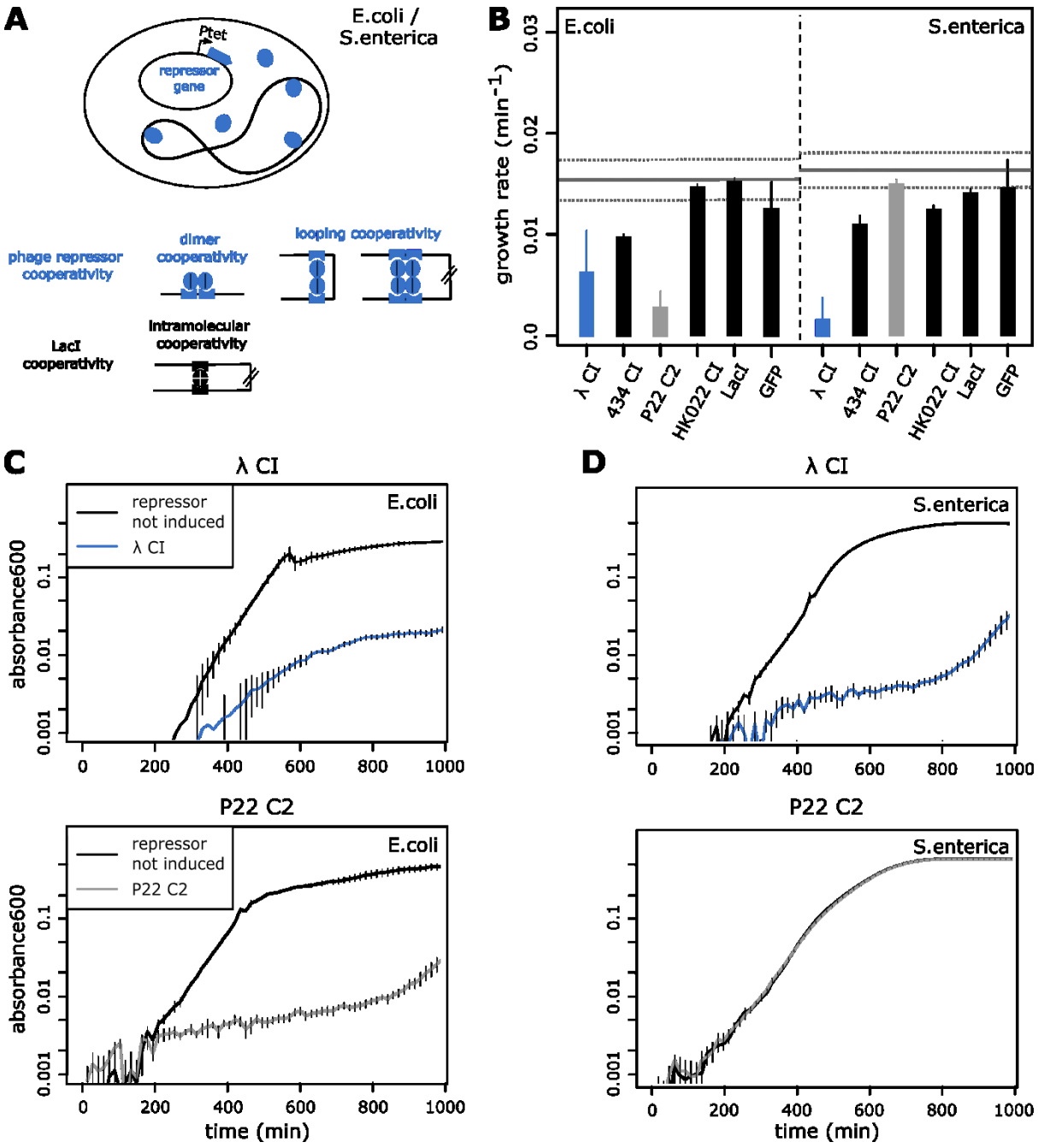


Figure 1. Growth reduction in the presence of repressors in minimal media with glucose. (A) The experimental model system with repressors being expressed from a plasmid and their binding cooperativity modes are shown. **(B)** Growth rates per minute (see Methods for calculation) are shown for *E. coli* and *S. enterica* cells in the absence (black line shows mean and dotted lines standard deviation of uninduced cells) or presence (25ng/ml aTc induction) of λ CI, 434 CI, P22 C2 and LacI repressors or a GFP reporter (bars). **(C,D)** Curves show mean absorbance$_{600}$ for *E. coli* (left) or *S. enterica* (right) cells in the presence (color) or absence (black) of **(C)** λ CI or **(D)** P22 C2. Error bars show standard deviation over 5 replicates (in some cases hidden by the mean) or over 10 replicates for *E. coli* with λ CI.

In addition to the phage repressors, we also chose a well-studied bacterial repressor from a different protein family, LacI, as it is one of the few *Escherichia coli* transcriptional regulators that is known to have only a single cognate target region at the *lac* operon (consisting of three binding sites) but nowhere else on the chromosome (Shimada et al., 2018). LacI uses facilitated diffusion for target search (Elf et al., 2007), which means that its non-cognate binding tendencies should not significantly affect bacterial growth. LacI is a tetrameric protein and cooperativity stems from the fact that one repressor tetramer can loop DNA by binding to one DNA site and then the other (Vilar & Saiz, 2005) — as opposed to cooperativity and looping resulting from binding of two separate repressor molecules for the phage repressors (Fig. 1A). The effect of LacI-type cooperativity on non-cognate binding remains unclear.

The four repressors (*λ cI, 434 cI*, *P22 c2* and *lacI*) were each cloned under the control of an inducible promoter (P_{tet}) onto a low-copy number plasmid (Fig. 1A, S1A). The plasmids were then transformed into *E. coli* and *S. enterica* strains whose genomes do not contain any of the operators for the four repressors (Methods). Our *E. coli* strain constitutively expresses TetR (for regulation of P_{tet}) and LacI, and *S. enterica* constitutively express TetR. We tested the effect of constitutive TetR and LacI expression on the growth of our strains and found no effect for TetR and a slight growth cost for LacI in *E. coli*, potentially because LacI is expressed at higher numbers than from its wildtype operon (Fig. S1B). We then investigated the impact of non-cognate binding of the four different TFs by measuring growth curves of *E. coli* and *S. enterica* and calculating growth rates as a fitness proxy.

Growth effects are highly repressor-specific

Effects on cellular growth due to the presence of repressors were determined via changes in growth rates (Methods). We observed a wide spectrum of growth behaviors across the four repressors and the two bacterial hosts. For cells grown in minimal M9 media (Methods), the standard media used if not specified otherwise, the induction of λ CI resulted in a strong reduction of growth in *E. coli* and an even more substantial reduction in *S. enterica* cells (Fig. 1B-D, S1C, Table S1). The induction of 434 CI repressor also reduced growth, though less than λ CI and with a similar effect in both hosts (Fig. 1B). On the other hand, P22 C2 showed no effect in its native host *S. enterica*, while nearly halting growth when expressed in *E. coli* (Fig. 1B-D, Table S1). There was no significant impact on growth in *S. enterica* with LacI and in both strains for expressing a fluorescent protein instead of a TF (Fig. 1B). Thus, three different repressors stemming from the same TF family showed a broad spectrum of growth effects in their native and non-native bacterial hosts. We further explored these growth effects and their causes by focusing on the two well-characterized repressors, λ CI and P22 C2, which are known to have different propensities for binding at DNA sequences far away from their target motif (Igler et al., 2018) (Box 1B).

We found that the severe reductions in growth rates translated into a fitness effect of repressor expression in direct competition experiments for both, λ CI and P22 C2. As a 'neutral' competitor, we used cells with LacI-expressing plasmids as LacI was the only repressor showing no significant growth defects in both bacteria when expressed from the plasmid — even in addition to the LacI that is constitutively expressed in *E. coli* cells (Fig. 1B, S1B). We added a constitutive YFP-*venus* marker on the LacI-expressing plasmids to determine relative fitness via CFU counts and fluorescence measurements (Fig. S2, Methods). In accordance with the growth rate measurements (Fig. 1,2), repressor expression led to

substantial fitness reductions with λ CI in *E. coli* and *S. enterica* as well as P22 C2 in *E. coli* relative to *lacI+venus* carriers (Fig. S2, Table S2). As the *venus* marker itself resulted in a slight fitness cost (Fig. S2, Methods), the observed increase in *lacI+venus* carrying cells indicates an even more pronounced fitness advantage than measured.

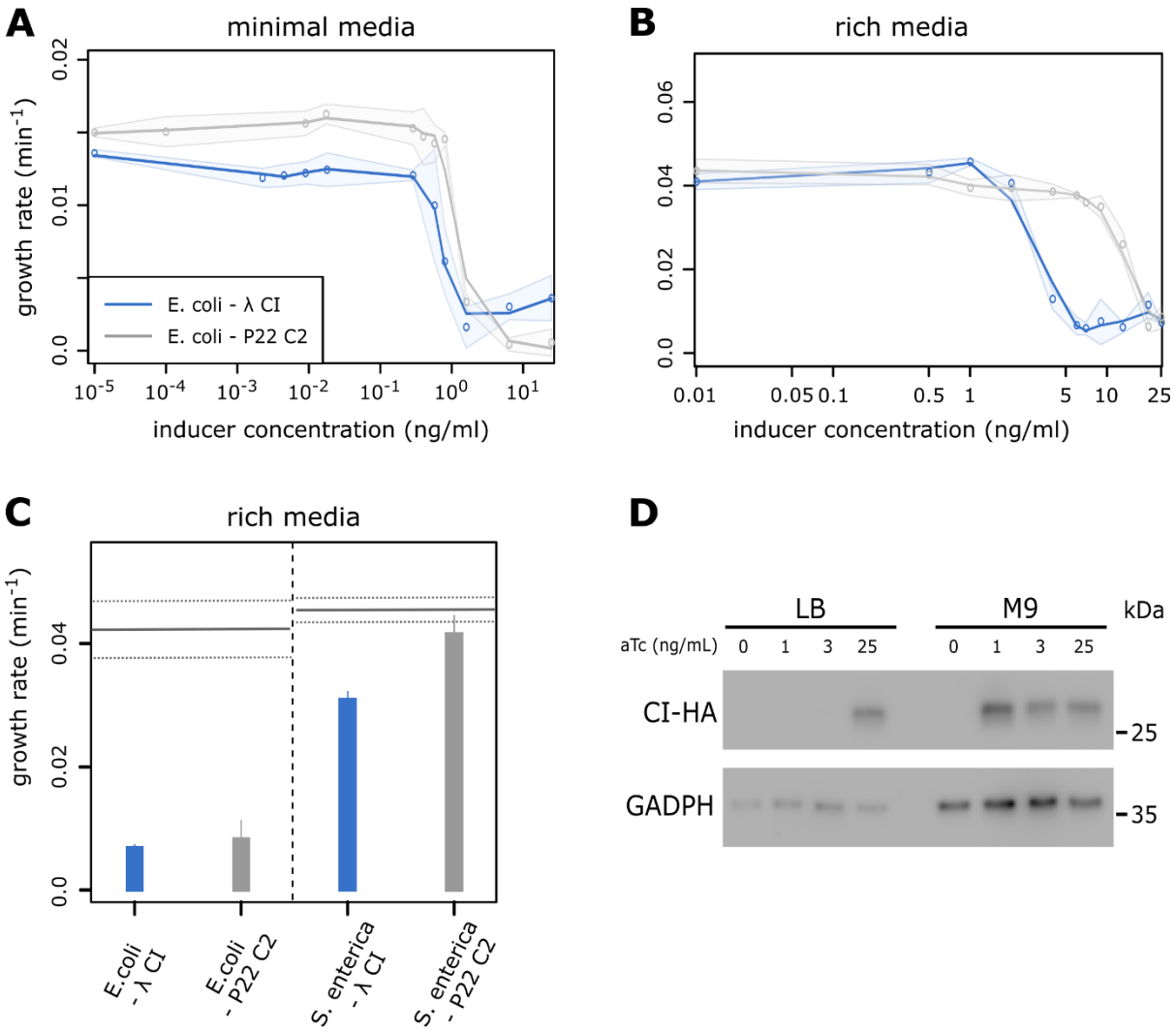


Figure 2. Effect of environment on repressor-dependent growth reduction.
(A,B) Growth rates (per minute) are shown for induction of λ CI or P22 C2 in *E. coli* cells over 11 different inducer concentrations (ng/ml aTc) in **(A)** M9 minimal or **(B)** LB rich media conditions (note the different scale of inducer concentration for rich media). Shaded areas give the standard deviation over 3 replicates and the solid line is a smoothed representation of calculated growth rates. (For *S. enterica* threshold values at which the growth rate effect is visible see Table S3.) **(C)** Growth rates are shown in the absence (black line shows mean and dotted lines standard deviation of uninduced cells) or presence (bars) of λ CI or P22 C2 in *E. coli* and *S. enterica* cells in rich media. Error bars show the standard deviation over 3 replicates. **(D)** Western blot shows how the λ CI-HA protein levels vary in *E. coli* between rich and minimal media for different levels of induction. In LB, λ CI-HA only visibly accumulates under 25ng/ml aTc induction, whereas in M9, 1ng/ml aTc is sufficient to result in accumulation of λ CI-HA. GADPH was used as a loading control. The molecular weight of the proteins is shown in kDa.

Growth effects are environment- and concentration-dependent

As the phage repressor levels are likely higher when expressed from our plasmid than in a lysogen, we tested the magnitude of growth reduction at varying repressor induction levels. In *E. coli*, decreasing the induction level of either repressor showed a gradual recovery of normal growth, with λ CI still having an effect at lower induction levels than P22 C2 (Fig. 2A, Table S3). In *S. enterica*, lower induction levels of λ CI than in *E. coli* resulted in growth reductions, while P22 C2 showed no growth changes at any concentration (Table S4).

As a next step, we tested the effect of environmental conditions by growing bacteria carrying λ CI or P22 C2 in rich media, which allows for faster growth. In rich media (LB), growth inhibition generally occurs at higher inducer levels than in minimal media (Fig. 2A,B), but the qualitative trends remain similar. P22 C2 only shows a growth effect in *E. coli* at higher inducer levels compared to λ CI, and no effect in *S. enterica* (Fig. 2B,C, Table S4). Interestingly, the growth reduction caused by λ CI in rich media occurs at much lower inducer levels in *S. enterica*, but the effect is stronger in *E. coli* (Fig. 2B,C, Table S4, S5).

To examine the cause for the growth reduction variability between different media conditions, we visualized the presence of λ CI tagged with HA via immunoblotting in cells sampled at an OD_{600} ~0.02 in minimal and rich media with different levels of inducer. In accordance with growth reduction observations, we found that λ CI-HA could already be detected at much lower inducer concentrations in M9 than compared to cells grown in LB (Fig. 2D). Lower concentrations of proteins in richer media are expected as faster growth leads to higher dilution of proteins and larger volume of cells (Klumpp et al., 2009), which suggests that repressor concentration is correlated with the growth reduction.

Growth reduction is caused by cooperative, low-affinity binding distributed across the genome

Given the surprisingly detrimental growth effect of the two repressors in several conditions, we set out to determine its cause. Transcriptional repressors are DNA-binding proteins and could therefore interfere with cellular programs through DNA binding at various non-cognate sites (Chakrabarti et al., 2011). To determine the role of TF binding in the observed growth reductions, we used the fact that λ CI is one of the best-studied TFs, and thus a range of mutants exist that change its binding and conformational properties.

We tested the expression of a λ CI mutant defective in DNA binding (Nelson & Sauer, 1986) and found that normal growth was almost completely restored in both *E. coli* and *S. enterica* cells (Fig. 3A, Table S1). Similar results for a λ CI mutant defective in cooperativity between repressor dimers – but not in DNA binding itself – additionally suggest an important contribution from DNA looping or some other form of repressor oligomerization (Fig. 3A, Table S1). This is intriguing as λ CI cooperativity and oligomerization are thought to be strengthened by binding to DNA in a sequence-dependent manner (Mazumder et al., 2017; Vilar & Saiz, 2005), which also appears to play a large role in the absence of cognate sites. Surprisingly, a λ CI mutant that cannot form dimers (Whipple et al., 1994), which should significantly interfere with the ability of λ CI to bind DNA (Weiss et al., 1987), completely restored growth in *E. coli* but not in *S. enterica* cells, potentially suggesting different or

additional mechanisms of λ CI-induced growth reduction in the two species. To exclude the possibility that repressor misfolding or aggregation was responsible for our observations, we over-expressed the chaperone gene *tig* (Saio et al., 2014) along with the repressors (Fig. S1D).

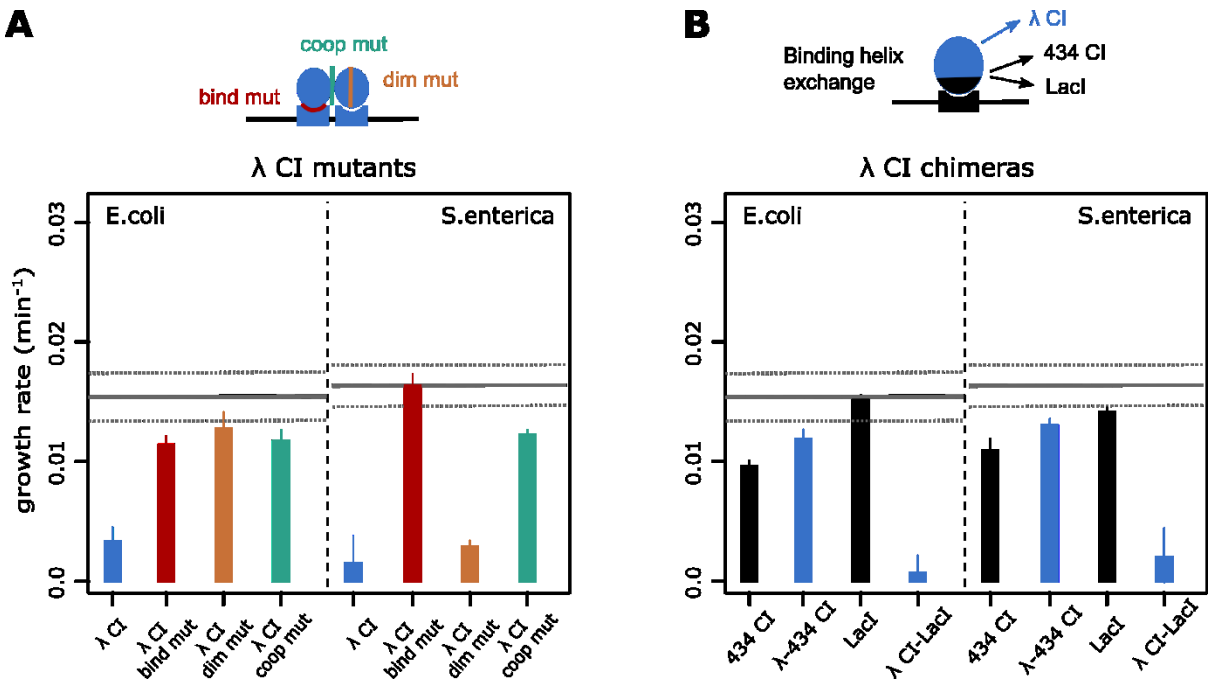


Figure 3. Effect of mutants and chimeric proteins on repressor-dependent growth reduction. Growth rates (per minute) are shown for *E. coli* and *S. enterica* cells in the absence (black line shows mean and dotted lines standard deviation of uninduced cells) or presence of a repressor are shown. **(A)** Comparison between λ CI wildtype (blue), a λ CI binding mutant (red), a λ CI dimerization mutant (orange) or a λ CI cooperativity mutant (turquoise) in minimal media. **(B)** Comparison of phage repressor 434 CI and the bacterial repressor LacI (black) to a chimera of the λ CI protein containing the binding affinity of either 434 CI (λ-434 CI) or LacI (λ CI-LacI) (blue) is shown for minimal media. Error bars show standard deviation over 5 replicates (in some cases hidden by the mean).

Hence, the ability to bind DNA, potentially in a cooperative manner, seems to be central to repressor-mediated growth effects. We tested this hypothesis by creating chimeric TFs, combining λ CI cooperativity with the binding affinity characteristics of another repressor. We replaced the DNA binding helix of λ CI (Methods) with: i) that of another phage repressor, 434 CI, which showed some growth defect; or ii) the bacterial repressor, LacI, which showed no growth defect when expressed from the plasmid. For 434 CI, it has been reported that changes in the cooperativity interface strongly interfere with its binding affinity and the structure of the TF-DNA complex (Donner et al., 1997; Guarnaccia et al., 2004). Our experiments agree with these findings, as the λ-434 CI chimera resulted in a reduction of the growth defect compared to λ-CI in *E. coli* and in *S. enterica* cells (Fig. 3B, Table S1), likely caused by a reduction in binding. In contrast, the growth reductions with the λ CI-LacI chimera were similar in magnitude to wildtype λ CI in both bacterial hosts (Fig. 3B, Table S1). The opposite behavior of LacI and λ CI-LacI could be explained by the chimera combining the high target binding strength of LacI with the strong intermolecular cooperativity and oligomerization potential of λ CI (Fig. 1A, Box 1). The additional binding stabilization through

cooperativity or oligomerization could then result in substantial non-cognate binding and strong interference with cell growth.

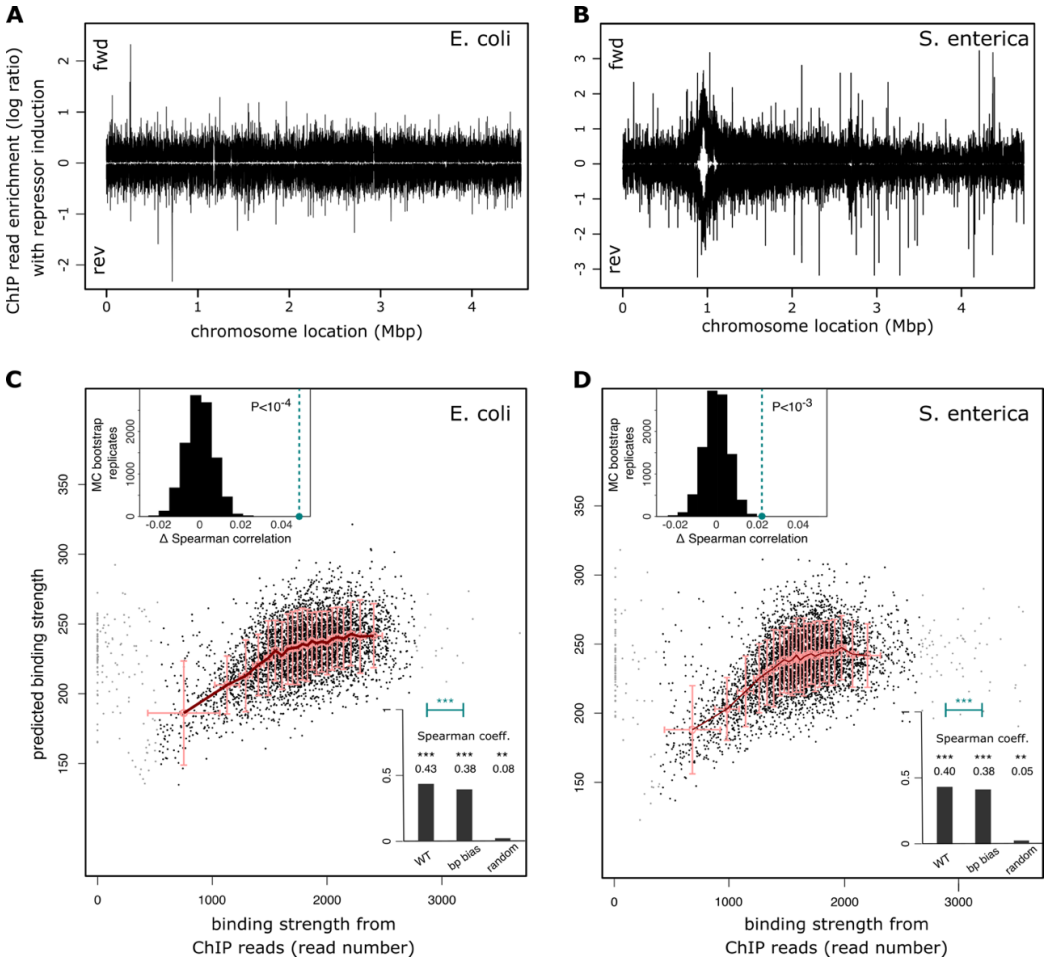


Figure 4. Distributed, low-affinity binding of λ CI across the genomes of *E. coli* and *S. enterica*. (A,B) Distributions of ChIP-sequencing reads for the regions found to be enriched in the experiment with λ CI over the experiment without λ CI are given for the forward and the reverse DNA strand across the **(A)** *E.coli* or **(B)** *S. enterica* genome. Enrichment was calculated within 100bp windows through the log ratio of the scaled coverage (scaled by number of mapped reads) between sample and control ChIP-seq experiments. Number of reads for significantly enriched regions are shown in Fig. S3. **(C,D)** Fit between binding strength predictions of a simple thermodynamic model using the energy matrix for λ CI binding and the ChIP-sequencing reads across 1000bp windows along the **(C)** *E. coli* or **(D)** *S. enterica* genome. In **(D)**, 33 out of 4856 data points showed more than 3600 reads and were omitted for clarity. Lower insets show the calculated Spearman correlations using either the wildtype energy matrix, one that only conserves the λ CI basepair bias (Fits are shown in Fig. S5) or one that has completely reshuffled entries (averaged over 100 permutations). Upper insets in **C** and **D**: Permutation test for the significance of the difference in Spearman correlation between binding predictions using the wildtype energy matrix (first bar in the lower inset) vs prediction with the energy matrix that only conserves λ CI basepair bias (second bar in the lower inset). Black histograms represent the Monte-Carlo-derived null distribution (10^4 random reassignments of ChIP reads to genomic regions), green dot and line show the true excess Spearman correlation. The correlations for *S. enterica* were not substantially affected by the strong binding peak in prophage regions shown in **(B)**.

To determine if the non-cognate binding effects involved either a few essential or many distributed regions of the chromosome, we performed ChIP-sequencing of λ CI in *E. coli* and *S. enterica*. As the data did not reveal strong peaks for specific genes in *E. coli*, we investigated overall binding to the genome by averaging data over 1000bp windows. This analysis indicated weak, low-affinity binding at numerous sites all over the *E. coli* chromosome (Fig. 4A, S3A, Table S6) when using a more lenient cutoff than typically used for strong, high-affinity binding, though completely sequence-independent binding should not be detectable via ChIP-sequencing, even with our more lenient cutoff (Fig. S4, see Methods section 'ChIP-sequencing analysis'). In *S. enterica* we found a broad peak (indicating substantial binding in several adjacent genes) and more distributed weak binding regions than in *E. coli*. Interestingly, this broad peak corresponds to prophage regions on the genome that seem to provide binding hotspots for λ CI (Fig. 4B, S3B, Table S6). As such binding hotspots were absent in *E. coli* but λ CI still showed a growth defect, we did not consider this finding necessary to qualitatively explain our results (although it could explain the stronger growth reduction in *S. enterica*). None of the peaks for either genome appears to encode an essential gene.

As ChIP-sequencing is typically used to determine high-affinity TF binding, there is no established methodology of assessing low-affinity binding in these assays. Here, we used a simple thermodynamic model to predict λ CI binding across the bacterial genome and correlate these model predictions with ChIP-Seq results (see Methods section 'Thermodynamic model'). The predictions showed a surprisingly high degree of correlation with the number of ChIP-Seq reads (Fig. 4C,D), especially since energy matrix models are believed to perform poorly for low affinity sites (Maerkl & Quake, 2007). Even more surprisingly, most (but not all) of this correlation could be accounted for by an energy matrix that conserved only the λ CI preference for the overall basepair composition (Fig. 4C,D, S5, see Methods section 'Thermodynamic model'). In contrast, a negative correlation was observed with the overall basepair preference of P22 C2 (Fig. S5), which indeed is opposite to that of λ CI (Box1). These results could not be explained by an inherent nucleotide bias of the ChIP-sequencing experiments (Fig. S6), which suggests the correct basepair composition bias in a genomic sequence could provide λ CI with sufficient recognition pattern to bind with low affinity. In agreement with this hypothesis, the GC bias of the stronger λ CI cognate operators (O_{R1}, O_{R2}, O_{L1} and O_{L2}) is 52.94%, which is very close to that of the *S. enterica* genome (52.2%) and somewhat higher than that of the *E. coli* genome (50.8%).

The basepair composition bias, while important, did not fully account for the agreement between model predictions based on full TF energy matrix and ChIP results. The sequence-dependent contribution to λ CI recognition, in excess of the basepair composition bias, remained highly significant in *E. coli* and *S. enterica* (as determined by Monte-Carlo permutation tests; Fig. 4C,D, S7). Overall, our results indicate substantial non-cognate, low-affinity binding due to sequence-dependence and basepair bias — which is similar to NAP binding patterns (Spurio et al., 1992; Yamada et al., 1991). Non-cognate binding is facilitated by repressor oligomerization (Pray et al., 1998; Sarkar-Banerjee et al., 2018), and distributed over the thousands of potential low-affinity λ CI binding sites, known to be present in the *E. coli* genome (Chakrabarti et al., 2011).

Low-affinity binding leads to arrest of cell division

The distributed non-cognate DNA binding demonstrated by the ChIP-sequencing data for λ CI is in agreement with the observation that increasing induction of repressor gradually intensifies the growth reduction (Fig. 2A,B). Hence, the ratio between repressor and potential binding sites, i.e., genomic DNA, might determine the magnitude of the growth defect. If cell doubling time is lower than the time needed for DNA replication, i.e., if the growth rate is less than ~0.017 doublings/min in *E. coli* and ~0.022 doublings/min in *S. enterica* (Cooper & Ruettinger, 1973), (which is the case in our experiments in minimal media: 0.0147($\pm$0.0008) doublings/min and 0.0163($\pm$0.002) doublings/min respectively), each bacterial cell contains on average only a single chromosome. At faster growth rates, replication cycles are overlapping and daughter cells inherit 2-8 origins at birth, together with partially replicated chromosomes (Nordström & Dasgupta, 2006). Hence, the richer the medium and the faster the growth, the more DNA will be available to titrate away potentially detrimental non-cognate binding TFs. This could be an additional factor next to growth-related expression differences explaining why much higher induction levels of repressor are required for growth reduction in rich media (Fig. 2): repressors are titrated more due to the presence of more DNA and repressor concentrations are generally lower due to faster dilution as well as larger cell volume.

We set out to test the titration hypothesis by introducing a medium-copy number plasmid carrying four cognate λ CI binding sites into *E. coli* cells with the inducible λ CI-plasmid, which should substantially reduce the number of free λ CI dimers available for non-cognate binding. Although the medium-copy plasmid by itself incurred a growth cost, the growth reduction caused by λ CI was significantly reduced, particularly at low expression levels (Fig. S8), which suggests that the titration of λ CI alleviates the growth reduction.

The titration hypothesis together with the Western Blot and ChIP-sequencing results implies that the overall ratio of chromosomal DNA to repressor proteins is an important factor determining the growth effects. This suggests that non-cognate binding might interfere with global cellular functions. Using microscopy, we found that λ CI-expressing *E. coli* cells generally formed long filaments with increased cell volumes, indicating an inhibition of cell division in these cells (Fig. 5A), which is also characteristic for antibiotics targeting DNA replications and other DNA-related functions (Bollenbach et al., 2009). In the presence of λ CI, induction of the stress response promoter *PsulA* is low (Fig. 5B, S9), making the stress response an unlikely cause for the substantial inhibition of cell division (Arends & Weiss, 2004). Furthermore, such a weak stress response will not lead to substantial self-cleavage (i.e., inactivation) of the repressor molecules (Sauer et al., 1982) - particularly because λ CI becomes a poor substrate for self-cleavage at higher concentrations (Phizicky & Roberts, 1980). Cell division can also be hindered by the presence of genomic DNA at mid-cell (Cambridge et al., 2014). Fluorescence microscopy indicates that many filamentous cells contain several chromosomal regions, which are enlarged compared to cells without the repressor and which are often located mid-cell (Fig. 5A). Mid-cell located DNA interferes with the establishment of a division septum, suggesting that it is not FtsZ-ring formation, but a subsequent step in the cell division cascade that is disrupted. Some filamentous cells manage to divide after growing to substantial length as division septa start forming between the chromosomal regions (Fig. 5A). Taken together, non-cognate TF binding could lead to inhibition of cell division by retaining chromosomal DNA at mid-cell.

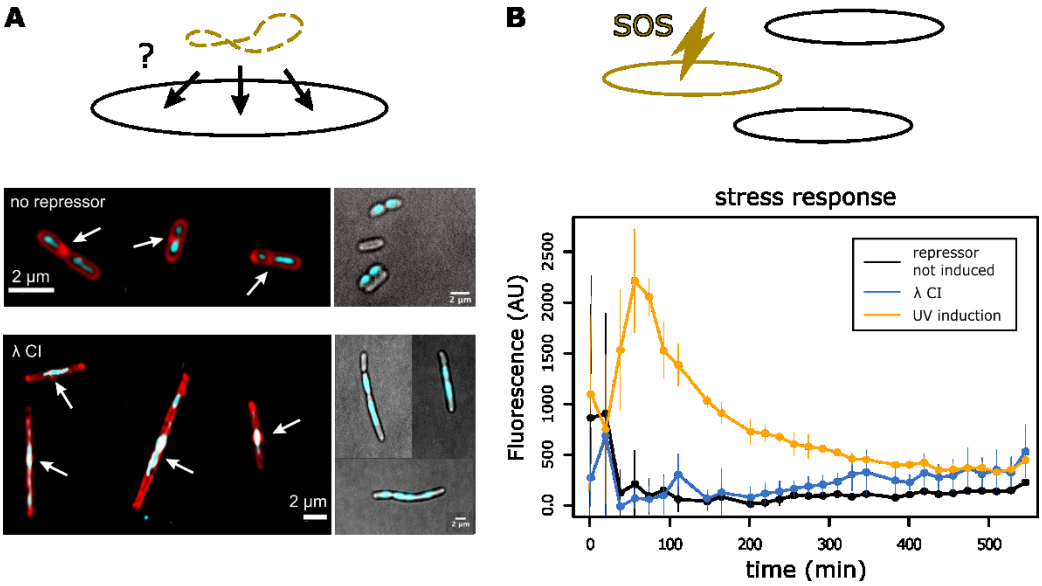


Figure 5. Induction of λ CI interferes with cell division but does not induce the SOS stress response. *E. coli* cells containing the λ CI repressor were **(A)** imaged using dye to visualize the cell membrane in red (Nile Red) and the chromosome in blue (Hoechst) or **(B)** tested for stress response (through the reporter P_{sulA}-*yfp*). **(A)** Cells grown in minimal media were imaged using fluorescence microscopy (see Methods) either in liquid media (left) or on an agar pad (right). Whereas chromosomes (blue) are positioned in the middle of the cell in the absence of λ CI (upper), the presence of λ CI (lower) leads to filamentous cells, which often contain several chromosomes, and their positioning overlaps with the middle of the cell including the area of division septa formation (thicker red dots). **(B)** Curves show mean fluorescence for cells that were not induced (black) or cells that were stressed by inducing λ CI expression (blue) or using UV exposure for 30sec (yellow). Error bars show standard error over 3 replicates. (Corresponding mean growth curves are shown in Fig. S9).

DISCUSSION

We investigated the consequences of non-cognate interactions in molecular recognition of DNA by proteins, using three different but related phage TFs and a bacterial repressor, which produced a wide range of effects on host cell growth, from high to no growth costs, and higher to lower costs in the non-native compared to the native host (Fig.1B, 2C). Taking advantage of the rich and well-established genetics and biochemistry of the classic bacteriophage repressor λ CI, we found that its growth cost likely results from cooperative, low-affinity binding, which interferes with growth by inhibiting cell division. Our data support the hypothesis that binding strategies of prokaryotic TFs are under selection to avoid low-affinity binding to the genomic background (Chakrabarti et al., 2011). These results may not generalize to eukaryotes, where weak binding to low-affinity sites can play a functional role (Kribelbauer et al., 2019), and the deleterious non-cognate aspects of such binding may be held in check by mechanisms such as silencing or proofreading (Grah et al., 2020; Zoller et al., 2022). These considerations highlight the fundamental differences in gene regulatory design between prokaryotes and eukaryotes, and therefore differing evolutionary constraints (Wunderlich & Mirny, 2008).

For prokaryotes – in contrast to eukaryotes – TF target sites are sufficiently long to allow recognition of single cognate operators with high affinity (Wunderlich & Mirny, 2008). Mismatches to the preferred target sequence lead to progressive loss of binding, but the speed of this loss can vary substantially between individual TFs (Igler et al., 2018). Among the four transcriptional regulators we analyzed here, λ CI combined several properties that together make it a promiscuous DNA binder and thus induce a high fitness cost due to low-affinity binding: (i) strong operator binding (offset) that allows for toleration of more mutations, (ii) low mismatch penalties for individual basepair changes (energy matrix) and (iii) strong cooperativity (Box 1B,C) (Igler et al., 2018). In contrast, for P22 C2, the lower offset and higher mismatch penalties make binding to non-cognate sites less likely (Box 1B,C) (Igler et al., 2018), producing a significant cost only in non-native host cells. LacI, which shows similar binding characteristics as λ CI (including similar basepair bias, see Fig. S5), did show a strong growth effect only if coupled with λ CI's intermolecular cooperativity. Together, the mutant and chimera experiments (Fig. 3) argue for a significant contribution of cooperativity – and likely oligomerization – to the potential for low-affinity binding, which supports the theoretical finding that TF cooperativity does not alleviate deleterious binding when it stabilizes cognate *as well as* non-cognate binding (Friedlander et al., 2015). LacI's lack of intermolecular cooperativity makes it less likely to bind non-cognate sites, suggesting an adaptation for high affinity to only few cognate sites — as opposed to most *E. coli* regulators (Shimada et al., 2018). The continuum of growth effects we observed for only four different TFs reveals that the complex combination of binding cooperativity, offset, and TF concentration determines the propensity for non-cognate binding of a TF and must be tuned to avoid fitness costs. Therefore, considering binding to low-affinity sites may be crucial in choosing TFs for synthetic biological systems to avoid global toxicity effects, as well as unwanted TF titration, which can affect target gene regulation (Brewster et al., 2014; Sasson et al., 2012).

The magnitude of the growth cost depends on a repressor's ability to bind non-cognate low-affinity DNA sites, as well as on a repressor's relative ratio to the total amount of DNA within the cell. Slow cell growth compounds the effect as cells contain less DNA that would provide potential sites for low affinity binding and repressor titration (Chakrabarti et al., 2011), while repressor density is higher than at fast growth (Klumpp et al., 2009). Hence, the slower growth rate in minimal media could allow for greater accumulation of repressor, which in turn slows growth, resulting in negative feedback between repressor accumulation and growth rate, which would explain the difference in repressor concentration between minimal and rich media. This phenomenon is reminiscent of growth bistability in drug resistant bacterial cells, which is caused by feedback between the growth rate and the speed of counteracting toxic agent (Greulich et al., 2015). In our system, the repressors can be seen as "toxic agents", which are "counter-acted" by dilution or titration if cells manage to start growing fast enough; otherwise, the cells are growth-arrested.

Additionally, stress tolerance could be higher under optimal growth conditions as found in rich media (Pleška et al., 2016). However, it does not seem likely that media-specific genes are targeted, as ChIP-sequencing generally revealed distributed, low-affinity binding all over the chromosome (Fig. 4). Rather, inhibition of cell division seems to result at least partially from nucleoid localization at mid-cell (Fig. 5). Clearance of the division site is impeded if sister chromosomes fail to segregate (Cambridge et al., 2014; de Boer, 2010), which may be caused

by the formation of “bridges” between cooperatively bound repressors, holding the sister chromosomes together. As found for λ CI, even non-cognate binding can increase and stabilize DNA loops through the formation of higher-order oligomers (Manzo et al., 2012; Sarkar-Banerjee et al., 2018). This underlines the importance of repressor cooperativity in the observed growth effects.

Generally, compaction of nucleoids seems to be a crucial factor for preventing the completion of division septa through nucleoid occlusion (Sun & Margolin, 2004). Intriguingly, a very similar growth effect has been found with H-NS, one of the NAPs responsible for chromosome organization and compaction in *E. coli* (Spurio et al., 1992; Yamada et al., 1991): H-NS overproduction drastically reduced cell viability, which seemed to be related to the formation of higher-order H-NS oligomers (Spurio et al., 1997) – a state that is favoring its ability to form bridges between DNA regions (Verma et al., 2019). As cell growth, shape, division and DNA replication are thought to be tightly linked in complex and poorly understood ways (Amir et al., 2019), a mechanistic explanation of the observed division inhibition is presently not possible, but prior studies on H-NS combined with our findings make the case for the existence of general constraints on DNA-binding proteins.

Our results suggest that the inherent ability of DNA-binding proteins to occupy non-cognate DNA regions can pose, in addition to potential regulatory interference (Sasson et al., 2012), a substantial challenge for host cell fitness overall – particularly considering facilitating conditions like cellular crowding (Zabet & Adryan, 2013), horizontal gene transfer (Baltrus, 2013) and mutations that alter the binding affinity of a protein. This challenge stems from the fundamental limits to molecular recognition that are set by the biophysics of molecular interactions (Friedlander et al., 2015) and could lead to various non-cognate effects on the physiology of the cell. The four transcription factors used here show a variety of non-cognate effects, which could indicate different selection pressures that have been acting on their binding and cooperativity characteristics as well as different potential for being tolerated in a particular environment. There might also be a variety of mechanisms underlying the growth phenotype, as suggested by the ChIP-sequencing results for λ CI: the additional strong peak region found in *S. enterica* could indicate an additional effect of higher-affinity sites and may explain the stronger growth defect in minimal media as compared to *E. coli* (especially at low repressor concentrations). A potential for high fitness costs, even in native environments as seen here with λ CI and 434 CI, can limit the number and binding affinity of promiscuous DNA-binding proteins in the cell (Friedlander et al., 2015). Less promiscuous binders such as P22 C2 are only detrimental in non-native environments. Hence, the influx of foreign genes through horizontal gene transfer could be considerably impaired through non-cognate binding effects, as horizontal gene transfer and non-cognate binding are likely to occur under slow growth conditions, where already low levels of transcription factors might be costly (Fig. 2, S2). As the phage repressors we used originate from temperate phages, high levels of interference with host cell growth can limit their potential host range regarding successful establishment of lysogeny. Considering that phage repressor concentrations are kept low during lysogenic cycles, the selection pressure to reduce low-affinity binding might generally be weak, which would explain the diversity in growth effects we observed with related phage repressors. Phage TFs expressed during lysogeny have however been suggested to be able to change bacterial cell physiology (Derdouri et al., 2023). More generally, experimentally uncovering the fundamental biophysical constraints imposed by the non-cognate interactions

of TFs is difficult, as TFs with many cognate targets need to recognize a diversity of sequences and by default affect many cellular functions, while single target TFs are rare (Shimada et al., 2018). This is what ultimately motivated our choice of focusing our experiments on phage repressors and LacI.

We have shown experimentally that non-cognate recognition of biomolecules can constitute a limiting factor for cellular function and evolution due to the fundamental biophysical constraints on protein-DNA interactions. However, these costs could be counter-balanced by increased TF robustness to target site mutations or higher evolvability, precisely because interactions can be more easily formed at low affinity (Friedlander et al., 2017; Igler et al., 2018; Shepherd et al., 2023; Taylor et al., 2022). For example, a TF could co-opt regulation of a non-cognate gene – even if only to a small degree – that provides an advantage in a certain environment, which can subsequently be refined by evolution. This opens up wider questions about the interplay of costs and benefits of low-affinity molecular interactions, especially when these interactions also serve as drivers of evolution (Friedlander et al., 2017).

METHODS

Plasmids and strains

Wild-type repressors:

The phage repressors λ CI, P22 C2, 434 CI or the bacterial repressor LacI were cloned under the control of a $P_{LtetO-1}$ promoter in a low copy number kan^R plasmid (pZS) by using the restriction enzymes KpnI and XbaI (Lutz & Bujard, 1997). In control experiments the phage repressor on the plasmid was replaced by a fluorescence marker gene (*gfp*). The low copy number plasmids used (pZS21) are under stringent replication control, which is linked to chromosome replication (Kües & Stahl, 1989).

The plasmids were then transformed into either MG1655 derived *E. coli* cells (Frag1B) or into LT2 derived *S. enterica* cells. *E. coli* cells constitutively expressed tetR (P_{N25} promoter) and LacI (P_{laci}^q promoter) together with a spectinomycin resistance from a cassette that was inserted twice into the chromosome at the λ attachment site. The *S. enterica* cells do not carry any lac repressor and have a tetracycline resistance cassette (tetra) inserted near the P22 attachment site (this strain was a gift from the M. Erhardt lab).

Titration of λ CI was tested by transforming *E. coli* cells containing the pZS21- *λ cI* plasmid with a compatible, medium-copy number pZA plasmid (20-30 copies) (Lutz & Bujard, 1997), which carries the natural λ CI operators O_{L1} (x2), O_{L2}, O_{L3}, i.e. 80-120 operators per cell.

For the competition experiments we introduced a constitutive fluorescent marker *venus-yfp* (Nagai et al., 2002) into the low-copy pZS plasmid carrying *lacI* (for a detailed description see below).

To test for misfolding of repressor proteins, we used a high copy number plasmid (pCA24N) containing the chaperone gene *tig* (Nishihara et al., 2000)), which is native to *E. coli* and has >95% sequence identity with the *tig* gene in *S. enterica*, under the control of a P_{Lac} promoter from the ASKA(-) library (Kitagawa et al., 2005), which was induced with 1mM IPTG.

To monitor induction of the stress response, we used a strain with a fast-maturing yellow fluorescent protein (YFP, (Nagai et al., 2002)) fused to the promoter of *sulA* (*PsulA-yfp*), which was placed on the chromosome using lambda red recombineering (Costantino & Court, 2003). *SulA* is strongly upregulated as a part of the stress response (Friedberg et al., 2005). The *PsulA-yfp* strain was then transformed with the pZS21- *λ cI* plasmid. We checked induction of the reporter by exposing cells to UV light for 30 seconds using the UV transilluminator function of the BioRad GelDoc XR+ station. These cells were diluted 1:1000 from overnight cultures in M9 medium supplemented with 1mM thiamine, 0.4% glucose and 25µg/ml kanamycin, grown for 2h before UV exposure, and then grown and measured in the plate reader after exposure.

λ CI mutants:

Based on previous studies, we cloned three different λ repressor mutants into the same low copy number plasmid (pZS) under the control of a $P_{LtetO-1}$ promoter: (i) a repressor mutant that is defective in its ability to bind DNA (N52D) (Nelson & Sauer, 1986); (ii) a repressor mutant that cannot form dimers (S228N) (Whipple et al., 1994) and hence can no longer effectively bind DNA; and (iii) a mutant that can dimerize but not form higher-order oligomers, i.e. that cannot bind cooperatively (Y210N) (Whipple et al., 1994). Mutations were introduced using site-directed mutagenesis. The function and stability of the proteins has been shown previously (Nelson & Sauer, 1986; Whipple et al., 1994), as well as that their production levels are not different from wildtype repressor (Whipple et al., 1994).

λ CI chimeras:

Chimeric λ CI repressors were constructed based on literature describing a chimeric λ CI-434 repressor (Marchetti et al., 1995) and a chimeric 434-P22 repressor (Wharton & Ptashne, 1985) by changing the recognition (i.e. DNA binding) helix of λ CI to either the one of the 434 CI repressor or of the LacI repressor. This was done by introducing the following changes for the λ CI-434 chimera: G44T, S46Q, G49E, A50Q; and for the λ CI-LacI chimera: G44S, Q45Y, S46Q, G49S, A50R. (Note, that Q45 was not changed in the λ CI-434 chimera because both repressors contain the amino acid Q at this position.)

Growth measurements

All cultures were grown overnight at 37°C in M9 medium supplemented with 1mM thiamine, 0.4% glucose and 25µg/ml kanamycin (except specified differently). Cultures were used to dilute (1:1000) 5 replicates without inducer and 5 replicates with different levels of aTc in 96 well plates and were grown at 37°C under shaking at 220 rpm. Populations were measured (absorbance$_{600}$) every 15min using Biotek H1 plate reader for 10h. Population growth in the plate reader was compared to population growth in a culture tube shaken in an incubator at 37°C with sampling via plating for CFU counts after 0, 300, 360, 420 and 480 min (2 biological and 3-6 technical replicates per time point). The chaperone gene was induced using 1mM IPTG.

Growth rates are hard to define under conditions with strong growth reductions, as growth no longer shows an exponential increase and we had to exclude the growth of mutant cells. Mutant growth occurred frequently in the induced cultures, being visible as sudden onset of fast exponential growth after an initial lag of several hours. These cells were no longer inducible when regrown in fresh media.

Accordingly, growth windows for growth rate calculation were chosen in the following way: Within an absorbance$_{600}$ range of about 5 doublings between 0.002 (as there was too much noise below that value) and 0.06 (which is about 3 doublings away from stationary phase), we used a sliding window approach to find an appropriate time frame for the growth rate calculation. We used a sliding window of 10 time points (with time points 15min apart) and calculated the Pearson coefficient on log(absorbance$_{600}$) data, which gave a better indication of the exponential part of growth than linear regression models. We used a cut-off of R^2>0.97 to define the exponential part of growth, and if this only occurred after more than 8h, we assume that this is due to mutant growth.

Using the derivative of the slope at each time point, we determined the approximate beginning of mutant growth and used the highest Pearson coefficient region before mutant growth as growth window. For experiments without mutant growth (Pearson coefficient <<1) and final absorbance$_{600}$ < 0.06, we calculated growth between absorbance$_{600}$ 0.002 and being one doubling away from final densities.

Growth rates were then calculated within the determined growth windows for each replicate individually using linear regression.

Competition assays

In order to distinguish between strains carrying a phage repressor (showing growth reduction) or the bacterial repressor LacI (not showing growth reduction), we used the pZS plasmid carrying *lacI* (under the control of $P_{LtetO-1}$), together with a constitutively expressed fluorescence marker (*venus*) cloned in the opposite direction upstream of $P_{LtetO-1}$. Venus was expressed from a mutated version of P_R, which abolishes *λ* CI binding affinity in the O_{R1} operator (Flashman, 1978; Sarai & Takeda, 1989) (the O_{R1} sequence used is:

TGCCTTAATACTGGATA) – and does not contain O_{R2} or O_{R3} - and is therefore constitutive ('$P_{constitutive}$'). The fluorescence marker was placed on the LacI-carrying plasmid because LacI showed no growth effect in *E. coli* or *S. enterica* (Fig. 1B, Table S1). For E. coli that means that there was no additional effect of expressing more LacI than from the chromosome. As chromosome expression of LacI is constitutive, induction of TF expression from plasmids still occurs under the same genomic background conditions and can be compared in direct competition. As there was no growth effect of LacI-carrying plasmids, bacterial cells did not show filamentation and hence there was no effect of cell morphology changes on fluorescence. There was a slight fitness cost due to the presence of the fluorescence marker on the plasmid in the presence of 0.8ng/ml aTc induction (Fig. S2), which however only strengthens our findings that LacI-expressing cells increase in competition with phage repressor-expressing cells despite this cost.

Overnight cultures of clonal *E. coli* cells carrying pZS21-*λ cI* or carrying pZS21-*lacI* were adjusted to ~10^8 cells and mixed 1:1 with a culture of the same cell density carrying pZS21-*lacI*-$P_{constitutive}$-*venus.* Cultures were grown in M9 minimal media supplied with 0.4% glucose, 1mM thiamine, 25 µg/ml kanamycin and 0.8ng/ml aTc at 37°C in a shaking incubator and sampled via serial dilution plating after 0, 300, 360, 420 and 480 minutes. Cells were distinguished based on the fluorescence marker. For the *λ* CI versus LacI+Venus competition, we used 3 biological replicates and 3-6 technical replicates each. For the control LacI versus LacI+Venus competition, we used 4 technical replicates each for one induced and one uninduced experiment. Selection coefficients w were calculated using $w=\ln[(B_t/B_0)/(V_t/V_0)]$, where B_t represents CFU counts of cells with *λ* CI or LacI and V_t represents cells carrying LacI+Venus at t=480min.

We further used fluorescence measurements as a proxy for abundance to compete a larger number of host strain (*E. coli* or *S. enterica*) – plasmid (pZS21-*lacI,* pZS21- *λ cI*, pZS21-*P22 c2*, pZS21-*lacI*-$P_{constitutive}$-*venus*) combinations. Strains containing an unmarked repressor plasmid were mixed 1:1 with a strain carrying pZS21-*lacI*-$P_{constitutive}$-*venus*, diluted 1:100 into fresh medium and grown in 96-wellplates for 10h. Fluorescence was measured every 30min and compared between cultures that were induced with 25ng/ml aTc and cultures that were not induced.

Immunoblotting

Overnight cultures of Frag1B::pZS21-*λ cI-HA* grown at 37°C under shaking at 220 rpm in M9 supplemented with 0.4% glucose, 1mM thiamine and 25 µg/ml kanamycin or LB supplemented with 25 µg/ml kanamycin were diluted in 50ml of fresh media to an OD of 0.005. Samples were induced with 0, 1, 3, or 25 ng/ml aTc and grown at 37°C under shaking at 220 rpm to an OD of 0.02-0.03. 40ml of culture was pelleted at 4°C at 8000g for 10 minutes, washed in 40 ml of PBS, repelleted under the same conditions, and resuspended in 1ml PBS. Samples were then pelleted at 4°C at 21130g for 10 minutes, resuspended in 200µL PBS and stored at -70°C. OD measurements were done using a Hitachi U-5100 Spectrophotometer. Samples were normalized to an original OD of 0.02 with PBS, and mixed in a 3:1 ratio with 4x Laemmli sample buffer and incubated at 95°C for 20 minutes. 5µL of each sample was separated by electrophoresis on a 12% Mini-PROTEAN® TGX™ Precast Protein Gel (Bio-Rad, 4561046), and transferred to a polyvinylidene fluoride (PVDF) membrane using the Trans-Blot Turbo Transfer System (Biorad). The transferred PVDF membrane was blocked at room temperature for one hour with gentle agitation in Tris-buffered saline containing 0.1% Tween 20 (TBST) and 1.5% skim milk, then stained with HA Tag Monoclonal Antibody (2-2.2.14)

(Invitrogen, 26183-HRP) at 4°C with gentle agitation overnight. The membrane was washed thrice for 10 minutes in TBST, treated with Supersignal™ West Fempto Maximum Sensitivity Substrate (Thermo Scientific, 34096), and visualized with an Amersham™ Imager 600 RGB (GE) with an exposure time of 30 sec. To visualize the loading controls, the membrane was stripped by incubating with Abcam mild stripping buffer (15 g/l glycine, 1 g/l SDS, 1% Tween 20, pH 2.2) twice for ten minutes at room temperature, washed with PBS twice for 10 minutes followed by washing with TBS twice for 10 minutes, reblocked at room temperature with gentle agitation in TBST + 1.5% skim milk, and restained with GAPDH Loading Control Monoclonal Antibody (GA1R) (Invitrogen, MA5-15738-HRP) for 2 hours at room temperature with shaking. The membrane was treated with Supersignal™ West Fempto Maximum Sensitivity Substrate (Thermo Scientific, 34096) and visualized with an Amersham™ Imager 600 RGB (GE) with 6 sec exposure.

ChIP-sequencing analysis

To perform ChIP-sequencing experiments, λ CI was cloned with an HA-Tag at the carboxy-terminal end and transformed into both host strains. HA-tagged λ CI showed the same growth phenotype as wild-type in minimal and rich media (Fig. S10, Table S7). Samples from strains grown in the presence or absence of λ CI were prepared according to (Waldminghaus & Skarstad, 2010):

For ChIP-seq, cells were grown at 37°C to an OD600 of about 0.1-0.2 in 50 ml LB before 27 µl of formaldehyde (37%) per ml medium were added (final concentration 1%). Crosslinking was performed at slow shaking (100 rpm) at room temperature for 20 min followed by quenching with 10 ml of 2.5 M glycine (final concentration 0.5 M). Cells were collected by centrifugation and washed twice with cold TBS (pH7.5). After resuspension in 1 ml lysis buffer (10 mM Tris (pH 8.0), 20% sucrose, 50 mM NaCl, 10 mM EDTA, 10 mg/ml lysozyme) and incubation at 37°C for 30 min followed by addition of 4 ml IP buffer, cells were sonicated at 4°C in 1.5 ml tubes with 48 cycles (30 sec. sonication, 30 sec. cooling) with Bioruptor Plus from Diagenode. After centrifugation for 10 min at 9000 g, 800 µl aliquots of the supernatant were stored at -20°C or processed directly. 800 µl of sonicated cell extract was incubated with 20 µl protein A/G agarose beads (Ultralink) and the different antibodies at 4°C overnight with slow mixing. Agarose beads were collected at the bottom of a 1.5 ml Eppendorf tube by centrifugation. The supernatant was then removed by pipetting. Washing was done by adding 700 µl buffer to the beads and rotation at room temperature for three minutes with subsequent collection of the beads by centrifugation as above. Washing was performed with the following buffers (IP buffer two times all others one time): IP buffer (50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton × 100, 0.1% Sodium deoxycholate, 0.1% SDS), IP buffer with 500 mM NaCl, wash buffer (10 mM Tris pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% Nonidet-P40, 0.5% Sodium deoxycholate) and TE. For elution, 100 µl elution buffer (50 mM Tris (pH 7.5), 10 mM EDTA, 1% SDS) was added to the column with the beads, incubated in a 65°C water bath for 10 min and centrifuged as above. DNAs were incubated with RNase A (50 µg/ml) for at least 90 min at 42°C. To reverse the cross link 80 µl TE and 20 µl proteinase K (20 mg/ml) were added and samples incubated for 2 h at 42 and 6 h at 65°C. DNA was purified with phenol/chloroform. The control DNA was taken from the supernatant resulting from centrifugation of the precipitated chromatin beads and processed further as the immunoprecipitated DNA after elution.

Library preparation and Illumina Sequencing was performed at the VBCF NGS Unit (www.vbcf.ac.at). The obtained sequencing data was analyzed using Galaxy and RStudio. Reads were aligned against a reference genome (E. Coli: NC_000913.2, Salmonella: AE006468.2) using bowtie (v1.1.2) and --wrapper basic-0 -q -p 6 -S -n 2 -e 70 -l 28 --maxbts 125 -k 1 -m 1 parameters. Alignment statistics are given in Table S6.

We performed peak calling by examining partially overlapping (50% overlap) sliding windows of 100bp each. Within each 100bp window, we calculated the log ratio of the scaled coverage (scaled by number of mapped reads) between sample and control ChIP-seq experiments while permitting a maximum of five mapped reads to share the same genomic coordinates. Enrichment, i.e. positive log ratios, are plotted in Fig. 4 (A,B). We calculated strand-wise p-values for enrichment by first resampling scaled read coverage within each fragment and then randomly partitioning them to calculate enrichments. Finally, we identified bound regions to be those with positive enrichment scores on both strands with a Benjamini-Hochberg false-discovery rate of less than 30% as we were looking for binding with low affinity and ChIP-binding data was previously found to be highly informative for a wide range of affinity profiles (Tanay, 2006). We estimated the false-discovery rates to assess significance of every peak. To account for local biases within the genome (Zhang et al., 2008), we utilized non-overlapping 2kb genomic fragments to estimate FDRs. Different choices of sliding windows (25bp, 50bp, 100bp), maximum number of reads at a specific genomic coordinate (3, 5, 10) as well as genomic fragment size (1000bp, 2000bp, 5000bp) identified similar sets of peaks. In Fig. S3, we plotted the read-depth across the genome for the regions that showed significant enrichment in this analysis and for comparison we plotted the read-depth for not significantly enriched regions. As control for our ChIP-sequencing procedure and analysis we used antibodies against SeqA, which gave the expected peaks as published previously (Waldminghaus & Skarstad, 2010).

We calculated the nucleotide composition of the sequences underlying enriched regions in ChIP-seq data for both bacterial species (Fig. S6). To test for sequence composition bias in these enriched regions, we sought to test if the sequence compositions of the enriched regions were significantly different compared to the rest of the genome. To this end, we randomly selected 50 genomic regions with at least 5kbp distance between them. We then calculated the nucleotide composition of these randomly selected regions and by repeating this procedure 1000 times, generated a null distribution for sequence composition of randomly selected genomic regions. Similarly, we calculated the di-nucleotide composition (with 1 bp overlap) of the same randomly selected genomic regions and compared it to that of the enriched regions.

Thermodynamic model

The unfiltered number of reads within 1000bp windows was compared with the predicted binding by calculating binding energy at each genome position (using a sliding window approach) from the λ CI offset (i.e. the energy difference between the repressor being bound with high affinity to a cognate operator and being free in solution (Koblan & Ackers, 1992)) and the energy penalty as given by the λ CI energy matrix (Sarai & Takeda, 1989). Lower energies result in stronger binding, meaning positive energy penalties decrease binding affinity (note that negative penalties could increase binding over the one seen with λ CI wildtype operator sites). Binding strength was calculated using $1/(1+\exp(E-\mu))$, with E being

the calculated binding energy, as described above and used in (Igler et al., 2018), and μ being the chemical potential, which we optimized to give the highest Spearman correlation fit (2.6 in *E. coli* and 2 in *S. enterica*). For comparison with the number of ChIP-sequencing reads, calculated binding strength was summed over the same genomic 1000bp regions (results shown are considering binding only to the forward strand, for ChIP reads and predictions, but the results remain the same when using both strands). In Fig. 4 (C,D) we plot a non-parametric, non-linear relationship estimate between the predicted binding energy and the ChIP-sequencing reads obtained from a series of conditional medians. To investigate the dependence of the correlation between the affinity predictions and the ChIP-sequencing reads on the structural versus the sequence information contained in the energy matrix, we repeated the analysis with i) a matrix of the same size that conserves only the ACGT bias of the λ CI energy matrix (each row contains the average value of that row) or ii) matrices that had completely reshuffled entries. For the latter the average correlation was taken over 100 permutations.

To assess the importance of sequence motif information versus nucleotide (GC) bias, we used a Monte-Carlo permutation test: We calculated the difference between Spearman correlations of ChIP reads with binding prediction using the wildtype energy matrix vs binding prediction using the energy matrix that only conserves λ CI basepair bias, for the true ChIP read assignment, and 10^4 random read assignments (null distribution). We found an overall strongly significant difference in *E. coli* and lower significance in *S. enterica* (Fig. 4C,D), even though the effect size was small. This means that while most of the measured ChIP signal can be accounted for by a TF model that predicts binding based on the nucleotide content of genomic fragments alone, there is a small but highly significant residual ChIP binding signal that requires the full binding site preference (energy matrix), not just single nucleotide bias, to be explained. Furthermore, we examined the influence of GC content by repeating the Monte-Carlo permutation test for genomic sequences of a specific GC %. Here, we found only a significant motif contribution for the 49% bin in *E. coli* (Fig. S7).

Additionally we used the offset and energy matrix for LacI (Barnes et al., 2019; Vilar & Saiz, 2005) and P22 C2 (Hilchey et al., 1997) to predict binding and calculate the Spearman correlation with the λ CI ChIP-sequencing reads (Fig. S5). The magnitudes of the Spearman correlations for LacI and P22 C2 predictions are similar to the ones with the λ CI basepair bias, but negative (inverse) for P22 C2. Basepair bias of the energy matrices was calculated as the sum of the average A and T preference minus the sum of the average G and C preference.

Microscope fluorescence measurements

Imaging of cell membranes and DNA positioning was done using a TIRF Leica DMI6000B (inverted) microscope with 1) an Andor iXon EM CCD camera (front illuminated, 8x8 square micron pixel size) and a 100x 1,47Na Oil HCX Plan Apo objective, giving an effective pixel size of 64nm/pixel; and 2) a Nikon Ti-2 microscope with a 100× oil immersion objective (Plan Apo λ, N.A. 1.45, Nikon) and a Nikon DS-Qi2 camera. Images were acquired using 405(20)nm and 561(10)nm laser excitation for blue (Hoechst) and red (NileRed) dyes respectively. The cells were grown overnight in M9 medium supplemented with 1mM thiamine, 0.4% glucose and 25μg/ml kanamycin, diluted 1:1000 in fresh media and grown to early exponential phase in the absence or presence of the inducer aTc. After addition of both dyes (Hoechst at 10ug/mL and NileRed at 1ug/mL), cells were shaken at room temperature for one hour and imaged in drops of the respective growth media. Images were deconvolved using Huygens Professional (version 4.5) and further analyzed using ImageJ.

Statistical analysis

We analyzed differences in growth rates between cultures using one-way ANOVA followed by pairwise comparisons using Tukey's HSD to account for multiple testing (RStudio, functions aov and TukeyHSD).

Spearman correlation was calculated for the fit between model predictions of binding strength and the number of obtained ChIP-sequencing reads per 1000bp window using Matlab R2019b. P-values for WT and basepair bias energy matrix predictions were P<.001 and for the average over random matrices P<.01. Peak calling on ChIP-sequencing data is described under 'ChIP-sequencing analysis'.

DATA AVAILABILITY

The datasets produced in this study are available in the Gene Expression Omnibus database and the computer code on GitHub (https://github.com/balaji-srinivasan-santhanam/peaks).

ACKNOWLEDGEMENT

We thank M. Erhardt for the donation of strains, T. Bergmiller, R. Chait, K. Jain, M. La Fortezza and the ETH Zurich ScopeM facility for their support with fluorescence microscopy, N. Gnyliukh for help with Western blot experiments, and T. Friedlander and J. Crocker for useful discussions. CI was supported by a DOC Fellowship of the Austrian Academy of Sciences, the Wellcome Trust [225565/Z/22/Z] and an ETH Zurich Postdoctoral Fellowship (19-2 FEL-74). Library preparation and sequencing of ChIP seq samples was performed by the Next Generation Sequencing Facility at Vienna BioCenter Core Facilities (VBCF), member of the Vienna BioCenter (VBC), Austria.

AUTHOR CONTRIBUTIONS

CI, CF, GT and CCG conceived the study together. CI designed the experiments and carried them out together with CF and BW (growth experiments) and TW (ChIP-sequencing). CI, FMP and BS analyzed the ChIP sequencing data. BW and BK performed and analyzed protein level experiments. CI wrote the initial draft of the manuscript and revised it with CCG, GT and BW.

CONFLICT OF INTEREST

Authors declare no competing financial interests.

REFERENCES

Amir, A., Männik, J., Woldringh, C. L., & Zaritsky, A. (2019). Editorial: The Bacterial Cell: Coupling between Growth, Nucleoid Replication, Cell Division, and Shape Volume 2. *Frontiers in Microbiology*, *10*(September), 1–3. https://doi.org/10.3389/fmicb.2019.02056

Arends, S. J. R., & Weiss, D. S. (2004). Inhibiting Cell Division in Escherichia coli Has Little If Any Effect on Gene Expression. *Journal of Bacteriology*, *186*(3), 880–884. https://doi.org/10.1128/JB.186.3.880-884.2004

Bakk, A., & Metzler, R. (2004a). In vivo non-specific binding of λ CI and Cro repressors is significant. *FEBS Letters*, *563*(1–3), 66–68. https://doi.org/10.1016/S0014-5793(04)00249-2

Bakk, A., & Metzler, R. (2004b). Nonspecific binding of the OR repressors CI and Cro of

bacteriophage lambda. *Journal of Theoretical Biology*, *231*(4), 525–533. https://doi.org/10.1016/j.jtbi.2004.07.007

Baltrus, D. A. (2013). Exploring the costs of horizontal gene transfer. *Trends in Ecology & Evolution*, *28*(8), 489–495. https://doi.org/10.1016/j.tree.2013.04.002

Barnes, S. L., Belliveau, N. M., Ireland, W. T., Kinney, J. B., & Phillips, R. (2019). Mapping DNA sequence to transcription factor binding energy in vivo. *PLOS Computational Biology*, *15*(2), e1006226. https://doi.org/10.1371/journal.pcbi.1006226

Bollenbach, T., Quan, S., Chait, R., & Kishony, R. (2009). Nonoptimal Microbial Response to Antibiotics Underlies Suppressive Drug Interactions. *Cell*, *139*(4), 707–718. https://doi.org/10.1016/j.cell.2009.10.025

Brewster, R. C., Weinert, F. M., Garcia, H. G., Song, D., Rydenfelt, M., & Phillips, R. (2014). The transcription factor titration effect dictates level of gene expression. *Cell*, *156*(6), 1312–1323. https://doi.org/10.1016/j.cell.2014.02.022

Burger, A., Walczak, A. M., & Wolynes, P. G. (2010). Abduction and asylum in the lives of transcription factors. *Proceedings of the National Academy of Sciences*, *107*(9), 4016–4021. https://doi.org/10.1073/pnas.0915138107

Cambridge, J., Blinkova, A., Magnan, D., Bates, D., & Walker, J. R. (2014). A Replication-inhibited unsegregated nucleoid at mid-cell blocks Z-ring formation and cell division independently of SOS and the SlmA nucleoid occlusion protein in Escherichia coli. *Journal of Bacteriology*, *196*(1), 36–49. https://doi.org/10.1128/JB.01230-12

Cepeda-Humerez, S. A., Rieckh, G., & Tkačik, G. (2015). Stochastic Proofreading Mechanism Alleviates Crosstalk in Transcriptional Regulation. *Physical Review Letters*, *115*(24), 1–5. https://doi.org/10.1103/PhysRevLett.115.248101

Chakrabarti, J., Chandra, N., Raha, P., & Roy, S. (2011). High-affinity quasi-specific sites in the genome: How the DNA-binding proteins cope with them. *Biophysical Journal*, *101*(5), 1123–1129. https://doi.org/10.1016/j.bpj.2011.07.041

Chen, Y., Golding, I., Sawai, S., Guo, L., & Cox, E. C. (2005). Population Fitness and the Regulation of Escherichia coli Genes by Bacterial Viruses. *PLoS Biology*, *3*(7), e229. https://doi.org/10.1371/journal.pbio.0030229

Cooper, S., & Ruettinger, T. (1973). Replication of deoxyribonucleic acid during the division cycle of Salmonella typhimurium. *Journal of Bacteriology*, *114*(3), 966–973.

Costantino, N., & Court, D. L. (2003). Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(26), 15748–15753. https://doi.org/10.1073/pnas.2434959100

Crocker, J., Preger-Ben Noon, E., & Stern, D. L. (2016). The Soft Touch. In *Current Topics in Developmental Biology* (1st ed., Vol. 117, pp. 455–469). Elsevier Inc. https://doi.org/10.1016/bs.ctdb.2015.11.018

de Boer, P. A. J. (2010). Advances in understanding E. coli cell fission. *Current Opinion in Microbiology*, *13*(6), 730–737. https://doi.org/10.1016/j.mib.2010.09.015

Derdouri, N., Ginet, N., Denis, Y., Ansaldi, M., & Battesti, A. (2023). The prophage-encoded transcriptional regulator AppY has pleiotropic effects on E. coli physiology. *PLOS Genetics*, *19*(3), e1010672. https://doi.org/10.1371/journal.pgen.1010672

Donner, A. L., Carlson, P. A., & Koudelka, G. B. (1997). Dimerization specificity of P22 and 434 repressors is determined by multiple polypeptide segments. *Journal of Bacteriology*, *179*(4), 1253–1261.

Elf, J., Li, G.-W., & Xie, X. S. (2007). Probing transcription factor dynamics at the single-

molecule level in a living cell. *Science (New York, N.Y.)*, *316*(5828), 1191–1194. https://doi.org/10.1126/science.1141967

Fattah, K. R., Mizutani, S., Fattah, F. J., Matsushiro, A., & Sugino, Y. (2000). *A comparative study of the immunity region of lambdoid phages including Shiga-toxin-converting phages : molecular basis for cross immunity* . 223–232.

Flashman, S. M. (1978). *NMutational Analysis of the Operators of Bacteriophage Lambda*. *73*, 61–73.

Flyvbjerg, H., Keatch, S. A., & Dryden, D. T. F. (2006). Strong physical constraints on sequence-specific target location by proteins on DNA molecules. *Nucleic Acids Research*, *34*(9), 2550–2557. https://doi.org/10.1093/nar/gkl271

Friedberg, E. C., Walker, G. C., Siede, W., & Wood, R. D. (2005). *DNA repair and mutagenesis*. American Society for Microbiology Press.

Friedlander, T., Prizak, R., Barton, N. H., & Tkačik, G. (2017). Evolution of new regulatory functions on biophysically realistic fitness landscapes. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00238-8

Friedlander, T., Prizak, R., Guet, C., & Barton, N. H. (2015). *Intrinsic limits to gene regulation by global crosstalk*.

Gerland, U., Moroz, J. D., & Hwa, T. (2002). Physical constraints and functional characteristics of transcription factor-DNA interaction. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(19), 12015–12020. https://doi.org/10.1073/pnas.192693599

Grah, R., Zoller, B., & Tkačik, G. (2020). Nonequilibrium models of optimal enhancer function. *Proceedings of the National Academy of Sciences*, *117*(50), 31614–31622. https://doi.org/10.1073/pnas.2006731117

Greulich, P., Scott, M., Evans, M. R., & Allen, R. J. (2015). Growth-dependent bacterial susceptibility to ribosome-targeting antibiotics. *Molecular Systems Biology*, *11*(3), 796. https://doi.org/10.15252/msb.20145949

Guarnaccia, C., Raman, B., Zahariev, S., Simoncsits, A., & Pongor, S. (2004). DNA-mediated assembly of weakly interacting DNA-binding protein subunits: In vitro recruitment of phage 434 repressor and yeast GCN4 DNA-binding domains. *Nucleic Acids Research*, *32*(17), 4992–5002. https://doi.org/10.1093/nar/gkh827

Hilchey, S. P., Wu, L., & Koudelka, G. B. (1997). *Recognition of Nonconserved Bases in the P22 Operator by P22 Repressor Requires Specific Interactions between Repressor and Conserved Bases *. 272*(32), 19898–19905.

Hopfield, J. J. (1974). Biosynthetic Processes Requiring High Specificity. *Pnas*, *71*(10), 4135–4139.

Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., & Sullivan, M. B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *The ISME Journal*, *11*(7), 1511–1520. https://doi.org/10.1038/ismej.2017.16

Igler, C., Lagator, M., Tkačik, G., Bollback, J. P., & Guet, C. C. (2018). Evolutionary potential of transcription factors for gene regulatory rewiring. *Nature Ecology & Evolution*, *2*(10), 1633–1643. https://doi.org/10.1038/s41559-018-0651-y

Jonge, W. J. De, Patel, H. P., Meeussen, J. V. W., & Lenstra, T. L. (2022). Biophysical Perspective Following the tracks : How transcription factor binding dynamics control transcription. *Biophysj*, *121*(9), 1583–1592. https://doi.org/10.1016/j.bpj.2022.03.026

Kao-Huang, Y., Revzin, A., Butler, A. P., O'Conner, P., Noble, D. W., & Von Hippel, P. H. (1977). Nonspecific DNA binding of genome-regulating proteins as a biological control

mechanism: measurement of DNA-bound Escherichia coli lac repressor in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(10), 4228–4232. https://doi.org/10.1073/pnas.74.10.4228

Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., & Mori, H. (2005). Complete set of ORF clones of Escherichia coli ASKA library (A complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Research*, *12*(5), 291–299. https://doi.org/10.1093/dnares/dsi012

Klumpp, S., Zhang, Z., & Hwa, T. (2009). Growth-rate dependent global effects on gene expression in bacteria. *Cell*, *139*(7), 1366–1375. https://doi.org/10.1016/j.cell.2009.12.001.Growth-rate

Koblan, K. S., & Ackers, G. K. (1992). Site-Specific Enthalpic Regulation of DNA Transcription at Bacteriophage λ OR. *Biochemistry*, *31*(1), 57–65. https://doi.org/10.1021/bi00116a010

Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J., & Mann, R. S. (2019). Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes. *Annual Review of Cell and Developmental Biology*, *35*(1), 357–379. https://doi.org/10.1146/annurev-cellbio-100617-062719

Kües, U., & Stahl, U. (1989). Replication of plasmids in gram-negative bacteria. *Microbiological Reviews*, *53*(4), 491–516. http://www.ncbi.nlm.nih.gov/pubmed/2687680

Lagator, M., Paixão, T., Barton, N. H., Bollback, J. P., & Guet, C. C. (2017). On the mechanistic nature of epistasis in a canonical cis-regulatory element. *ELife*, *6*, 1–16. https://doi.org/10.7554/eLife.25192

Lutz, R., & Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research*, *25*(6), 1203–1210. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=146584&tool=pmcentrez&rendertype=abstract

Maerkl, S. J., & Quake, S. R. (2007). A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. *Science*, *315*(5809), 233–237. https://doi.org/10.1126/science.1131007

Manzo, C., Zurla, C., Dunlap, D. D., & Finzi, L. (2012). The effect of nonspecific binding of lambda repressor on DNA looping dynamics. *Biophysical Journal*, *103*(8), 1753–1761. https://doi.org/10.1016/j.bpj.2012.09.006

Marchetti, A., Abril-Marti, M., Illi, B., Cesareni, G., & Nasi, S. (1995). Analysis of the Myc and Max interaction specificity with λ repressor-HLH domain fusions. *Journal of Molecular Biology*, *248*(3), 541–550. https://doi.org/10.1006/jmbi.1995.0241

Mazumder, A., Batabyal, S., Mondal, M., Mondol, T., Choudhury, S., Ghosh, R., Chatterjee, T., Bhattacharyya, D., Pal, S. K., & Roy, S. (2017). Specific DNA sequences allosterically enhance protein-protein interaction in a transcription factor through modulation of protein dynamics: Implications for specificity of gene regulation. *Physical Chemistry Chemical Physics*, *19*(22), 14781–14792. https://doi.org/10.1039/c7cp01193h

Mirny, L., Slutsky, M., Wunderlich, Z., Tafvizi, A., Leith, J., & Kosmrlj, A. (2009). How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *Journal of Physics A: Mathematical and Theoretical*, *42*(43), 434013. https://doi.org/10.1088/1751-8113/42/43/434013

Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K., & Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological

applications. *Nature Biotechnology*, *20*(1), 87–90. https://doi.org/10.1038/nbt0102-87

Nelson, H. C. M., & Sauer, R. T. (1986). Interaction of mutant λ repressors with operator and non-operator DNA. *Journal of Molecular Biology*, *192*(1), 27–38. https://doi.org/10.1016/0022-2836(86)90461-4

Nishihara, K., Kanemori, M., Yanagi, H., & Yura, T. (2000). Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. *Applied and Environmental Microbiology*, *66*(3), 884–889. https://doi.org/10.1128/AEM.66.3.884-889.2000.Updated

Nordström, K., & Dasgupta, S. (2006). Copy-number control of the Escherichia coli chromosome: A plasmidologist's view. *EMBO Reports*, *7*(5), 484–489. https://doi.org/10.1038/sj.embor.7400681

Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L., & Adhya, S. (2005). Switches in bacteriophage lambda development. *Annual Review of Genetics*, *39*, 409–429. https://doi.org/10.1146/annurev.genet.39.073003.113656

Perkins, M. L., Crocker, J., & Tkačik, G. (2024). *Chromatin enables precise and scalable gene regulation with factors of limited specificity*. *bioRxiv*.

Phizicky, E. M., & Roberts, J. W. (1980). Kinetics of recA protein-directed inactivation of repressors of phage λ and phage P22. *Journal of Molecular Biology*, *139*(3), 319–328. https://doi.org/10.1016/0022-2836(80)90133-3

Pleška, M., Qian, L., Okura, R., Bergmiller, T., Wakamoto, Y., Kussell, E., & Guet, C. C. (2016). Bacterial Autoimmunity Due to a Restriction-Modification System. *Current Biology : CB*, *26*(3), 404–409. https://doi.org/10.1016/j.cub.2015.12.041

Pray, T. R., Burz, D. S., & Ackers, G. K. (1998). Cooperative non-specific DNA binding by octamerizing λcI repressors: a site-specific thermodynamic analysis. *Journal of Molecular Biology*, *282*(5), 947–958. https://doi.org/10.1006/jmbi.1998.2056

Priest, D. G., Cui, L., Kumar, S., Dunlap, D. D., Dodd, I. B., & Shearwin, K. E. (2014). Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and repressors. *Proceedings of the National Academy of Sciences*, *111*(1), 349–354. https://doi.org/10.1073/pnas.1317817111

Ptashne, M. (2004). *A Genetic Switch: Phage Lambda Revisited* (Third Edit). Cold Spring Harbor Laboratory Press.

Saio, T., Guan, X., Rossi, P., Economou, A., & Kalodimos, C. G. (2014). Structural Basis for Protein Antiaggregation Activity of the Trigger Factor Chaperone. *Science*, *344*(6184), 21–23. https://doi.org/10.1126/science.1250494

Sarai, A., & Takeda, Y. (1989). Lambda repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(17), 6513–6517. https://doi.org/10.1073/pnas.86.17.6513

Sarkar-Banerjee, S., Goyal, S., Gao, N., Mack, J., Thompson, B., Dunlap, D., Chattopadhyay, K., & Finzi, L. (2018). Specifically bound lambda repressor dimers promote adjacent non-specific binding. *PLOS ONE*, *13*(4), e0194930. https://doi.org/10.1371/journal.pone.0194930

Sasson, V., Shachrai, I., Bren, A., Dekel, E., & Alon, U. (2012). Mode of Regulation and the Insulation of Bacterial Gene Expression. *Molecular Cell*, *46*(4), 399–407. https://doi.org/10.1016/j.molcel.2012.04.032

Sauer, R. T., Ross, M. J., & Ptashne, M. (1982). Cleavage of the lambda and P22 repressors by recA protein. *The Journal of Biological Chemistry*, *257*(8), 4458–4462.

Shahein, A., López-Malo, M., Istomin, I., Olson, E. J., Cheng, S., & Maerkl, S. J. (2022). Systematic analysis of low-affinity transcription factor binding site clusters in vitro and in vivo establishes their functional relevance. *Nature Communications*, *13*(1), 1–17. https://doi.org/10.1038/s41467-022-32971-0

Shepherd, M. J., Pierce, A. P., & Taylor, T. B. (2023). Evolutionary innovation through transcription factor rewiring in microbes is shaped by levels of transcription factor activity, expression, and existing connectivit. *PLoS Biology*, *21*(10 October), 1–26. https://doi.org/10.1371/journal.pbio.3002348

Shimada, T., Ogasawara, H., & Ishihama, A. (2018). Single-target regulators form a minor group of transcription factors in Escherichia coli K-12. *Nucleic Acids Research*, *46*(8), 3921–3936. https://doi.org/10.1093/nar/gky138

Spurio, R., Dürrenberger, M., Falconi, M., La Teana, A., Pon, C. L., & Gualerzi, C. O. (1992). Lethal overproduction of the Escherichia coli nucleoid protein H-NS: ultramicroscopic and molecular autopsy. *MGG Molecular & General Genetics*, *231*(2), 201–211. https://doi.org/10.1007/BF00279792

Spurio, R., Falconi, M., Brandi, A., Pon, C. L., & Gualerzi, C. O. (1997). The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending. *EMBO Journal*, *16*(7), 1795–1805. https://doi.org/10.1093/emboj/16.7.1795

Stracy, M., Schweizer, J., Sherratt, D. J., Kapanidis, A. N., Uphoff, S., & Lesterlin, C. (2021). Transient non-specific DNA binding dominates the target search of bacterial DNA-binding proteins. *Molecular Cell*, *81*(7), 1499-1514.e6. https://doi.org/10.1016/j.molcel.2021.01.039

Sun, Q., & Margolin, W. (2004). Effects of perturbing nucleoid structure on nucleoid occlusion-mediated toporegulation of FtsZ ring assembly. *Journal of Bacteriology*, *186*(12), 3951–3959. https://doi.org/10.1128/JB.186.12.3951-3959.2004

Tanay, A. (2006). Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Research*, *16*(8), 962–972. https://doi.org/10.1101/gr.5113606

Taylor, T. B., Shepherd, M. J., Jackson, R. W., & Silby, M. W. (2022). Natural selection on crosstalk between gene regulatory networks facilitates bacterial adaptation to novel environments. *Current Opinion in Microbiology*, *67*, 102140. https://doi.org/10.1016/j.mib.2022.02.002

Verma, S. C., Qian, Z., & Adhya, S. L. (2019). Architecture of the Escherichia coli nucleoid. In *PLoS Genetics* (Vol. 15, Issue 12). https://doi.org/10.1371/journal.pgen.1008456

Vilar, J. M., & Saiz, L. (2005). DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Current Opinion in Genetics & Development*, *15*(2), 136–144. https://doi.org/10.1016/j.gde.2005.02.005

von Hippel, P. H., Revzin, A., Gross, C. A., & Wang, A. C. (1974). Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(12), 4808–4812. https://doi.org/10.1073/pnas.71.12.4808

Waldminghaus, T., & Skarstad, K. (2010). ChIP on Chip: surprising results are often artifacts. *BMC Genomics*, *11*(1), 414. https://doi.org/10.1186/1471-2164-11-414

Weiss, M. A., Pabo, C. O., Karplus, M., & Sauer, R. T. (1987). Dimerization of the operator binding domain of phage lambda repressor. *Biochemistry*, *26*(3), 897–904. https://doi.org/10.1021/bi00377a034

Wharton, R. P., & Ptashne, M. (1985). Changing the binding specificity of a represser by

redesigning an α-helix. *Nature*, *316*(6029), 601–605. https://doi.org/10.1038/316601a0

Whipple, F. W., Kuldell, N. H., Cheatham, L. a, & Hochschild, a. (1994). Specificity determinants for the interaction of lambda repressor and P22 repressor dimers. *Genes & Development*, *8*(10), 1212–1223. https://doi.org/10.1101/gad.8.10.1212

Wunderlich, Z., & Mirny, L. (2008). Fundamentally different strategies for transcriptional regulation are revealed by analysis of binding motifs. *Nature Precedings*, *2*(i). https://doi.org/10.1038/npre.2008.2688.2

Yamada, H., Yoshida, T., Tanaka, K., Sasakawa, C., & Mizuno, T. (1991). Molecular analysis of the Escherichia coli has gene encoding a DNA-binding protein, which preferentially recognizes curved DNA sequences. *Molecular and General Genetics MGG*, *230*(1–2), 332–336. https://doi.org/10.1007/BF00290685

Zabet, N. R., & Adryan, B. (2013). The effects of transcription factor competition on gene regulation. *Frontiers in Genetics*, *4*(OCT), 1–10. https://doi.org/10.3389/fgene.2013.00197

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Shirley, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9). https://doi.org/10.1186/gb-2008-9-9-r137

Zoller, B., Gregor, T., & Tkačik, G. (2022). Eukaryotic gene regulation at equilibrium, or non? *Current Opinion in Systems Biology*, *31*, 100435. https://doi.org/10.1016/j.coisb.2022.100435

SUPPLEMENTARY FIGURES

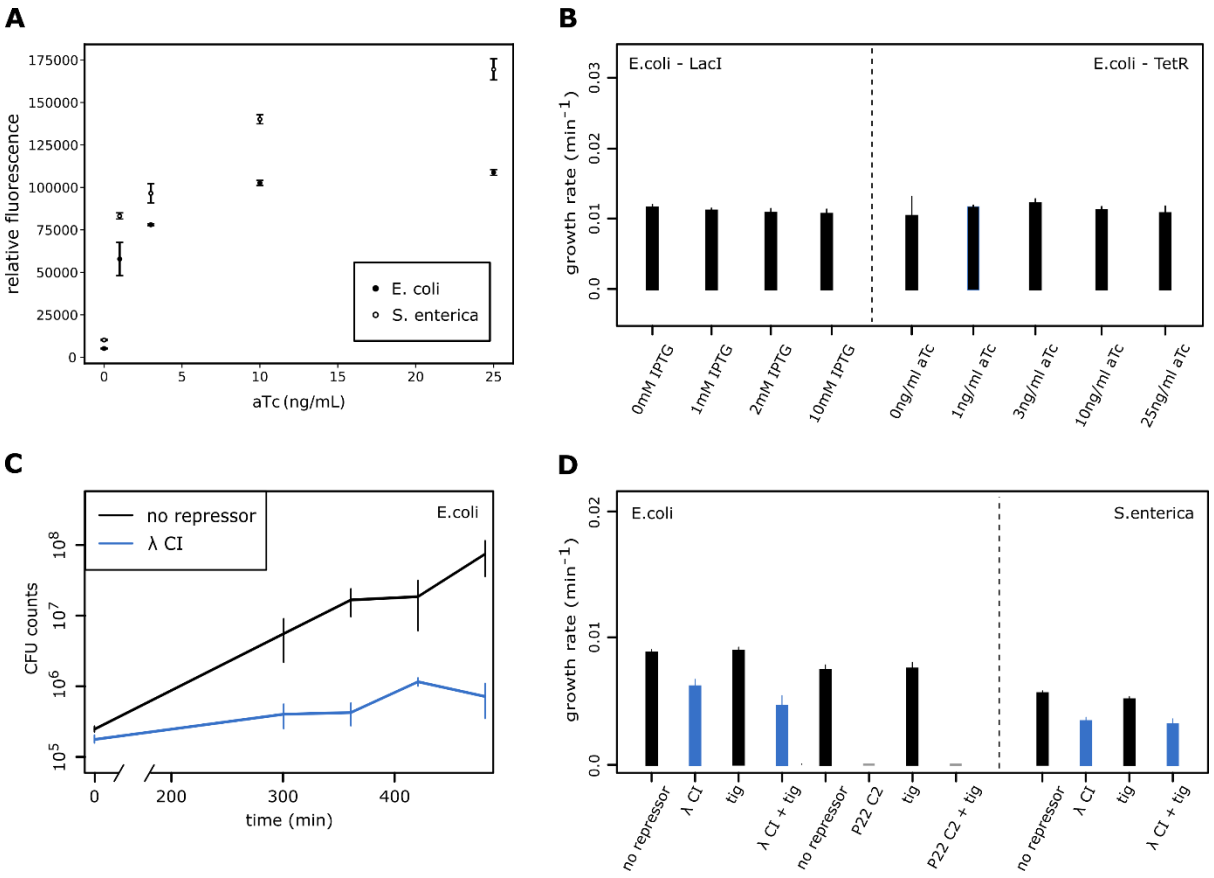


Figure S1. Controls for growth effects.
(A) *E. coli* or *S. enterica* cells containing pZS21 plasmids carrying an inducible fluorescence gene (pZS21-GFP) were grown in minimal media with glucose and induced with 0,1,3,10 or 25 ng/ml aTc. Fluorescence was measured after 11h of induction and normalized by absorbance600. Error bars show standard deviation over 3 replicates. **(B)** Growth rates (per minute) for *E. coli* cells grown in minimal media that are not induced or induced with 1, 2 and 10mM IPTG (LacI binding), and not induced or induced with 1, 3, 10 and 25ng/ml aTc (TetR binding). 0mM IPTG / 0ng/ml aTc shows growth in the presence LacI and TetR binding. **(C)** CFU counts for *E. coli* cells grown in minimal media in a shaking incubator, either expressing (25ng/ml aTc, blue) or not expressing (black) λ CI**,** sampled at 0, 300, 360, 420 and 480 min. Error bars give standard error over 3-8 replicates. **(D)** Growth rates (per minute) for *E. coli* or *S. enterica* cells grown in minimal media (after overnight dilution 1:100) in the absence (black) or presence (25ng/ml aTc) of a repressor, λ CI (blue) or P22 C2 (grey), without or with (1mM IPTG) a chaperone gene. Additional expression of a chaperone (Tig) from a high copy number plasmid (Methods) that prevents formation of aggregates does not alleviate repressor-mediated growth rate reductions (repressor alone versus repressor & tig) in *E. coli* or *S. enterica* cells. Error bars show standard deviation over 3 replicates.

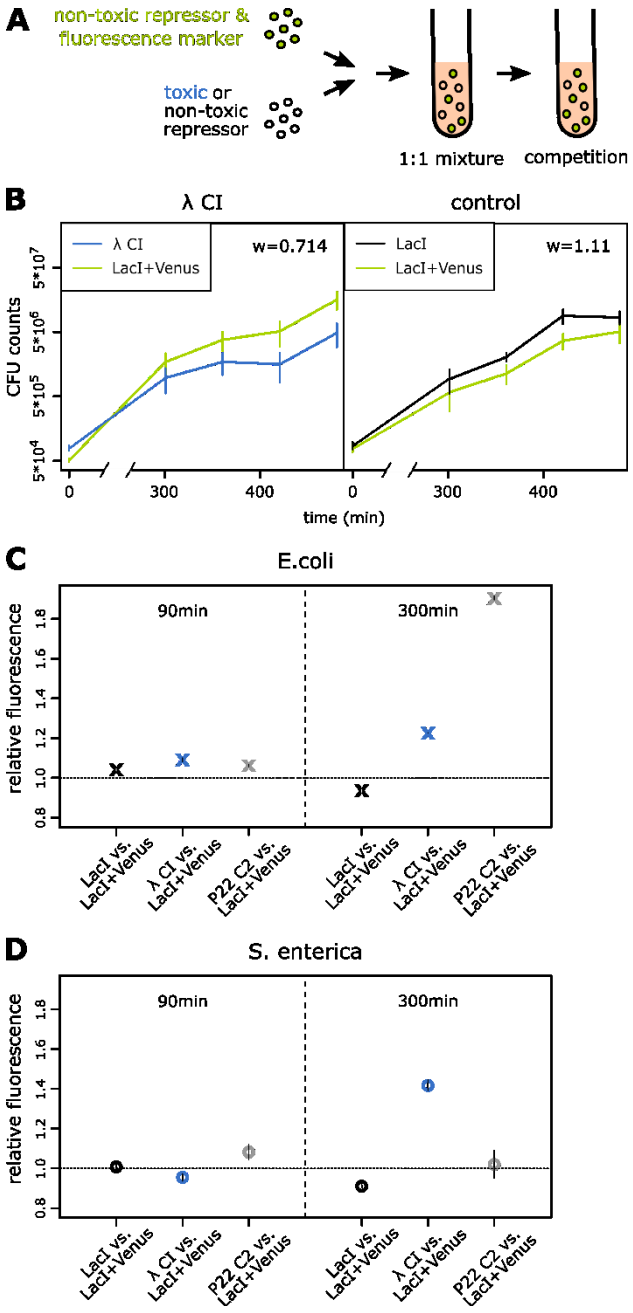


Figure S2. Competition assays reveal the fitness cost of repressor expression.
(A) Cells containing plasmids with a repressor that is affecting growth when expressed (λ CI, P22 C2) or with a repressor not affecting growth when expressed (LacI) were mixed 1:1 with cells containing a plasmid with the repressor not affecting growth and a separately, constitutively expressed Venus marker (LacI+Venus). Note that LacI is also present on the chromosome of *E. coli*, where it has a minimal fitness cost (Fig. S1), but that additional LacI expression is not costly (Fig. 1B). Competition was performed in minimal media with glucose and **(B)** 0.8ng/ml aTc or **(C,D)** 25ng/ml aTc. **(B)** Colony forming units (CFU) for competitions between *E. coli* cells carrying LacI+Venus and *E. coli* cells carrying **(left)** λ CI or **(right)** LacI were determined at the beginning and after 300, 360, 420 and 480min. Relative fitness w was calculated as log ratio between population growth of cells carrying λ CI or LacI and that of LacI+Venus (Methods). The venus marker carried a cost as relative growth of *E. coli* without the marker is 1.11. Expression of λ CI was even more costly, as fitness of λ CI -carrying cells

relative to LacI+Venus -carrying cells was w=0.74. Error bars show standard error over 4 to 14 replicates. **(C,D)** Competition was performed in minimal media with glucose and 25ng/ml aTc and fluorescence was used as a measure of the relative change in the cells carrying LacI-Venus. Relative fluorescence was calculated for **(C)** *E. coli* or **(D)** *S. enterica* between induced and non-induced samples of cell mixtures of LacI+Venus together with LacI (control), λ CI or P22 C2 after **(left)** 90min or **(right)** 300min of competition; error bars show relative errors.

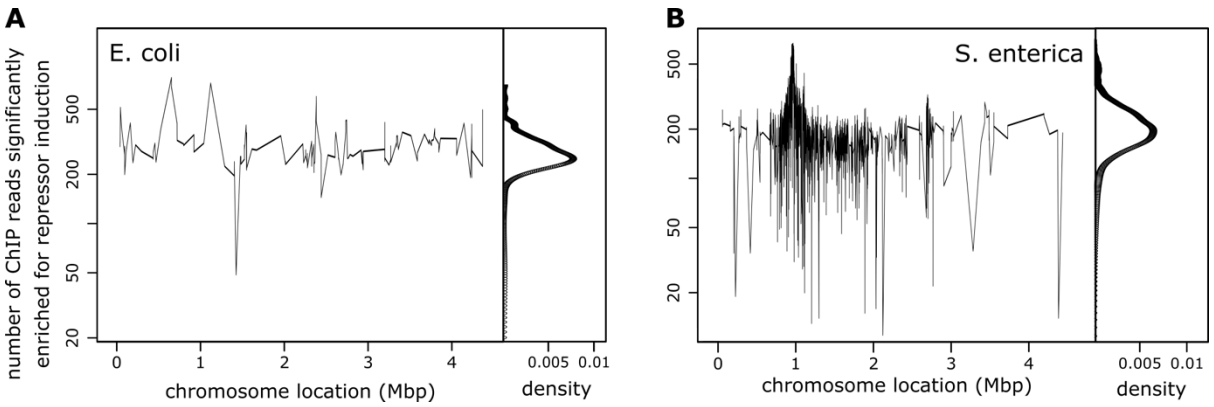


Figure S3. ChIP-sequencing read-depth significantly enriched for repressor induction across the *E. coli* and *S. enterica* genomes.

(A,B) Distributions of ChIP-sequencing reads for the regions found to be significantly enriched when using a lenient peak-calling threshold (Benjamini-Hochberg false-discovery rate of less than 30%) in the experiment with λ CI over the experiment without λ CI across the **(A)** *E.coli* or **(B)** *S. enterica* genome. On the right density plots of enriched read numbers are given. None of the apparent peaks in **(A)** or **(B)** encodes for an essential gene, nor one obviously involved in cell division.

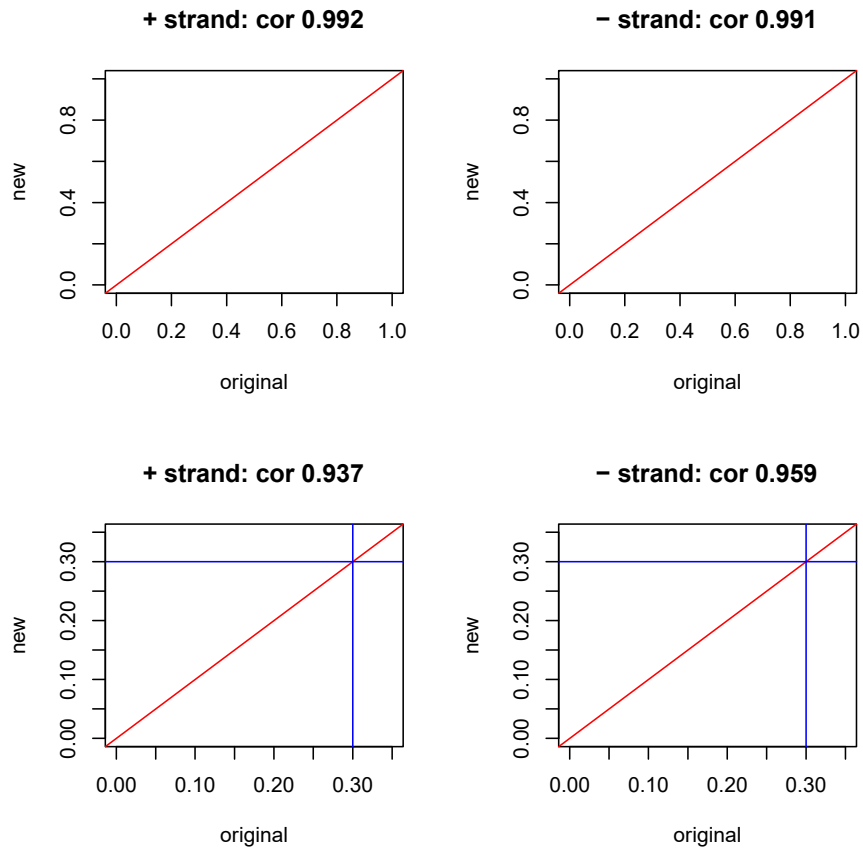


Fig. S4. Correlation between different FDR calculation approaches.
FDR calculation was compared between using the induced *E. coli* – λ CI sample only (original) and using both induced (ChIP) and uninduced (control) samples (new) by randomly partitioning them and assigning experiment labels to the two partitions. This process is repeated for the forward (+) and the reverse (–) strand (left and right plots respectively). Scatterplots show that FDR with random label swaps as well as FDR calculated using our original method are largely concordant (correlation > 0.9).

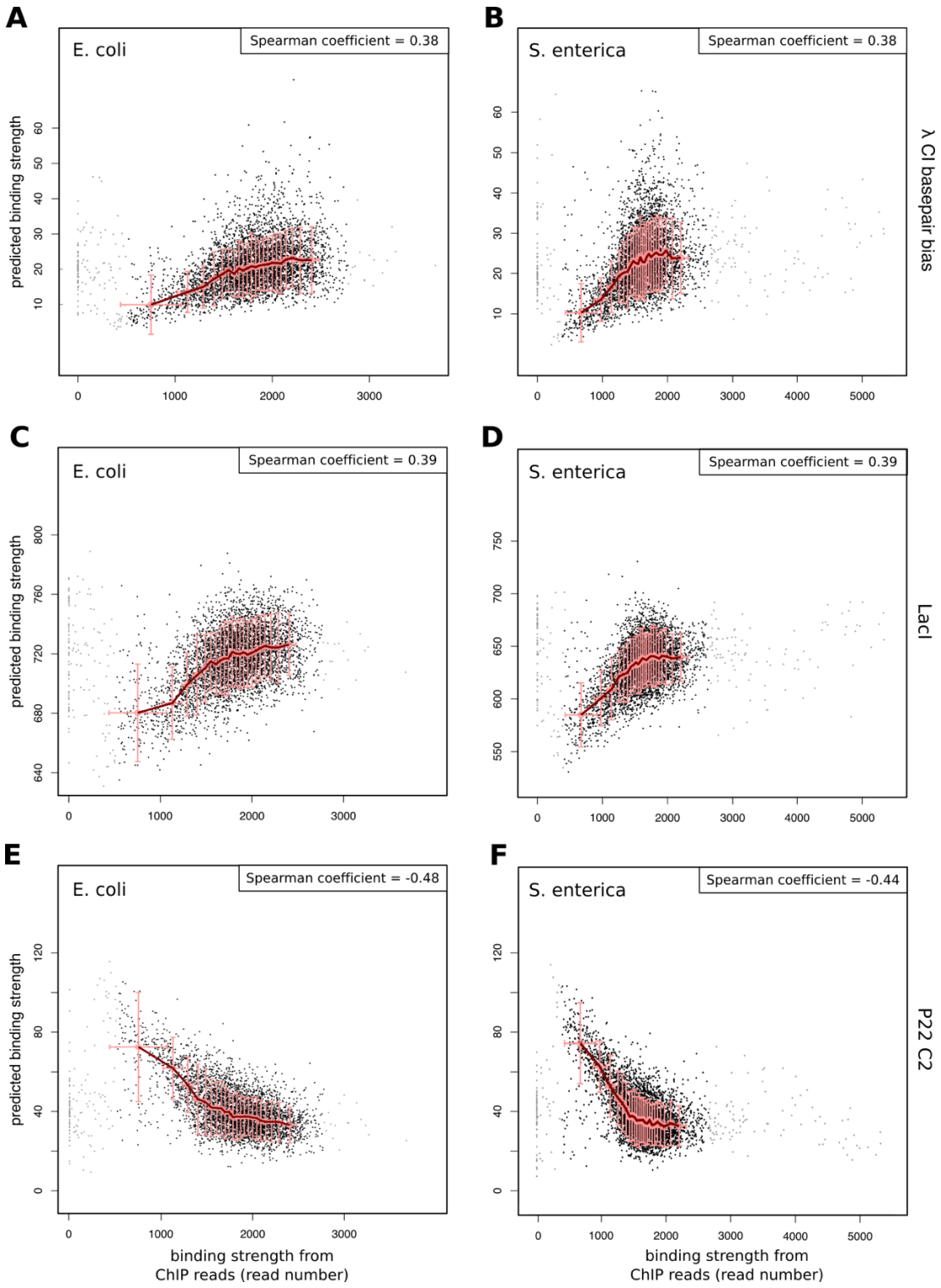


Fig. S5. Fit of λ CI basepair bias, LacI WT and P22 C2 WT energy predictions with the λ CI ChIP-sequencing data.

Fit between binding strength predictions of a simple thermodynamic model using an energy matrix that conserves **(A,B)** only the λ CI basepair bias or the WT energy matrices for **(C,D)** LacI and **(E,F)** P22 C2 binding (together with the corresponding WT offsets) and the λ CI ChIP-sequencing reads across 1000bp windows along the **(A,C,E)** *E. coli* or **(B,D,F)** *S. enterica* genome. The Spearman coefficients calculated for the correlations are shown. The magnitudes of the Spearman correlations for LacI and P22 C2 predictions are similar to the

ones with the λ CI basepair bias, but negative (inverse) for P22 C2. As LacI has a similar basepair bias as λ CI, but P22 C2 an inverse one, this further supports that DNA sequences with the correct basepair bias provide enough recognition motif for λ CI to bind with low affinity. Note that Spearman correlation only considers the qualitative fit, but the range of predicted binding strengths for λ CI energy predictions is substantially larger than for either LacI or P22 C2.

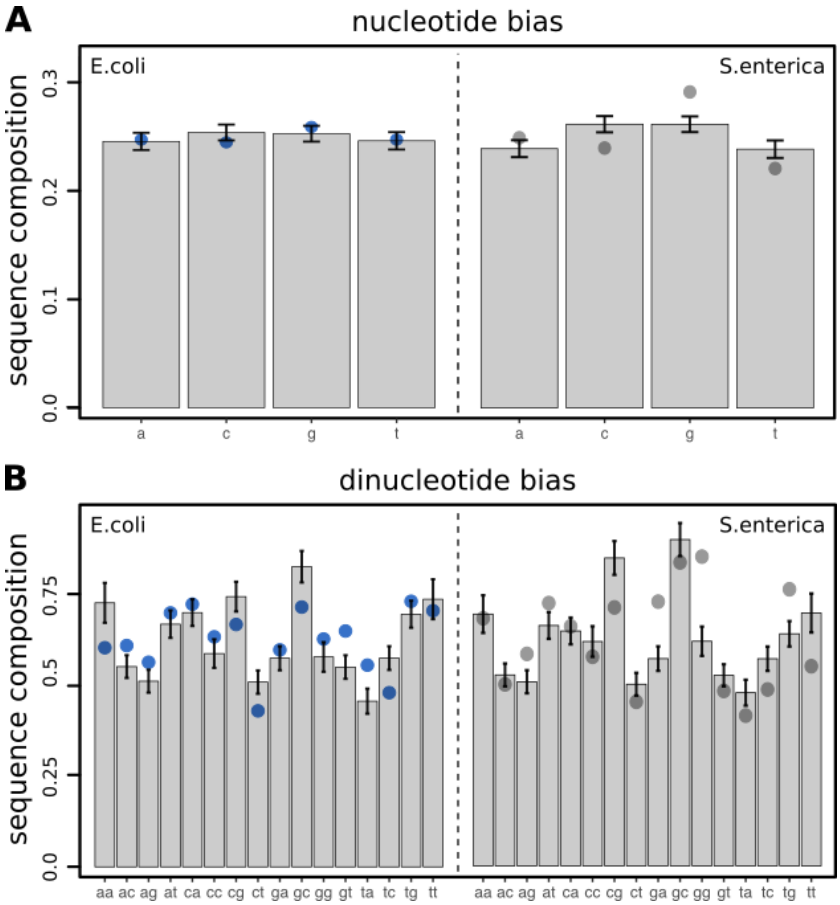


Fig. S6. Nucleotide bias in ChIP-sequencing data.
Bar plots show mean and standard deviation of the background sequence composition (i.e. for regions that were not enriched in ChIP-sequencing data) as compared to the sequence composition in enriched regions (blue and grey dots for *E. coli* and *S. enterica* respectively).

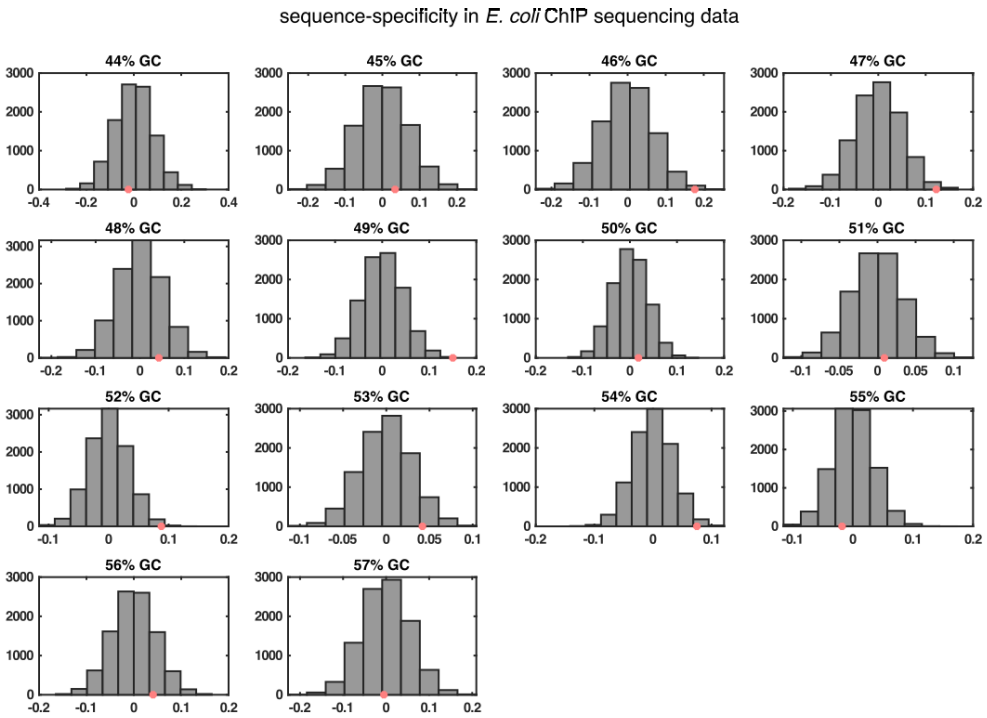


Fig. S7. Sequence-dependence of λ CI binding to the *E. coli* genome.
1000bp fragments of the *E. coli* genome were divided into bins of equal GC percentage (ranging from 44-57%). Histograms show the difference in Spearman correlation between the wildtype energy matrix and the one that conserves basepair bias for 10000 random assignments of ChIP reads to genome regions, whereas red dots show the difference for the real ChIP read assignment. We only included bins of a certain GC composition when they contained at least 100 genome fragments.

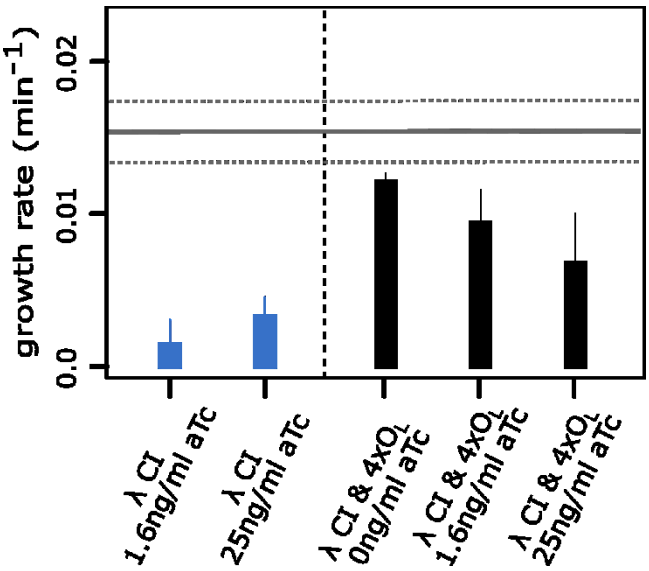


Figure S8. Titration of λ CI by target binding sites on a medium-copy plasmid.
Growth rates (per minute) are shown for *E. coli* cells containing the low-copy plasmid carrying λ CI (blue) or containing the λ CI – plasmid together with a medium-copy plasmid carrying four natural λ CI binding sites (black). λ CI was induced at low (1.6ng/ml aTc induction) or high (25ng/ml aTc induction) concentrations. The black line shows mean and dotted lines standard deviation of uninduced cells with the λ CI – plasmid only, which shows that the medium-copy plasmid itself is slightly reducing growth. Error bars show standard deviation over 3 replicates.

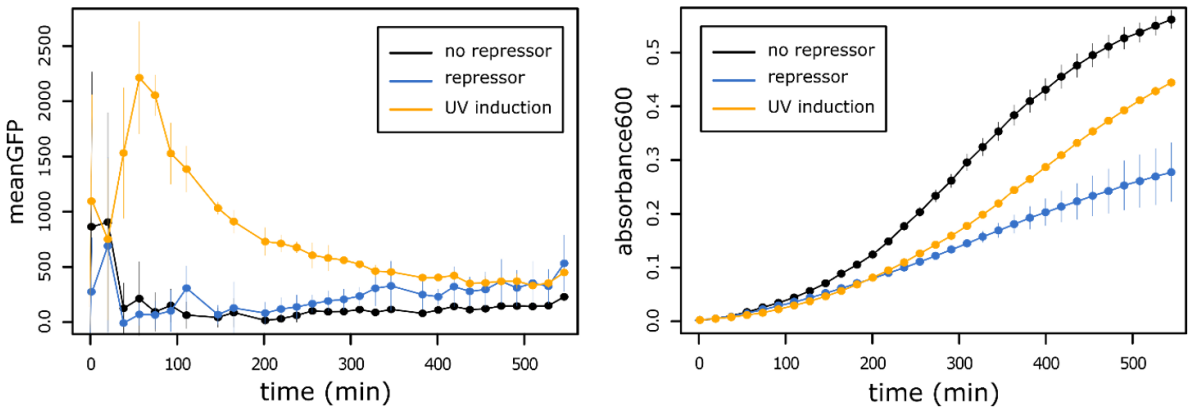


Figure S9. Stress response induction in *E. coli* with λ CI.
E. coli cells containing a stress response reporter gene (P_{sulA}-yfp) and λ CI were grown in minimal media with glucose. Cells were either induced for a certain stressor (25ng/ml aTc for λ CI expression (blue), or UV exposure for 30sec (yellow)), or not induced (black) and growth ($absorbance_{600}$) was measured over several hours. UV induction showed a spike in the stress response (mean absolute gfp fluorescence) and an initial decrease in growth ($absorbance_{600}$), whereas λ CI only led to a much lower stress response induction at later time points and showed an increasing growth reduction. Error bars show standard error over 3 replicates.

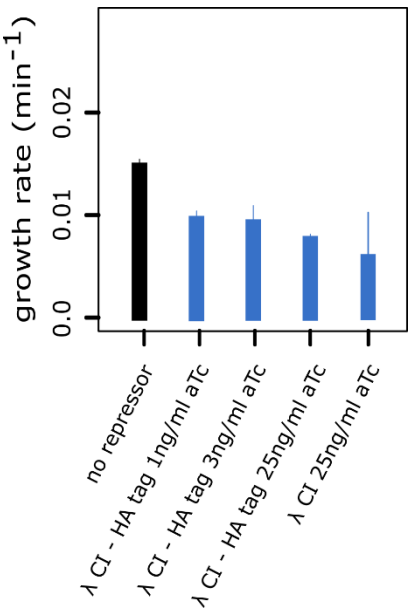


Figure S10. HA-tagged λ CI produces the same growth patterns as wildtype repressor. Growth rates (per minute) for *E. coli* in minimal media without (black) or with induction of λ CI-HA using 1ng/ml, 3ng/ml or 25ng/ml aTc (blue). For comparison induction of λ CI using 25ng/ml aTc in is shown as well. Error bars show standard deviation over 3 replicates.

3.3 Postamble

In the previous paper, we found that transcription factors with a combination of cooperativity and low-affinity binding have the ability to slow growth. We showed that high levels of expression of phage repressors can interfere with host growth, especially when expressed in a naïve host which has not coevolved with the phage, suggesting limits to the composability of phages, as their systems interfere with that of naïve hosts. This suggests that a period of coevolution is required to fully integrate the phage system with its host, which, while feasible for natural systems, is antithetical to human-designed ones: a key aspect of systems design is predictability. Synthetic biologists have designed badly-functioning systems and evolved them towards better performance, but requiring such a tuning step defeats the purpose of composable design, especially if such tuning steps are unreliable.

One thing to note is that phage lambda is unable to undergo either lysis or lysogeny with *S. enterica* as a host, even when the binding receptor is provided. Interestingly, this appears to be solely related to an incompatibility with the *S. enterica* NusA protein, which is involved in N-mediated antitermination; complementing *S. enterica* with the *E. coli nusA* gene fully restores phage lambda's ability to undergo lysis and lysogeny (Harkki & Pavla, 1984). This suggests that *Salmonella* is an unlikely natural host for phage lambda, despite only being blocked by a few amino acids in a single protein, although it is possible that the diversity of *nusA* in *Salmonella* is sufficient to allow replication of phage lambda in some strains. Additionally, it has been shown that overexpression of lambdoid *Salmonella* phage N analogs can complement function of phage lambda's N protein (Franklin & Doelling, 1989), which could lead to possible scenarios in which phage lambda successfully infects *Salmonella* in the presence of other *Salmonella* phages.

Despite being absent in *S. enterica*, *lacI* can sometimes be found in strains of *S. enterica* on plasmids, or in the chromosome flanked by insertion sequences (Leonard *et al.*, 2015), indicating that *S. enterica* may have transient exposure to LacI and thus some amount of coevolution with it.

We found that high expression of phage repressors can lead growth inhibition, generally with a greater effect on non-cognate hosts. This supports our hypothesis that prophages maintain a low level of phage repressor to prevent off-target effects, and that the host coevolves to prevent the prophage from interfering with its cellular processes. But it is interesting to compare this effect with the expression of LacI, which showed no impact on growth rate in either *E. coli* or *S. enterica*.

So why are phage repressors not like LacI? The impact on growth reduction appears to be tied to the cooperativity of binding, as abolishing cooperativity removes the impact, and adding λCI cooperativity to LacI results in a similar growth reduction as λCI. Perhaps phage repressor cooperativity, which is known to play key regulatory roles, is essential to the functioning of the phage repressor and cannot be replaced by LacI-style cooperative binding; it has been shown that replacing λCI with TetR, a repressor with no known cooperativity, can dramatically alter the behavior of the lambda switch (Atsumi & Little, 2006). Despite having less

interference with host cells, the cooperative properties of LacI may not be sufficient for it to function as a phage repressor.

4 Impact of the Host on Decision-Making

4.1 Preamble

A large amount of research has been done on in the lambda switch in model organisms examining the mechanistic interactions between the phage and the host, often using knockouts to elucidate the role of specific proteins in the interaction. However, little research has been done on examining the variability of the rate of lysogeny between bacterial strains. A composable phage would be robust to small changes in its host, with a minimal impact on its lysogenization rate, whereas large changes in lysogeny rate between strains would suggest that the phage is sensitive to small changes in the host, and may be easily interfered with by the genomic background of its host.

It is not necessarily that simple, of course. Given that we are working with a natural system, it is difficult to disentangle sources of information from unexpected interference, as we do not know precisely what sources of information the system should respond to. It is possible that the small changes in the host genomic background accurately report critical changes to important inputs to the system, but it is also possible that these changes are merely interfering with proper functioning.

To this end I wished to screen strains of *E. coli* to identify variations their susceptibility to lysogeny by phage lambda, in order to evaluate the composability of phage lambda; however, priorities changed as the project evolved, which have been documented in this chapter.

Initially I wanted to design fluorescent reporters for both phages and lysogens so that their populations could be tracked under chemostatic conditions, as such a design would be novel and could be used to feed information into the mathematical models developed in Chapter 2.

4.2 Design of a Fluorescent Phage as a Lytic Reporter

Fluorescent phages have been previously designed and used to observe lytic events (Alvarez *et al.*, 2007). For sustained replication of fluorescent phages, they required the modification of the phage to replace the capsid decoration protein gpD with a fluorophore fusion; however, phage lambda with only fluorescently-tagged gpD are unstable, and coexpression of wildtype gpD and fluorescently-tagged gpD rescues stability (Zanghi *et al.*, 2005). Previous papers used a modified fluorescent phage lambda replicating in a host containing a plasmid which expressed wildtype gpD (Zeng *et al.*, 2010), but I noted that one could skip the phage modification step and put the fluorescently-labelled gpD on the plasmid instead, which would simplify the process.

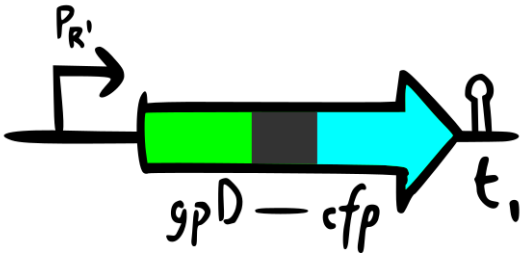


Figure 3. Design of the fluorescent phage reporter construct.

The fluorophore can be exchanged by cutting with BamHI and HindIII. The construct was cloned into a pZA3 backbone with a p15A origin and chloramphenicol resistance.

On a medium copy pZA3 backbone (Lutz & Bujard, 1997), I cloned a 792bp fragment downstream of the *λq* that contains the late lytic promoter $P_{R'}$ upstream of a copy of *λgpD* fused to a fluorophore via a linker previously used for gpD fusions $(GL(GSGG)_3TA$, Alvarez *et al.*, 2007). The plasmid is designed with a number of cut sites for easy swapping of components, and I generated gpD-fluorophore fusions for both CFP and mCherry.

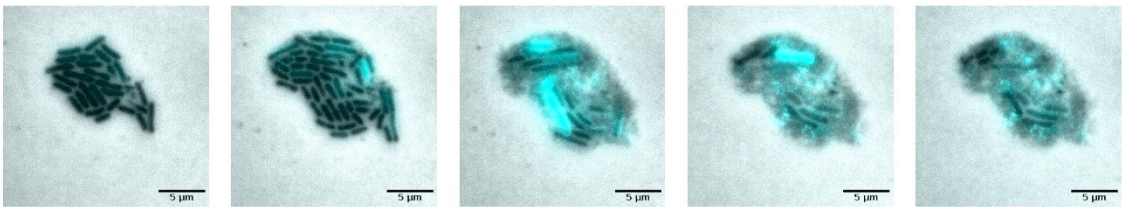


Figure 4. Timelapse of infection by fluorescent phage.

The fluorescent phage construct is capable of producing fluorescence in cells prior to lysis, and producing fluorescent puncta post-lysis. Each successive image is 150 minutes.

To visualize the growing cells, we applied 10µl of mid-log cells diluted to OD_{600} 0.01 mixed with phage lambda in a 10:1 ratio to an M9 agar pad, and took pictures every 3 minutes on an Olympus IX83 inverted fluorescence microscope, with a 100ms exposure time for CFP.

We were able to visualize lytic cells, which began fluorescing prior to lysis, and were able to capture video of multiple rounds of infection taking place (Figure 4). We also identified cells that fluoresced without lysing, suggesting that abortive infection was taking place in some cells. In some experiments we were also able to identify fluorescent puncta that appear after cell lysis, which may be aggregates or fluorescent phage, but we were unable to get these fluorescent points to reliably occur. It is possible that the fluorophores that we chose were not suitable for the application, as it requires fluorescence both intracellularly and extracellularly, and thus requires a fluorophore that is robust to large changes in environmental condition, such as pH.

Despite the promising results in the microscope, we were unable to observe any difference in fluorescence between infected and uninfected cultures either LB or M9 media with either the Chi.Bio chemostat device or with a BioTek Synergy H1 plate reader. At this point I decided to

abandon work with the fluorescent phage construct, as I would be unable to use fluorescent phages if they were undetectable in bulk culture, and optimizing and screening single cells under the microscope using the agar pad would be difficult and impractical, as, despite our best effort, a large fraction of the cells on the agar pad appeared uninfected.

Future work with this construct would do well in optimizing the choice of fluorophore, swapping to a defenseless strain of *E. coli* (ΔmcrBC, ΔsymER, ΔEcoKI, Δmrr, ΔmcrA, Δlit, ΔCRISPR–Cas Type I-E) that has been developed in the lab to increase susceptibility to phage λ, and potentially switching from agar pads to grooved agar pads or microfluidic devices to aid in visualization. Furthermore, the stability of the phage may be sensitive to the ratio of fluorescent gpD to nonfluorescent gpD (Shao *et al.*, 2015), and could use further optimization.

4.3 Design of a Fluorescent Lysogeny Reporter

In parallel I worked on a construct that would allow for the visualization of lysogens under chemostatic conditions. Although a fluorescent reporter for phage λ lysogeny exists (Zeng e*t al.*, 2010), it reports on the expression of P_{RE}, which is only active during the establishment of lysogeny. Although that reporter functioned for their use case of visualizing the lysis-lysogeny decision immediately after infection, it would be unsuitable for long-term identification and tracking of lysogens. Additionally, current identification of lysogens requires either testing susceptibility to phage or using modified phages that confer antibiotic resistance to the lysogen (Pleška et *al.*, 2018), both of which have their issues with construction complexity and evolution. Thus, I designed my own construct.

In lysogens, λCI is under the regulation of the P_{RM} promoter; however this promoter is under negative autoregulation by λCI which maintains λCI at a low level, and I was concerned that this would result in undetectable levels of fluorescence while active. Testing P_{RM} also showed that it was sufficiently leaky to brightly fluoresce in colonies on plates. Instead I decided to use the P_R promoter, which is fully repressed during lysogeny, and a double repressor system, so that fluorescence increases upon lysogeny, which is easier to detect compared to a decrease in fluorescence.

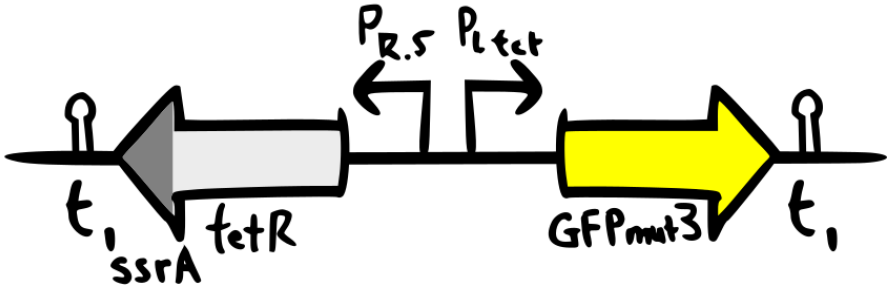


Figure 5. Design of the lysogeny reporter construct.

The promoter $P_{R.5}$ is a modified λP_R promoter with a predicted expression of half the wildtype (sequence GATAAATATCTAACACCGTGCGTGTTGCCGATTTTACCTCTGGCGGTGATAATGGTTGCATGTACTA, Lagator et al*., 2022, personal communication). The construct was cloned into a pZS*1 backbone with a pSC101* origin and ampicillin resistance.*

On a low copy number pS*1 backbone (Lutz & Bujard, 1997) I cloned two outward-facing genes: *tetR* with an SsrA degradation tag under the P_R promoter, and *GFPmut3* under the P_{tet} promoter. The P_R promoter was modified based off a thermodynamic energy model to reduce its predicted expression by half to reduce the constitutive expression of TetR (Lagator *et al.*, 2022, personal communication). The start codon of *tetR* was also replaced with GTG to reduce translation.

I also swapped out the t_0 terminator with the terminator of *rpoC*, as the t_0 originates from phage λ and overlaps with *oop*, an antisense miRNA which suppresses CII expression (Kobiler *et al.*, 2002). Due to directionality it is unlikely to impact the system, but spurious transcription in the opposite direction could impact the lysis-lysogeny decision, and I removed it out of an abundance of caution.

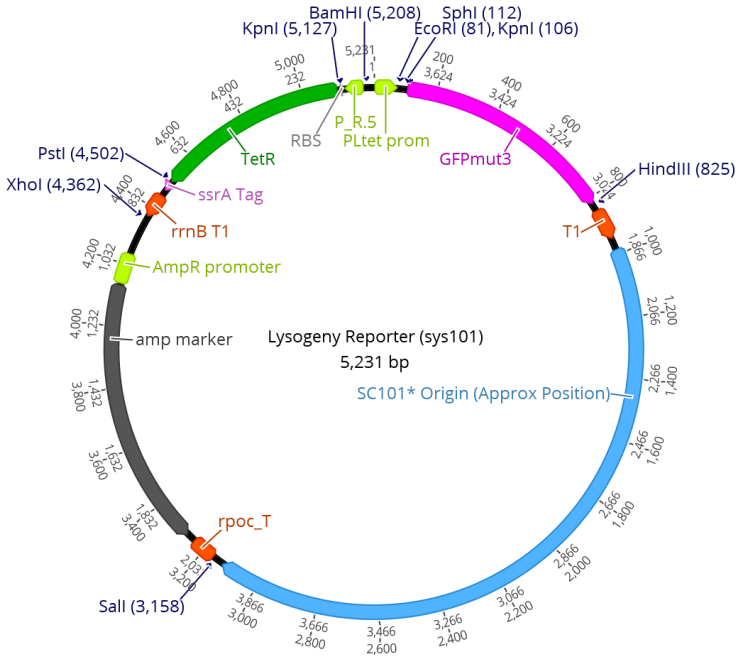


Figure 6. Plasmid map of the lysogeny reporter construct (sys101).

Notable cut sites are annotated. Created in Geneious version 9.1 created by Biomatters.

The plasmid is also designed for the easy replacement of components, as it is anticipated that the replacement of the P_R promoter should be sufficient to adapt the lysogeny reporter to other phage repressors (Figure 6).

With λKAN, a modified phage lambda that confers kanamycin resistance upon lysogeny (Pleška et *al.*, 2018), I was able to verify that lysogenic colonies were fluorescent. Without antibiotic selection, the lysogenic reporter allows for the discrimination between lysogens and uninfected cells; however, the number of uninfected survivors usually greatly outnumbers the number of lysogens, even at high multiplicity of infection, which makes it difficult to quantify the number of lysogens with this method. It may be possible to use flow cytometry to quantify the number of lysogens, but the amount of time after infection before

fluorescence is detectable will have to be quantified, which may result in complications in counting the number of lysogens immediately after infection.

4.4 Visualization of Lysogeny within Plaques

Usage of the lysogen reporter was not restricted to screening individual colonies. Using the lysogen reporter in plaque and spot assays allows for identification of fluorescence at the center of cloudy plaques.

I transformed the lysogeny reporter construct into a number of *E. coli* laboratory strains (MG1655, Frag1, C, B, W), and tested their susceptibility to phage lambda with both spot and plaque assays. I also tested strains of MG1655 with all insertion sequences deleted (ΔIS) and all insertion sequences deleted, and interrupting restored genes with a stop codon (ΔIS*), which were developed in the lab at the same time.

E. coli B and W were both resistant to phage lambda, which was expected, as it is known that *E. coli* B has a mutation in the *malB* region which prevents phage lambda binding (Ronen & Raanan-Ashkenazi, 1971), whereas *E. coli* W hosts prophages which confer resistance to phage lambda through restriction systems (Kerszman *et al.*, 1967).

I screened the sensitive strains at three temperatures (30°C, 37°C, 45°C) and visually compared the intensity of fluorescence and plaque morphology. Fluorescence in plaque centers greatly decreased at 30°C and 45°C across all strains, making it difficult to compare between strains.

I additionally found that, at 30°C in spot assays, our λ_{vir} strain would form fluorescent plaques on all strains, indicating that the virulence mutation of λ_{vir} could be overcome at lower temperatures.[4]

[4] The λ_{vir} strain in our lab has not been sequenced. It does not infect lysogens, therefore it is unlikely to be a mutation in the operator region. It is possible that a mutation in *cII* has made it unstable at 37°C, but still capable of folding at 30°C. An additional possibility is that the slower growth rates at 30°C allows for sufficient buildup of mutant CII to trigger lysogeny. In any case, the term “virulent” is ambiguously used within phage biology, and sometimes also refers to ultravirulent or intemperate mutants which can overcome homoimmunity, of which our strain is not.

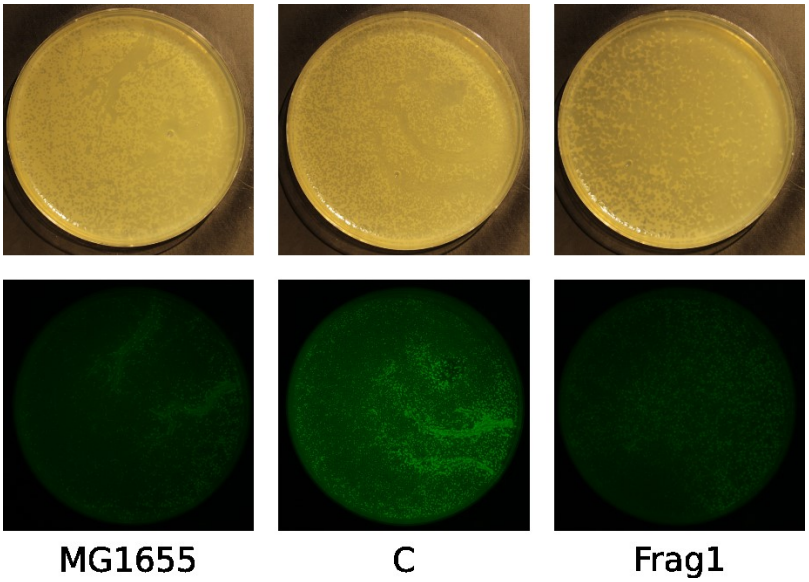


Figure 7. Fluorescence of plaques with the lysogeny reporter.

Plaque assays of strains transformed with the lysogeny reporter. Exposure settings: Bright Field 100ms, Fluorescence 1300ms, with 500/20nm excitation and 535/25nm emission filters.

Under all conditions, *E. coli* C exhibits markedly higher fluorescence within the plaques compared to MG1655 (Figure 7). Frag1 also exhibits slightly higher fluorescence within the plaques, but was more variable in response. *E. coli* C with the lysogeny reporter also generally had a higher background fluorescence even when uninfected, which I initially assumed was due to metabolic effects and ignored, as lysogens are dramatically more fluorescent, but will revisit in section 4.6.3.

At 45°C, the ΔIS* strain would occasionally undergo complete killing, even when not exposed to phage, indicating some degree of temperature sensitivity independent of phage infection.

The lysogeny reporter in MG1655 was also briefly screened on a number of other lambdoid phages, and found to result in fluorescent cloudy plaques in HK106 and HK544, indicating that the lysogeny reporter can cross-react with similar phages. Further examination of phage immunity groups can be found in section 4.6.3.

4.5 E. coli *C has High Lysogen Formation at High Phage Concentrations Compared to MG1655*

Following the phenotypic differences observed in *E. coli* C and Frag1 compared to MG1655, I decided to examine the probability of lysogeny of the three strains.

Cells were grown in LB to an OD_{600} to 0.1. 1mL of culture was mixed with phage and incubated at 4°C for 30 minutes to allow for phage binding, then washed twice at 4°C in LB media. The cells were resuspended in LB and incubated at 37°C for 15 minutes to allow for lysogenization and the expression of kanamycin resistance, and then plated. A sample was also plated on a lawn of cells suspended in phage soft agar (1% tryptone, 0.1% yeast extract, 0.8% NaCl, 0.7% agar, 0.01% glucose, 0.2mM $CaCl_2$) to visualize infective centers as plaques.

Attempts to measure the probability of lysogeny solely with my lysogeny reporter construct and λPaPa proved to be difficult, as, although it was possible to observe lysogens in the final plates, they were greatly outnumbered by surviving non-lysogenized cells, making them unfeasible to count. The surviving cells, when regrown, were still sensitive to phage, indicating that they were either never exposed to phage, despite using large phage-to-bacteria ratios, or that they physiologically resistant to the phage. Plating overnight cultures of phage and cells show a countable number of lysogens, but would introduce issues of evolution and competition dynamics. Thus, I returned to using the λKAN phages and plating on kanamycin selective LB plates (50μg/mL), which kills off any non-lysogenized cells and allows for easier counting of lysogens.

Preliminary experiments showed that at low phage to cell ratios (<1), there was little difference in the lysis-lysogeny rate between the three strains; however, at 7.5:1 and 15:1, there was a dramatic increase in the number of lysogens in *E. coli* C, reaching 263 and 912 lysogens respectively (compared to 0 and 1 for MG1655). Both C and Frag1 had a higher number of infected centers compared to MG1655, suggesting that this may be due to an increase in binding. No lysogens of Frag1 were recovered from any lysis-lysogeny measurement, but cloudy plaques with fluorescent centers show that Frag1 is capable of forming lysogens under different conditions.

We measured the lysogeny rate of *E. coli* C and MG1655 at different phage to bacteria ratios and different growth stages. In early log phase (3.3h), *E. coli* C has a lower rate of lysogeny at both 1:10 (0.2% vs 4%) and 10:1 (6.6% vs 18%) phage to bacteria ratios compared to MG1655, but appears to have a greater fold-change in lysogeny probability. At late log phase (5.5h), C had a lower rate of lysogeny at a 1:10 phage to bacteria ratio (1.2% vs. 5.5%), but had a higher rate of lysogeny at a 10:1 phage to bacteria ratio (17% vs 10.7%), suggesting a steeper response to the ratio of phage to bacteria.

Additionally, at a 1:10 phage to bacteria ratio, *E. coli* C shows no reduction in free phage after a 30 minute incubation at 4°C, compared to a 47% reduction in free phage for MG1655, suggesting that the increased lysogeny rates is not due to increased binding.

As an aside, most phage titres are quantified by plating a known volume on a lawn of susceptible bacteria, and counting the number of plaques that form. This gives the number of "plaque forming units"; however, this is dependent on the strain the phage is being plated on, as even susceptible strains may contain defense systems that are partially active against the phage, which may undercount the true number of functional phage particles. This can be an issue when infecting different strains with the same multiplicity of infection, as what one calculates is the ratio of plaques to cells, as opposed to the ratio of phages to cells.

4.6 Screening of Environmental Strains

Although I screened various laboratory strains of *E. coli* to find differences in phage lambda lysogeny rates, I had access to a number of environmental strains with which I thought I could also test my lysogeny reporter on. I resequenced the environmental strains to obtain complete genomic assemblies, and developed a pipeline to predict and annotate prophages.

Table 1. Overview of Strains Used

	Species	*Origin*	*Isolation Location*	*Predicted Prophages (+extrachromosomal)*	*Source*
MG1655	*E. coli*	Human Feces	California, USA	9 (+0)	Laboratory Strain
B	*E. coli*			7 (+0)	Laboratory Strain
C	*E. coli*			5 (+0)	Laboratory Strain
Frag1	*E. coli*			8 (+0)	Laboratory Strain, via Le *et al.*, 2006
W	*E. coli*	Cemetery Soil	New Jersey, USA	11 (+4)	Laboratory Strain, Waksman, 1943 via Archer *et al.*, 2011
TW15838	*Escherichia* clade I	Freshwater sediment	Australia	19 (+1)	Luo *et al.*, 2011
TW09231	*Escherichia* clade III	Freshwater beach	Michigan, USA	19 (+8)	Luo *et al.*, 2011
TW09276	*Escherichia* clade III	Freshwater beach	Michigan, USA	7 (+2)	Luo *et al.*, 2011
TW11588	*Escherichia* clade IV	Soil	Puerto Rico, USA	13 (+0)	Luo *et al.*, 2011
TW14182	*Escherichia* clade IV	Freshwater beach	Michigan, USA	19 (+8)	Luo *et al.*, 2011
TW15844 (H605)	*Escherichia* clade IV	Human Feces	Australia	3 (+1)	Broad Institute
TW09308	*Escherichia* clade V	Freshwater beach	Michigan, USA	28 (+0)	Luo *et al.*, 2011
TW15839 (E1118)	*Escherichia* clade V	Freshwater	Australia	9 (+0)	Broad Institute
TW15820 (B156)	*E. albertii*	Magpie Feces	Australia	9 (+7)	Broad Institute

The definition of *Escherichia coli* as a species is historically defined by phenotypic traits (lactose fermentation, indole production, lack of citrate utilization), and mostly sampled from the gut of warm-blooded animals; however, phylogenetically similar environmental bacteria cluster similarly, yet being monophyletically distinct, from classical *E. coli*. These isolates are known as the cryptic clades (clades I-V, Walk *et al.*, 2009), with active discussion on whether they should be part of *E. coli* or put into separate species (Cobo-Simón *et al.*, 2022). Nevertheless, the cryptic clades are closely related to *E. coli*, and may serve as hosts to related phages.

4.6.1 Prophage Prediction in Environmental Strains

Genetic sequences were run through Prokka (Seemann, 2014) to annotate the bacterial genes, and run in parallel through PhageBoost (Sirén *et al.*, 2021) to identify prophage regions. Prophage regions were then further run through Pharokka (Bouras *et al.*, 2023) to annotate phage genes. Genetic sequences were also run in ISEScan (Xie & Tang, 2017) to identify insertion sequences, the results of which are not relevant to this thesis.

Sequences of environmental strains often also returned extrachromosomal regions, assembling into putative plasmids, free phage and phage episomes. These extrachromosomal regions were also run through the same process to predict prophages.

I found many predicted prophages in all environmental strains, ranging from as low as 3 to as many as 19 (Table 1); however, these predictions include partial prophages. Not all predicted regions are full prophages, nor is it likely that all of these prophages can still form viable phage particles, as many of these predicted prophages show signs of disruption by insertion sequences.

4.6.2 Environmental Strains are All Resistant to Phage Lambda and LamB is Likely Not Responsible

All environmental tested strains were found to be resistant to phage lambda, in both spot and plaque assays. No environmental strains showed any plaques in either the plaque or spot assays.

This was not unexpected, given the variety of mechanisms by which bacteria can become resistant to phages. Modifications in the phage lambda receptor LamB could prevent binding and thus confer resistance, but resistance could be additionally conferred by any number of defense systems, and a search using DefenseFinder (Tesson *et al.*, 2022) against the genomic sequences show a wide number of defense systems that could defend against phage infection and prevent a plaque from forming.

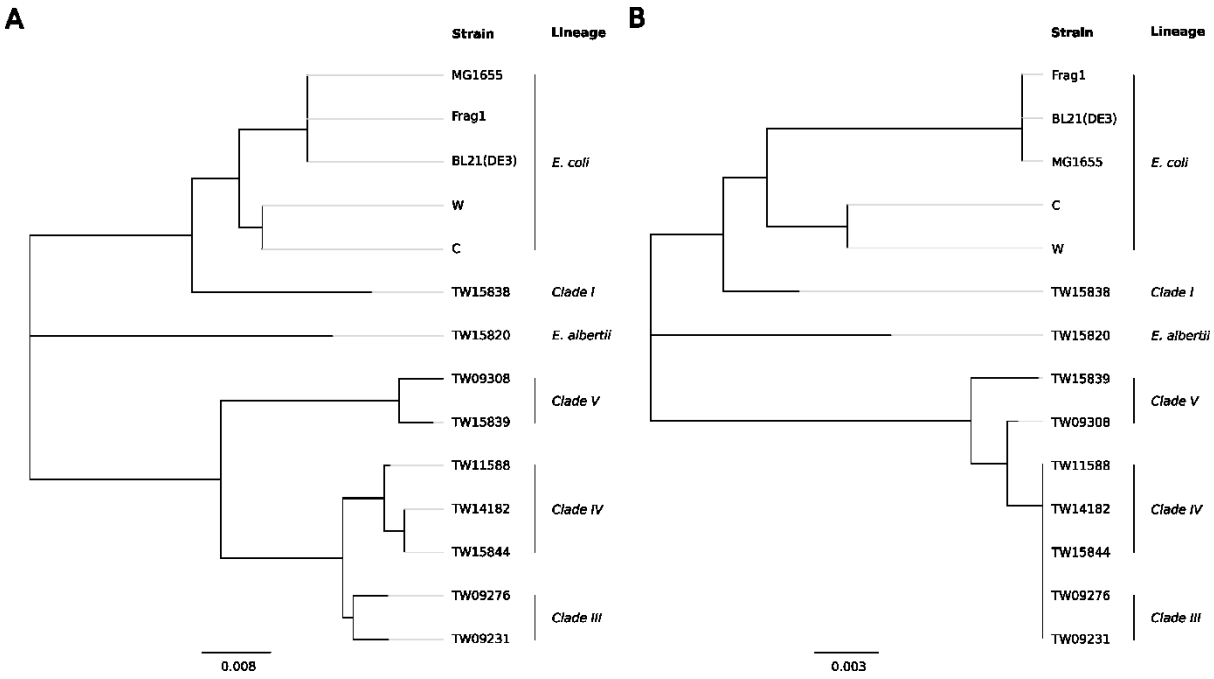


Figure 8. Phylogenetic tree of the* lamB *across various strains.

Phylogenetic comparison of the lamB *gene (A) and protein (B) across various laboratory and environmental strains.*

I compared the *lamB* gene of the environmental and laboratory strains, and found that *lamB* clustered based off *E. coli* clade, and most clustered away from laboratory strain *lamB*, which forms a single group (Figure 8A). *lamB* of TW15838, our sole member of clade I, clusters more closely to the laboratory strain *lamB*, which matches the phylogenetic relationship of the clades (Walk *et al.*, 2009). Interestingly, the LamB protein sequence was identical across all

sequenced clade III and clade IV bacteria despite varying at the nucleotide level and being isolated from geographically different areas (Figure 8B).

As most of the environmental strains shared identical LamB proteins, which differed from the LamB protein in the susceptible laboratory strains, I decided to test the ability the most common environmental LamB protein to complement the phage susceptibility of a *lamB* knockout strain. The *ΔlamB* region of a Keio Knockout Collection strain, which is based on BW25113 (Baba *et al.*, 2006), was transduced into MG1655 by P1 along with the kanamycin selection marker, and phage lambda was shown to no longer form plaques on MG1655*ΔlamB*. Phage lambda also does not form plaques on BW25113*ΔlamB*, but does on BW25113*ΔcirA* (which was chosen at random from the Keio Collection), suggesting that the parent strain BW25113 is susceptible.

For complementation, I designed a plasmid based off the pZA32 backbone (Lutz & Bujard, 1997), and cloned in a representative environmental *lamB* from TW15844 downstream of $P_{LlacO-1}$. As MG1655 has a functional lac operon, there was some concern that native levels of LacI would suppress expression, but transformation of MG1655 with pZA32-venus showed high levels of fluorescence. Plasmid sequencing showed a 23bp deletion of one of the operator sites in $P_{LlacO-1}$, but despite this deletion, the plasmid was able to restore susceptibility to phage lambda in both MG1655*ΔlamB* and BW25113*ΔlamB*, suggesting that binding to environmental LamB is not responsible for their immunity to phage lambda.

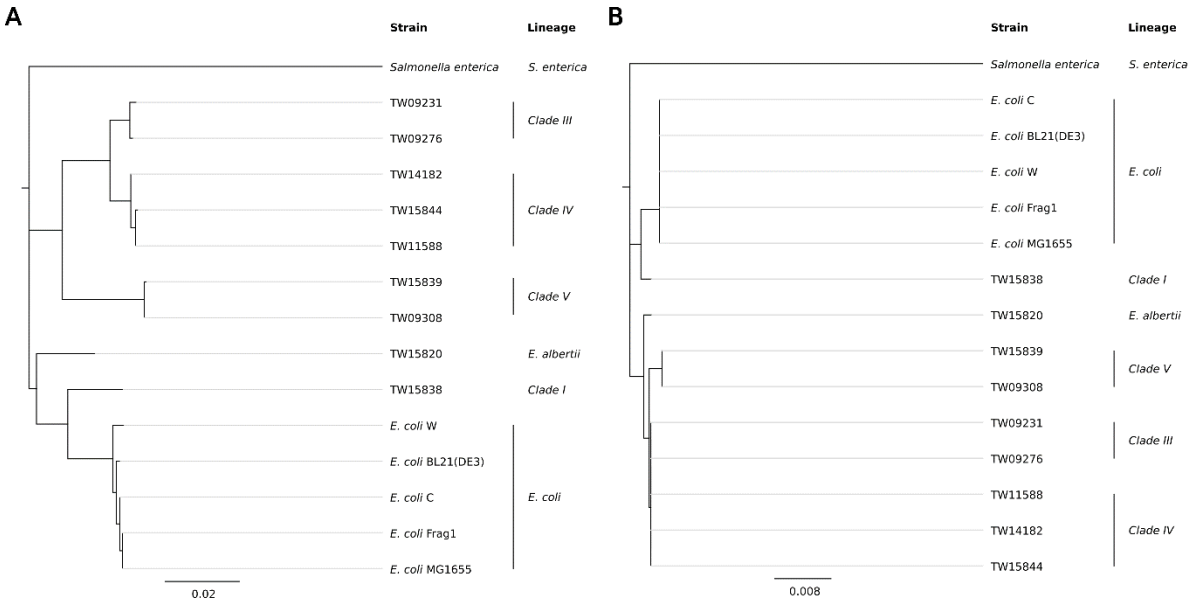


Figure 9. Phylogenetic tree of* nusA *across various strains.

Phylogenetic comparison of the nusA *gene (A) and protein (B) across various laboratory and environmental strains.* S. enterica *has been included for reference.*

Given the incompatibility of *Salmonella* NusA with phage lambda (Harkki & Pavla, 1984), I did a brief bioinformatic comparison between the *nusA* gene and proteins between the environmental strains and the laboratory strains (Figure 9). The NusA proteins clustered based off clade in a similar manner to that of LamB, and once again clade III and IV had identical NusA proteins. The greatest difference was of four amino acids between the

laboratory strains and clade III and IV (V64I, Y82F, K361R and V440M). For comparison, *S. enterica* NusA differs from *E. coli* MG1655 NusA by 31 amino acids, including a 5aa insertion.

A brief search of the environmental genomes for restriction enzymes identified *Pvu*II in TW09308, located adjacent to predicted phage integrase genes, to which phage lambda is susceptible.

4.6.3 Cross-Reaction of Lysogeny Reporter with Putative Prophage in TW15838

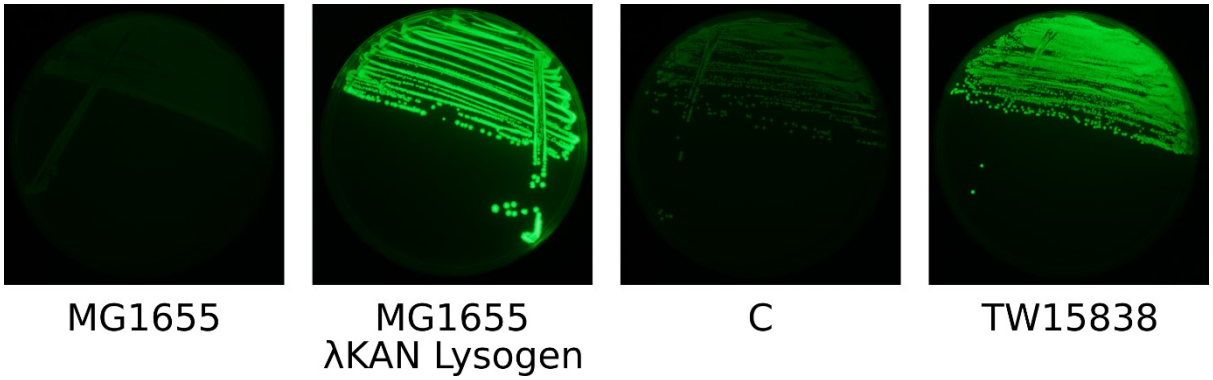


Figure 10. Background fluorescence of the lysogeny reporter.

All strains were transformed with the lysogeny reporter. The λKAN lysogen is included to show the increased fluorescence in response to lysogeny. E. coli *C has a higher background fluorescence than MG1655, but does not reach comparable levels to that of the lysogen, whereas TW15838 reaches markedly high fluorescence levels comparable to the lysogen. Exposure settings: 1300ms, with 500/20nm excitation and 535/25nm emission filters.*

In preparation for measuring the lysogeny rate of the environmental strains, I transformed all the environmental strains with the lysogeny reporter, and noted that the strain TW15838 began to brightly fluoresce even without exposure to phage lambda, comparable to the level as that of confirmed lysogens (Figure 10), which suggests that TW15838 is already lysogenized by a prophage that cross-reacts with the λP_R promoter.

The bioinformatics analysis shows that the TW15838 *lamB* gene and protein both more closely resembles that of the laboratory strains than the other environmental strains, and that a predicted lambdoid prophage TW15838_phage4 is already present at the lambda attachment site. Although prophages have been detected at the lambda attachment site in environmental strains (Kuhn & Campbell, 2001), they have not been sequenced. TW15838_phage4 does not map closely to any single phage, which is expected due to their recombinatorial ability, but the lambda switch region only differs from the lambdoid phage BP-4795 by 2bp (Creuzburg *et al.*, 2005). Similar to BP-4795, this prophage also contains the Stx1 genes which code for the Shiga toxin, indicating that it is likely related to the Stx1 phages.

The integrase of TW15838_phage4 is nearly identical to that of phage lambda, with 4 point mutations that translate to a 3 amino acid difference. This likely explains its choice of integration site.

I attempted to induce prophages from TW15838 with 60s UV exposure in a biosafety cabinet and exposure to 0.5µg/mL mitomycin C, but no plaques were formed on a MG1655 or

defenseless MG1655. A number of prophages (across all environmental strains) include regions which match predicted insertion sequences, which suggests that those phages may be disrupted and unable to form functional phage particles.

To check whether the CI of TW15838_phage4 was responsible for activating the lysogeny reporter, I cloned the *ci* gene from TW15838_phage4 into a pZA32 backbone. Cotransforming both this CI-expression vector and the lysogeny reporter into MG1655 results no increase in fluorescence, which suggests that the TW15838_phage4 CI cannot cross react with λP_R; however, spot assays on strains with the CI-expression vector show a decreased cloudiness in the plaque, especially when cotransformed with the lysogeny reporter, implying that it may interfere with phage lambda in other ways.

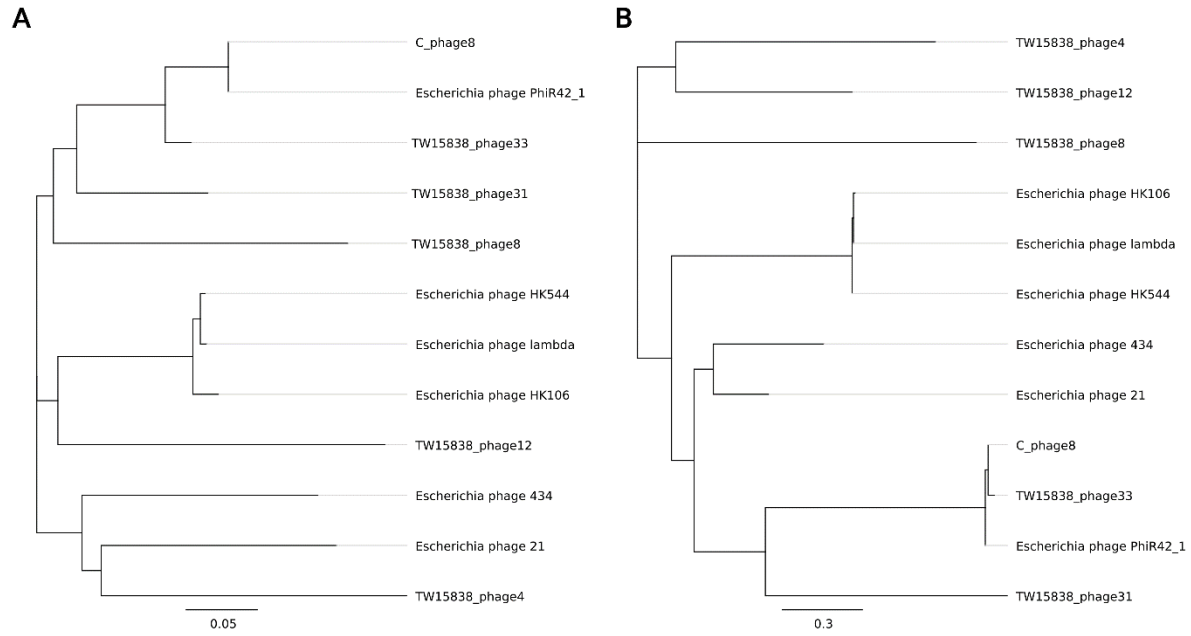


Figure 11. Phylogenetic tree of select* cI *of lambdoid phages and prophages.

Phylogenetic comparison of the cI *gene (A) and protein (B) across various lambdoid phages and prophages. Included in the tree are phage lambda; phages HK106 and HK544, which cross-react with the lysogeny reporter; phage 434 and phage 21, classically studied lambdoid phages known not to react with P_R, all predicted lambdoid prophages of TW15838 and* E. coli *C; and phage PhiR42_1, an extremely close match to prophages in previously mentioned strains.*

TW15838 has a 19 predicted prophages in its genome, five of which appear to be lambdoid, containing a genetic structure similar to that of the lambda switch. A protein alignment of all five putative CI proteins shows a large amount of diversity and none appear to very homologous to λCI (Figure 11).

Measurement issues made me reconsider the higher background fluorescence of *E. coli* C. *E. coli* C only has a single predicted lambdoid prophage, which matches to Escherichia phage PhiR42_1 (Laub & Brenes, 2025) with two major deletions (97bp of the end of *cII* extending through DNA replication protein gene *o* and continuing another 1.2kb; and a large number of terminase, portal, capsid and tail fibre proteins), and an insertion of two transposases. Pertinently, the CI proteins are identical, and are a single amino acid away from CI from a different prophage in TW15838, and may serve as a target to confirm mild cross-reactivity

with λP_R; however, as phage lambda is capable of replicating on *E. coli* C, the PhiR42_1 CI is unlikely to be sufficient to confer immunity, unless the CI is not being expressed.

This raises interesting implications into polylysogens and incomplete immunity. Most environmental strains are polylysogens, and my experiments suggest that there may be some crosstalk between similar but non-identical phage repressors. Though providing partial immunity, each subsequent lysogen may provide more immunity to similar repressors, until the cell is fully resistant to phages with similar repressors. This leads into interesting discussion about competition between phages, as although more lysogens may contribute to greater resistance to phages, during induction lysogens must compete over host resources (Silpe *et al.*, 2023).

Alternatively, a separate system may be interfering with the lysogeny reporter in TW15838. The P_R promoter may be inhibited by alternative means, or the TetR protein might be rapidly degraded in this strain. Further research is required to whether these are playing a role, but the lysogeny reporter system is definitely being perturbed.

4.7 Live Lysogeny Measurements in a Plate Reader

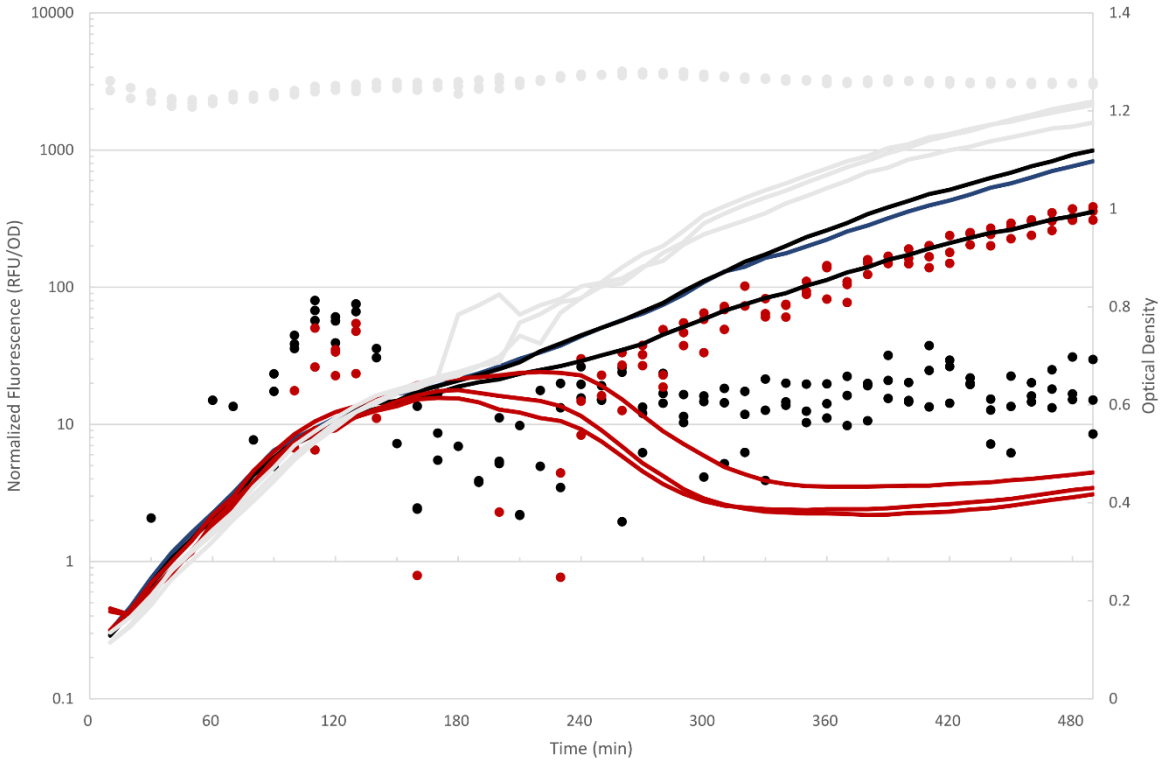


Figure 12. Plate reader measurements of the lysogeny reporter.

Plates were initially seeded with OD 0.1 E. coli *MG1655 cells (black), λKAN lysogens of MG1655 (gray) and MG1655 mixed with λKAN at a 1:1 ratio (red). Normalized fluorescence (dots, left axis) for both MG1655 and infected cells increases at 120 minutes. In the infected cells, normalized fluorescence also increases as the optical density (lines, right axis) declines at 240 minutes and recovers.*

An additional advantage of my lysogeny reporter is that it can report on the number of lysogens in the population without sampling or selection. I infected strains with my lysogeny

reporter with lambda phage in a plate reader and was able to detect an increase in fluorescence after the collapse of the population (Figure 12), which corresponds with when the most lysogens are theorized to be formed (Lang *et al.*, 2020); however, it is unclear how soon after the initiation of lysogeny fluorescence becomes detectable, nor is it clear how fluorescence per lysogen increases over time until it reaches its maximum. Additionally, we also observe an increase in fluorescence in both infected and uninfected cells at 120 minutes, which may be due to physiological properties of the cell.

It is theoretically possible to extract more quantifiable information out of these assays but determining an instantaneous probability of lysogeny would be difficult as any increase in fluorescence is a combination of new lysogenic events and replication of preexisting lysogens, in combination with the delayed response of the lysogeny reporter to the initiation of lysogeny. Future work can be done to model the population-level fluorescence as a combination of all these factors.

5 Conclusions

Biological systems are difficult to work with. They are not the minimal systems that exist in engineering, but systems with millions of years of evolutionary history, evolving, coevolving and diverging with and from each other.

At the outset, I chose to examine temperate phages for their theoretical capacity for composability. They would encounter many different hosts in the environment, and should make their decisions to the best of their ability within those environments.

In Chapter 2 I have shown what sources of information in the environment these temperate phages should respond to, and revealed that the number of lysogens is an important signal in environments with sorptive scavenging to lysogens.

In Chapter 3, I showed that limits exist to the expression of the phage repressors, which might otherwise interfere with the host, and suggest that hosts are coevolving with their phages to reduce interference, suggesting that phage repressors are not composable and that hosts must adapt to the phages they encounter.

In Chapter 4 I found that environmental hosts have a high amount interference with the phage lambda system, but many of these, instead of being background changes that only incidentally interfere, appear to be targeted systems: other prophages and potentially defense systems, specifically evolved to interfere with invading phage systems, or closely-related relatives whose components interfere.

Synthetic systems have an advantage in this regard, as they are likely completely novel to the organism they are introduced to; however, care must be chosen with the components that make up these systems, to ensure that they do not interfere with, or that they are not targeted for interference by, the host.

I was unable to determine if any of our environmental strains were capable of replicating phage lambda, as I was unable to determine the precise cause of resistance. I was able to show that the LamB receptor was still capable of binding to phage lambda and thus not the reason why the environmental strains are immune, but whether the failure to replicate is due to a fundamental incompatibility with the replication machinery, as is with phage lambda in *Salmonella*, or due to specific defense systems or prophages, is still unclear. What is clear, however, is that prophages are common amongst environmental strains, and that lambdoid switches are common and can potentially cross react with other incoming lambdoid phages, even when they do not confer complete immunity.

In *E. coli* C I found a large change in the number of lysogens formed compared to MG1655, suggesting that phage lambda is sensitive to changes in the host; however, I have not yet been able to distinguish the cause of this effect, and whether it is caused by an effect on the phage decision-making machinery itself or on the fidelity of the information that the phage genome receives.

Although I have not been able to conclusively prove whether phages are composable or not, the evidence I have gathered suggests that they are not, and that phages can be sensitive to

changes in their hosts. This does not bode well for synthetic biology, as it suggests that systems must be specifically designed to suit their hosts and to not cross-interfere. Synthetic systems designed for a small range of hosts or systems may be feasible, but a lack of composability, even in natural systems thought to be so, implies that scaling will always be a limitation for synthetic biology without development towards membranes, timing, or other forms of system separation.

I hope this thesis has proved useful in elucidating the properties and complexities in dealing with even the most well-studied organisms, and that the tools that I have developed can prove useful in the future for greater examination of phage-bacteria interactions.

References

Alvarez LJ, Thomen P, Makushok T, Chatenay D. Propagation of fluorescent viruses in growing plaques. *Biotechnol Bioeng*. 2007 Feb 15;96(3):615-21. doi: 10.1002/bit.21110. PMID: 16900526.

Archer CT, Kim JF, Jeong H, Park JH, Vickers CE, Lee SY, Nielsen LK. The genome sequence of E. coli W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. *BMC Genomics*. 2011 Jan 6;12:9. doi: 10.1186/1471-2164-12-9. PMID: 21208457; PMCID: PMC3032704.

Atsumi S, Little JW. A synthetic phage lambda regulatory circuit. *Proc Natl Acad Sci U S A*. 2006 Dec 12;103(50):19045-50. doi: 10.1073/pnas.0603052103. Epub 2006 Nov 29. PMID: 17135356; PMCID: PMC1748174.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol.* 2006;2:2006.0008. doi: 10.1038/msb4100050. Epub 2006 Feb 21. PMID: 16738554; PMCID: PMC1681482.

Bode W. Lysis inhibition in Escherichia coli infected with bacteriophage T4. *J Virol*. 1967 Oct;1(5):948-55. doi: 10.1128/JVI.1.5.948-955.1967. PMID: 4912240; PMCID: PMC375373.

Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald PJ, Vreugde S. Pharokka: a fast scalable bacteriophage annotation tool. *Bioinformatics*. 2023 Jan 1;39(1):btac776. doi: 10.1093/bioinformatics/btac776. PMID: 36453861; PMCID: PMC9805569.

Bruce JB, Lion S, Buckling A, Westra ER, Gandon S. Regulation of prophage induction and lysogenization by phage communication systems. *Curr Biol*. 2021 Nov 22;31(22):5046-5051.e7. doi: 10.1016/j.cub.2021.08.073. Epub 2021 Sep 24. PMID: 34562385; PMCID: PMC8612742.

Castle SD, Stock M, Gorochowski TE. Engineering is evolution: a perspective on design processes to engineer biology. *Nat Commun*. 2024 Apr 29;15(1):3640. doi: 10.1038/s41467-024-48000-1. PMID: 38684714; PMCID: PMC11059173.

Cobo-Simón M, Hart R, Ochman H. Escherichia Coli: What Is and Which Are? *Mol Biol Evol*. 2023 Jan 4;40(1):msac273. doi: 10.1093/molbev/msac273. PMID: 36585846; PMCID: PMC9830988.

Creuzburg K, Recktenwald J, Kuhle V, Herold S, Hensel M, Schmidt H. The Shiga toxin 1-converting bacteriophage BP-4795 encodes an NleA-like type III effector protein. J Bacteriol. 2005 Dec;187(24):8494-8. doi: 10.1128/JB.187.24.8494-8498.2005. PMID: 16321954; PMCID: PMC1317009.

Eswarappa SM, Karnam G, Nagarajan AG, Chakraborty S, Chakravortty D. lac repressor is an antivirulence factor of Salmonella enterica: its role in the evolution of virulence in Salmonella. *PLoS One*. 2009 Jun 4;4(6):e5789. doi: 10.1371/journal.pone.0005789. PMID: 19495420; PMCID: PMC2686271.

Franklin NC, Doelling JH.1989.Overexpression of N antitermination proteins of bacteriophages lambda, 21, and P22: loss of N protein specificity. *J Bacteriol 171*: https://doi.org/10.1128/jb.171.5.2513-2522.1989

Frost LS, Leplae R, Summers AO, Toussaint A. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol. 2005 Sep;3(9):722-32. doi: 10.1038/nrmicro1235. PMID: 16138100.

Harkki, A., Palva, E.T. Application of phage lambda technology to *Salmonella typhimurium*. *Mol Gen Genet* **195**, 256–259 (1984). https://doi.org/10.1007/BF00332756

Haudiquet M, de Sousa JM, Touchon M, Rocha EPC. Selfish, promiscuous and sometimes useful: how mobile genetic elements drive horizontal gene transfer in microbial populations. *Philos Trans R Soc Lond B Biol Sci*. 2022 Oct 10;377(1861):20210234. doi: 10.1098/rstb.2021.0234. Epub 2022 Aug 22. PMID: 35989606; PMCID: PMC9393566.

He L, Miguel-Romero L, Patkowski JB, Alqurainy N, Rocha EPC, Costa TRD, Fillol-Salom A, Penadés JR. Tail assembly interference is a common strategy in bacterial antiviral defenses. *Nat Commun*. 2024 Aug 30;15(1):7539. doi: 10.1038/s41467-024-51915-4. PMID: 39215040; PMCID: PMC11364771.

Kobiler O, Koby S, Teff D, Court D, Oppenheim AB. The phage lambda CII transcriptional activator carries a C-terminal domain signaling for rapid proteolysis. *Proc Natl Acad Sci U S A*. 2002 Nov 12;99(23):14964-9. doi: 10.1073/pnas.222172499. Epub 2002 Oct 23. PMID: 12397182; PMCID: PMC137528.

Kuhn J, Campbell A. The bacteriophage lambda attachment site in wild strains of Escherichia coli. J Mol Evol. 2001 Dec;53(6):607-14. doi: 10.1007/s002390010247. PMID: 11677620.

Lagator M, Sarikas S, Steinrueck M, Toledo-Aparicio D, Bollback JP, Guet CC, Tkačik G. Predicting bacterial promoter function and evolution from random sequences. *Elife*. 2022 Jan 26;11:e64543. doi: 10.7554/eLife.64543. PMID: 35080492; PMCID: PMC8791639.

Lang M, Pleška M, Guet CC. Population dynamics of decision making in temperate bacteriophages. *biorXiv*. 2020 Mar 20. doi: 10.1101/2020.03.18.99691.

Le TT, Emonet T, Harlepp S, Guet CC, Cluzel P. Dynamical determinants of drug-inducible gene expression in a single bacterium. Biophys J. 2006 May 1;90(9):3315-21. doi: 10.1529/biophysj.105.073353. Epub 2006 Feb 3. PMID: 16461398; PMCID: PMC1432126.

Leonard SR, Lacher DW, Lampel KA. Acquisition of the *lac* operon by *Salmonella enterica*. *BMC Microbiol*. 2015 Aug 25;15:173. doi: 10.1186/s12866-015-0511-8. PMID: 26303940; PMCID: PMC4549013.

Luo C, Walk ST, Gordon DM, Feldgarden M, Tiedje JM, Konstantinidis KT. Genome sequencing of environmental Escherichia coli expands understanding of the ecology and speciation of the model bacterial species. *Proc Natl Acad Sci U S A*. 2011 Apr 26;108(17):7200-5. doi: 10.1073/pnas.1015622108. Epub 2011 Apr 11. PMID: 21482770; PMCID: PMC3084108.

Lutz R, Bujard H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res*. 1997 Mar 15;25(6):1203-10. doi: 10.1093/nar/25.6.1203. PMID: 9092630; PMCID: PMC146584.

Mukherjee S, Bassler BL. Bacterial quorum sensing in complex and dynamically changing environments. *Nat Rev Microbiol*. 2019 Jun;17(6):371-382. doi: 10.1038/s41579-019-0186-5. PMID: 30944413; PMCID: PMC6615036.

Obuchowski, M., Shotland, Y., Koby, S., Giladi, H., Gabig, M., Wegrzyn, G., & Oppenheim, A. B. (1997). Stability of CII is a key element in the cold stress response of bacteriophage lambda infection. *Journal of Bacteriology*, *179*(19), 5987–5991. doi: 10.1128/jb.179.19.5987-5991.1997

Pleška M, Lang M, Refardt D, Levin BR, Guet CC. Phage-host population dynamics promotes prophage acquisition in bacteria with innate immunity. *Nat Ecol Evol*. 2018 Feb;2(2):359-366. doi: 10.1038/s41559-017-0424-z. Epub 2018 Jan 8. PMID: 29311700.

Ptashne, M. (2004). *A genetic switch: phage lambda revisited*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Ronen A, Raanan-Ashkenazi O. Temperature sensitivity of maltose utilization and lambda resistance in Escherichia coli B. *J Bacteriol*. 1971 Jun;106(3):791-6. doi: 10.1128/jb.106.3.791-796.1971. PMID: 4934064; PMCID: PMC248693.

Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014 Jul 15;30(14):2068-9. doi: 10.1093/bioinformatics/btu153. Epub 2014 Mar 18. PMID: 24642063.

Shao Q, Hawkins A, Zeng L. Phage DNA dynamics in cells with different fates. *Biophys J*. 2015 Apr 21;108(8):2048-60. doi: 10.1016/j.bpj.2015.03.027. PMID: 25902444; PMCID: PMC4407255.

Shao Q, Trinh JT, Zeng L. High-resolution studies of lysis-lysogeny decision-making in bacteriophage lambda. J Biol Chem. 2019 Mar 8;294(10):3343-3349. doi: 10.1074/jbc.TM118.003209. Epub 2018 Sep 21. PMID: 30242122; PMCID: PMC6416446.

Silpe JE, Duddy OP, Johnson GE, Beggs GA, Hussain FA, Forsberg KJ, Bassler BL. Small protein modules dictate prophage fates during polylysogeny. *Nature*. 2023 Aug;620(7974):625-633. doi: 10.1038/s41586-023-06376-y. Epub 2023 Jul 26. PMID: 37495698; PMCID: PMC10432266.

Sirén K, Millard A, Petersen B, Gilbert MTP, Clokie MRJ, Sicheritz-Pontén T. Rapid discovery of novel prophages using biological feature engineering and machine learning. *NAR Genom Bioinform*. 2021 Jan 6;3(1):lqaa109. doi: 10.1093/nargab/lqaa109. PMID: 33575651; PMCID: PMC7787355.

St-Pierre, F., & Endy, D. (2008). Determination of cell fate selection during phage lambda infection. *Proceedings of the National Academy of Sciences*, *105*(52), 20705–20710. doi: 10.1073/pnas.0808831105

Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, Bernheim A. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun. 2022 May 10;13(1):2561. doi: 10.1038/s41467-022-30269-9. PMID: 35538097; PMCID: PMC9090908.

Walk ST, Alm EW, Gordon DM, Ram JL, Toranzos GA, Tiedje JM, Whittam TS. Cryptic lineages of the genus Escherichia. Appl Environ Microbiol. 2009 Oct;75(20):6534-44. doi: 10.1128/AEM.01262-09. Epub 2009 Aug 21. PMID: 19700542; PMCID: PMC2765150.

Weitz JS, Mileyko Y, Joh RI, Voit EO. Collective decision making in bacterial viruses. *Biophys J*. 2008 Sep 15;95(6):2673-80. doi: 10.1529/biophysj.108.133694. Epub 2008 Jun 20. PMID: 18567629; PMCID: PMC2527279.

Winstanley C, O'Brien S, Brockhurst MA. Pseudomonas aeruginosa Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections. *Trends Microbiol*. 2016 May;24(5):327-337. doi: 10.1016/j.tim.2016.01.008. Epub 2016 Mar 3. PMID: 26946977; PMCID: PMC4854172.

Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics*. 2017 Nov 1;33(21):3340-3347. doi: 10.1093/bioinformatics/btx433. PMID: 29077810.

Zeng L, Skinner SO, Zong C, Sippy J, Feiss M, Golding I. Decision making at a subcellular level determines the outcome of bacteriophage infection. *Cell*. 2010 May 14;141(4):682-91. doi: 10.1016/j.cell.2010.03.034. PMID: 20478257; PMCID: PMC2873970.

Zanghi CN, Lankes HA, Bradel-Tretheway B, Wegman J, Dewhurst S. A simple method for displaying recalcitrant proteins on the surface of bacteriophage lambda. *Nucleic Acids Res*. 2005 Oct 13;33(18):e160. doi: 10.1093/nar/gni158. PMID: 16224099; PMCID: PMC1258178.

Zeng L, Golding I. Following cell-fate in E. coli after infection by phage lambda. *J Vis Exp*. 2011 Oct 14;(56):e3363. doi: 10.3791/3363. PMID: 22025187; PMCID: PMC3227188.

A. Declaration of the use of Generative AI and AI tools

The word “AI” is vague and unhelpful; it can and has been used to refer to any kind of work done by a computer. Whether or not AI tools have been used depends upon the definition of AI, of which none were provided; hence, no truthful statement can be made on the usage of AI without a definition thereof.

But I must make statements. I have not used Generative AI; all written content in the thesis was generated by myself. Content in the included papers are collaborative works between all authors and, to the best of my knowledge, contain no generated AI content. Most figures were drawn by hand with marker and paper, and scanned into the computer. Scans were then vectorized using Inkscape via the Potrace algorithm. Phylogenetic trees were generated by Geneious 9.1.

Annotation programs prokka, pharokka and ISEScan all use hidden markov models to annotate genetic features. PhageBoost uses a machine learning classifier to predict prophage sequences.

ChatGPT, PubMed and Google Search were used in literature review to find relevant papers; all papers were thoroughly read by me before I decided to include them in this work.

The thesis was written in Microsoft Word 2013, which has a built in spelling and grammar check. Modern Microsoft Editor uses AI to check spelling and grammar; I can make no claims on the usage of AI in Microsoft Word 2013 for spelling and grammar checking.