

A FORWARD-GENETIC SCREEN IDENTIFIES NEW HOST FACTORS
INFLUENCING BACTERIOPHAGE LAMBDA'S LYSIS-LYSOGENY DECISION

A THESIS
SUBMITTED TO THE DEPARTMENT OF BIOENGINEERING
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
BACHELOR OF SCIENCE WITH HONORS

Nicolas Quach
April 2018

I certify that I have read this thesis and that, in my opinion, it is fully adequate in scope and quality as a thesis for the degree of Bachelor of Science.

(Markus Covert) Principal Adviser

I certify that I have read this thesis and that, in my opinion, it is fully adequate in scope and quality as a thesis for the degree of Bachelor of Science.

(Markus Covert)

I certify that I have read this thesis and that, in my opinion, it is fully adequate in scope and quality as a thesis for the degree of Bachelor of Science.

(Megan Brennan)

Approved by the Stanford University Bioengineering Undergraduate Curriculum Committee

Abstract

In order to survive, organisms must integrate internal and external signals via complex signaling pathways to intelligently navigate their environment. Bacteriophage lambda has long been a classic model for understanding such pathways, and how they allow cells to effect changes in physiology in order to adapt to varying circumstances. When infecting its host *Escherichia coli*, phage lambda may either make more viruses (lysis), or integrate itself into the host genome (lysogeny). To better understand the extent to which host physiology influences the lysis-lysogeny decision, I performed a forward genetic screen to identify genes whose deletions bias the infected cell towards lysis or lysogeny. This screen revealed numerous host-encoded genes linked to the lysis-lysogeny decision, including transcription factors and genes in metabolic pathways. These results demonstrate previously unknown links between host physiology and viral decision-making, shedding new light on this classic model system.

Acknowledgments

I would like to first thank David Van Valen for his help with experimental design, data collection, and figure creation, as well as for writing a number of Python analysis scripts and developing the Bayesian analysis to infer the lysis ratio and multiplicity of infection. I would also like to thank Katie Bodner and Yu Tanouchi (Covert Lab) for helping optimize the phage cloning protocol. Next, I would like to thank Amanda Miguel and Handuo Shi of the Huang lab for providing technical assistance with microscopy and SLIP. More generally, I must give special thanks to KC Huang for generously providing his lab's microscope for use in this library screen.

I would also like to especially thank my friends Tabitha Walker, Pujun Bhatnagar, and Amanda Spyropoulos for their support during the late nights and long microscopy sessions that inevitably occurred during this project, and my parents for their continual encouragement throughout this process.

Finally, both Markus Covert and David Van Valen have been amazing mentors and advisors to me for these past three years. Their guidance has been instrumental to this thesis, and more generally to my development as a scientist.

Contents

Abstract	iii
Acknowledgments	iv
1 Introduction	1
2 Background	3
2.1 The Ptashne model of the lysis-lysogeny decision	3
2.1.1 Overview of infection	3
2.1.2 Phage protein cII regulates the lysis-lysogeny decision	5
2.2 Genetically identical cells produce different cell fates	8
2.3 Host factors influence infection	13
2.3.1 Summary	16
3 A high-throughput screen for host factors	17
3.1 Creation of a phage-encoded cell-fate reporter: λ NQ009	17
3.1.1 Designing a phage-encoded lysis-lysogeny reporter	18
3.1.2 Bacteriophage engineering by yeast gap repair cloning	19
3.1.3 Construction of λ NQ009	20
3.2 Validation of λ NQ009	23
3.2.1 Phenotypic validation by live-cell imaging of infection events	23
3.2.2 λ NQ009 exhibits a normal lysis-lysogeny ratio	23
3.3 Development of a high-throughput lysis-lysogeny assay with single-cell resolution	25
3.3.1 A high-throughput procedure for infection, fixation, and imaging: the modified SLIP protocol	25
3.3.2 Image segmentation with deep learning	26
3.3.3 Image analysis and infection classification	27
3.3.4 Bayesian analysis to infer the probability of lysis	28
3.3.5 Bayesian analysis to infer the multiplicity of infection	29

3.4	Validation of the high throughput lysis-lysogeny assay	31
3.4.1	Validation by varying the multiplicity of infection	31
3.4.2	Validation by varying the carbon source	33
3.5	Screening the Keio Collection	34
4	New host factors affect lysis-lysogeny	35
4.1	Hits are consistent with previously known host-virus interactions	35
4.2	Comparing hits with genes identified in the Blasche and Maynard screen	41
4.3	New host factors whose deletion bias towards lysis	42
4.4	New host factors whose deletion bias towards lysogeny	43
4.5	Discussion	44
5	Conclusion	47
A	Supplemental Materials	50
A.1	Materials	50
A.1.1	Media	50
A.1.2	Primers and sequences	52
A.2	Methods	54
A.2.1	Phage fragment amplification	54
A.2.2	Colony PCR of λ NQ009 yeast transformants	55
A.2.3	Rebooting phage from yeast genomic DNA	56
A.2.4	Live cell microscopy to verify phage phenotype	56
A.2.5	High throughput infection and fixation of the Keio Collection	56
A.2.6	Modified SLIP sample preparation and imaging	57
A.2.7	Propagation of phage from single plaques via plate lysis	57
A.2.8	Population-level lysis-lysogeny assay	58
A.2.9	Constructing and training convolutional neural networks for image segmentation	58
A.2.10	Classifying cell-fate outcomes using a Gaussian mixture model	60
A.2.11	Measuring the effect of MOI on lysis ratio	60
A.2.12	Measuring the effect of carbon source on lysis ratio	61
A.3	Additional Statistical Analyses	61
A.3.1	Estimating the variance the lysis ratio in bulk lysis-lysogeny assays	61
A.4	Dataset Accession	62
Bibliography		63

List of Tables

1.1	Global and US burden of phage-mediated horizontal gene transfer	2
2.1	Previously known protein-protein interactions between lambda phage and its host	15
3.1	Primer and templates for constructing λ NQ009.	21
3.2	Size and homology length of DNA fragments for constructing λ NQ009.	22
3.3	Features of the DNA fragments for constructing λ NQ009.	22
3.4	Macroscopic plate assay data for λ NQ009 and λ borkan	24
4.1	Strains exhibiting a lytic-biasing phenotype.	38
4.2	Strains exhibiting a lysogeny-biasing phenotype.	39
A.1	Primer sequences for constructing λ NQ009. Bolded sequences indicate homology overhang arms	53
A.2	Primer sequences for verifying fragment junctions in λ NQ009.	54
A.3	Phage fragment thermocycling protocol	54
A.4	YAC fragment thermocycling protocol	55
A.5	Junction PCR thermocycling protocol	55
A.6	Junction PCR Primers	55
A.7	Architecture of the batch-normalized feature net 61x61 conv-net	59
A.8	Training sets for the GMM infection classifier	60

List of Figures

2.1	Phage lambda's genome	4
2.2	Establishment of lysogeny	6
2.3	Establishment of lysis	7
2.4	cII lifetimes	8
2.5	Phage lambda's regulatory network	9
2.6	Competing theories for cell-fate selection.	9
2.7	Cell-fate heterogeneity arises from noisy subcellular decision.	12
2.8	Genes required for lambda phage infection	14
3.1	Design of the phage-encoded cell-fate reporter: λ NQ009.	19
3.2	Phage engineering by yeast gap-repair cloning	20
3.3	Fragments for cloning λ NQ009	21
3.4	Timelapse montages of λ NQ009 infection	23
3.5	Bulk assay results	24
3.6	Summary of the image analysis pipeline	26
3.7	Lysis ratio decreases with increasing MOI	32
3.8	Carbon source affects the lysis-lysogeny decision	33
3.9	Histograms of the lysis ratio and MOI of knockout strains in the Keio Collection	34
4.1	Scatterplot of fluorescence data for lysis-biasing hits	36
4.2	Scatterplot of fluorescence data for lysogeny-biasing hits	37
4.3	Summary of lysis-biasing and lysogeny biasing hits	40

Chapter 1

Introduction

In all forms of life, the ability to adapt to changing circumstances is paramount for survival. Organisms must integrate information about their internal state as well as environmental signals in order to intelligently navigate their environment. To do so, these internal and external factors are incorporated into complex signaling and regulatory networks [1, 2]. Bacteriophage lambda has long been a classic model system for studying such regulatory pathways [2, 3]. During infection of its host *E. coli*, phage lambda may choose to undergo one of two life cycles: lysis or lysogeny [1]. In lysis, the virus takes over the host replication machinery and produces about 100 new viral particles before lysing the cell [1]. In lysogeny, the virus integrates its genome into the host chromosome and enters a dormant state, sparing its host [1]. How exactly lambda phage makes this crucial decision has been the subject of much study over the past 60 years [2].

More broadly, bacteriophages like phage lambda play an important role in bacterial ecosystems, and thus have under-appreciated impacts on human health and disease [4]. Indeed, studies have shown that the presence of bacteriophages in microbial communities facilitate the evolution of antibiotic resistance [5]. Several *in vitro* co-evolution studies have demonstrated that phage-infected bacteria develop a hypermutator phenotype, assisting the host organism in evolving antibiotic resistance [5, 6]. Furthermore, bacteriophages also serve as a reservoir for antibiotic resistance genes [7]. These genes may then be transferred to hosts via infection, facilitating the spreading of resistant phenotypes throughout the microbial community [8]. In addition to harboring antibiotic resistance genes, phages also encode genes that confer virulence to previously benign bacteria, such as *Staphylococcus aureus*, *Vibrio cholerae*, *Escherichia coli*, *Shigella dysenteriae*, *Corynebacterium diphtheriae*, *Pseudomonas aeruginosa*, *Enterococcus faecalis*, and *Clostridium botulinum* [9–13]. In particular, *V. cholerae* and *E. coli* have a strong connection with their phages; cholerae toxin is located on a prophage, while Shiga toxin in the O157:H7 strains of *E. coli* is encoded in the Stx phage [10–13]. Overall, phage-induced virulence as well as phage-mediated horizontal transfer of resistance genes

Table 1.1: Global and US burden of phage-mediated horizontal gene transfer

Global Burden	US Burden
Cholera: <ul style="list-style-type: none">• ~3-5 million infections/year• ~100,000-200,000 deaths/year	Antibiotic resistance: <ul style="list-style-type: none">• ~2 million infection/year• ~20,000 deaths/year
Shigella: <ul style="list-style-type: none">• ~165 million infections/year• ~1 million deaths/year	Shigella: <ul style="list-style-type: none">• ~300,000-400,000 infections/year• ~70-600 deaths/year
Diphtheria: <ul style="list-style-type: none">• ~30,000 infections/year• ~3,000 deaths/year	Pseudomonas: <ul style="list-style-type: none">• ~51,000 infections/year• ~400 deaths/year
	Enterotoxigenic <i>E. coli</i>: <ul style="list-style-type: none">• ~260,000 infections/year• ~30 deaths/year

produce an enormous US and global disease burden, accounting for millions of deaths every year (see Table 1.1) [9, 14–19]. It is clear that a better understanding of phage-bacteria interactions is needed to address this threat to global health.

Despite significant progress in our understanding of the mechanisms governing lambda phage’s lysis-lysogeny decision, there still remain a number of unanswered questions [1, 20]. In particular, the role of the host in influencing this crucial decision is ill-characterized [21]. Indeed, prior studies on the role of the host during infection are often limited by their inability to quantify phenotypes, preventing an in-depth look into the host-virus interaction [20].

With all viruses, infection is a complex interplay between the viral genetic circuitry and the host machinery that the virus must commandeer to replicate. As a result, a complete understanding of infection requires a thorough appreciation of both the virus and the host [20]. In this thesis, I explore the host-virus interaction between lambda phage and its host *E. coli* by uncovering new host-encoded factors influencing the lysis-lysogeny decision. To do so, I first engineered a strain of lambda phage capable of reporting the outcome of its decision at the resolution of single cells. Then, I used this strain to perform a forward genetic screen to identify the *E. coli* genes whose absence causes a perturbation in the ratio of lytic to lysogenic outcomes. From the results of this genome-wide screen, I glean new insights into how *E. coli* and phage lambda cooperate to produce an infection outcome.

Chapter 2

Background

2.1 The Ptashne model of the lysis-lysogeny decision

Over the years, a number of key discoveries have been made concerning the genetic circuitry controlling the lysis-lysogeny decision [1, 3]. From these findings, models for the cell-fate decision were created. Among the most famous of these models is one proposed by Mark Ptashne in 1986 [1]. Using numerous pieces of biochemical and biological evidence, Ptashne stitches together a compelling story for how phage lambda's regulatory networks give rise to its "genetic switch". Here, I present this model as a launching point from which to begin my exploration into lambda phage and its lysis-lysogeny decision.

2.1.1 Overview of infection

When phage lambda infects its host, the virus must choose between its two developmental pathways: lysis or lysogeny. At the core of this regulatory network is the competition of two regulatory proteins, Cro and CI, for binding to the overlapping pR and pRM promoter operator regions. If CI accumulates to significant levels to shut off pR activity, and therefore Cro production, lysogeny is established. Otherwise pR remains active and lysis ensues [1].

The genes driving the regulatory network that produce the two outcomes are generally classified into immediate-early, delayed-early, or late genes. Immediate-early genes are expressed as soon as phage DNA has entered the cell. It is thought that the virus has not yet committed to either lysis or lysogeny at this point. Once expressed, these immediate-early genes drive the expression of delayed-early genes which produce the regulators of lysis-lysogeny. Finally, once a decision has been made, the late genes are expressed to produce the desired developmental outcome [1, 23].

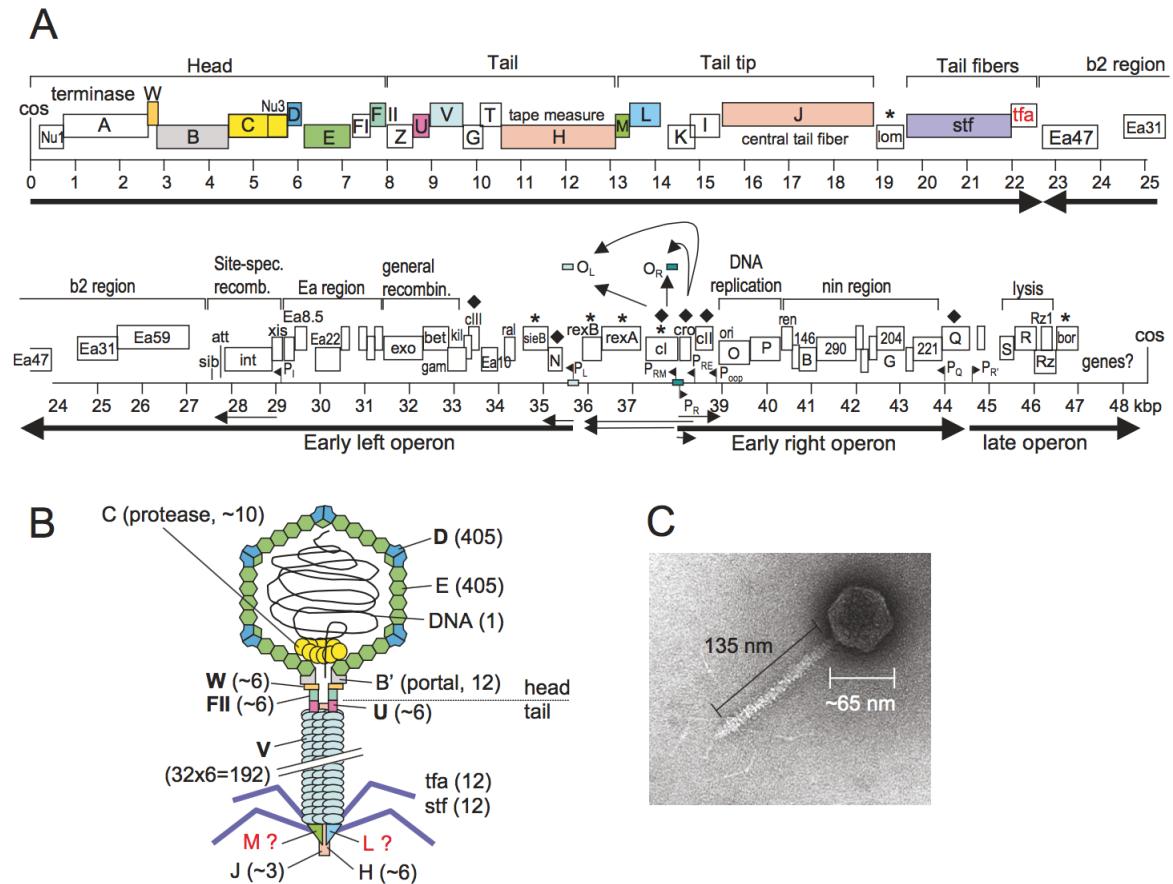


Figure 2.1: A) Diagram of phage lambda's genome in its linear form. B) Organization of phage lambda's capsid proteins C) Electron microscope image of phage lambda. Figure reproduced from reference [22].

2.1.2 Phage protein cII regulates the lysis-lysogeny decision

Upon entry into the cell, transcription of phage DNA begins via strong promoters pR and pL, producing regulatory proteins N and Cro¹ [1, 24]. At this point, transcriptional terminators tR1 and tL1 prevent host RNA polymerase from transcribing genes downstream of *N* and *cro*² [26, 27].

N protein acts as an antiterminator when bound with Nus host factors; when N accumulates to sufficient levels, transcription proceeds past tR1 and tL1 allowing the delayed-early genes to be expressed [28]. Among these genes are *cII* and *cIII* whose products are thought to be the key regulators of the lysis-lysogeny decision [1, 29].

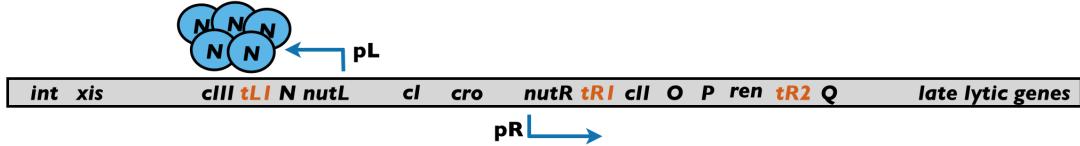
Current data suggests that when CII concentrations within the cell exceed a certain threshold, the operons responsible for lysogeny activate [30, 31]. In particular, CII acts as a transcriptional activator for promoters pRE, pI, and pAQ [1, 32, 33]. Upon activating pAQ, antisense mRNA is produced against the *Q* gene, switching off Q production and preventing expression of genes responsible for effecting lysis. Activation of pI results in high levels of protein Int, which integrates the phage genome into the host chromosome at particular sites known as *attB* sites [1, 32]. Finally, turning on pRE allows expression of the lambda repressor CI, which shuts down all other operons except for its own (pRM) and the operon controlling phage integrase (pI), establishing lysogeny [1] (see Figure 2.2 for graphical summary of the transcriptional cascade).

Although CII is required to establish lysogeny, it is degraded rapidly within the cell by host proteases FtsH (HflB), HflK and HflC. In wild type *E. coli*, CII may have a half-life as short as 1.5 mins at 37°C, enabling rapid fluctuations in its concentration [34]. This inherent *in vivo* instability stems from a C-terminal degradation tag recognized by host proteases HflC, HflK, and FtsH [35]. Strains containing deletion of *hflC* and *hflk* display increased CII lifetimes during infection, resulting in an increase frequency of lysogeny (see Figure 2.4) [36]. CII accumulation therefore requires another component: phage protein CIII, which inhibits these proteases [37]. According to the model of lambda phage infection proposed by Ptashne, should CII not accumulate to sufficient levels to activate pRE and establish lysogeny, phage lambda will default to lysis [1]. In this case, the other gene products expressed upon N antitermination—proteins O, P and Q—begin the cascade towards lysis [1]. Proteins O and P begin phage genome replication, while protein Q serves as an antiterminator allowing expression of the late operon, which contains genes encoding structural proteins used to generate new particles [1, 38]. Once new viral particles self-assemble, phage genes *S*, *R*, *Rz*, and *Rz1* are expressed to breach both the inner and outer cell membrane of the host, causing

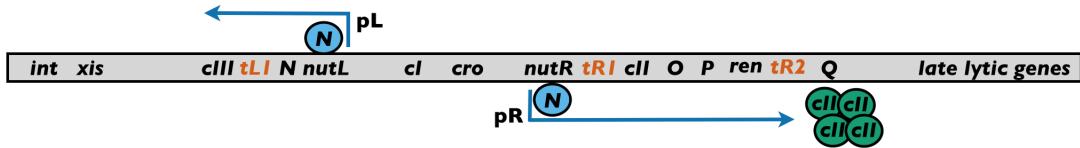
¹To differentiate proteins and genes with the same name, the following notation will herein be used: proteins will be denoted in plain script while genes encoding the proteins will be italicized (e.g. Cro = protein, *cro* = gene).

²Note that terminators tR1 and tL1 are not 100% effective, and allow some expression of delayed-early genes located downstream of *N* and *cro* [25].

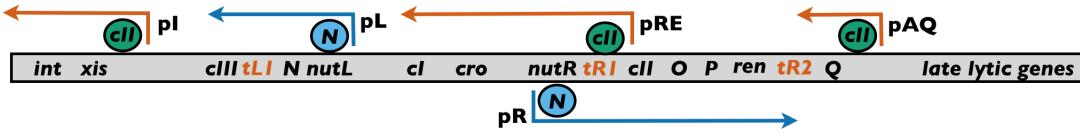
Early expression at pL and pR produces N:



Antitermination by N turns on cII:



cII expression activates pI, pRE, and pAQ:

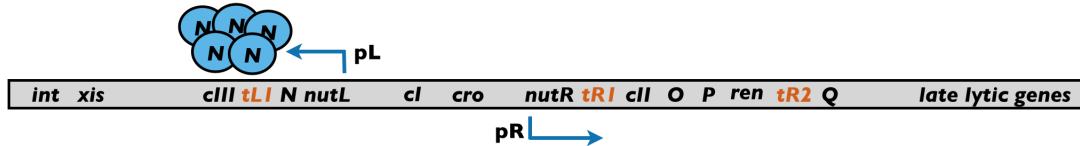


Expression at pI, pRE and pAQ induces lysogeny:

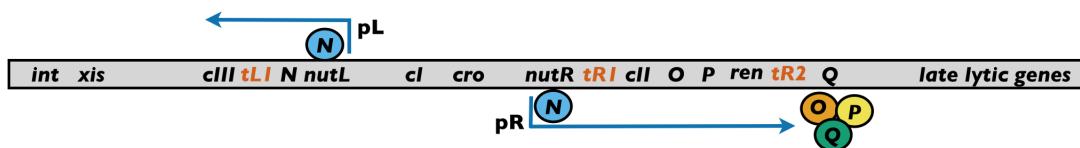


Figure 2.2: Transcriptional cascade for the establishment of lysogeny. According to the Ptashne model, N induces expression of CII, which accumulates to sufficient levels to induce production of CI, resulting in lysogeny [1].

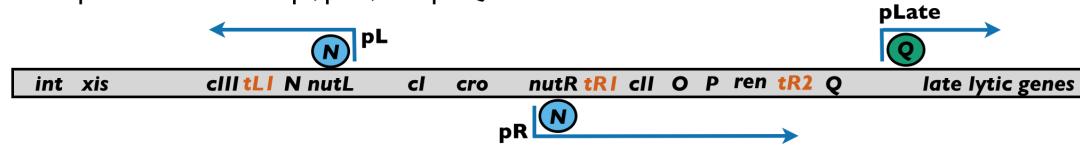
Early expression at pL and pR produces N:



Antitermination by N turns on O, P, and Q:



cII expression activates pI, pRE, and pAQ:



O and P initiate phage genome replication

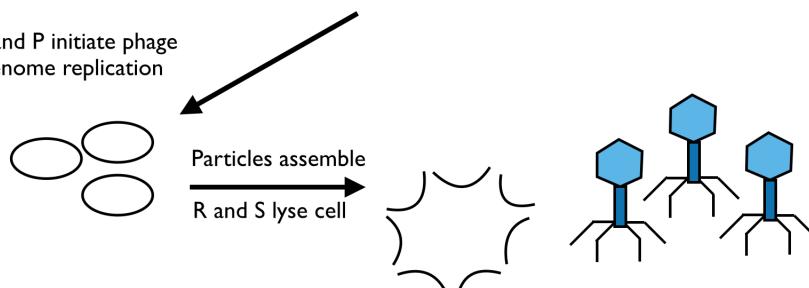


Figure 2.3: Transcriptional cascade towards lysis. According to the Ptashne model, should CII not accumulate to sufficient levels to induce lysogeny, antiterminator N turns on expression of lytic genes such as O, P and Q, which in turn pushes the phage towards a lytic outcome [1].

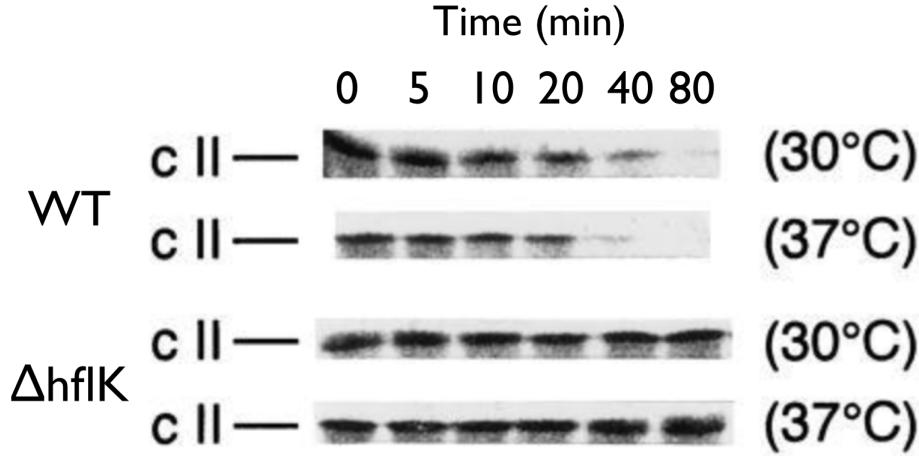


Figure 2.4: CII lifetimes in wild-type and *hflC* and *hflK* deficient strains. In the absence of host protease HflK, CII lifetimes increase significantly. Figure reproduced from reference [36].

new particles to be released into the environment to find a new host to infect (see Figure 2.3 for a graphical summary of the transcriptional cascade towards lysis) [1].

In summary, the Ptashne model of lambda's lysis-lysogeny decision places CII as the central “switchman,” where high CII concentrations produce lysogenic outcomes, while low CII concentrations default to lysis. In this way, CII would integrate multiple inputs into a single output from which a decision can be made [1, 37]. These inputs may come from the virus, the environment, or the host, as illustrated in Figure 2.5 [2]. Note that although there is evidence to support this model where phage lambda incorporates input signals to choose an outcome, it is unclear what phage lambda is sensing in the host. Known host-encoded inputs will be discussed in more detail in Section 2.3.

2.2 Genetically identical cells produce different cell fates

When genetically identical cells grown in identical conditions are infected by a single phage particle, different cell fates are still observed. In other words, at the single-cell level, phage lambda's decision appears “noisy” [39]. How exactly such variability in cell-fate arises remains one of the most perplexing questions of this host-virus interaction.

Such variability in cell-fate during infection of seemingly identical cells with single phages may arise from two possible sources: the viral regulatory network, or the host machinery being repurposed (see Figure 2.6). In terms of the former, one compelling hypothesis points to biochemical stochasticity

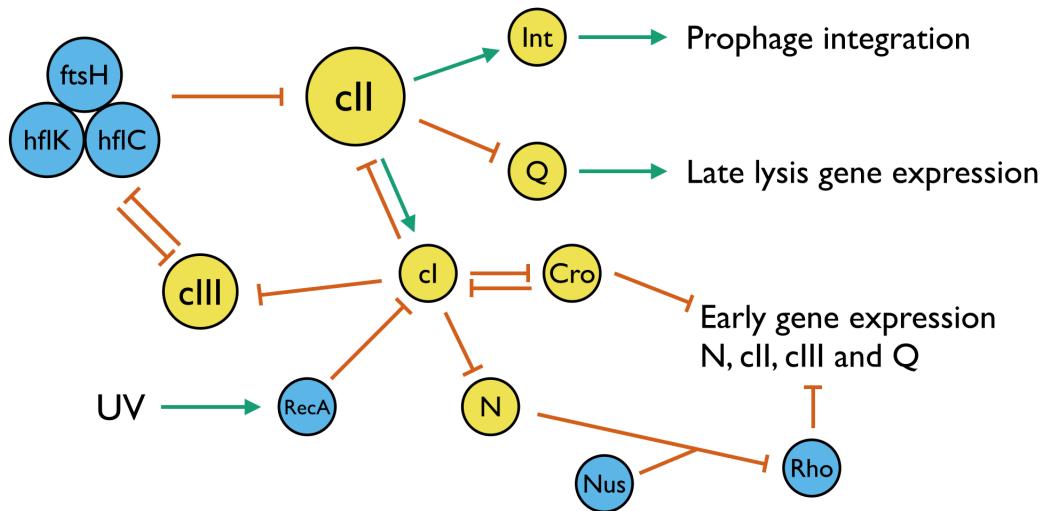


Figure 2.5: Main components of phage lambda's regulatory network according to the Ptashne model. In this model, CII is the “switchman” for the lysis-lysogeny decision, and all inputs are integrated at the level of CII expression to determine an outcome [1].

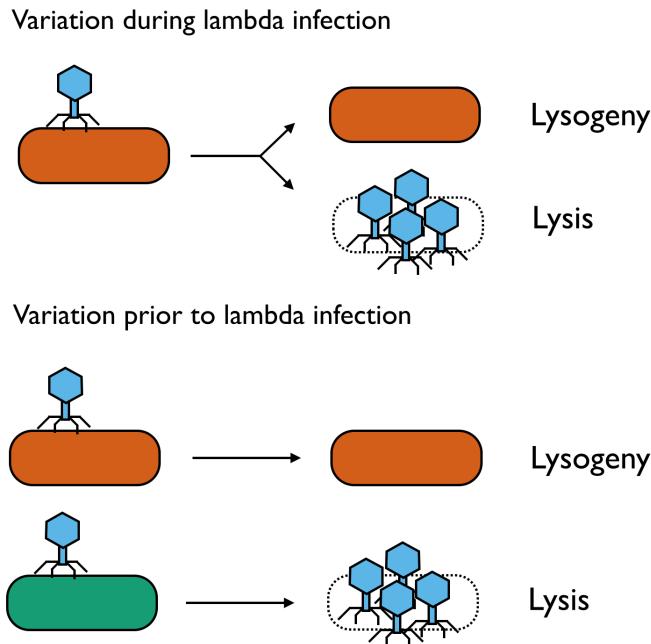


Figure 2.6: Variation in cell-fate outcome may result from variability during infection, or from undetected variations prior to infection.

in viral gene expression as the source of such differences in cell-developmental outcome in otherwise identical cells (see Figure 2.6). In this model, random fluctuations in the rates of the many biological processes that take place within a cell result in significant heterogeneity within a population [39].

In particular, fluctuations in the rates of gene expression can create erratic patterns of protein production in a cell [40]. Indeed, it has been shown that both transcription and translation occurs in a burst-like, non-Poissonian manner, resulting in significant variations in protein levels throughout time [41]. These random bursts of activity have striking consequences, as discovered by Taniguchi et al. during their measurements of the *E. coli* proteome and transcriptome. By tagging genes at their native locus with YFP, Taniguchi et al. were able to quantify the protein expression level while simultaneously performing RNA fluorescence in-situ hybridization (RNA-FISH) to quantify RNA copy numbers. From these measurements, they find that there is no correlation between mRNA copy number and protein levels [41]. Although this lack of correlation can be partially explained by the drastically different lifetimes of mRNA and protein within a cell, the biological noise introduced by bursts in transcription and translation can have significant effects on complex regulatory circuits [40, 41]. Indeed, they may even cause otherwise identical cells to display vastly different phenotypes [39], as Arkin et al. demonstrated with their numerical simulation of phage lambda's decision network.

Using a detailed stochastic chemical kinetics simulation, Arkin et al. showed that the outcome of the lysis-lysogeny decision may arise from spontaneous fluctuations of regulatory molecules [40]. Important to the model is the recognition that in cells, the abundances of key regulatory proteins lie far below the domain where the laws of continuous chemical kinetics are valid [40]. In this case, precise behavior can no longer be predicted from the bulk averaging of individual reactions, and developmental outcomes become non-deterministic [40].

In particular, the Arkin model predicts that random bursts in CIII production during the expression of delayed-early genes are responsible for inducing lysogeny. Indeed, if CIII levels do not attain a certain level, the phage regulatory network will fail to induce lysogeny, and instead, will default to the lytic pathway. In this way, identically infected cells accrue random differences that cascade through the phage genetic circuitry to produce distinct outcomes. Cell fates under this hypothesis are therefore completely due to stochastic fluctuations in gene expression [40].

Although the Arkin model provides a compelling argument for phage-side stochasticity as the main determinant of the lysis-lysogeny decision, a competing theory points to undetected differences between host cells within a seemingly homogeneous population as the source of the “noisy” cell-fate

decision [23, 42]. Under this “hidden variables” hypothesis, the lysis-lysogeny decision is fully deterministic, and if all such “hidden variables” were accounted for, this view holds that the cell-fate outcome for any particular cell can be predicted with full certainty (see Figure 2.6).

Evidence is mounting in favor for the existence of these “hidden variables.” For example, by making careful measurements of the cell-cell difference within a population of genetically identical cells, St. Pierre and Endy were able to account for some variability in the cell-fate decision [43]. Using counterflow centrifugal elutriation, St. Pierre separated cells by size, then infected cells of a specific size range with an engineered strain of lambda phage to track infected cells [43]. By following the infection via live-cell microscopy, cell-fate statistics were measured with single cell resolution [43]. From this data, St. Pierre and Endy showed that a two-fold increase in cell volume at the time of infection resulted in a 4-5 fold decrease in the probability of lysogeny [43]. Thus, cell size appears to be one such “hidden variable.” Perturbations in host genes responsible for cell morphology are therefore expected to produce perturbations in the lysis-lysogeny decision.

Building upon this result, Zeng et al. provide a more nuanced view on these parallel hypotheses [39]. They propose that the inherent stochasticity of chemical reactions may play a smaller role than previously thought, and that observed cell-fate heterogeneity arises from noise originating from the level of individual phages [39]. To reach this conclusion, Zeng et al. followed lambda phage infection at the single cell and single phage level, using a fluorescently labeled phage and a plasmid-based lysogeny reporter driven by the promoter pRE. With this setup, individual phages infecting individual cells can be tracked via fluorescence microscopy [39]. From the images, Zeng et al. were able to quantify the contribution of subcellular parameters—such as cell volume, multiplicity of infection and location of infection—to the observed cell-fate heterogeneity [39]. Their results indicate that heterogeneity in phenotype arises at the transition from the single-phage level to the single-cell level [39]. In other words, it is a cascade of “noisy” subcellular decisions, such as individual phages choosing a developmental pathway, that are integrated deterministically to produce a whole cell phenotype (see Figure 2.7) [39].

In a recent follow up experiment, Trinh et al. used a four-color fluorescent reporter system to investigate how exactly these individual phage decisions are integrated into a whole cell phenotype [44]. Most interestingly, they find that during lysogeny, phages infecting the same cell cooperate to produce the whole cell phenotype, while in lysis, phages compete for replication resources [44]. DNA replication therefore plays an important role for lysis, since the phage that replicate its genome the fastest will win the larger share of progeny upon cell lysis [44]. Thus, interactions between multiple phages also play an important role in the lysis-lysogeny decision.

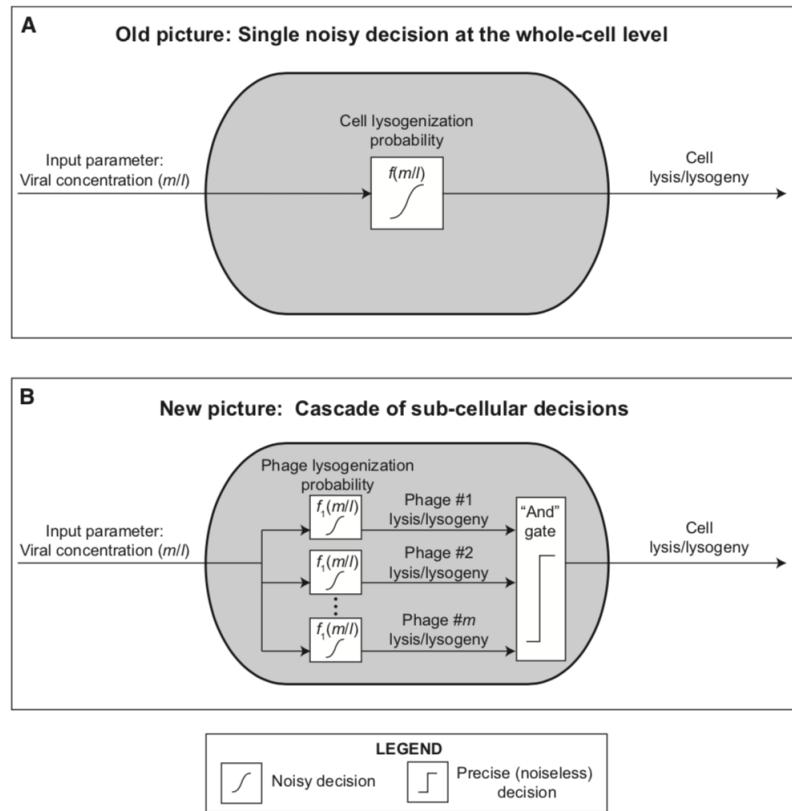


Figure 2.7: Cell-fate heterogeneity arises from the noise-free integration of noisy, subcellular decision. Inputs are integrated at the level of individual phages, and the decision of each individual phage is summed in a noise-free operation to produce a cell-fate outcome. Figure reproduced from reference [39].

In all, it is likely that the true origin of cell-fate heterogeneity in lambda phage infection lies within a complex interplay between host-side and phage-side stochastic gene expression. In other words, infection outcomes may indeed be non-deterministic as the Arkin model predicts, but “hidden variables” may also play a significant role in biasing the decision towards one outcome over the other [23].

2.3 Host factors influence infection

Over the years, a number of host factors have been linked to lambda phage infection and its lysis-lysogeny decision [20]. Among the most well-known of these factors are the *E. coli* proteases HflC, HflK, and HflD [45]. In the cell, HflC and HflK binds with ATP-dependent metallopeptidase FtsH to form a proteolytic complex that degrades cII, facilitating lytic outcomes in lambda infection [45]. HflD plays a similar role, though it directly binds with cII and inhibits its DNA binding ability [46]. Another set of host factors involved in infection are the subunits of Integration Host Factor (IHF), encoded by *E. coli* genes *ihfA* and *ihfB* [47]. IHF forms sharp bends in DNA required for the phage genome to be spliced into the host chromosome [47]. IHF is also thought to play a role in phage DNA maturation during lysis [48]. Other important host factors for transcription and replication include DNA chaperone DnaJ which helps unwind phage DNA, and nusB which helps form an antitermination complex with phage protein N [49, 50].

Another important set of host-encoded proteins essential for lambda phage infection are those involved in the initial stages of infection: injection of phage DNA into the cell. To infect its host, lambda phage must bind to maltose-porin protein LamB located on the outer membrane [51]. Once bound, phage DNA passes through the pore into the intermembrane space, where it encounters inner membrane maltose PTS permease³ (PtsM) to be transported into the cytoplasm [52]. The expression of *lamB* is controlled by a number of transcription factors, namely *malI*, *malT*, and *cyaA* [20]. In addition to these transcription factors, genes involved in lipopolysaccharide (LPS) synthesis, like *gmhA*, *gmhB*, *hldD*, *hldE*, *rfaC*, *rfaD*, *rfaH*, and *rfaP*, have also been shown to affect LamB activity and thus infectivity by lambda phage [20, 53].

In an effort to uncover new host factors involved in lambda phage infection, Maynard et al. performed a screen of the “Keio Collection,” a library of 3,985 *E. coli* strains, each containing an in-frame single-gene knockout [20, 54]. Here, the authors sought to identify genes whose absence significantly reduced the infectivity of lambda phage by looking for gene knockouts displaying impaired plaque formation [20]. In all, Maynard et al. found 57 genes required for efficient lambda phage infection, of which 19 were previously reported and discussed in the previous paragraph [20]. The rest of the 38 genes uncovered by the genetic screen are involved in a number of cellular processes,

³PtsM is encoded by the gene *manZ*

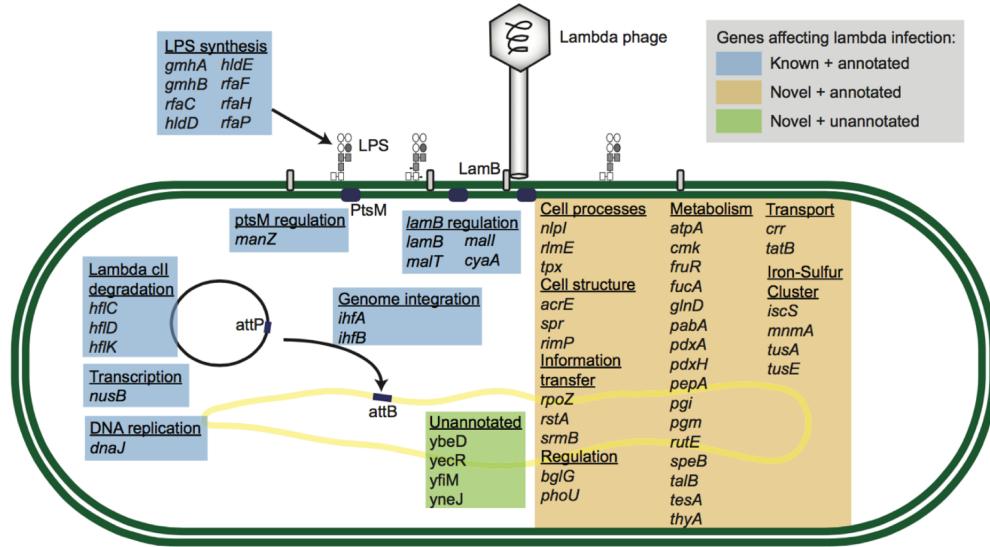


Figure 2.8: Genes required for effective lambda phage infection, identified by a forward genetic screen of the Keio collection by Maynard et al. Figure reproduced from reference [20].

such as central metabolism, structure, and regulation [20]. The results of the Maynard screen are summarized in Figure 2.8. In this thesis, I also screened the Keio Collection, this time looking for genes whose absence causes a perturbation in the lysis-lysogeny decision.

Although the Maynard screen greatly advances our understanding of lambda phage infection, its scope is still not wide enough to capture the full complexity of the host-virus interactions. For one, the use of plaque morphology reduces the screen's ability to draw strong mechanistic conclusions, since plaque morphology is neither a precise nor quantitative phenotype [20]. Furthermore, the Maynard screen cannot detect genes that influence the host-virus interaction but do not produce a change in plaque morphology or growth curve dynamics [20]. In an attempt to surpass these shortcomings, Blasche et al. performed a yeast two-hybrid screen to identify protein-protein interactions (PPI's) between the host and the virus [55]. In this screen, the authors screened 3974 *E. coli* open reading frames⁴ (ORFs) against 68 lambda ORFs. Altogether, Blasche et al. found 62 new interactions (two of which are part of the 31 previously published host-virus PPIs, summarized here in Table 2.1) between host factors and viral proteins [55].

From these 62 new interactions, Blasche et al. observed that lambda phage proteins preferentially interact with host proteins centrally connected to a complex host-side network of PPIs. This suggests that these host-virus PPIs allow the virus to commandeer the cell at various levels of cellular organization [55].

⁴This corresponds to about 94% of all ORFs in the *E. coli* genome [55].

Table 2.1: Previously known protein-protein interactions between lambda phage and its host *E. coli*. Figure reproduced from reference [55].

Function	λ	Host	Description
Transcription	CI	RecA	RecA degrades CI
Transcription	CI	RpoA	
Transcription	CI	RpoD	
Transcription	CII	ClpYQ	ClpYQ degrades CII <i>in vitro</i>
Transcription	CII	ClpAP	ClpAP degrades CII <i>in vitro</i>
Transcription	CII	HflD	HflD makes CII more vulnerable to FtsH
Transcription	CII	HflB	HflB protease degrades CII
Transcription	CII	RpoA	
Transcription	CII	RpoD	
Transcription	CIII	HflB	CIII inhibits HflB; HflB degrades CIII
Transcription	gpN	NusA	Transcriptional regulation
Transcription	gpN	Lon	Lon degrades gpN
Transcription	gpQ	σ^{70}	gpQ makes RNAP insensitive to <i>cis</i> termination
Head	gpB	GroE	Genetic interaction
Head	gpE	GroE	Genetic interaction
Tail	gpJ	LamB	LamB is the <i>E. coli</i> receptor
Tail	Stf	OmpC	OmpC is a secondary <i>E. coli</i> receptor
Recombination	Xis	Lon	Xis is degraded by <i>E. coli</i> Lon protease
Recombination	Xis	FtsH	Xis is degraded by <i>E. coli</i> FtsH protease
Recombination	Xis	Fis	Both required for excision
Recombination	Int	IHF	Both catalyze recombination at attP/attB
Recombination	Gam	SbcC	SbcCD is a dsDNA exonuclease + ssDNA endonuclease
Recombination	Gam	RecB	Gam inhibits RecBCD
Recombination	NinB	SSB	NinB also binds ssDNA
Recombination	Ral	Hsd	Ral inhibits restriction enzyme complex HsdMSR (by binding HsdM or S)
Replication	gpO	ClpXP	ClpXP degrades gpO
Replication	gpO	DnaK	
Replication	gpO	RpoB	
Replication	gpP	DnaA	
Replication	gpP	DnaB	
Replication	gpP	DnaK	

Although the Blasche screen is, in some ways, more comprehensive than the Maynard screen, it too suffers from certain shortcomings. For one, the screen failed to find the majority of the previously published host-viral PPI's, which the authors suggest may be a result of the weak nature of these PPI's [55]. Indeed, host proteases and their phage substrates interact weakly and transiently, and thus may not be detected using a yeast two-hybrid screen approach [55]. Additionally, certain interactions may require multiple proteins to stabilize a quaternary structure or may require other cofactors/coenzymes to function—conditions that were not present in the screen [55].

In all, the combined knowledge gained from the Maynard and Blasche screens provides a richer understanding of the complex interaction between lambda phage and its host *E. coli*. The hits from these screens will be compared later with the results of my screen of the Keio collection, described in more detail in the next chapter.

2.3.1 Summary

In this chapter, I have summarized the Ptashne model for lambda phage infection, from viral DNA injection to the lysis-lysogeny decision. In this model, the concentration of phage protein cII during infection is thought to play a key role in determining the outcome of the cell-fate decision, where high concentrations of cII produce lysogenic outcomes, while low concentrations of cII induce lysis. Although the model is backed by considerable biochemical evidence, it remains incomplete. For one, it cannot account for the observation that genetically identical cells infected by a single phage can produce disparate outcomes. Indeed, there is no consensus as to the true nature of the lysis-lysogeny decision—whether it is inherently stochastic in nature or whether so-called “hidden variables” control the cell-fate in a deterministic fashion. In the next chapter, I will describe my efforts towards uncovering new host factors that mediate lambda phage's lysis-lysogeny decision.

Chapter 3

Development of a high-throughput screen for host factors involved in the lysis-lysogeny decision

3.1 Creation of a phage-encoded cell-fate reporter: λ NQ009

The problem of detecting shifts in cell-fate statistics has been considered since the beginning of the study of lambda phage and its life cycle [1,3,23]. Several solutions have been proposed and executed, but each suffers from certain drawbacks. Classical measurements have been made at a population level. Such bulk assays generally comprise of infecting a number of cells, diluting the infection mixture, then using various plating techniques to visualize the number of lytic and lysogenic events that transpired [23]. Although relatively easy to conduct, I found this approach to be inaccurate and to have issues with reproducibility. Another way of detecting perturbations in the lysis-lysogeny decision is to use plaque morphology as an indicator of the cell-fate distribution. Again, although these methods are straightforward to conduct, plaque morphology is not a quantitative phenotype. In order to accurately measure the cell-fate statistics, it is necessary to collect data at the single cell level [39]. To accomplish this measurement, a lysis-lysogeny reporter must be encoded either on the host or the phage. To efficiently screen a large knock-out library such as the Keio Collection, a phage-encoded reporter is the best solution. In this section, I detail the design and creation of a phage-encoded lysis-lysogeny: λ NQ009.

3.1.1 Designing a phage-encoded lysis-lysogeny reporter

In order to create an optimal phage-encoded lysis-lysogeny reporter, there are several design requirements that must be met. For one, a reporter with high temporal resolution is preferred, since lysogenic cells will continue to divide after a decision is made. A reporter that indicates a lysogenic outcome must ideally report the result before the cell divides; otherwise lysogeny counts will be artificially inflated. As for lytic counts, accurate measurements require that cells not burst if the lytic life cycle is chosen. By preventing bursting, cells choosing lysis can be imaged and analyzed at any time after the cell-fate decision has been made. Finally, the reporter must be carefully designed to have optimal signal-to-noise in order to correctly assign cell-fate outcomes to individual cells.

In order to meet these three design criteria, several engineering choices were made. To achieve high temporal resolution, I used the fluorescent proteins mKate2 and mNeonGreen to report on cell-fate outcomes because of their spectral compatibility and fast maturation times [56, 57]. Next, to prevent lytic cells from bursting, I introduced an amber mutation into the *S* gene, encoding the phage holin responsible for cell membrane lysis and release of viral particles [1]. When infecting non-amber suppressing host strains, infected cells choosing lysis will produce new, infectious particles, but will not burst. Finally, to optimize the signal-to-noise ratio on these reporters, I had to assign mKate2 and mNeonGreen to a particular cell-fate outcome. During my first attempt at producing a phage-encoded lysis-lysogeny reporter system, an important insight was revealed about the importance of copy number and its relationship with the strength of the reporter signal. During lysogeny, approximately one copy of the phage genome exists within the cell, whereas during lysis there may be up to 100 copies in the cytoplasm [1]. As a result, the lysogeny reporter signal will inherently be weaker than that of the lysis reporter. Thus, to improve signal detection for the lysogeny reporter, mKate2 was used instead of mNeonGreen, since its signal will not be confounded by cellular autofluorescence [56].

Several other design choices were made to enhance the ease of phage manipulation and phage cloning. For one, the *bor* gene was partially deleted and replaced with a cassette conferring kanamycin resistance, allowing lysogens to be easily selected for. The *bor* gene is not thought to be involved in the lysis-lysogeny decision, and its deletion has been shown not to disrupt the phage regulatory network [23]. Along with *bor* replacement, the cI857ts allele was included, rendering the lambda repressor temperature sensitive; at temperatures above 40°C, cI875ts misfolds and phage lambda produces only lytic outcomes. At lower temperatures, this temperature sensitive repressor behaves like wild-type, producing the normal mix of lytic and lysogenic outcomes [23].

Embedding these design choices into a single phage, I created the reporter phage for my library screen: λ NQ009. To report on lysis, I inserted a cassette encoding mNeonGreen driven by the strong

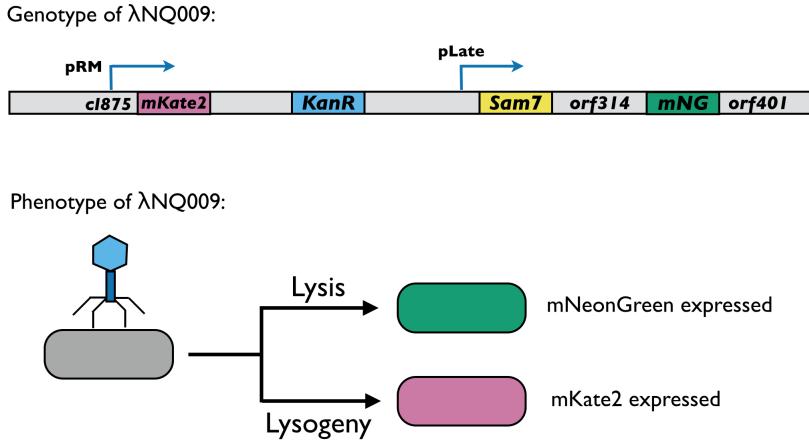


Figure 3.1: Design of the phage-encoded cell-fate reporter: λ NQ009. Upon lysis, cells exhibit green fluorescence due to mNeonGreen expression, while in lysogeny, red fluorescence is observed due to mKate2 expression.

ribosome binding site (RBS) BBa_B0034 into the region between *orf401* and *orf314* (part of the late lytic transcript) [57–59]. As for lysogeny, I placed a cassette expressing mKate2 also driven by BBa_B0034 immediately downstream of *cI*, separated by a 15 bp spacer sequence (catatcggtcacgaa) derived from the untranslated region of the lambda genome between genes *S* and *orf64* [56] [58,59]. Thus, if λ NQ009 establishes lysogeny in an infection, mKate2 is cotranscriptionally expressed with *cI*, and thus lysogens will fluoresce red. If instead λ NQ009 chooses lysis, mNeonGreen will be expressed along with the other late lytic genes that produce the lytic phenotype. The design of λ NQ009 is summarized in Figure 3.1.

3.1.2 Bacteriophage engineering by yeast gap repair cloning

Over the past ten years, a number of methods have been devised to modify bacteriophage genomes [60]. In particular, recombineering of prophages has become the method of choice for engineering lambda phage [61]. First applied in mycobacteriophages, recombineering involves electroporating double-stranded DNA containing the desired change and homology to the region of interest into the prophage to be edited, along with a plasmid encoding RecE/RecT-like proteins that promote recombination [60,61]. After electroporation, cells are recovered and recombinants containing the desired edit are selected. The prophage is then induced into the lytic cycle, generating new recombinant viral particles [61].

In this thesis, I opted to use an alternate strategy based off yeast homologous recombination. Originally developed by Ando et al. to engineer the slightly shorter T7 phage genome, this strategy was

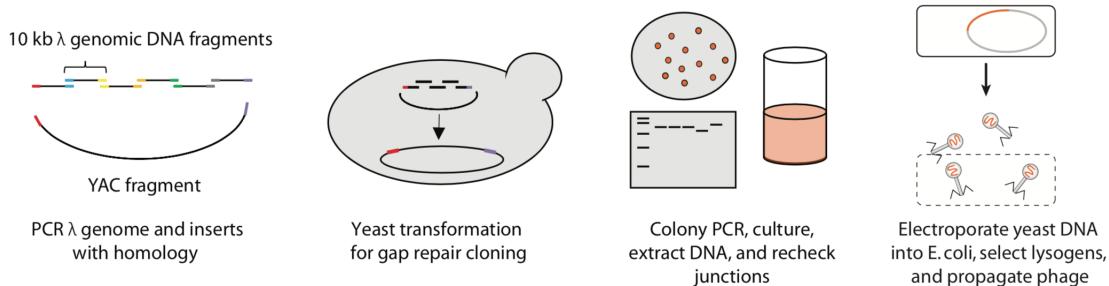


Figure 3.2: Work flow for lambda phage engineering by yeast gap-repair cloning. The phage genome is first copied in roughly 10kb fragments, and the fragments transformed into *S. cerevisiae* to subclone the fragments into a yeast artificial chromosome. Yeast colonies are then checked for the presence of the cloned construct. Next, the engineered phage genome is extracted and electroporated into *E. coli* to produce phage particles containing the new construct.

recently adapted by the Covert lab for phage lambda. This method presents three main advantages over recombineering: 1) multiple genes can be inserted “scarlessly” 2) the use of selection markers is not necessary, and 3) the efficiency of recombination is significantly higher¹ [62]. In this approach, the entire lambda phage genome is amplified in approximately 10kb fragments containing homology arms. These fragments along with DNA inserts encoding the desired changes and a linearized yeast artificial chromosome plasmid (YAC) are then transformed into *Saccharomyces cerevisiae* and the pieces stitched together using the yeast native homologous recombination system. Yeast colonies are then screened for the correct construct using polymerase chain reaction (PCR), then the yeast genomic DNA is extracted and electroporated into *E. coli*. Once inside bacteria, the engineered phage genome “reboots” and new phage particles can be produced and isolated (Figure 3.2) [60, 62].

3.1.3 Construction of λNQ009

The first stage of engineering λNQ009 consisted of amplifying the phage genome in multiple fragments and copying the fluorescent reporter cassettes off the appropriate templates (see Figure 3.3). Templates and primers for each of the nine fragments are listed in Tables 3.1. Primer sequences can be found in Supplemental Materials, in Table A.1, with homology overhangs highlighted in bold, should they exist. Over the course of several phage cloning iterations, the length of the homology arm overhangs were optimized empirically. I found that for efficient recombination, primers for fragment amplification should have a minimum of 30 bp of overhang containing homology to the adjacent fragment. Thus overall, there should be a total of at least 60 bp of homology at the joining ends of two adjacent fragments. As for the features of each DNA fragment, Tables 3.2 and 3.3 summarize the important characteristics of each fragment. For all fragments, Phusion polymerase

¹We have been able to achieve upwards of 50% cloning efficiency.

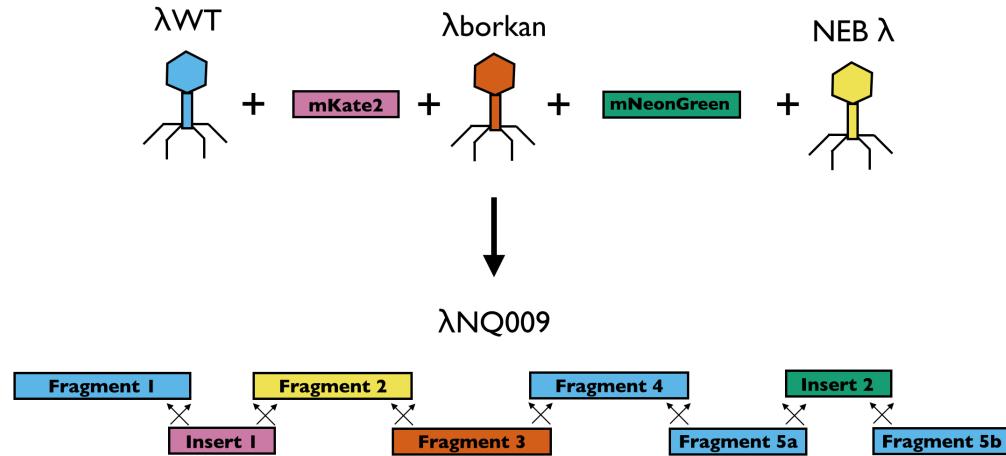


Figure 3.3: Fragments for cloning λ NQ009. Fragments from wild-type λ phage, λ borkan and genomic λ phage DNA from New England Biolabs were incorporated to create the phage-encoded cell-fate reporter.

Table 3.1: Primer and templates for constructing λ NQ009.

Fragment	Template	Primers
1	BW25113: λ WT	pRS415-R-BW802721-F/mKate2-hom:Frag2a-R
I1	MG1655: λ KB003	Frag2b-hom:mKate2-F/Frag2a-hom:mKate2-R
2	NEB λ	mKate2-hom:Frag2b-F/L46231-46260-R L46149-46178-F/L7841-7870-R
3	BW25113: λ borkan	L7791-7820-F/L17811-17840-R
4	BW25113: λ WT	L17761-17790-F/mNG-hom:Frag5a-R
5a	BW25113: λ WT	Frag5a-hom:mNG-F/Frag5b-hom:mNG-R
I2	pD441-SR-BBa_M50003	mNG-hom:Frag5b-F/pRS415-F-BW802845-R
5b	BW25113: λ WT	pRS415-F60/pRS415-R60
YAC	pRS415	

(New England Biolabs) was used to amplify DNA because it has an exceptionally low error rate², and thus is unlikely to introduce unwanted mutations. All PCR amplification of the phage fragments were performed according to the protocol specified in section A.2.1 in Supplemental Materials.

Once the phage fragments, fluorescent reporter cassettes, and YAC fragment were amplified, at least 1 μ g of each of the DNA pieces was transformed into yeast strain BY4714 using published methods [64]. Cells containing the spliced genome were selected by plating on Synthetic Defined media deficient in leucine (SD-Leu) plates³ and grown at 30°C for 3 days. Colonies were then picked

²Phusion polymerase has an error rate of $4.4 - 9.5 \times 10^{-7}$ errors/bp depending on the buffer used [63]. Overall, since λ NQ009's genome is 47,244 bp long, the number of expected mutations introduced by Phusion is approximately 0.02.

³If splicing occurs with the YAC fragment, the transformant will gain the ability to synthesize leucine, and thus survive on the leucine deficient SD-Leu plate.

Table 3.2: Size and homology length of DNA fragments for constructing λ NQ009.

Fragment	Fragment Size (bp)	5' Homology Size (bp)	3' Homology Size (bp)
1	9626	60 to Fragment YAC	60 to Fragment I1
I1	786	60 to Fragment 1	75 to Fragment 2
2	9079	75 to Fragment I1	112 to Fragment 3
3	10224	112 to Fragment 2	80 to Fragment 4
4	10050	80 to Fragment 3	80 to Fragment 5a
5a	3173	80 to Fragment 4	60 to Fragment I2
I2	796	60 to Fragment 5a	60 to Fragment 5b
5b	4157	60 to Fragment I2	60 to Fragment YAC
YAC	3031	60 to Fragment 5b	60 to Fragment 1

Table 3.3: Features of the DNA fragments for constructing λ NQ009.

Fragment	Notable Features
1	None
I1	mKate2, 15bp spacer sequence
2	cI857, Sam7
3	KanR cassette, bor deletion
4	None
5a	None
I2	mNeonGreen
5b	None

and checked via colony PCR (detailed in section A.2.2) for presence of the nine expected junctions between the six phage fragments, two fluorescent reporter cassettes, and YAC fragment. Once all nine junctions were confirmed by PCR, yeast colonies containing the spliced phage genome were propagated in 20mL cultures of SD-Leu media. After three days of culture at 30°C, yeast genomic DNA was isolated using YeaStar Genomic DNA kit (Zymo). The genomic DNA preparations were then checked again for all nine junctions via PCR to ensure that no unwanted recombination occurred during propagation.

Once the junctions were again confirmed, the yeast genomic DNA was electroporated into cells and the phage rebooted according to the protocol in section A.2.3. The engineered phage was then propagated via plate lysis (detailed in section A.2.7). The phage stock was then titered using existing protocols and stored at 4°C [65].

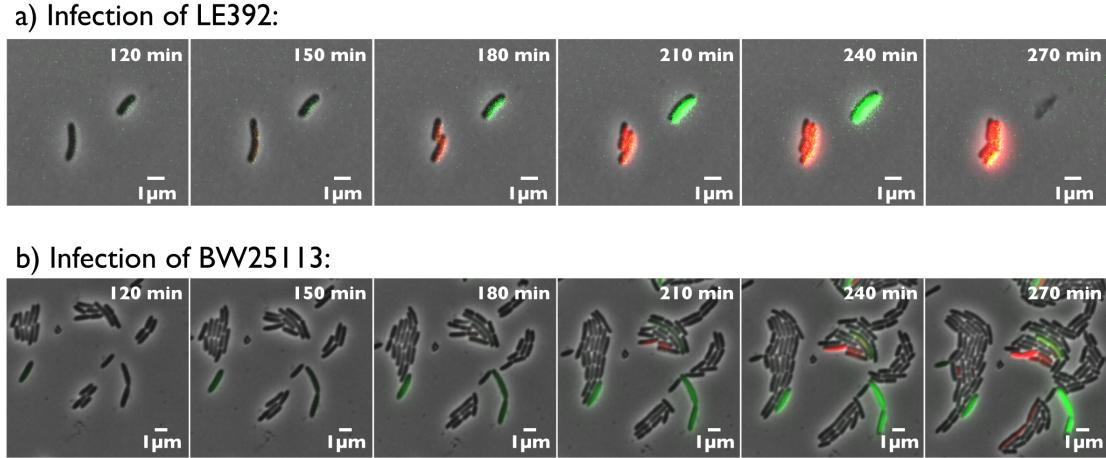


Figure 3.4: a) Timelapse of λ NQ009 infecting LE392. Lytic cells display increasing green fluorescence followed by lysis, while lysogenic cells fluoresce red and continue to divide. b) Timelapse of λ NQ009 infecting BW25113. Lytic cells display increasing green fluorescence without subsequent lysis, indicating the presence of the amber mutation in the S holin gene. Lysogenic cells continue to display ramping red fluorescence and divide normally post-infection.

3.2 Validation of λ NQ009

3.2.1 Phenotypic validation by live-cell imaging of infection events

In order to verify that λ NQ009 displayed the expected phenotype, I imaged live infection of the amber suppressing strain LE392 and the non-suppressor strain BW25113 with the engineered phage via wide field epifluorescence microscopy (see protocol in section A.2.4). Infection of LE392 led to lytic cells displaying increasing green fluorescence until they lyse, while cells undergoing lysogeny began to fluoresce red, confirming that the reporter cassettes function as expected (Figure 3.4a). Infection of BW25113 displayed similar fluorescence dynamics, but no cells lyse, confirming the presence of the amber mutation in the S gene in λ NQ009 (Figure 3.4b).

3.2.2 λ NQ009 exhibits a normal lysis-lysogeny ratio

In order to verify that the alterations made in λ NQ009 have not disrupted the phage lysis-lysogeny circuits, I performed a population-level lysis-lysogeny assay on λ NQ009 and on λ borkan, based off of the macroscopic plate methods describe by St. Pierre in his 2009 thesis [23]. In this assay, a liquid culture of the amber-suppressing host strain LE392 grown in M9 minimal media supplemented with 0.4% maltose (M9-maltose) was infected with phage at a certain MOI, and the infection allowed to proceed until cell-fate decisions have been made. The infected culture was then split into three subsamples, where either the number of lytic, lysogenic, or total infection events present were tallied via

Table 3.4: Macroscopic plate assay data for λ NQ009 and λ borkan

Phage	Lytic Events ($10^3/\mu\text{L}$)	Lysogenic Events ($10^3/\mu\text{L}$)	Total Infection ($10^3/\mu\text{L}$)
λ borkan	2958 ± 434	249 ± 34	4100 ± 90
λ NQ009	2350 ± 368	72 ± 14	3117 ± 169

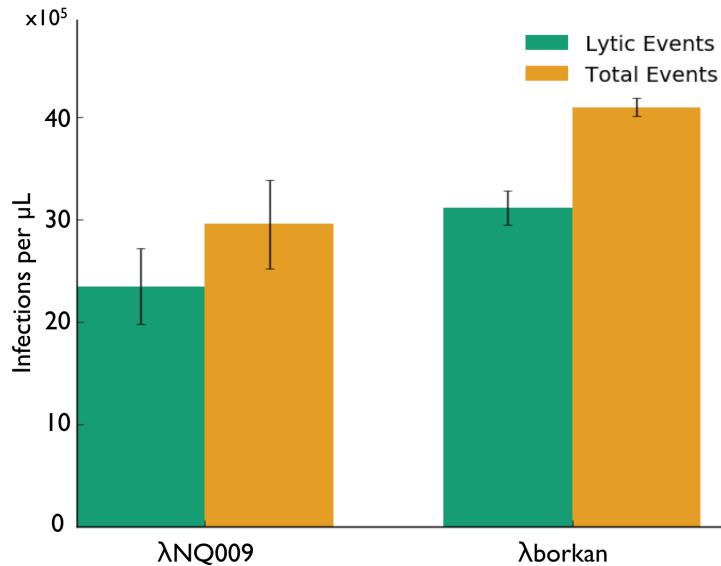


Figure 3.5: Population-level lysis-lysogeny assay for LE392 infected by λ borkan and λ NQ009 at low MOI in M9-maltose. Both phages exhibit similar lysis ratios, and thus we can conclude that the regulatory networks in λ NQ009 are not significantly perturbed. Error bars represent the standard error.

plate counts of serial dilutions. From the plate counts, the population cell-fate outcome distribution could be roughly estimated (detailed protocol described Supplemental Materials in Section A.2.8). The resulting data is plotted in Figure 3.5, and displayed in Table 3.4. The lysis ratios (defined here as $\frac{\text{lytic}}{\text{total}}$ ⁴) of the engineered phage and the wild-type analog λ borkan were then compared. For λ NQ009, the calculated lysis ratio was 0.7539 ± 0.0109 ⁵, while for λ borkan, I obtained a ratio of 0.7215 ± 0.0115 . From these ratios, I therefore conclude that λ NQ009's lysis-lysogeny network is not perturbed significantly compared to wild-type.

⁴This definition of lysis ratio was chosen over $\frac{\text{lytic}}{\text{lytic} + \text{lysogenic}}$, due to inconsistency of growing and counting lysogens in replicate experiments. Lysogen growth may be impaired in this assay.

⁵See Section A.3.1 for the derivation of the standard error for the lysis ratio obtained in the bulk assay

3.3 Development of a high-throughput lysis-lysogeny assay with single-cell resolution

3.3.1 A high-throughput procedure for infection, fixation, and imaging: the modified SLIP protocol

The functional phage-encoded lysis-lysogeny reporter allows us to measure the lysis ratio with single cell resolution. To uncover new host factors, I measured the lysis ratios of each strain in the entire Keio collection, a comprehensive library containing in-frame single gene knockouts of all non-essential genes in *E. coli* [54]. To do so, I developed a standardized protocol for infecting strains in 96-well plates, and preserving the infection mixtures for future imaging.

Knockout strains were first inoculated from glycerol stocks into $200\mu\text{L}$ LB supplemented with $20\mu\text{g}/\text{mL}$ kanamycin sulfate to form preliminary liquid overnight cultures. The cultures were then back-diluted 1:50 into $200\mu\text{L}$ of MOPS-minimal media supplemented with 0.4% maltose and $20\mu\text{g}/\text{mL}$ kanamycin sulfate (MOPSMK), and allowed to grow overnight to allow cells to adjust to the new media conditions. MOPS minimal media was used to further reduce cellular autofluorescence, and maltose was chosen as the carbon source to increase the infectivity of cells by increasing host production of the phage receptor, LamB [1,51]. From this secondary overnight culture, a third culture of the strains was inoculated 1:50 into $200\mu\text{L}$ of fresh MOPSMK and grown until cells were in exponential phase. The log-phase cells were then infected with $\lambda\text{NQ}009$ and the infection allowed to proceed for three hours before cells were fixed using 4% formaldehyde in phosphate buffered saline (PBS). See Supplemental Materials section A.2.5 for a more detailed protocol for infection and fixation.

The library of strains were then imaged in high-throughput with single cell resolution. In order to efficiently image the approximately 4000 infected strains, I used a modified version of the Strain Library Imaging Protocol (SLIP), a high-throughput microscopy workflow described by Shi et al. to automatically obtain many images of infected cells [66]. With normal SLIP, bacterial cultures in 96-well plates are transferred onto large agar pads using replicator pins, then imaged on a microscope controlled by a MATLAB script [66]. Here, I further optimized SLIP by binding cells to the surface of 96-well imaging plates (Thermo Fisher Scientific) to improve the ease of imaging. Unlike normal SLIP, this modified version forgoes transferring cells onto freshly made agar pads and instead binds fixed cells onto poly-lysine treated imaging plates. This allows the samples to be prepared in large batches, days in advance of imaging, saving time. Images of cells obtained using modified SLIP are albeit of lower quality compared to traditional SLIP (e.g. poor contrast, presence of improperly bound cells, and optical aberrations), but are still of sufficient quality to obtain the desired measurements because of our novel image analysis algorithm.

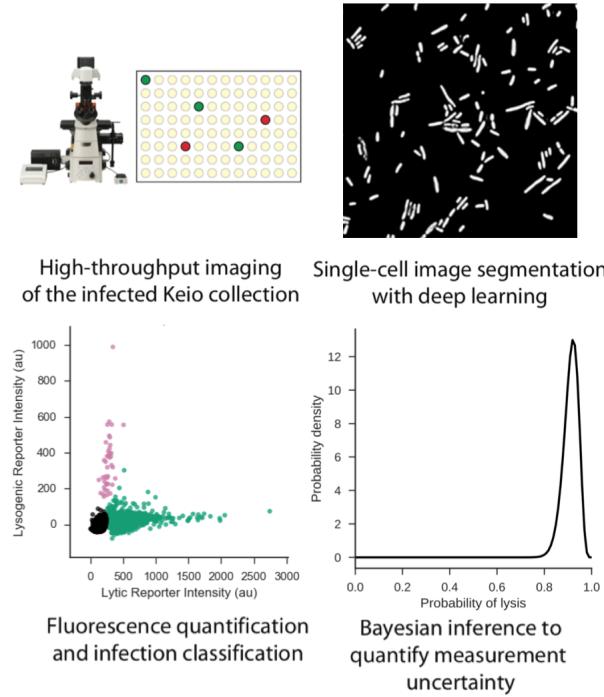


Figure 3.6: Summary of the image analysis pipeline. Knockout strains are first infected and fixed on 96-well plates. The fixed cells are then imaged in high-throughput using a modified SLIP protocol. These images are segmented using convolutional neural networks, then the masks were used to extract fluorescence data for all cells. From these values, cells are classified as lytic, lysogenic, or uninfected. Bayesian analysis was then used to infer the lysis ratio for each knockout strain.

3.3.2 Image segmentation with deep learning

To analyze the images obtained via modified SLIP, I first needed to classify the pixels in each image as either background, cell edge, or cell interior—a problem known as image segmentation [67]. Here, I used DeepCell, a segmentation algorithm utilizing deep convolutional neural networks developed by the Covert lab, to perform high throughput segmentation of the phase channel images of each knockout strain [67]. Deep convolution neural networks (conv-nets) are a type of supervised machine learning method particular well suited for image classification, as they have substantial representational power to adequately encode the relationship between images and labels. A conv-net generally consists of two components: a dimensionality reducer and a classifier. The dimensionality reducer iteratively applies a series of three operations—convolutions, a transfer function (in this case the rectified linear unit (ReLU) function), and downsampling—to produce a low dimension representation of the image [67]. This low-dimension representation is then fed into the classifier, which outputs scores for each possible class. From these scores, the image can be classified [67]. In all, compared to

other segmentation algorithms, DeepCell is more accurate and requires less manually curated training data to perform segmentation [67]. Details on network architecture, training dataset creation, and model architecture can be found in section A.2.9.

With the trained neural network, images of each infected strain were segmented, and binary segmentation masks were created, allowing me to proceed to the next steps of the analysis pipeline: data post-processing, quality control, and infection classification.

3.3.3 Image analysis and infection classification

From the segmentation masks, several post-processing steps were performed before they could be used to analyze the fluorescence data. Data-post processing involved the following steps: image stitching or image cropping, and mask quality control.

Some images produced via modified SLIP exhibited some overlap between adjacent field of views. In order to prevent double counting of cells during the analysis, it was necessary to either crop the overlapping regions of the images, or crop and stitch images together. To crop images, 100 pixels from each edge were removed. To stitch images together, I used the fast-Fourier transform cross-correlation method to obtain the registration parameters to stitch images together [68].

To filter the masks for incorrectly segmented cells, segment debris, or other segmentation artifacts, the regions labeled as cells were screened by the following constraints: area, eccentricity, minor axis length, major axis length, and solidity. First, all regions with area less than 20 pixels ($0.0839 \mu\text{m}^2$) were removed, as these regions are likely falsely segmented debris in the field of view. Next, regions with properties further than two standard deviations away from the mean area, eccentricity, minor axis length, major axis length, and solidity were removed. The remaining regions were considered to be segmented cells.

From the regions of interest identified as properly segmented cells, the mean fluorescence values for both the EGFP and mCherry channels can be calculated for each cell in each image for each knockout strain. I then used a Gaussian mixture model (GMM) classifier to classify cells into three categories: lytic, lysogenic, or uninfected. In general, the GMM classifier was trained using the expectation-maximization algorithm on the set of all mean fluorescence values of the 96 well plate the model was to classify [69]. In certain cases, training on the full set of fluorescence values yielded poor classification results, in which case several empirically chosen wells were used instead for training. See section A.2.10 for more details on the GMM classifier. Now that the infected cells of each knockout strain have been classified, I moved on to the next step of the analysis: inferring the probability of lysis and the multiplicity of infection.

3.3.4 Bayesian analysis to infer the probability of lysis

Once the images had been segmented and cells had been classified by the mean fluorescence values into lytic and lysogenic classes, I used Bayesian inference to estimate the true probability of lysis for every strain, and quantify the uncertainty of the estimate. Under the Bayesian paradigm of statistical analysis, probability quantifies the credibility of a parameter value given a statistical model and the observed data [70].

I begin by specifying a statistical model for the observed cell-fate outcomes. For every infection, the outcome can be modeled by a Bernoulli distribution:

$$p(\gamma|\theta) = \theta^\gamma(1-\theta)^{1-\gamma}$$

where θ is the probability of lysis and γ is the outcome (lysis = 1, lysogeny = 0). Assuming that infections are independent from each other and that θ remains constant between infections, the probability of observing a certain set of outcomes in N infections can be written as:

$$p(\{\gamma_i\}|\theta) = \prod_{i=1}^N p(\gamma_i|\theta) = \prod_{i=1}^N \theta^{\gamma_i}(1-\theta)^{1-\gamma_i}$$

If there are x cells undergoing lysis within this set of outcomes, this product can be rewritten as:

$$\prod_{i=1}^N \theta^{\gamma_i}(1-\theta)^{1-\gamma_i} = \theta^x(1-\theta)^{N-x}$$

Now that the model has been defined, a prior distribution must be chosen. Because there is no prior information about the true lysis ratio of the knockout strains, I chose a uniform distribution spanning from 0 to 1. This particular uniform distribution can be written as a beta distribution: $\beta(1,1)$. Here, the beta distribution is defined here as $\beta(a,b) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}$, where $B(a,b) = \int_0^1 \theta^{a-1}(1-\theta)^{b-1} d\theta$ is the beta function, and serves to normalize the numerator. Using Bayes' Theorem, credibility can be reallocated based on the observed data, and the posterior distribution can be calculated:

$$\begin{aligned} p(\theta|x, N) &= \frac{p(x, N|\theta)p(\theta)}{p(x, N)} \\ &= \frac{\theta^x(1-\theta)^{N-x} \cdot \beta(1, 1)}{p(x, N)} \\ &= \frac{\theta^x(1-\theta)^{N-x}}{p(x, N)} \\ &= \frac{\theta^{(x+1)-1}(1-\theta)^{(N-x+1)-1}}{p(x, N)} \end{aligned}$$

$$\begin{aligned}
&= \frac{\theta^{(x+1)-1}(1-\theta)^{(N-x+1)-1}}{B(x+1, N-x+1)} \\
&= \beta(x+1, N-x+1)
\end{aligned}$$

Note that the denominator $p(x, N)$ was never computed because after simplification of the numerator, the expression appeared to be the unnormalized beta distribution $\beta(x+1, N-x+1)$. The denominator $p(x, N)$ must therefore be $B(x+1, N-x+1)$, and thus the posterior distribution is $\beta(x+1, N-x+1)$. With this posterior distribution, the lysis ratio can be estimated by using the fraction of lytic cells:

$$\hat{\theta} = \frac{x}{N}$$

As for the uncertainty of the estimate, the variance of the posterior can be used:

$$Var(\theta) = \frac{(x+1)(N-x+1)}{(N+2)^2(N+3)}$$

Thus, by knowing the number of infected and lytic cells in each well, the true lysis ratio of each knockout strain can be inferred, and the uncertainty of the estimate quantified using the formulas derived in this section.

3.3.5 Bayesian analysis to infer the multiplicity of infection

The number of viruses infecting a cell, known as the multiplicity of infection (MOI), is a known parameter that affects the observed lysis-lysogeny ratio; developing a method of determining the MOI is therefore desirable to control for deviations in lysis ratio due to MOI effects. Normally, the MOI is estimated by the average phage input (API), which is the ratio of the total number of phage over the total number of cells. However, API often overestimates MOI, since not all phage may bind, and phage may not infect the population evenly (i.e. there may be regions with high concentration of phage, leading to locally higher MOI). Thus, I opt to infer the MOI using Bayesian analysis.

To build a model how N_p phage particles infecting N_c bacterial cells to produce a population containing M uninfected cells and $N_c - M$ infected cells, we begin by determining the probability of observing one uninfected cell given one phage:

$$P(M = 1 | N_p = 1) = 1 - \frac{1}{N_c}$$

Here, I am assuming that there are no free phages, and that each infection is independent to one another. Using the latter assumption, the probability of observing one uninfected cells given N_p phages, denoted ρ , is:

$$P(M = 1 | N_p) = \rho = \left(1 - \frac{1}{N_c}\right)^{N_p}$$

Extending the model further, the probability of observing M uninfected cells given N_p phages can be modeled by the binomial distribution:

$$P(M|N_p) = \binom{N_c}{M} \rho^M (1 - \rho)^{N_c - M}$$

Given this model, the Bayesian posterior distribution for inferring ρ given the uniform prior $\beta(1, 1)$ can be derived:

$$\begin{aligned} P(\rho|M, N_c) &= \frac{P(M, N_c|\rho) \cdot P(\rho)}{P(M, N_c)} \\ &= \frac{\binom{N_c}{M} \rho^M (1 - \rho)^{N_c - M} \cdot \beta(1, 1)}{P(M, N_c)} \\ &= \frac{\binom{N_c}{M} \rho^{(M+1)-1} (1 - \rho)^{(N_c - M + 1) - 1}}{P(M, N_c)} \\ &= \beta(M + 1, N_c - M + 1) \end{aligned}$$

Now that ρ has been estimated, the MOI can be calculated by rewriting MOI in terms of ρ . I begin by finding an expression for N_p in terms of ρ :

$$\begin{aligned} \rho &= (1 - \frac{1}{N_c})^{N_p} \\ \ln \rho &= N_p \ln(1 - \frac{1}{N_c}) \\ N_p &= \frac{\ln \rho}{\ln(1 - \frac{1}{N_c})} \end{aligned}$$

Substituting this expression into the definition of MOI (denoted by μ):

$$\mu(\rho) = \frac{N_p}{N_C} = \frac{\ln \rho}{[\ln(1 - \frac{1}{N_c})]}$$

If we let $\gamma = \ln(1 - \frac{1}{N_c})$, then MOI can be rewritten as:

$$\mu(\rho) = \frac{\ln \rho}{\gamma}$$

In order to find the posterior distribution for μ given the posterior distribution for ρ , a change of stochastic variable is performed. Let P_ρ be the posterior for ρ , and let P_μ be the posterior for μ . Additionally, let $f(\rho) = \frac{\ln \rho}{\gamma}$, where $f(\rho)$ is the relationship between random variables μ and ρ . Given these definitions, the desired change of variable is given by:

$$P_\mu(\mu) = P_\rho(f^{-1}(\mu)) \left[\frac{df^{-1}}{dy} \right]$$

$$\begin{aligned}
&= P_\rho(e^{\gamma\mu}) \left| \gamma e^{\gamma\mu} \right| \\
&= \left| \gamma e^{\gamma\mu} \right| \beta(M+1, N_c - M + 1) \Big|_{\theta=e^{\gamma\mu}}
\end{aligned}$$

The MOI posterior distribution is therefore:

$$P(\mu | N_c, M) = \left| \gamma e^{\gamma\mu} \right| \frac{e^{\gamma\mu M} (1 - e^{\gamma\mu})^{N_c - M}}{B(M+1, N_c - M + 1)}$$

where $B(\alpha, \beta)$ is the beta function.

3.4 Validation of the high throughput lysis-lysogeny assay

Before proceeding with the actual screen of the Keio collection, the pipeline was evaluated on its ability to correctly detect shifts in lysis ratio. To do so, I used the protocols described in the previous sections to examine factors known to affect the lysis-lysogeny decision, namely MOI and carbon source.

3.4.1 Validation by varying the multiplicity of infection

To evaluate the effect of MOI on the lysis ratio, I first infected log phase BW25113⁶ *E. coli* cells growing in MOPS minimal media supplemented with 0.4% maltose with either 0, 1, 2, 5, 10, 20, 50, or 100 μ L of λ NQ009 phage stock (titered at 3×10^{10} pfu/mL). Each infection condition was repeated as eight replicates. The infection was allowed to proceed for three hours at 30°C before being fixed with 4% formaldehyde in PBS. The fixed cells were then imaged using the modified SLIP protocol, and the images analyzed using the analysis pipeline described in the previous sections. See section A.2.11 for a more detailed protocol.

The assay described above was performed twice and the results were displayed in Figure 3.7. As the fraction of cells that become infected (and thus MOI) increases, there is a corresponding decrease in lysis ratio. The data shows good agreement (in both trend and shape) with previously reported experiments examining the effect of MOI on the lysis-lysogeny decision. Furthermore, the data from the two trials roughly fall along the same sigmoid curve, indicating that the screening protocol and analysis pipeline produce consistent results between trials.

⁶BW25113 is the parent strain of the Keio Collection [54].

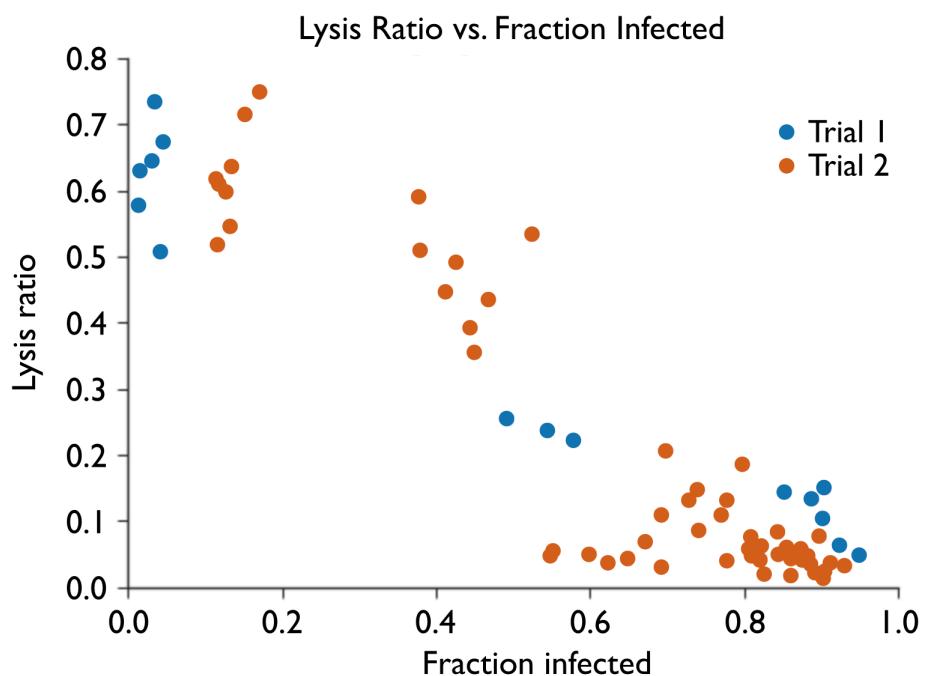


Figure 3.7: Plot of lysis ratio versus fraction infected for log phase BW25113 cells infected with λ NQ009 at various MOIs. With increasing MOI, the lysis ratio decreases, a trend consistent with literature. Data from the two trials lies roughly on the same sigmoidal curve, indicating that the assay is consistent within trials.

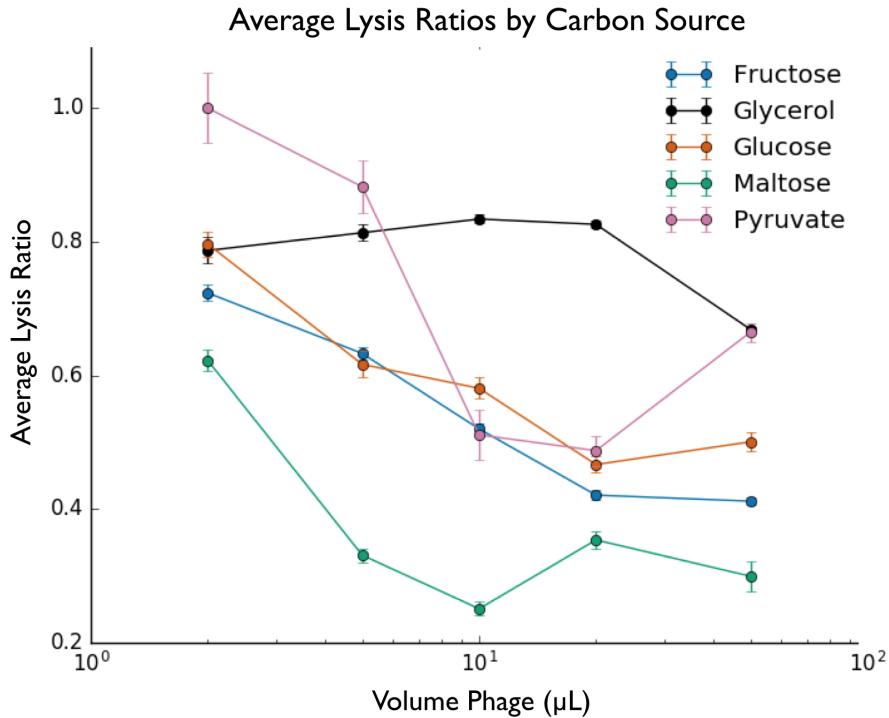


Figure 3.8: Plot of average lysis ratio versus volume of phage added. Carbon source affects the lysis-lysogeny decision, resulting in different curves. Error bars represent one standard deviation of the average lysis ratio posterior distribution.

3.4.2 Validation by varying the carbon source

In addition to varying the MOI, I also varied the carbon source within the growth media to elicit changes in the lysis ratio [71]. Here, I also infected log phage BW25113 cells grown in MOPS minimal media supplemented with 0.4% glucose, fructose, glycerol, maltose or sodium pyruvate with various amounts of λ NQ009 (1, 2, 5, 10, 20, or 50 μ L of phage stock titered at 3×10^{10} pfu/mL). Each infection condition was duplicated. The infection was allowed to proceed for three hours at 30°C, before the cells were fixed with 4% formaldehyde in PBS. These cells were then imaged and analyzed using the imaging and analysis pipeline. A more detailed protocol for this carbon source assay can be found in Section A.2.12.

From the data, we can see that the screening protocol and data analysis protocol is able to detect the perturbations in lysis ratio caused by changing the carbon source within the growth media (Figure 3.8). Furthermore, the data obtained here for maltose is consistent with the lysis ratio estimated using the bulk assay. Finally, for all the carbon sources, the data continues to exhibits the

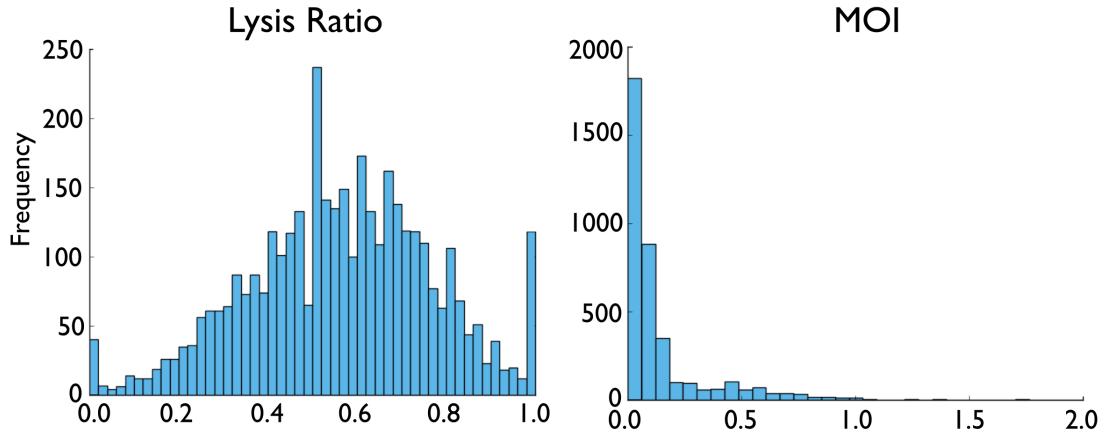


Figure 3.9: Left: Histogram of the lysis ratios of all strains of the Keio Collection. The ratios are distributed roughly normally, with peaks at either extreme. Right: Histogram of the MOI's for all strains of the Keio collection. Almost all strains display an $\text{MOI} < 1$, ensuring that the observed deviations in lysis ratio is a result of the gene deletion.

same trend with varying MOI: lysis ratio decreases with increasing phage volume (and thus increasing MOI). In all, the data obtained by varying the MOI and the carbon source indicate that the screening protocol and image analysis pipeline described in previous sections is consistent, accurate, and sensitive enough to detect changes in lysis ratio.

3.5 Screening the Keio Collection

The Keio collection was screened in various stages. First, each of the 95 plates of the Keio collection was infected using the protocol described in Section 3.3.1. These plates were then imaged and analyzed using DeepCell and other custom Python scripts to create a preliminary list of genes involved in the lysis-lysogeny decisions (see Figure 3.9). The MOI for each strain was also inferred, ensuring that large deviations in lysis ratio were not due to MOI effects (see Figure 3.9). These preliminary hits were then arrayed in a 96 well format and compiled into new glycerol stocks plates. These new plates containing possible hits were then grown in duplicate plates, and infected using the same infection protocol. These plates were then imaged using modified SLIP, and passed through the analysis pipeline. The data collected from these plates was then aggregated with data from the first round of screening, and cell-fate statistics calculated. A final list of hits was then created using these statistics.

Chapter 4

New host factors affect lysis-lysogeny

From the Keio collection’s 3985 in-frame single gene knockout strains, I identified 180 genes whose deletion biases infection towards lysis, and 194 genes whose deletion biases infection towards lysogeny. These particular strains were selected because they exhibited lysis ratios ≥ 0.9 or ≤ 0.2 for lysis-bias and lysogeny-bias respectively¹. From these 374 preliminary hits, 63 knockout strains were confirmed to be lysis-biasing, and another 47 knockout strains were confirmed to be lysogeny-biasing. Scatterplots of the pooled fluorescence data for these final hits are shown in Figures 4.1 and 4.2. The genes identified are listed in Tables 4.1 and 4.2, and summarized in Figure 4.3. For the verification round of screening, a strain was identified as a “true” hit if its pooled lysis ratio² was ≥ 0.9 or ≤ 0.2 and the measurement contained data for at least 30 infected cells.

4.1 Hits are consistent with previously known host-virus interactions

A number of genes identified in this screen have known interactions with lambda phage during infection. As expected, *hflC* and *hflK*, whose gene products facilitate protease FtsH degrade cII, were identified as lysogeny-biasing hits [45]. Another known host factor whose deletion was found to be lysis-biasing in this screen was antiterminator *nusB*. NusB is known to play an important role in phage transcriptional dynamics during infection, and is involved in N-mediated antitermination

¹Strains identified as preliminary hits had to display an OD600 reading of 0.04 or greater on the plate reader after the 4 hour back-dilution period during the infection. Strains also had to have a minimum of 30 cells in the measurement in order to be included as a hit. These additional criteria were imposed for quality-control.

²Data on infection events from the preliminary screen was pooled with the verification round data, then cell-fate statistics were calculated.

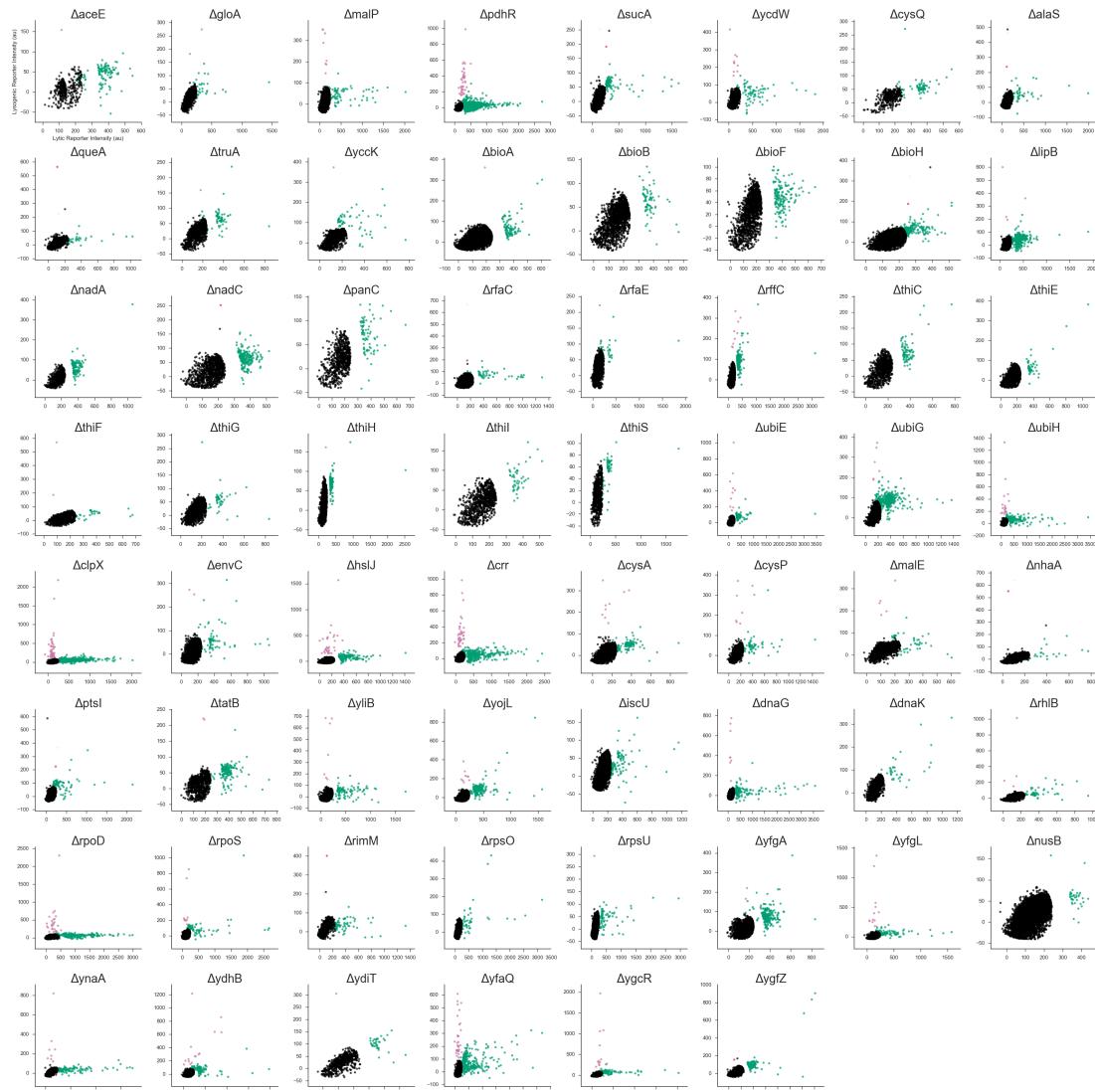


Figure 4.1: Scatterplot of fluorescence data for lysis-biasing hits. The data displayed was pooled between data from the initial screen and data from the validation screen. In all, 63 knockout strains were identified as lysis-biased.



Figure 4.2: Scatterplot of fluorescence data for lysogeny-biasing hits. The data displayed was pooled between data from the initial screen and data from the validation screen. In all, 47 knockout strains were identified as lysogeny-biased.

Table 4.1: Strains exhibiting a lytic-biasing phenotype.

Gene	Functional group	Annotation	Maynard hit?	Blasche hit?
bioA	Biosynthesis	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase	-	-
bioB	Biosynthesis	Biotin synthase	-	-
bioC	Biosynthesis	Malonyl-ACP methyltransferase	-	-
bioF	Biosynthesis	8-amino-7-oxononanoate synthase	-	-
bioH	Biosynthesis	Pimeloyl-ACP methyl ester esterase	-	-
lipB	Biosynthesis	Lipoil(octanoyl) transferase	-	-
nadA	Biosynthesis	Quinolate synthase	-	-
nadC	Biosynthesis	Quinolinate phosphoribosyltransferase	-	-
panC	Biosynthesis	Pantothenate synthetase	-	-
rfaC	Biosynthesis	ADP-heptose:LPS heptosyltransferase I	+	-
rfaE	Biosynthesis	Fused heptose 7-phosphate kinase/heptose 1-phosphate adenyltransferase	-	-
rffC	Biosynthesis	dTDP-4-amino-4,6-dideoxy-D-galactose acyltransferase	-	-
thiC	Biosynthesis	HMP-P-synthase	-	-
thiE	Biosynthesis	Thiamine phosphate synthase	-	-
thiF	Biosynthesis	ThiS adenylyltransferase	-	-
thiG	Biosynthesis	1-deoxy-D-xylulose 5-phosphate:thiol sulfurtransferase	-	-
thiH	Biosynthesis	2-iminoacetate synthase	-	-
thiI	Biosynthesis	Sulfurtransferase	-	-
thiS	Biosynthesis	Sulfur carrier protein	-	-
ubiE	Biosynthesis	C-methyltransferase	-	-
ubiG	Biosynthesis	O-methyltransferase	-	-
ubiH	Biosynthesis	2-acetaprenyl-6-methoxyphenol hydroxylase	-	-
clpX	Cell processes	ATP-dependent protease	-	+
envC	Cell processes	Murrain hydrolase activator	-	-
hsfJ	Cell processes	Lipoprotein implicated in novobiocin resistance	-	-
yfgA	Cell structure	Transmembrane component of cytoskeleton	-	-
yfgL	Cell structure	Outer membrane protein assembly factor	-	-
dnaG	DNA/RNA manipulation	DNA primase	-	-
dnaK	DNA/RNA manipulation	Chaperone protein	-	+
nusB	DNA/RNA manipulation	Transcription antiterminator	+	-
rhlB	DNA/RNA manipulation	RNA helicase	-	-
rpoD	DNA/RNA manipulation	RNA polymerase σ^{70}	-	+
rpoS	DNA/RNA manipulation	RNA polymerase σ^{38}	-	+
iscU	Iron-sulfur cluster	Scaffold protein for Fe-S cluster assembly	-	-
aceE	Metabolism	Pyruvate dehydrogenase	-	-
cysQ	Metabolism	3'(2'),5'-bisphosphate nucleotidase	-	-
gloA	Metabolism	Glyoxalase I	-	-
malP	Metabolism	Maltodextrin phosphorylase	-	-
pdhR	Metabolism	Pyruvate dehydrogenase complex regulator	-	-
sucA	Metabolism	2-oxoglutarate decarboxylase	-	-
ycdW	Metabolism	glyoxylate/hydroxypyruvate reductase A	-	-
ynaA	Prophage	Putative tail protein of Rac prophage	-	-
rimM	Ribosome	Ribosome maturation factor	-	-
rpsO	Ribosome	30S ribosomal subunit protein S15	-	-
rpsU	Ribosome	30S ribosomal subunit protein S21	-	-
crr	Transport	Glucose-specific PTS enzyme IIA	+	-
cysA	Transport	ATP-binding subunit of sulfate/thiosulfate uptake system	-	-
cysP	Transport	Thiosulfate binding protein	-	-
malE	Transport	Maltose ABC transporter periplasmic binding protein	-	-
nhaA	Transport	Na:H antiporter	-	-
ptsI	Transport	PTS enzyme I	-	-
tatB	Transport	Twin arginine protein translocation system	+	-
yliB	Transport	Glutathione ABC transporter periplasmic binding protein	-	-
yojL	Transport	Flavin transferase	-	-
alaS	tRNA synthesis	Alanine-tRNA ligase	-	-
queA	tRNA synthesis	tRNA preQ1-34-SAM-ribosyltransferase-isomerase	-	-
truA	tRNA synthesis	tRNA pseudouridine synthase	-	-
yccK	tRNA synthesis	sulfur transfer protein	-	-
ydhB	Unknown	Putative LysR transcriptional regulator	-	-
ydiT	Unknown	Putative 4Fe-4S ferredoxin protein	-	-
yfaQ	Unknown	Unknown function	-	-
ygcR	Unknown	Putative flavoprotein	-	-
ygfZ	Unknown	Putative folate binding protein	-	-

Table 4.2: Strains exhibiting a lysogeny-biasing phenotype.

Gene	Functional group	Annotation	Maynard hit?	Blasche hit?
ilvL	Biosynthesis	ilv operon leader peptide	-	-
leuA	Biosynthesis	2-isopropylmalate synthase	-	-
leuL	Biosynthesis	leu operon leader peptide	-	-
cedA	Cell cycle	Cell division modulator	-	-
hflC	Cell processes	Regulator of FtsH protease	+	-
hflK	Cell processes	Regulator of FtsH protease	+	-
ygiI	Cell processes	c-di-GMP binding protein involved in biofilm dispersal	-	-
hupB	Cell regulation	Transcriptional dual regulator	-	-
dbpA	DNA/RNA manipulation	RNA helicase	-	-
rhlE	DNA/RNA manipulation	ATP-dependent RNA helicase	-	-
tatD	DNA/RNA manipulation	3'-5' ssDNA/RNA exonuclease	-	-
rsxE	Iron sulfur cluster	SoxR reducing system protein	-	-
astD	Metabolism	aldehyde dehydrogenase	-	-
atpC	Metabolism	ATP synthase F1 complex subunit ϵ	-	+
dgoA	Metabolism	2-dehydro-3-deoxy-6-phosphogalactonate aldolase	-	-
dhaH	Metabolism	Dihydroxyacetone kinase	-	-
feaB	Metabolism	Phenylacetaldehyde dehydrogenase	-	-
fucO	Metabolism	L-1,2-propanediol oxidoreductase	-	-
glpK	Metabolism	Glycerol kinase	-	-
gntr	Metabolism	Gluconate operon repressor	-	-
nrdE	Metabolism	Ribonucleoside-diphosphate reductase II, α subunit	-	-
intG	Prophage	Putative defective phage integrase	-	-
ymfE	Prophage	Uncharacterized protein from prophage e14	-	-
ynfO	Prophage	Putative protein from Qin prophage	-	-
rplK	Ribosome	50S ribosomal subunit protein L11	-	-
yhhF	Ribosome	16S rRNA m ² G966 methyltransferase	-	-
hokB	Toxin/Antitoxin	Toxin in HokB-SokB type I toxin-antitoxin system	-	-
agaV	Transport	N-acetyl-D-galactosamine specific PTS enzyme IIB component	-	-
cstA	Transport	Peptide transporter induced by carbon starvation	-	-
fruB	Transport	Fructose-specific PTS multiphosphoryl transfer protein	-	-
gspC	Transport	Type II secretion system protein	-	-
mallM	Transport	Maltose regulon periplasmic protein	-	-
mhpT	Transport	3-hydroxyphenylpropionate/3-hydroxycinnamate:H ⁺ symporter	-	-
mtr	Transport	Tryptophan:H ⁺ symporter	-	-
tatE	Transport	twin arginine protein translocation system	-	-
ugpA	Transport	Sn-glycerol-3-phosphate ABC transporter membrane subunit	-	-
xylF	Transport	Xylose ABC transporter periplasmic binding protein	-	-
msyB	Stress response	Acidic protein, suppressor of heat sensitivity	-	-
osmY	Stress response	Periplasmic chaperone	-	-
sodB	Stress response	Superoxide dismutase	-	-
ycgH	Unknown	Putative transporter component	-	-
ydgH	Unknown	Unknown function	-	+
yfaS	Unknown	Putative macroglobulin family protein	-	-
yhaC	Unknown	Unknown function	-	-
ymcA	Unknown	Putative lipoprotein	-	-
ynjD	Unknown	Putative ABC transporter ATP binding protein	-	-
yoaC	Unknown	Unknown function	-	-

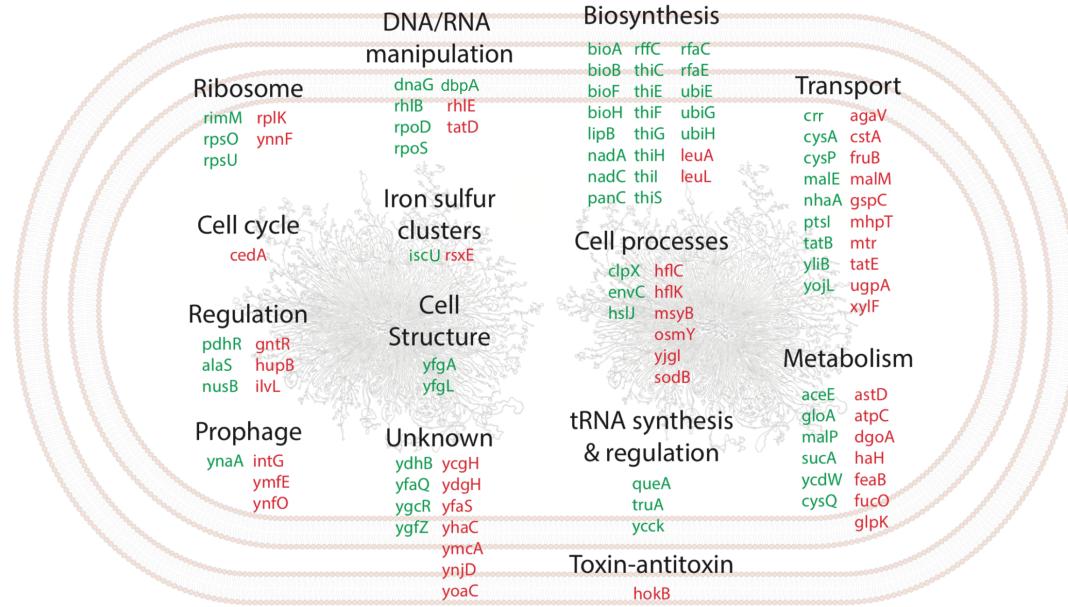


Figure 4.3: Summary of lysis-biasing and lysogeny biasing hits. Hits found in this library screen cluster in several functional groups.

of phage transcripts [20, 50]. In terms of known protein-protein interactions, my screen found host protease ClpX, host chaperone protein DnaK, and sigma factor RpoD (also known as σ^{70}). ClpX is known to degrade phage protein O, and thus disrupts phage replication initiation [55, 72]. Deletion of *clpX* causes a bias towards lysis as expected, since deletion of the protease should facilitate phage genome replication. DnaK is known to interact with phage proteins O and P, and is essential for phage genome replication [73]. It is unclear how deletion of *dnaK* results in a lysis-biased phenotype, and more work must be done to examine how DNA replication occurs in the $\Delta dnaK$ mutant. As for RpoD, it has numerous known interactions with several phage proteins. RpoD interacts with phage repressor CI to regulate *cI* expression and maintain lysogeny [74]. RpoD also plays a crucial role in inducing lysogeny by effecting CII activity on the promoters pRE, pAQ and pI [75]. Deletion of *rpoD* can therefore be expected to exhibit a bias towards lysis; this is exactly what is seen in my data.

4.2 Comparing hits with genes identified in the Blasche and Maynard screen

As mentioned previously, two important genome-wide screens have been conducted: one looking for protein-protein interactions (Blasche screen), and another searching for host factors necessary for infection (Maynard screen) [20, 55]. A number of hits found in the Blasche and Maynard screens also appear in my screen.

In all, three Blasche screen hits appear in this screen: *rpoS* (lysis-biased), *atpC* (lysogeny-biased), and *ydgH* (lysogeny-biased) [55]. *rpoS* encodes a stress-mediating sigma factor (σ^{38}), and was found to interact with phage antiterminator N. *atpC* encodes the ϵ subunit of the ATP synthase F1 complex, and interacts with phage protein P involved in phage DNA replication. As for *ydgH*, it encodes a putative protein with unknown function, and interacts with the phage packaging protein Fi [55].

In terms of the Maynard screen, six hits appear in this screen: *hflC*, *hflK*, *crr*, *nusB*, *rfaC*, and *tatB*. *hflC* and *hflK* have already been discussed, and are known to bias infection towards lysogeny, thus producing the abnormal plaque morphology seen in the Maynard screen [45]. *nusB* has also been previously discussed, and plays a key role in phage transcription antitermination [20, 50]. Deletion of *nusB* was shown to reduce infectivity in the Maynard screen, but was found to bias infection towards lysis in this screen [20]. These results may be compatible if the reduction of infectivity is epistatic to the lysis-biasing effect (which would presumably produce large plaques rather than the smaller plaques observed in the Maynard screen). As for *crr*, it encodes glucose-specific PTS enzyme IIA, which is an intermediate phosphotransfer protein involved in the uptake and phosphorylation of glucose [76]. Crr also plays an important role in the regulation of carbon metabolism in *E. coli* [77]. Here again, reduced infectivity must be epistatic to the lysis-biasing effect of the *crr* knockout. Next, *rfaC* encodes ADP-heptose:LPS heptosyltransferase I, an enzyme involved in lipopolysaccharide (LPS) synthesis [78]. Knockout of this gene causes reduced LamB activity and produces a lysis-biasing effect, though it is unclear how LPS synthesis affects the lysis-lysogeny decision [53, 78]. Finally, *tatB* encodes the inner membrane component of the twin arginine translocation (Tat) complex for the export of folded proteins [79]. More work must be done to determine how *tatB* interacts with the phage lysis-lysogeny network.

Overall, there is an unusually small amount of overlap between the genes found in this screen, the Maynard screen, and the Blasche screen. Given that all three screens aimed to probe the effect of host factors on phage infection, one might expect genes identified in one screen to appear in the other two. In particular, *hflD* was not identified as a hit in this screen, even though Maynard et al. found that $\Delta hflD$ displayed impaired plaque formation [20]. Here, strain $\Delta hflD$ displayed some

lysogeny bias, but not enough to meet the lysis ratio criterion defined earlier, indicating that the discrepancy between my screen and the Maynard screen is, in this case, due to the inclusion criteria. Along with the criteria for identifying hits, there are several other factors that may contribute to the large discrepancies between screens. For one, the Maynard screen was performed in rich media, while in this screen, cells were grown in minimal media. The use of rich media down-regulates a number of biosynthetic operons, reducing the magnitude of physiological perturbation upon gene knockout, explaining the absence of genes responsible for synthesis of key coenzymes such as thiamine (*thiC*, *thiE*, *thiF*, *thiG*, *thiH*, *thiI*, *thiS*) and biotin (*bioA*, *bioB*, *bioC*, *bioF*, *bioH*) in the Maynard screen. Furthermore, impaired plaque formation may be more related to host infectivity rather than the lysis-lysogeny decision. In this case, discrepancies between the Maynard screen results and the hits found in this screen *should* be expected, since the two screens explore different infection phenomena.

As for discrepancies between my results and those of the Blasche screen, these are likely due to fundamental methodological differences in the screening approach. Because the Blasche screen only considers interactions between two proteins in isolation, it likely misses a number of host factors since proteins often require cofactors and coenzymes to bind [55]. Furthermore, its biochemical approach cannot detect host perturbations that propagate via system-level effects (i.e. the deleted gene does not directly affect the phage genetic networks, but the deletion causes physiological changes that does). In all, there are a number of potential explanations for the poor consistency between screens, but further experiments must be conducted to validate these hypotheses.

4.3 New host factors whose deletion bias towards lysis

Interestingly, the largest functional group of new host factor identified in this screen as affecting lysis belong to important biosynthetic pathways. Of the 63 lysis-biasing knockout strains, five belong to the biotin biosynthetic pathway (*bioA*, *bioB*, *bioC*, *bioF*, *bioH*), seven belong to the thiamine biosynthetic pathway (*thiC*, *thiE*, *thiF*, *thiG*, *thiH*, *thiI*, *thiS*), three are involved in ubiquinone biosynthesis (*ubiE*, *ubiG*, *ubiH*), two participate in NADH biosynthesis (*nadA* and *nadC*), two help synthesize LPS (*rfaC* and *rfaE*), and another two (*panC* and *rffC*) aid in panthoate biosynthesis and enterobacterial common antigen (ECA) biosynthesis respectively. Along with biosynthetic pathways, numerous new host factors appear to be part of metabolism. Six genes (*aceE*, *gloA*, *malP*, *sucA*, *ycdW*, *cysQ*) encode key enzymes in central metabolism, another nine genes encode proteins that serve as transporters (*crr*, *cysA*, *cysP*, *malE*, *nhaA*, *ptsI*, *tatB*, *yliB*, *yojL*), and one gene (*pdhR*) encodes a key regulator of carbon metabolism. Another important functional class represented in the hits is DNA/RNA manipulation genes. Six genes (*dnaG*, *dnaK*, *nusB*, *rhlB*, *rpoD*, *rpoS*) are directly involved in DNA/RNA replication, transcription, and regulation, while another three (*rimM*, *rpsO*,

rpsU) are involved in ribosome synthesis and maturation. In addition to the DNA/RNA manipulation enzymes, the screen found several other genes involved in tRNA synthesis and modification (*alaS*, *queA*, *truA*, *yccK*).

Several other interesting genes emerged in this group of hits. IscU is a scaffold protein for iron-sulfur cluster assembly [80]. YfgA is bitopic inner membrane protein involved in the maintenance of cell shape through interaction with the MreB cytoskeleton [81]. Deletion of *yfgA* results in long and misshapen cells in the log phase of growth, and spheroidal cells in stationary phase [81]. These filamentous cells that are produced during the log phase of growth have a larger volume than the typical cell, and are therefore expected to display a lysis-biasing phenotype based off the work by St. Pierre et al. on cell volume as a predictor of cell-fate outcome [23]. *yfgL* encodes an outer membrane protein assembly factor important for assembly and insertion of β -barrel outer membrane proteins [82]. EnvC is a divisome associated factor that activates peptidoglycan hydrolases required for septal splitting [83]. HslJ is lipoprotein implicated in novobiocin resistance [84]. Finally, *ynaA* encodes a putative tail protein for the defective Rac prophage found in the *E. coli* K12 genome [85].

The remaining lysis-biasing hits comprise of five essentially uncharacterized genes: *ydhB*, *ydiT*, *yfaQ*, *ygcR*, and *ygfZ*. From computational analyses, YdhB has similarity to helix-turn-helix transcription factors, and is thought to be a LysR-type DNA-binding transcription factor [86]. YdiT is predicted to be a 4Fe-4S iron sulfur cluster ferrodoxin protein, and may play a role in electron transport between anaerobic fatty acid oxidation and the respiratory chain [87]. YgcR is thought to be a flavoprotein [88]. YgfZ is a folate-dependent protein with unknown function, though it is thought to play a role in iron-sulfur cluster assembly or repair during oxidative stress [89].

4.4 New host factors whose deletion bias towards lysogeny

Unlike the lytic-biasing hits, there are only a few lysogeny-biasing hits involved in biosynthetic pathways, namely *ilvL*, *leuA*, and *leuL*, which are involved in isoleucine and leucine synthesis. Instead, the largest functional class represented among the lysogeny-biasing hits is metabolism. Within this class, eight genes (*astD*, *atpC*, *dgoA*, *dhaH*, *feaB*, *fucO*, *glpK*, *gntR*) encode key enzymes in central metabolism, and ten genes (*agaV*, *cstA*, *fruB*, *gspC*, *malM*, *mhpT*, *mtr*, *tatE*, *ugpA*, *xylF*) encode proteins that serve as transporters.

Like the lytic-biasing hits, the lysogeny biasing contain a number of DNA/RNA manipulation enzymes (*dbpA*, *rhlE*, *tatD*), as well as genes involved in ribosome synthesis and regulation (*rplK* and *yhhF*). DbpA and RhlE are both ATP-dependent RNA helicases, while TatD is a 3'-5' ssDNA/RNA exonuclease [90–92]. As for *rplK* and *yhhF*, they encode the 50S ribosomal subunit protein L11 and

the 16S rRNA m²G966 methyltransferase respectively [93, 94].

Another functional cluster represented in the lysogeny-biasing hits are genes involved in the *E. coli* stress response (*sodB*, *msyB*, and *osmY*). *sodB* encodes an iron-containing superoxide dismutase protein, which aids in responding to oxidative stress [95]. MsyB is an acidic protein that may improve protein solubility and may confer heat resistance to certain mutants [96]. Finally, *osmY* encodes a periplasmic chaperone protein that is thought to confer resistance to hyperosmotic stress [97].

Several other interesting genes were found among the lysogeny-biasing hits. Three genes (*intG*, *ynfO*, *ymfE*) are thought to be genes from defective prophages in the *E. coli* K12 genome [85]. Poorly studied gene *yfaS* is thought to have been acquired by horizontal gene transfer from the eukaryotic metazoan genome [98].

hokB encodes a toxin that depolarizes the membrane when expressed, and is part of the HokB-SokB type I toxin-antitoxin system [99]. *cedA* plays a key role in the regulation of cell division, as CedA is a cell division modulator [100]. CedA has been found to interact with DnaK, a chaperone protein also identified as a hit in this screen [101]. RsxE is a SoxR 2Fe-S iron-cluster reducing system [102]. *yggI* encodes c-di-GMP binding protein involved in biofilm dispersal [103].

4.5 Discussion

In all, I have measured the lysis ratio for all 3985 knockout strains of the Keio Collection, and have identified 110 genes whose deletion causes a significant and reproducible disruption of the lysis-lysogeny decision. Of these hits, a large proportion belong to metabolic and biosynthetic pathways. The large presence of these metabolic factors in the identified hits provide a possible way for environmental information may be integrated into the cell-fate decision. These results may also speak to the metabolic necessities for infection; if the host machinery is disrupted, the transcription factors required to induce lysogeny may not accumulate to sufficient levels, causing the infection to default to lysis. This hypothesis may explain why a large number of genes involved in the biosynthesis of key coenzymes like thiamine and biotin appear as lytic-biasing hits.

In addition to speaking to the metabolic necessities of infection, the hits uncovered also seem to point to new interactions between host and virus. For one, numerous host enzymes responsible for DNA/RNA manipulation appear as hits, suggesting that phage DNA transcription and replication play a more active role in the lysis-lysogeny decision than previously thought. Indeed, if DNA replication only occurred downstream of the lysis-lysogeny decision (as is thought to happen according to the Ptashne model), then deletion of such genes would have no impact on the cell-fate outcome. This is clearly not the case, since the screen revealed nine DNA/RNA manipulation genes exhibiting

marked perturbations in lysis ratio when deleted. In addition to the previously known interactions between phage replication protein O and P and several host proteins, DNA replication may hold a new pathway with which the virus may interface with the host [55].

Another potential avenue for host-virus interaction revealed by the hits is membrane potential. One of the most interesting hits found in this library screen is HokB, the toxin component of the HokB/SokB toxin-antitoxin system [99]. When overexpressed (or expressed without the antitoxin SokB), HokB causes membrane depolarization and arrest of respiration [99]. Thus when HokB is deleted, it might be expected that the cell membrane experiences hyperpolarization. Here, the $\Delta hokB$ strain exhibited a strong bias towards lysogeny, suggesting that membrane potential may play a role in the lysis-lysogeny decision. Perhaps this is one of the many environmental factors that the phage is “sensing” in order to make the best decision. Further evidence of this potential interaction lies in the numerous hits involved in the electron transport chain (ETC), such as genes encoding an ATP synthase subunit (*atpC*), iron-sulfur clusters (*iscU*, *rsxE*), biosynthetic enzymes for key ETC coenzymes ubiquinone (*ubiE*, *ubiG*, *ubiH*), and citric acid cycle coenzymes thiamine (*thiC*, *thiE*, *thiF*, *thiG*, *thiH*, *thiI*, *thiS*), biotin (*bioA*, *bioB*, *bioC*, *bioF*, *bioH*), and NADH (*nadA*, *nadC*) [55, 80, 102].

Given these possible new pathways for the phage and the host to interact, there are several follow-up experiments I would perform. In order to investigate the role of DNA replication in the lysis-lysogeny decision, I would introduce a *parS* site onto λ NQ009’s genome to allow the phage genome copy number to be observed in real-time via live-cell microscopy. The *parS/parB* system is a DNA-protein interaction used during the segregation of bacterial chromosomes during division [104]. When fused with a fluorescent protein, multiple ParB proteins bind to a *parS* site, creating fluorescent foci that can be seen via microscopy [104]. Using this system to monitor the number of phage genomes present in a cell, I will be able to visualize any changes in DNA replication dynamics in the hits found in this screen.

As for the membrane potential hypothesis, several possible experiments can be conducted. For one, a measurement of the lysis ratio in wild-type *E. coli* at various membrane potentials would be important to determine whether membrane depolarization is sufficient to produce the perturbed lysis ratio. Membrane depolarization could be achieved using any number of antimicrobial molecules, such as gramicidin or valinomycin, or uncoupling agents such as dinitrophenol [105–107]. In terms of measuring the cell membrane potential, voltage sensitive fluorescent dyes such as thioflavin T can be used to obtain quantitative potential measurements during live infection [108]. In addition to determining the sufficiency of membrane depolarization to bias the lysis-lysogeny decision, it would be interesting to measure the cell membrane potential of all the hits found in this library screen (with

priority given to the genes known to be involved in the electron transport chain or in maintaining membrane potential).

It is important to note that although only 110 genes were identified here as “hits,” many other genes may influence the lysis-lysogeny decision in more subtle ways. The criteria for classifying strains as “hits” was chosen somewhat arbitrarily; as a result, the list host factors involved in the cell-fate decision that have a more subtle impact may be missed. Instead, this list only comprises of genes whose deletion produces a “large” shift in lysis ratio, where the cutoffs for the magnitude of the shift were chosen empirically. For a deeper understanding of the host-virus interaction, a more thorough examination of the dataset generated is needed³.

Although this forward-genetic screen greatly advances our understanding of the host-virus interaction and the lysis-lysogeny decision, there are several experimental limitations that should be noted. For one, due to its design, this library screen was only able to identify genes whose deletion affects lysis-lysogeny without significantly reducing infectivity. If a gene deletion disrupts both lysis-lysogeny and infectivity, the reduced infectivity phenotype will be epistatic to the effects on the cell-fate decision. Another limitation of the library screen comes from the use of single-gene knockout strains. Proteins may have a pleiotropic effect, and thus deletion of a single gene may result in the modulation of several pathways. Thus, extending mechanistic conclusions from this dataset alone may present some challenges. Another limitation of this screen is the varying growth rate of mutants. Because minimal media was used for this screen instead of rich media, a number of mutants displayed growth rate defects of varying severity. As a result, for strains displaying significant growth defects, there may not have been enough cells for phages to infect.

In summary, I have measured the lysis ratio for every strain in the Keio collection, revealing novel insight into the effects of host factors on the lysis-lysogeny decision. From this dataset, 110 host factors were implicated in the lysis-lysogeny decision. These genes thus represent a rich pool of leads for future investigation, and continue to push us to reconsider our understanding of the host-virus interaction.

³The cell-fate measurements dataset is freely available, as is the raw dataset. See Section A.4 for more information.

Chapter 5

Conclusion

For over 60 years, bacteriophage lambda has been providing important insight into how biological systems make important cell-fate decisions [1–3]. Despite decades of work, a complete understanding of lambda phage and its decision to commit to one of its two mutually exclusive life cycles—lysis or lysogeny—continues to elude biologists [1, 2]. In particular, the role of the host in the lysis-lysogeny decision remains understudied [20]. In this thesis, my aim was to uncover new insight into the complex interplay between the host and the virus during infection by screening the Keio Collection for mutants displaying perturbations in cell-fate statistics. In order to do so, I designed and engineered a strain of lambda phage, λ NQ009, capable of reporting infection outcomes at the single cell level. Using this engineered strain, I then developed a protocol to infect all 3985 knockout strains of the Keio Collection in high-throughput, and imaged each infected strain using a modified version of the SLIP imaging protocol [66]. To analyze these images, I employed DeepCell, an algorithm based on deep convolutional neural networks, to segment cells within each field of view [67]. From the segmentation masks produced by DeepCell, I then extracted fluorescence data for each cell, and classified the cell as lytic, lysogenic or uninfected using a Gaussian-mixture model classifier. Once cells had been classified into the three classes, the lysis ratio and multiplicity of infection for each strain was inferred using Bayesian analysis. Utilizing this approach, the lysis ratio for all 3985 strains of the Keio collection was measured. Of these strains, 110 were found to have large, reproducible shifts in lysis ratio, thus implicating these genes in the host-virus interaction. A large number of these genes are involved in central metabolism, providing new clues into how lambda phage integrates environmental information into its cell-fate decision. In particular, the hits suggest that phage DNA replication within the host plays an active role in lysis-lysogeny, and that host membrane potential may be one of the key physiological parameters that lambda phage “senses” during infection.

As a whole, this screen is the first of its kind to look for host genes involved in the lysis-lysogeny decision. It is also one of the first genome-wide screens to incorporate single cell imaging as its output.

As a result, novel methodologies, such as automated fluorescence microscopy, a deep learning based method for single cell image segmentation, and phage cloning, had to be created in order for this screen to be feasible to conduct. It is apparent from my results that the story of lambda phage's cell-fate decision is more complicated than our current understanding may suggest; indeed, the host plays a larger role than previously thought, and any model that aims to accurately represent the lysis-lysogeny decision must account for these host factors. The results presented here provide more evidence supporting the deterministic "hidden variables" view on infection. In the context of other work done on the lysis-lysogeny decision, my results provide possible mechanistic leads for known host-mediated effects. For example, in his 2009 thesis, Francois St. Pierre identifies cell volume as a reliable marker for cell fate, where large cells tend to produce more lytic events [23]. Under one hypothesis, St Pierre posits that physiologic and metabolic differences may mediate this correlation between cell volume and cell fate [23]. Given the disproportionate number of metabolic genes implicated in this screen, his hypothesis may very well be correct, and we now have leads on which exactly are involved in controlling lysis-lysogeny.

Despite the advances made in this genetic screen, a number of questions still remain. How directly do host factors interact with the phage regulatory networks? How do these factors convey information to the lysis-lysogeny genetic circuitry? And what benefit does the phage gain from mediating such complex host interactions? In order to begin answering some of these questions, there are several follow-up experiments that can be performed. For one, measuring host factor levels of wild-type *E. coli* grown in various carbon sources can be conducted to examine the effect of specific metabolic changes on the cell-fate decision. To do this, strains from the yellow fluorescent protein (YFP) fusion library created by Taniguchi et al. could be infected with a phage-encoded cell-fate reporter like λ NQ009, and the resulting infection followed via live-cell fluorescence microscopy. Each strain in this library contains a particular gene tagged with a C-terminal YFP tag at the native chromosomal locus [41]. The YFP-fusion mutants would allow measurement of specific host factors in a single cell during infection, and allow correlations between host factor dynamics/variation and cell fate outcomes to be determined [41]. As for the virus-side of infection, it would be interesting to examine whether CII dynamics are altered during infection of the knockout mutants identified in this screen. Such experiment would provide insight into the entry point of host-encoded information into the lysis-lysogeny decision circuitry, and determine whether CII is indeed the "switchman" for the cell-fate decision. Finally, given that several transcription factors have been implicated here as host factors, it would be of interest to scan the lambda phage genome for potential binding sites for these host-encoded transcription factors. If any binding sites are identified, gel-shift assays could be performed to verify the predictions. Alternatively, strains of lambda phage could be engineered to contain knockouts of these putative binding sites, and verify if the infection phenotype is still affected by the particular host transcription factor.

In all, I have designed and conducted a forward genetic screen to identify new host factors involved in lambda phage's lysis-lysogeny decision. Contrary to current understanding, my results suggest that the host plays a significant role in the cell-fate decision [23,39,42]. Indeed, the 110 genes implicated here along with the lysis ratio dataset produced in the screen will provide a rich ground for further exploration into the host-mediated mechanisms governing the lysis-lysogeny decision. As for the true nature of lambda phage infection, it is likely that both stochastic and deterministic factors are at play during the cell-fate decision. Regardless of model, it is clear that the effect that host factors play in host-virus interaction play an important role and should be the subject of future studies.

Appendix A

Supplemental Materials

A.1 Materials

A.1.1 Media

MOPS supplemented with maltose and kanamycin (MOPSMK)

For 1L of media:

- 100 mL MOPS media buffer (Teknova)
- 10 mL 0.132M K₂HPO₄ (Teknova)
- 20 mL 20% maltose (filter sterilized)
- 869 mL sterile H₂O
- 0.4 mL 50 mg/mL kanamycin sulfate

Store at 4°C.

Cell fixative: 4% formaldehyde in PBS

For 40 mL:

- 10 mL 16% methanol-free formaldehyde (Sigma-Aldrich)
- 30 mL 1x phosphate-buffered saline pH 7.4 (Thermofisher Scientific)

SM buffer with gelatin

For 1L:

- 5.8g NaCl
- 2g MgSO₄ · H₂O
- 50 mL 1M Tris pH 7.5
- 5 mL 2% (w/v) gelatin
- H₂O to 1L

Mix reagents together, then autoclave to sterilize.

TB agar

For 100 mL:

- 1g tryptone
- 0.5g NaCl
- 1.5g agar
- H₂O to 100 mL

Autoclave to sterilize.

TB soft agar

For 100 mL:

- 1g tryptone
- 0.5g NaCl
- 0.7g agar
- H₂O to 100 mL

TB media

For 100 mL:

- 1g tryptone
- 0.5g NaCl
- H₂O to 100 mL

Autoclave to sterilize.

Synthetic Defined media supplement with leucine

For 500 mL:

- 13.35g SD Base (Clontech)
- 0.325g DO supplement (Clontech)
- 10 mL 1mg/mL histidine (filter sterilized)
- 10 mL 1mg/mL uracil (filter sterilized)
- 9g agar (for plates only)
- 500 mL water

Mix SD base, DO supplement, water, and agar (if plates are to be made). Autoclave at 121°C for 15 min. Once cooled, add the histidine and uracil solutions.

M9 minimal media supplemented with 0.4% maltose

For 1L:

- 200 mL 5x M9 minimal salts
- 34 mL 10 mg/mL thiamine
- 20 mL 20% maltose
- 20 mL 10% Casamino acids
- 2 mL 1M MgSO₄
- 100 μL 1M CaCl₂
- H₂O to 1L

A.1.2 Primers and sequences**Primers****Sequences**

The sequence map of λNQ009 cloned into pRS415 can be found on Benchling: <https://benchling.com/s/seq-PVk2sKSm54v6nLjJ2sMO>

Table A.1: Primer sequences for constructing λ NQ009. Bolded sequences indicate homology overhang arms

pRS415-R-BW802721-F	TATAACGTTTTGAACACACATGAACAAAGGACTGAAAA TGTGTTCACAGGTTGCTCCGGG
mKate2-hom:Frag2a-R	TGCGACCTTCCGTAAA ACTGGTCACCGCTCGGCAAGGT-GTTCTGGTCGGCGCATAGCT
Frag2b-hom:mKate2-F	AGTCAGTGGCCTGAAGAGACGTTGGCTGACATATCGGTCA CGAAAAAAGAGGAGAAATACTAGATGTCGGAATTG
Frag2a-hom:mKate2-R	AGCTATGCCGACCAGAACACCTTGCCGAGCGGTGA CCCAGTTTGACGGAAGGTCGCA
mKate2-hom:Frag2b-F	CAATTCCGACATCTAGTATTCTCCTTTTCGTGACCGAT ATGTCAGCAAACGTCTTCAGGCCACTGACT
L46231-46260-R	AGACTGCTTGATGTGCAACCGACGA
L46149-46178-F	GCTGCGCTCGATGCAAAATACACGAAGGAG
L7841-7870-R	TTTCCTCACCGATGGTCAGCGTGTCTCCAC
L7791-7820-F	CTCCAGCCGTCCTGTTGTCCGGACTGA
L17811-17840-R	ATATTGATACTGGCGGCTATCCAGTACAGC
L17761-17790-F	ATTGCGGATATCAGACAGGTTGAAACCAAGC
mNG-hom:Frag5a-R	A C C A T C T A G T A T T T C T C C T T T C T C T A G A GGGTATTGTTGTCCTTGAGCGCGGTTG
Frag5a-hom:mNG-F	CAACCGCGCTCAGGGGAACAAACAATACCCCTCTAGAG AAAGAGGAGAAATACTAGATGGT
Frag5b-hom:mNG-R	CGGCCAGTACAAAAGCGGTGTTGCAATCTTACT-TATATAACTCATCCATACCCATTAC
mNG-hom:Frag5b-F	GTAATGGGTATGGATGAGTTATAAGTAAAGATTGCG AACACCGCTTTGTACTGGCCG
pRS415-F-BW802845-R	ACATTCAAATATGTATCCGCTCATGAGACAGGCGCAATGC-CATCTGGTATCACTAAAGG
pRS415-F60	CCTTTAAGTGATACCAGATGGCATTGCGCCTGTCTCAT GAGCGGATACATATTGAATGT
pRS415-R60	CCCGGAGCAACCTGTGAACACATTTCACTGTCAT- GTGTGTTCAAAAACGTTATA

Table A.2: Primer sequences for verifying fragment junctions in λ NQ009.

InsertCheck-1-F3	AGCGCCCCTGTGTGTTCTCGTTATGTTGAG
InsertCheck-1-R3	AAGCATTCAAGGTCTTCCTTCGAAGGGGATC
NQ009-j2aI-F	CAACCTGCAGGTGATGATTATCAGCCAGCA
NQ009-j2aI-R	TACCTCGCTCCAGGATGGTGCCTGATCTA
NQ009-jI2b-F	GTCTTCATAGGTGGTAACACGCTCCCAGGT
NQ009-jI2b-R	GATGGGGCAGTCAGGCCCTGGTGCTTATT
InsertCheck-3-F	GCGGCAATTACTGACATGCAGATGCGTCAG
KanR1	CCTGATTGCCGACATTATCG
InsertCheck-4-F2	ACCGTGATTCTGGATACGTCTGAACGGTC
InsertCheck-4-R2	AAGCCAGAGATGACAACCTCCGCCATCATC
InsertCheck-5-F	TGAGTTCTGCTCCGTCTGACCGTAACAG
InsertCheck-5-R	CACTCTTCGAAAACCTCCTCCAGTCTGCTG
NQ009-j5aI-F	GGATAGTGCAGCTCAGCAGTGCAACCAACA
NQ009-j5aI-R	AGGAGACATCCATCAGGGTACGGCAAGTA
NQ009-jI5b-F	TAATTCTTGACCGCAGCAGATTGGTGCCG
NQ009-jI5b-R	GATTGTCCAGCCTCGCATATCAGGAAGCAC
InsertCheck-6-F	CCTTTAAGTGTACCCAGATGGCATTGCGCC
InsertCheck-6-R	CTTGCCTGTAACCTACACGCGCCTCGTATC

Table A.3: Phage fragment thermocycling protocol

Step	Temperature (°C)	Time
Initial denaturation	98	2 min
Denaturation	98	10 sec
Anneal	72	30 sec
Extension	72	5 min
Final extension	72	5 min

A.2 Methods

A.2.1 Phage fragment amplification

In a PCR tube, 2.5 μ L of each primer was added. The primers were diluted to 10 μ M with deionized water. 1 μ L of template (about 30-100 ng of DNA), along with 32.5 μ L of deionized water, 10 μ L of HF or GC buffer (New England Biolabs), 1 μ L of 10mM dNTP's, and 0.5 μ L of Phusion polymerase was then added. The mixture was then placed on a thermocycler for 40 cycles using the program detailed in Table A.3 for a phage fragment or Table A.4 for a YAC fragment. Once amplified, the PCR reactions were purified using Zymo-5 DNA Clean-up Kit, eluting the fragments in 15 μ L of water to yield DNA concentrations around 200 ng/ μ L (measured by a NanoDrop 1000 spectrophotometer).

Table A.4: YAC fragment thermocycling protocol

Step	Temperature (°C)	Time
Initial denaturation	98	2 min
Denaturation	98	10 sec
Anneal	68	30 sec
Extension	72	90 sec
Final extension	72	5 min

Table A.5: Junction PCR thermocycling protocol

Step	Temperature (°C)	Time
Initial denaturation	95	5 min
Denaturation	95	30 sec
Anneal	55	1 min
Extension	72	1 min
Final extension	72	5 min

A.2.2 Colony PCR of λ NQ009 yeast transformants

In a PCR tube, colonies were picked and resuspended in 40 μ L sterile water. 18 μ L of the suspension was transferred into a clean PCR tube, then 2 μ L 200mM sterile aqueous sodium hydroxide solution (made fresh) was added. The PCR tube was then heated on a thermocycler at 95°C for 10 minutes, then centrifuge at >12,000g for 10 minutes. The supernatant will be used as template for the PCR reactions.

For the PCR reactions, 12.5 μ L of primer mix containing 2 μ M of each primer was mixed with 12.5 μ L GoTaq Green PCR master mix (Promega), then 1 μ L of template was added. The mixture was then cycled for 35 cycles using the program detailed in Table A.5. The primer pairs to check for the specific λ NQ009 fragment junctions are detailed in Table A.6, and the sequence of these primers are given in Table A.2 in Section A.1.2.

Table A.6: Junction PCR Primers

Junction	Primers
YAC-Fragment 1	InsertCheck-1-F3/R3
Fragment 1-Insert 1	NQ009-j2aI-F/R
Insert 1-Fragment 2	NQ009-jI2b-F/R
Fragment 2-Fragment 3	InsertCheck-3-F/KanR1
Fragment 3-Fragment 4	InsertCheck-4-F2/R2
Fragment 4-Fragment 5a	InsertCheck-5-F/R
Fragment 5a-Insert 2	NQ009-j5aI-F/R
Insert 2-Fragment 5b	NQ009-jI5b-F/R
Fragment 5b-YAC	InsertCheck-6-F/R

A.2.3 Rebooting phage from yeast genomic DNA

After fragment junctions are confirmed, 2.5 μ L of yeast genomic DNA preparation is electroporated at 2.0 kV, 200 ohm and 25 μ F, in a 1mm gap cuvette into commercial Mega DH10B electrocompetent cells (Thermo Fisher Scientific). Immediately after electroporation, the cells are allowed to recover at 37°C in 1 mL Recovery Media for one hour. After recovery, 50 μ L of chloroform is added to the culture and mixed thoroughly for 1 minute to lyse cells and release any phage particles. After a 10 minute incubate at room temperature, the mixture was then centrifuged at 12000g for one minute, and 200 μ L of supernatant mixed with 200 μ L of plating cells. To produce the plating cells, an overnight culture of amber-suppressor strain LE392 was back-diluted 1:100 into TB supplemented with 0.2% (w/v) maltose and allowed to grow at 37°C for three hours. The culture was then chilled on ice for 15 minutes to stop growth, then spun down at 3000g for 10 minutes. The cell pellet was then resuspend in enough ice cold TB supplemented with 10mM magnesium sulfate to obtain an OD600 of 2. The mixture was incubated on ice for 30 minutes to allow phage particles to bind to the plating cells, then placed at 37°C for DNA ejection. Finally, the cells were plated with 3 mL of TB soft agar on TB plates, and incubated at 37°C overnight. If rebooting is successful, the plate should yield plaques after overnight incubation.

A.2.4 Live cell microscopy to verify phage phenotype

To image infection, bacterial overnight cultures are infected with λ NQ009 at MOI 0.5, then incubated at 30°C for 15 minutes before being spotted onto a MOPS-maltose agar pad. Such agar pads are made by melting 1.5% agarose melted into MOPS minimal media supplemented with 0.4% maltose. This pad is then sealed in a silicone gasket and topped with a glass cover slip. The sample is then imaged on a TI-Nikon Eclipse confocal epifluorescent microscope at 60x magnification with a 1.5x tube magnifier under Phase, Cherry and FITC channels. The exposure times for each of the channels are as follows: Phase 20 ms, Cherry 150 ms, FITC 150 ms. Images were taken in five minute intervals for five hours at 30°C.

A.2.5 High throughput infection and fixation of the Keio Collection

Each plate of the collection is first inoculated from glycerol stocks into 200 μ L of LB supplemented with 20 μ g/mL kanamycin sulfate (LB/kan) and incubated at 37°C overnight. (Note that for all incubation periods, the plates were sealed with a BreathEasy membrane and agitated with a plate shaker at 1050 rpm.) This overnight culture was used to inoculate a new 96-well plate containing 200 μ L of MOPS-minimal media supplemented with 0.4% maltose and 20 μ g/mL kanamycin sulfate (MOPSMK) by back-diluting by 1:50. The new plate was incubated overnight at 37°C to allow cells to adjust to the new media. This second overnight culture was then back-diluted 1:50 into a new plate containing 200 μ L of MOPSMK media, and allowed to grow at 37°C for four hours. After the

back-dilution period, $5\mu\text{L}$ of high titer λ NQ009 stock¹ was added to each well. The infected plate was incubated at 30°C for three hours in the dark, then spun down at 4000 rpm in an Eppendorf 5810R centrifuge. The supernatant was removed, then $50\mu\text{L}$ of 4% formaldehyde in phosphate-buffered saline (PBS) was added to fix the cells. The cells were incubated at room temperature for 30 minutes, then spun down at 4000 rpm in an Eppendorf 5810R centrifuge. The supernatant was removed and the cells resuspended in $100\mu\text{L}$ PBS. The fixed cells were then stored at 4°C in the dark.

A.2.6 Modified SLIP sample preparation and imaging

To bind cells on the imaging plate, the imaging plate was incubated with $50\mu\text{L}$ of 70-150 kDa polylysine (Sigma Aldrich) for 1 hour at room temperature, then rinsed the wells with $200\mu\text{L}$ of PBS. $50\mu\text{L}$ of infected cells were then introduced into the wells of the plate, and allowed to incubate for 15 minutes at room temperature to bind to the bottom of the wells. The cell mixture was then removed and $50\mu\text{L}$ of PBS added to each well to prevent cells from drying out. The plate was then centrifuged at 4000 rpm on an Eppendorf 5810R centrifuge for 20 minutes to ensure that cells are lying flat against the glass, then imaged.

To perform the imaging, a Nikon TI-E epifluorescent confocal microscope at 100x magnification was used. All hardware was controlled automatically by the SLIP custom MATLAB script. For each strain, either 4x4 or 5x5 grids of images were obtained under Phase, EGFP and mCherry channels. An XCite 180 LED lamp was used as the excitation light source. The imaging parameters were as follows: Excitation strength = 80%. Phase exposure time = 20ms. EGFP exposure time = 200ms. mCherry exposure time = 200ms. Images were then saved as TIFF files.

A.2.7 Propagation of phage from single plaques via plate lysis

Cut out the desired plaque using a sterilized razor blade, and resuspend the plaque in $100\mu\text{L}$ SM-gelatin buffer. Incubate at 4°C for two hours. While the plaque is incubated, dilute an overnight culture of the propagation strain 1:100 in 20 mL of TB, and incubate at 37°C with shaking for 3 hours. Add 1 mL of this culture to the resuspended plaque, and incubate the mixture on ice for 30 minutes to allow phage particles to bind. Then, incubate the mixture at 30°C to 10 minutes to allow DNA ejection. Plate $350\mu\text{L}$ of the mixture with 3mL of TB soft agar on an LB plate. Incubate the plated mixture at 37°C overnight. After overnight incubation, add 5 mL of SM-gelatin buffer to the plate, and incubate at 4°C on a plate rocker for at least 2 hours. Collect the liquid, then add 1 mL of SM-gelatin to the plate and incubate at room temperature for 5 minutes on a rocker plate to collect a second harvest. Collect the liquid and add it to the first harvest. Add $100\mu\text{L}$ of chloroform

¹Phage stock was titered at 3×10^{10} pfu/mL

to the liquid, mix thoroughly for 1 minute, then incubate on ice for 30 minutes. Centrifuge at 4000g for 10 minutes, then transfer the supernatant into a clean tube to obtain the phage stock.

A.2.8 Population-level lysis-lysogeny assay

To perform the population level lysis-lysogeny assay, 100 μ L of LE392 overnight culture in M9-maltose was infected with phage at MOI 0.2. The mixture was then incubated on ice for 30 minutes, then placed at 30°C for 10 minutes. Then, the mixture was spun down at 12000g for 1 minute, and the pellet resuspended in 100 μ L of fresh M9-maltose to remove unbound phage.

To determine the number of total infections, the resuspended infection mixture was serial diluted 1:10 five times into 40 μ L of LE392 overnight culture. Each dilution was then plated with 500 μ L of TB soft agar in a well of a 6-well plate containing TB agar. The plate was then incubated at 37°C overnight. The number of plaques was then manually counted to determine the number of total infections.

To determine the number of lytic infections, the resuspended infection mixture was serial diluted 1:10 five times into 40 μ L of LE392 overnight culture. Each dilution was then plated with 500 μ L of TB soft agar in a well of a 6-well plate containing TB agar. The plate was then incubated at 30°C overnight. The number of plaques was then manually counted to determine the number of lytic infections.

To determine the number of lysogenic infections, the resuspended infection mixture was serial diluted 1:10 five times into 40 μ L of fresh M9-maltose. Each dilution was then infected with 10 μ L of high-titer (10^{11} pfu/mL) killer phage stock². Each dilution was then plated with 500 μ L of TB soft agar in a well of a 6-well plate containing TB agar. The plate was then incubated at 30°C overnight. The number of colonies was then counted manually to determine the number of lysogenic infections. Assays for the number of lytic, lysogenic and total infections were repeated in triplicate to allow the standard error to be calculated for each infection event type.

A.2.9 Constructing and training convolutional neural networks for image segmentation

To segment the modified SLIP images, a nine weight-layer batch-normalized convolutional neural network (conv-net) architecture was used (defined by Table A.7) [67]. To attain the best segmentation accuracy, the conv-net was trained *de novo* on a training set derived from images obtained via modified SLIP. This training set was created by manually segmenting five phase-channel SLIP

²Killer phage is a strain of lambda phage that cannot lysogenize due to a mutation in the lambda repressor CI. Here, we used the strain λ c126. Killer phage is used to select lysogens, which are immune to superinfection.

Table A.7: Architecture of the batch-normalized feature net 61x61 conv-net

Layer	Type	Filter size
Layer 1	Convolution	64x2x3x3
Layer 2	Batch Normalization	
Layer 3	Activation: ReLU	
Layer 4	Convolution	64x64x4x4
Layer 5	Batch Normalization	
Layer 6	Activation: ReLU	
Layer 7	Max Pooling	2x2
Layer 8	Convolution	64x64x3x3
Layer 9	Batch Normalization	
Layer 10	Activation: ReLU	
Layer 11	Convolution	64x64x3x3
Layer 12	Batch Normalization	
Layer 13	Activation: ReLU	
Layer 14	Max Pooling	2x2
Layer 15	Convolution	64x64x3x3
Layer 16	Batch Normalization	
Layer 17	Activation: ReLU	
Layer 18	Convolution	64x64x3x3
Layer 19	Batch Normalization	
Layer 20	Max Pooling	2x2
Layer 21	Convolution	200x64x4x4
Layer 22	Batch Normalization	
Layer 23	Activation: ReLU	
Layer 24	Fully Connected	200x200
Layer 25	Batch Normalization	
Layer 26	Activation: ReLU	
Layer 27	Fully Connected	3x200

Table A.8: Training sets for the GMM classifier for certain plates of the Keio Collection.

Plate	Training set
29	Plate 29, H12
31	Plate 31, G1
33	Plate 33, H11
35	Plate 35, A1
37	Plate 29, H12
43	Plate 29 H12
77	Plate 77, A1-A3
91	Plate 91, G12

images, using ImageJ and a Wacom Intuos Draw graphics tablet (Model# CTL490DW) to manually outline each cell with the freehand selection tool. These manually segmented images were then converted into a training dataset by sampling 61x61 pixel images from the segmented image, and labels generated from the identity of the center pixel of the 61x61 sample image. Data augmentation was performed on this training set according to previously reported protocols [67]. From this training set, the conv-net was trained for 80 epochs using the stochastic gradient descent (SGD) optimizer with Nesterov momentum. A learning rate of 0.01 with decay rate of 0.95 per epoch, momentum of 0.9, and regularization of 10^{-5} . Additionally, for training, data was batched into 256 training pairs. Computation was accomplished with the Stanford Sherlock Cluster, using a NVIDIA GTX Titan Black GPU-based server. All code can be accessed on Github: https://github.com/nquach/lysis_lysgeny_screen

A.2.10 Classifying cell-fate outcomes using a Gaussian mixture model

To classify the mean fluorescence values into lytic, lysogenic, and uninfected events, a two dimensional, three component Gaussian mixture model (GMM) was employed. For the most part, the GMM was trained using the expectation-maximization algorithm on the set of all fluorescence values within the plate it was to classify (e.g. classifier for Plate 9 used all fluorescence values from Plate 9 as training data). When training on the full set yielded poor classification, data from empirically chosen wells were used instead for training. These exceptions are listed in Table A.8. Once trained, the GMM was used to generate the class probabilities and assign a label to each data point. Data points with labels assigned with less than 0.95 confidence were removed. All computation was done using the Stanford Sherlock Cluster. Custom Python scripts for classification can be found on Github: https://github.com/nquach/lysis_lysgeny_screen

A.2.11 Measuring the effect of MOI on lysis ratio

To measure the effect of MOI on the lysis ratio, a 5 mL LB overnight culture of *E. coli* BW25113 was inoculated and grown for 16 hours at 37°C, with shaking at 120 rpm. The overnight culture

was then back-diluted 1:50 into 5 mL MOPS minimal media supplemented with 0.4% maltose, and the new culture was incubated at 37°C with shaking at 120 rpm for 16 hours. This new overnight culture was then back-diluted 1:50 into 200 μ L fresh MOPS minimal media supplemented with 0.4% maltose in a 96 well plate, and the cultures incubated at 37°C on a plate shaker at 1050 rpm for three hours. The cultures were then infected with 0, 1, 2, 5, 10, 20, 50, 100 μ L of phage stock (titered at 3.0×10^{10} pfu/mL). The infections were incubated in the dark at 30°C on a plate shaker at 1050 rpm for three hours, then the cells were fixed using 4% formaldehyde in PBS using the fixing protocol detailed in section A.2.5. The fixed cells were then imaged using the modified SLIP protocol (detailed in section A.2.6). The images obtained were then analyzed using the analysis pipeline and custom Python scripts to obtain the plots in section 3.4. The data and scripts are accessible via Github: https://github.com/nquach/lysis_lysogeny_screen.

A.2.12 Measuring the effect of carbon source on lysis ratio

To measure the effect of carbon source on the lysis ratio, a 5 mL LB overnight culture of *E. coli* BW25113 was inoculated and grown for 16 hours at 37°C, with shaking at 120 rpm. The overnight culture was then back-diluted 1:50 into 5 mL MOPS minimal media supplemented with 0.4% of either fructose, glucose, glycerol, maltose, or sodium pyruvate, and the new culture was incubated at 37°C with shaking at 120 rpm for 16 hours. This new overnight culture was then back-diluted 1:50 into 200 μ L fresh MOPS minimal media supplemented with 0.4% of the corresponding carbon source in a 96 well plate, and the cultures incubated at 37°C on a plate shaker at 1050 rpm for three hours. The cultures were then infected with 0, 0.5, 1, 5, 10, or 20 μ L of phage stock (titered at 3.2×10^{10} pfu/mL). The infections were incubated in the dark at 30°C on a plate shaker at 1050 rpm for three hours, then the cells were fixed using 4% formaldehyde in PBS using the fixing protocol detailed in section A.2.5. The fixed cells were then imaged using the modified SLIP protocol (detailed in section A.2.6). The images obtained were then analyzed using the analysis pipeline and custom Python scripts to obtain the plots in section 3.4. The data and scripts are accessible via Github: https://github.com/nquach/lysis_lysogeny_screen.

A.3 Additional Statistical Analyses

A.3.1 Estimating the variance the lysis ratio in bulk lysis-lysogeny assays

In the bulk lysis-lysogeny assay, the number of measured lytic events, X, and measured total events, Y, can be treated as random variables. Given X and Y, the lysis ratio ρ is given by:

$$\rho(X, Y) = \frac{X}{Y}$$

If the true mean and variance of X and Y (denoted here as μ_X , σ_X^2 , and μ_Y , σ_Y^2) are known, the variance of ρ can be estimated using the δ method:

$$\begin{aligned} Var(\rho) &\approx Var\left[\rho(\mu_X, \mu_Y) + \frac{\partial\rho(\mu_X, \mu_Y)}{\partial X}\Delta X + \frac{\partial\rho(\mu_X, \mu_Y)}{\partial Y}\Delta Y\right] \\ &\approx \left[\frac{\partial\rho(\mu_X, \mu_Y)}{\partial X}\right]^2 \sigma_X^2 + \left[\frac{\partial\rho(\mu_X, \mu_Y)}{\partial Y}\right]^2 \sigma_Y^2 \\ &\approx \frac{\sigma_x^2}{\mu_y^2} + \frac{\mu_x^2 \sigma_y^2}{\mu_y^4} \\ &\approx \frac{1}{\mu_y^4} (\mu_y^2 \sigma_x^2 + \mu_x^2 \sigma_y^2) \end{aligned}$$

A.4 Dataset Accession

The full dataset from this library screen, as well as the dataset of lysis ratios can be accessed at https://github.com/nquach/lysis_lysogeny_screen. Analysis scripts can also be found via the link.

Bibliography

- [1] M. Ptashne. *A Genetic Switch*. Cold Spring Harbor Laboratory Press, 3rd edition, 2004.
- [2] S. R. Casjens and R. W. Hendrix. Bacteriophage lambda: Early pioneer and still relevant. *Virology*, 479-480:310–330, 2015.
- [3] D. L. Court and A. B. Oppenheim. A new look at bacteriophage λ genetic networks. *J. Bacteriol.*, 189(2):298–304, 2007.
- [4] P. Manrique, M. Dills, and M. J. Young. The human gut phage community and its implications for health and disease. *Viruses*, 9(6):E141, 2017.
- [5] C. Pal, M. D. Macia, A. Oliver, I. Schachar, and A. Buckling. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature*, 450:1079–1081, 2007.
- [6] J. L. Martinez and F. Baquero. Mutation frequencies and antibiotic resistance. *Antimicrob. Agents Chemother.*, 44(7):1771–1777, 2000.
- [7] M. Colomer-Lluch, J. Jofre, and M. Muniesa. Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples. *PLoS ONE*, 6(3):e17549, 2011.
- [8] M. Colomer-Lluch, L. Imamovic, J. Jofre, and M. Muniesa. Bacteriophages carrying antibiotic resistance genes in fecal waster from cattle, pigs, and poultry. *Antimicrob. Agents Chemother.*, 55(10):4908–4911, 2011.
- [9] P. L. Wagner and M. K. Waldor. Bacteriophage control of bacterial virulence. *Infect. Immun.*, 70(8):3985–3993, 2002.
- [10] S. M. Faruque, I. B. Bin Naser, M. J. Islam, A. S. G. Faruque, and et al. Seasonal epidemics of cholera inversely correlate with prevalence of environmental cholera phages. *Proc. Natl. Acad. Sci. USA*, 102(5):1702–1707, 2005.
- [11] S. M. Faruque and J. J. Mekalanos. Phage-bacterial interactions in the evolution of toxigenic *Vibrio cholerae*. *Virulence*, 3(7):556–565, 2012.

- [12] S. M. Chekabab, G. Jubelin, C. M. Dozois, and J. Harel. PhoB activates Escherichia coli O157:H7 virulence factors in response to inorganic phosphate limitation. *PLoS ONE*, 9(4):e94285, 2014.
- [13] X. Zhang, A. D. McDaniel, L. E. Wolf, G. T. Keusch, M. K. Waldor, and D. W. Acheson. Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. *J. Infect Dis.*, 181(2):664–670, 2000.
- [14] T. Frieden and et al. Antibiotic resistance trends in the united states. *CDC*, 2013.
- [15] CDC. National shiga toxin-producing Escherichia coli (STEC) surveillance overview. *USD-HHS*, 2012.
- [16] WHO. Cholera. *Weekly epidemiological record*, 89(31):345–356, 2014.
- [17] WHO. Diphtheria vaccine. *Weekly epidemiological record*, 81(3):321–32, 2006.
- [18] K. L. Kotloff, J. P. W., B. Ivanoff, J. D. Clemens, D. L. Swerdlow, P. J. Sansonetti, G. K. Adak, and M. M. Levine. Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. *Bulletin of the World Health Organization*, 77(8):651–666, 1999.
- [19] R. C. Matos and et al. Enterococcus faecalis prophage dynamics and contributions to pathogenic traits. *PLoS Genet.*, 9(6):e1003539, 2013.
- [20] N. D. Maynard, E. W. Birch, J. C. Sandhvi, L. Chen, M. V. Gutschow, and M. W. Covert. A forward-genetic screen and dynamic analysis of lambda phage host-dependencies reveals an extensive interaction network and a new anti-viral strategy. *PLOS Genet.*, 6(7):e1001017, 2010.
- [21] D. I. Friedman, E. R. Olson, C. Georgopoulos, K. Tilly, and I. Herskowitz. Interactions of bacteriophage and host macromolecules in the growth of bacteriophage lambda. *Microbiol. Rev.*, 48:299–325, 1984.
- [22] S. V. Rajagopala, S. Casjens, and P. Uetz. The protein interaction map of bacteriophage lambda. *MBC Microbiol.*, 11:213–228, 2011.
- [23] F. St. Pierre. *Determination of cell fate selection in phage lambda infection*. PhD thesis, Massachusetts Institute of Technology, 2009.
- [24] H. Echols, L. Gree, and A. B. Oppenheim. Role of the cro gene in bacteriophage lambda development. *J. Mol. Biol.*, 80(2):203–216, 1973.

- [25] D. Court, C. Brady, M. Rosenberg, D. L. Wulff, M. Mahoney, and S. U. Izumi. Control of transcription termination: a Rho-dependent termination site in bacteriophage lambda. *J. Mol. Biol.*, 138(2):231–254, 1980.
- [26] M. Rosenberg, D. Court, H. Shimatake, C. Brady, and D.L. Wulff. The relationship between function and DNA sequence in an intercistronic regulatory region in phage lambda. *Nature*, 272(5652):414–423, 1978.
- [27] R. Lutz and H. Bujard. Regulation of lambda exonuclease synthesis: role of the N gene product and lambda repressor. *J. Mol. Biol.*, 49(2):515–519, 1970.
- [28] J. W. Roberts. Termination factor for RNA synthesis. *Nature*, 224(5225):1168–1174, 1969.
- [29] A. D. Kaiser. Mutations in a temperate bacteriophage affecting its ability to lysogenize Escherichia coli. *Virology*, 3(1):42–61, 1957.
- [30] I. Herskowitz and D. Hagen. The lysis-lysogeny decision of phage lambda: explicit programming and responsiveness. *Annu. Rev. Genet.*, 14:399–445, 1980.
- [31] P. Kourilsky. Lysogenization by bacteriophage. I. Multiple infection and the lysogenic response. *Mol. Gen. Genet.*, 122(2):183–195, 1973.
- [32] A. Oppenheim and A. B. Oppenheim. Regulation of the int gene of bacteriophage lambda:activation by the cII and cIII gene products and the role of the pI and pL promoters. *Mol. Gen. Genet.*, 165(1):39–46, 1978.
- [33] I. Herskowitz and D. Hagen. The lysis-lysogeny decision of phage lambda: explicit programming and responsiveness. *Ann. Rev. Genet.*, 14:399–445, 1980.
- [34] A. Rattray, S. Altuvia, G. Mahajna, A. B. Oppenheim, and M. Gottesman. Control of bacteriophage lambda CII activity by bacteriophage and host functions. *J. Bacteriol.*, 159(1):238–242, 1984.
- [35] O. Kobiler, S. Koby, D. Teff, and A. B. Oppenheim. The phage lambda CII transcriptional activator carries a c-terminal domain signaling for rapid proteolysis. *Proc. Natl. Acad. Sci. USA.*, 99(23):14964–14969, 2002.
- [36] A. Kihara, Y. Akiyama, and K. Ito. Host regulation of lysogenic decision in bacteriophage lambda: transmembrane modulation of FtsH (HflB), the cII degrading protease, by HflKC (HflA). *Proc. Natl. Acad. Sci. USA.*, 94(11):5544–5549, 1997.
- [37] H. Echols and L. Green. Establishment and maintenance of repression by bacteriophage lambda: the role of the cI, cII, and cIII proteins. *Proc. Natl. Acad. Sci. USA*, 68(9):2190–2194, 1971.

- [38] I. Herskowitz and E. R. Signer. A site essential for expression of all late genes in bacteriophage lambda. *Proc. Natl. Acad. Sci. USA*, 82(10):3134–3138, 1985.
- [39] L. Zeng, S. O. Skinner, C. Zong, J. Sippy, M. Feiss, and I. Golding. Decision making at the subcellular level determines the outcome of bacteriophage infection. *Cell*, 141:682–691, 2010.
- [40] A. Arkin, J. Ross, and H. H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. *Genetics*, 149(4):1633–1648, 1998.
- [41] Y. Taniguchi, P. J. Choi, G. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538, 2010.
- [42] P. Kourilsky. Lysogenization by bacteriophage lambda and the regulation of lambda repressor synthesis. *Virology*, 45(3):853–857, 1971.
- [43] F. St. Pierre and D. Endy. Determination of cell fate selection during lambda phage infection. *Proc. Natl. Acad. Sci. USA*, 105(52):20705–20710, 2008.
- [44] J. T. Trinh, T Szekely, Q. Shao, G. Balazsi, and L. Zeng. Cell fate decision emerge as phages cooperate or compete inside their host. *Nat. Commun.*, 8(14341):1–13, 2017.
- [45] C. Herman, T. Ogura, T. Tomoyasu, S. Hiraga, and Y. Akiyama. Cell growth and lambda phage development controlled by the same essential Escherichia coli gene, ftsH/hflB. *Proc. Natl. Acad. Sci. USA*, 90:10861–10865, 1993.
- [46] A. Kihara, Y. Akiyama, and K. Ito. Revisiting the lysogenization control of bacteriophage lambda. identification and characterization of a new host component, HflD. *J. Biol. Chem.*, 276:13695–13700, 2001.
- [47] H. I. Miller and D. I. Friedman. An E. coli gene product required for lambda site-specific recombination. *Cell*, 20:711–719, 1980.
- [48] M. E. Ortega and C. E. Catalano. Bacteriophage lambda gpNu1 and Escherichia coli IHF proteins cooperatively bind and bend viral DNA: implications for the assembly of a genome-packaging motor. *Biochemistry*, 45:5180–5189, 2006.
- [49] J. Osipiuk, C. Georgopoulos, and M. Zylicz. Initiation of lambda DNA replication. the Escherichia coli small heat shock proteins DnaJ and GrpE, increase DnaK’s affinity for the lambda P protein. *J. Biol. Chem.*, 268:4821–4827, 1993.
- [50] A. Das. How the phage lambda N gene product suppresses transcription termination: communication of RNA polymerase with regulatory proteins mediated by signals in nascent RNA. *J. Bacteriol.*, 174:6711–6716, 1992.

- [51] L. Randall-Hazelbauer and M. Schwartz. Isolation of the bacteriophage lambda receptor from *Escherichia coli*. *J. Bacteriol.*, 116:1436–1446, 1973.
- [52] J. Elliott and W. Arber. *E. coli* k-12 pel mutants, which block phage lambda DNA injection, coincide with ptsM, which determine a component of a sugar transport system. *Mol. Gen. Genet.*, 161:1–8, 1978.
- [53] L. L. Randall. Quantitation of the loss of the bacteriophage lambda receptor protein from the outer membrane of lipopolysaccharide-deficient strains of *Escherichia coli*. *J. Bacteriol.*, 123:41–46, 1975.
- [54] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, 2:2006.0008, 2006.
- [55] S. Blasche, S. Wuchty, S. V. Rajagopala, and P. Uetz. The protein interaction network of bacteriophage lambda and its host, *Escherichia coli*. *J. Virol.*, 87(23):12745–12755, 2013.
- [56] D. Schcherbo, C. S. Murphy, G. V. Ermakova, E. A. Solovieva, T. V. Chepurnykh, and et al. Far-red fluorescent tags for protein imaging in living tissues. *Biochem. J.*, 418(3):567–574, 2009.
- [57] N. C. Shaner, G. G. Lambert, A. Chammas, Y. Ni, P. J. Cranfill, M. A. Baird, B. R. Sell, J. R. Allen, R. N. Day, M. Israelsson, M. W. Davidson, and J. Wang. A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nature Methods*, 10:407–409, 2013.
- [58] E. H. Szybalski and W. Szybalki. A comprehensive molecular map of bacteriophage lambda. *Gene*, 7(3):217–270, 1979.
- [59] Registry of standard biological parts.
- [60] D. P. Pires, S. Cleto, S. Sallankorva, J. Azeredo, and T. K. Lu. Genetically engineered phages: a review of advances over the last decade. *Microbiol. Mol. Biol. Rev.*, 80:523–543, 2016.
- [61] L. J. Marinelli, G. F. Hatfull, and M. Piuri. Recombineering: a powerful tool for modification of bacteriophage genomes. *Bacteriophage*, 2(1):5–14, 2012.
- [62] H. Ando, S. Lemire, D. P. Pires, and T. K. Lu. Engineering modular viral scaffolds for targeted bacterial population editing. *Cell Syst.*, 1:187–196, 2015.
- [63] P. McInerney, P. Adams, and M. Z. Hadi. Error rate comparison during polymerase chain reaction by DNA polymerase. *Mol. Biol. Int.*, page 287430, 2014.

- [64] R. D. Gietz and R. H. Schiestl. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, 2(1):31–34, 2007.
- [65] K. Lech and R. Brent. *Current Protocols in Molecular Biology*. John Wiley and Sons, Inc, 3rd edition, 1990.
- [66] H. Shi, A. Colavin, T. K. Lee, and K. C. Huang. Strain library imaging protocol for high-throughput, automated single-cell microscope of large bacterial collections array in multiwell plates. *Nat. Protoc.*, 12:429–438, 2017.
- [67] D. A. Van Valen, T. Kudo, K. M. Lane, D. N. Macklin, N. T. Quach, M. M. DeFelice, and et al. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS Comput. Biol.*, 12(11):e1005177, 2016.
- [68] A. Alba, R. M. Aguilar-Ponce, J. F. Vigueras-Gomez, and E. Arce-Santana. Phase correlation based image alignment with subpixel accuracy. *Advances in Artificial Intelligence*, 7629:171–182, 2013.
- [69] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 3rd edition, 2006.
- [70] G. D'Agostini. Bayesian inference in processing experimental data: principles and basic applications. *Reports on Progress in Physics*, 66(9), 2003.
- [71] F. Miao, S. K. Drake, and D. S. Kompala. Characterization of gene expression in recombinant Escherichia coli cells infected with phage lambda. *Biotechnol. Prog.*, 9(2):153–159, 1993.
- [72] O. Kobiler, A. B. Oppenheim, and C. Herman. Recruitment of host ATP-dependent proteases by bacteriophage lambda. *J. Struct. Biol.*, 146:72–78, 2004.
- [73] K. Liberek, C. Georgopoulos, and M. Zylicz. Role of the Escherichia coli DnaK and DnaJ heat shock proteins in the initiation of bacteriophage lambda dna replication. *Proc. Natl. Acad. Sci. USA*, 85:6632–6636, 1988.
- [74] M. Li, H. Moyle, and M. M. Susskind. Target of the transcriptional activation function of phage lambda cI protein. *Science*, 263:75–77, 1994.
- [75] B. Kedzierska, D. J. Lee, G. Wegrzyn, S. J. Busby, and M. S. Thomas. Role of the RNA polymerase alpha subunits in CII-dependent activation of the bacteriophage lambda pE promoter: identification of important residues and positioning of the alpha C-terminal domains. *Nucleic Acids Res.*, 32:834–841, 2004.
- [76] A. Buhr, K. Flukiger, and B. Erni. The glucose transporter of Escherichia coli. overexpression, purification, and characterization of functional domains. *J. Biol. Chem.*, 269(38):23437–23443, 1994.

- [77] M. H. Saier and S. Roseman. Sugar transport. the crr mutation: its effect on repression of enzyme synthesis. *J. Biol. Chem.*, 251(21):6598–6605, 1976.
- [78] J. L Kadrmas and C. R. Raetz. Enzymatic synthesis of lipopolysaccharide in Escherichia coli. purification and properties of heptosyltransferase I. *J. Biol. Chem.*, 273(5):2799–2807, 1998.
- [79] F. Sargent, E. G. Bogsch, N. R. Stanley, M. Wexler, C. Robinson, B. C. Berks, and T. Palmer. Overlapping functions of components of a bacterial Sec-independent protein export pathway. *EMBO J.*, 17(13):3640–3650, 1998.
- [80] J. H. Kim, J. R. Bothe, T. R. Alderson, and J. L. Markley. Tangled web of interactions among proteins involved in iron-sulfur cluster assembly as unraveled by NMR, SAXS, chemical crosslinking, and functional studies. *Biochim. Biophys. Acta.*, 1853(6):1416–1428, 2015.
- [81] D. Shiomi, M. Sakai, and H. Niki. Determination of bacterial rod shape by novel cytoskeletal membrane protein. *EMBO J.*, 27(23):3081–3091, 2008.
- [82] T. Wu, J. Malinverni, N. Ruiz, S. Kim, T. J. Silhavy, and D. Kahne. Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli. *Cell*, 121(2):235–245, 2005.
- [83] T. Uehara, K. R. Parzych, T. Dinh, and T. G. Bernhardt. Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis. *EMBO J.*, 29(8):1412–1422, 2010.
- [84] M. Lilic, M. Jovanovic, G. Jovanovic, and D. J. Savic. Identification of the CysB-regulated gene, hslJ, related to the Escherichia coli novobiocin resistance phenotype. *FEMS Microbiol. Lett.*, 224(2):239–246, 2003.
- [85] Ecocyc.
- [86] S. Eberhardt, G. Richter, W. Gimbel, T. Werner, and A. Bacher. Cloning, sequencing, mapping and hyperexpression of the ribC gene coding for riboflavin synthase of Escherichia coli. *Eur. J. Biochem.*, 242(3):712–719, 1996.
- [87] J. W. Campbell, R. M. Morgan-Kiss, and J.E. Cronan Jr. A new Escherichia coli metabolic competency: growth on fatty acids by a novel anaerobic beta-oxidation pathway. *Mol. Microbiol.*, 47(3):793–805, 2003.
- [88] S. Amitai, I. Kolodkin-Gal, M. Hananya-Meltabashi, A. Sacher, and H. Engelberg-Kulka. Escherichia coli MazF leads to the simultaneous selective synthesis of both death proteins and survival proteins. *PLoS Genet.*, 5(3):e1000390, 2009.

- [89] J. C. Waller, S. Alvarez, V. Naponelli, A. Lara-Nunez, I. K. Blaby, V. Da Silva, M. J. Ziemak, T. J. Vickers, and et al. A role for tetrahydrofolates in the metabolism of iron-sulfur clusters in all domains of life. *Proc. Natl. Sci. USA*, 107(23):10412–10417, 2010.
- [90] C. M. Diges and O. C. Uhlenbeck. Escherichia coli Dbpa is a 3'-5' rna helicase. *Biochemistry*, 44(21):7903–7911, 2005.
- [91] C. Jain. The E. coli RhlE RNA helicase regulates the function of related RNA helicases during ribosome assembly. *RNA*, 14(2):381–389, 2007.
- [92] Y. C. Chen, C. L. Li, Y. Y. Hsiao, Y. Duh, and H. S. Yuan. Structure and function of TatD exonuclease in DNA repair. *Nucleic Acids Res.*, 42(16):10776–10785, 2014.
- [93] C. N. Chang and N. Chang. Methylation of the ribosomal proteins in Escherichia coli. nature and stoichiometry of the methylated amino acids in 50S ribosomal proteins. *Biochemistry*, 14(3):468–477, 1975.
- [94] D. V. Lesnyak, J. Osipiuk, T. Skarina, P. V. Sergiev, A. A. Bogdanov, A. Edwards, A. Savchenko, A. Joachimiak, and O. A. Dontsova. Methyltransferase that modifies guanine 966 of the 16S rRNA: functional identification and tertiary structure. *J. Biol. Chem.*, 282(8):5880–5887, 2007.
- [95] L. Britton and I. Fridovich. Intracellular localization of the superoxide dismutases of Escherichia coli: a reevaluation. *J. Bacteriol.*, 131(3):815–820, 1977.
- [96] C. Ueguchi and K. Ito. Multicopy suppression: an approach to understanding intracellular functioning of the protein export system. *J. Bacteriol.*, 174(5):1454–1461, 1992.
- [97] H. H. Yim and M. Villarejo. osmY, a new hyperosmotically inducible gene, encodes a periplasmic protein in Escherichia coli. *J. Bacteriol.*, 174(11):3637–3644, 1992.
- [98] A. Budd, S. Blandin, E. A. Levashina, and T. J. Gibson. Bacterial a2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome Biol.*, 5(6):R38, 2004.
- [99] N. Verstraeten, W. J. Knapen, C. I. Kint, V. Lievens, B. Van den Bergh, L. Dewachter, J. E. Michiels, and et al. Obg and membrane depolarization are part of a microbial bet-hedging strategy that leads to antibiotic tolerance. *Mol. Cell.*, 59(1):9–21, 2015.
- [100] T. Katayama, M. Takata, and K. Sekimizu. CedA is a novel Escherichia coli protein that activates the cell division inhibited by chromosomal DNA over-replication. *Mol. Microbiol.*, 26(4):687–697, 1997.

- [101] P. Sharma, A. K. Tomar, and B. Kundu. Identification of functional interactome of a key cell division regulatory protein CedA of *E. coli*. *Int. J. Biol. Macromol.*, 106:763–767, 2018.
- [102] M. S. Koo, J. H. Lee, S. Y. Rah, W. S. Yeo, J. W. Lee, K. L. Lee, Y. S. Koh, S. O. Kang, and J. H. Roe. A reducing system of the superoxide sensor SoxR in *Escherichia coli*. *EMBO J.*, 22(11):2614–2622, 2003.
- [103] Q. Ma, Z. Yang, M. Pu, W. Peti, and T. K. Wood. Engineering a novel c-di-GMP-binding protein for biofilm dispersal. *Environ. Microbiol.*, 13(3):631–642, 2011.
- [104] B. E. Funnell. ParB partition proteins: complex formation and spreading at bacterial and plasmid centromeres. *Front. Mol. Biosci.*, 3:44, 2016.
- [105] D. A. Kelkar and A. Chattopadhyay. The gramicidin ion channel: a model membrane protein. *Biochim. Biophys. Acta.*, 1768(9):2011–2025, 2007.
- [106] B. Kleuser, H. Rieter, and G. Adam. Selective effects by valinomycin on cytotoxicity and cell cycle arrest of transformed versus nontransformed rodent fibroblasts in vitro. *Front. Cell Dev. Biol.*, 45(7):3022–3028, 1985.
- [107] G. B. Pinchot. The mechanism of uncoupling of oxidative phosphorylation by 2,4-dinitrophenol. *J. Biol. Chem.*, 242(20):4577–4583, 1967.
- [108] A. Prindle, J. Liu, M. Asally, S. Ly, J. Garcia-Ojalvo, and G. M. Suel. Ion channels enable electrical communication within bacterial communities. *Nature*, 527(7576):59–63, 2015.